CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

# How Family Size Affects Parents' Labour Market Outcomes: A Meta-Analysis

Master's thesis

Author: Bc. Nigina Baydadaeva
Study program: Economics and Finance
Supervisor: doc. PhDr. Zuzana Havránková, Ph.D.
Year of defense: 2024

## 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, July 31, 2024

Nigina Baydadaeva

# Abstract

The wage gap between parents and non-parents is a subject of extensive research. Numerous scholars have thoroughly discussed the "motherhood penalty" and "fatherhood premium" as phenomena illustrating the varying labour market impacts of parenthood on women and men. To what extent can we trust these findings, and does publication bias musk the true impact of family size on parents' labour market performance? This thesis aims to investigate the impact of family size on parents' labour market outcomes through a comprehensive meta-analytical review. We collect 1542 estimates from 89 primary studies, comprising 1323 estimates for females, 209 for males, and 10 gender-neutral estimates. Subsequently, we convert all estimates into Partial Correlation Coefficients (PCCs). Our analysis indicates a slightly negative publication bias for women, and a substantial positive bias for men. After correcting for publication bias, the effect of parenthood on males appears negligible, while the effect on mothers remains negative. Furthermore, we employ Bayesian Model Averaging to explore the heterogeneity among PCCs. Consequently, we control for 32 additional study characteristics for the female sample, and 28 for the male sample. The findings reveal that after controlling for additional research parameters, there is no further evidence of publication bias in the female sample. However, the results of the robustness checks suggest that there might be negative publication bias present in the subsample of studies that control for endogeneity. For the male sample, the heterogeneity analysis confirms the presence of positive publication bias, suggesting that the effect of fatherhood on labour market outcomes is exagerated.



| | |
|---|---|
| **JEL Classification** | F12, F21, F23, H25, H71, H87 |
| **Keywords** | meta-analysis, publication bias, family size, labour market outcome, parenthood |
| **Title** | How Family Size Affects Parents' Labour Market Outcomes: A Meta-Analysis |

# Abstrakt

Mzdový rozdíl mezi rodiči a bezdětnými je předmětem rozsáhlého výzkumu. Řada studií rozsáhle diskutuje o „penalizaci mateřství“ a „prémii otcovství“ jako o fenoménech ilustrujících různorodé dopady rodičovství na trh práce pro ženy a muže. Do jaké míry můžeme těmto zjištěním důvěřovat a zkresluje publikační zkreslení skutečný dopad velikosti rodiny na pracovní výkon rodičů? Tato práce si klade za cíl prozkoumat vliv velikosti rodiny na pracovní výsledky rodičů prostřednictvím komplexního meta-analytického výzkumu. Shromažďujeme 1542 odhadů z 89 primárních studií, z nichž 1323 odhadů se týká žen, 209 mužů a 10 odhadů je genderově neutrálních. Následně všechny odhady převádíme na parciální korelační koeficienty. Naše analýza naznačuje mírné negativní publikační zkreslení u žen a výrazné pozitivní publikační zkreslení u mužů. Po opravě o publikační zkreslení se zdá, že vliv rodičovství na muže je zanedbatelný, zatímco vliv na matky zůstává negativní. Dále používáme bayesovské průměrování modelů k prozkoumání heterogenity mezi parciálními korelačními koeficienty. V důsledku toho kontrolujeme 32 dalších charakteristik studií pro ženský vzorek a 28 pro mužský vzorek. Výsledky ukazují, že po kontrole dalších výzkumných parametrů již v ženském vzorku není žádný důkaz o přítomnosti publikačního zkreslení. Nicméně výsledky kontrol robustnosti naznačují, že v podvzorku studií, které kontrolují endogenitu, může být přítomno negativní publikační zkreslení. Pro mužský vzorek analýza heterogenity potvrzuje přítomnost pozitivního publikačního zkreslení, což naznačuje, že vliv otcovství na pracovní výsledky je nadhodnocený.

**Klasifikace JEL** F12, F21, F23, H25, H71, H87

**Klíêová slova** meta-analýza, publikační zkreslení, velikost rodiny, pracovní výsledky, rodičovství

**Název práce** Jak velikost rodiny ovlivňuje pracovní výsledky rodičů: Meta-analýza

# Acknowledgments

The author expresses deep gratitude to doc. PhDr. Zuzana Havránková, Ph.D., for the profound guidance and consistent mentorship throughout the research process. The author also extends the heartfelt thanks to family, friends, and close ones for their love, unwavering support and patience during the academic journey.

Typeset in LaTeXusing the IES Thesis Template.

**Bibliographic Record**

Baydadaeva, Nigina: *How Family Size Affects Parents' Labour Market Outcomes: A Meta-Analysis*. Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2024, pages 101. Advisor: doc. PhDr. Zuzana Havránková, Ph.D.

# Contents

**List of Tables** viii

**List of Figures** x

**Acronyms** xi

**1 Introduction** 1

**2 Theoretical background** 4

2.1 Family size and parental employment: An introductory survey . 4

2.2 Evolution of patterns in family size . . . . . . . . . . . . . . . . . 10

2.3 Approaches to Estimation of Family Size Impact . . . . . . . . . 10

**3 Data Overview** 15

3.1 Data Compilation . . . . . . . . . . . . . . . . . . . . . . . . . . 15

3.2 Data Processing . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

3.3 Descriptive Analysis . . . . . . . . . . . . . . . . . . . . . . . . . 20

**4 Publication Bias** 26

4.1 Visual Test: Funnel Plot . . . . . . . . . . . . . . . . . . . . . . 27

4.2 Linear Tests . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

4.3 Non-Linear Tests . . . . . . . . . . . . . . . . . . . . . . . . . . 35

4.4 Caliper Test . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37

4.5 Robustness check . . . . . . . . . . . . . . . . . . . . . . . . . . 40

**5 Heterogeneity** 42

5.1 Explanatory variables . . . . . . . . . . . . . . . . . . . . . . . . 42

5.1.1 Data parameters . . . . . . . . . . . . . . . . . . . . . . . 43

5.1.2 Publication specifics . . . . . . . . . . . . . . . . . . . . . 43

5.1.3 Methodology . . . . . . . . . . . . . . . . . . . . . . . . 44

5.1.4 Design of the analysis . . . . . . . . . . . . . . . . . . . . . 44
5.1.5 Variable specifications . . . . . . . . . . . . . . . . . . . . . 45
5.1.6 Set of controls . . . . . . . . . . . . . . . . . . . . . . . . . 46
5.2 Estimation methods . . . . . . . . . . . . . . . . . . . . . . . . . 48
5.2.1 Overview of Bayesian Model Averaging . . . . . . . . . . . 48
5.2.2 Application of Bayesian Model Averaging . . . . . . . . . . 51
5.2.3 Frequentist Model Averaging . . . . . . . . . . . . . . . . . 51
5.2.4 Frequentist Check . . . . . . . . . . . . . . . . . . . . . . . 52
5.3 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52
5.3.1 Results: female sample . . . . . . . . . . . . . . . . . . . . 53
5.3.2 Results: male sample . . . . . . . . . . . . . . . . . . . . . 57
5.4 Robustness check . . . . . . . . . . . . . . . . . . . . . . . . . . 60

**6 Best practice estimate 62**

**7 Conclusion 64**

**Bibliography 77**

# List of Tables


3.1 Primary studies used in the meta-analysis . . . . . . . . . . . . 17
3.2 Descriptive statistics for PCCs. Female sample . . . . . . . . . . 21
3.3 Descriptive statistics for PCCs. Male sample . . . . . . . . . . . 24
3.4 Descriptive statistics for weighted PCCs . . . . . . . . . . . . . 25

4.1 Linear Tests' results: female sample . . . . . . . . . . . . . . . . . 32
4.2 IV's results: female sample . . . . . . . . . . . . . . . . . . . . . . . 33
4.3 Linear Tests' results: male sample . . . . . . . . . . . . . . . . . . 34
4.4 IV's results: male sample . . . . . . . . . . . . . . . . . . . . . . . . 34
4.5 Non-linear Tests' results: female sample . . . . . . . . . . . . . 36
4.6 Non-linear Tests' results: male sample . . . . . . . . . . . . . . 37
4.7 Caliper Test results: female sample . . . . . . . . . . . . . . . . . 39

5.1 Descriptive statistics of variables used for heterogeneity . . . . . 47
5.2 Summary of results for BMA, FMA, and Frequentist Check: female sample . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56
5.3 Summary of results for BMA, FMA, and Frequentist Check: male sample . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59

6.1 Best practice estimates for female and male samples . . . . . . . 63

A.1 Primary studies used in the meta-analysis: sample of female respondents . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . I
A.2 Primary studies used in the meta-analysis: sample of male respondents . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . II
B.1 Linear Tests' results: female sample, subsample Earnings . . . . VI
B.2 Linear Tests' results: female sample, subsample Endogeneity . . VI
B.3 IV's results: female sample, subsample Earnings . . . . . . . . . VII
B.4 IV's results: female sample, subsample Endogeneity . . . . . . . VII
B.5 Non-linear Tests' results: female sample, subsample Earnings . . VIII

B.6 Non-linear Tests' results: female sample, subsample Endogeneity VIII

# List of Figures


3.1 Distribution of PCCs for the female sample . . . . . . . . . . . . 22
3.2 Forest plot of PCCs for the female sample across countries . . . 23
3.3 Distribution of PCCs for the male sample . . . . . . . . . . . . 24

4.1 Funnel plot of partial correlation coefficients for the female sample. 28
4.2 Funnel plot of partial correlation coefficients for the male sample. 29
4.3 Distribution of t-statistics of PCCs: female sample . . . . . . . 38
4.4 Distribution of t-statistics of PCCs: male sample . . . . . . . . 39

5.1 Model inclusion of the BMA estimation: female sample . . . . . 53
5.2 Model inclusion of the BMA estimation: male sample . . . . . . 57

A.1 Forest plot of PCCs for the female sample across studies . . . . III
A.2 Forest plot of PCCs for the male sample across countries . . . . IV
A.3 Forest plot of PCCs for the male sample across studies . . . . . V
C.4 Correlation matrix between variables of variables capturing heterogeneity: female sample . . . . . . . . . . . . . . . . . . . . . . IX
C.5 Correlation matrix between variables of variables capturing heterogeneity: male sample . . . . . . . . . . . . . . . . . . . . . . . X
C.6 BMA estimation across different priors: female sample . . . . . XI
C.7 BMA estimation across different priors: male sample . . . . . . XI
C.8 Model inclusion of the BMA estimation: female sample, subsample Earnings . . . . . . . . . . . . . . . . . . . . . . . . . . . . XII
C.9 BMA estimation across different priors: female sample, subsample Earnings . . . . . . . . . . . . . . . . . . . . . . . . . . . . XII
C.10 Model inclusion of the BMA estimation: female sample, subsample Endogeneity . . . . . . . . . . . . . . . . . . . . . . . . . . XIII
C.11 BMA estimation across different priors: female sample, subsample Endogeneity . . . . . . . . . . . . . . . . . . . . . . . . . . . XIII

# Acronyms

**BMA** Bayesian Model Averaging

**EU** European Union

**FAT** Funnel Asymmetry Test

**FE** Fixed Effects

**FMA** Frequentist Model Averaging

**GLS** Generalized Least Squares

**GMM** Generalized Method of Moments

**IV** Instrumental Variables

**MMA** Mallows Model Average

**OECD** Organisation for Economic Co-operation and Development

**OLS** Ordinary Least Squares

**PCC** Partial Correlation Coefficient

**PET** Precision Estimate Test

**PEESE** Precision-Effect Estimate with Standard Error

**PIP** Posterior Inclusion Probability

**PMP** Posterior Model Probability

**RePEc** Research Papers in Economics

**SE** Standard Error

**TFR** Total Fertility Rate

**TSLS (2SLS)** Two-Stage Least Squares

**UIP** Unit Information Prior

**VIF** Variance Inflation Factor

**WAAP** Weighted Average of Adequately Powered

**WLS** Weighted Least Squares

# Chapter 1

# Introduction

The transition to parenthood induces significant changes in individuals' lives, not only reshaping their personal aspirations and family dynamics, but also affecting their labour market behaviour and economic stability. The conventional caregiving role of women, coupled with the existing gender wage gap, may place mothers at a significant disadvantage in the labour market. In contrast, fathers are often perceived as benefiting from a "fatherhood premium", where their wages are positively influenced by increasing family size. Yet, does the evidence substantiate the existence of the motherhood penalty and fatherhood premium, or are these findings commonly distorted by authors' intentions to publish widely expected results?

Female labour force participation has notably increased in recent years (Agüero & Marks 2011; Lundborg *et al.* 2021), a trend often attributed to declining childbirth rates (Browning 1992). Women are more likely to adjust their labour supply in response to changes in family size (Cools *et al.* 2017), primarily due to the disproportionate distribution of parental responsibilities (Agüero & Marks 2008) and the conflicts between the roles of workers and mothers (Mason & Palan 1981). Some scholars suggest that mothers may face discrimination compared to their childless colleagues and be perceived as less committed to their jobs (Correll *et al.* 2007), further experiencing missed career promotions (Aisenbrey *et al.* 2009), wage penalties (Budig & Hodges 2010) and unwanted occupation changes to more flexible and family-friendly positions (Gangl & Ziefle 2009).

Fathers, on the other hand, tend to spend more time at work than their childless counterparts and do not encounter the same role incompatibility due to their traditional roles as providers and breadwinners (Cukrowska-Torzewska

2015). From the "good provider" perspective, men tend to work more as they become fathers (Kaufman & Uhlenberg 2000). However, the impact of children on men's labour market outcomes is more heterogenous than on women, with some authors reporting a negative effect (Lundberg & Rose 2000), while others suggest no effect (Carlin & Flood 1997) or even posit the existence of premiums (Kaufman & Uhlenberg 2000; Baranowska-Rataj & Matysiak 2022) in comparison to non-fathers.

At the heart of this thesis is the examination of the relationship between family size and parents' labour market outcomes. Building upon the literature review by Clarke (2018), we aim to perform a systematic review by implementing the meta-analytical regressions on two separate samples: female and male. This approach seeks to mitigate potential publication selectivity, thereby revealing the true nature of the relationship between family size and labour market outcomes.

To conduct our analysis, we compiled 1542 estimates from 89 primary studies, of which 1323 estimates pertain to females, 209 to males, and 10 estimates capture the relationship irrespective of respondents' gender. We chose not to restrict ourselves to specific estimates of family size and labour market outcomes, thereby collecting all available estimates that provide insights into the effect of interest. Consequently, both family and labour-related variables in our dataset manifest various forms. Therefore, the analysis and discussion of the results are structured around Partial Correlation Coefficients (PCCs), the standardization method employed in recent meta-analyses (Doucouliagos 2005; Zigraiova & Havranek 2016; Cazachevici & Horvath 2020).

As this thesis aims to determine and cleanup the publication bias, it is essential to introduce the reader to the context. Publication bias typically occurs when statistically significant, favourable or widely expected results are more likely to be published than those non-significant and unfavourable (Rosenthal 1979). Journal editors and reviewers often prefer significant findings, and authors tend to produce and submit studies that are more likely to be accepted. This decision-making process can distort the overall body of empirical evidence. However, underrepresenting of insignificant outcomes does not necesserily result from deliberate actions. Unintentional biases in reporting practices and research design can also contribute to this issue.

In this paper, we conduct several analyses to identify publication bias. Firstly, we perform the visual test, then we continue with testing linear and non-linear models, and we relax the exogeneity assumption. In the second part of the thesis, we introduce distinct study characteristics possibly affecting the heterogeneity in PCCs and present the subjective best practice estimate. We perform the analysis on the female and male sample separately and compare the results at the end of each section.

Although the funnel plot for the female sample does not clearly indicate publication bias, the results from linear and Caliper tests suggest the presence of negative publication selectivity in primary studies. However, after controlling for an additional 32 study characteristics, we find no further evidence of publication bias in our main body of our analysis for the female sample. Specifically, our findings indicate that the negative effect of motherhood is more pronounced in more recent data and is associated with studies employing simplistic methodologies that ignore the potential endogeneity in the relationship. For the male sample, our analysis suggests the presence of positive publication bias. We observe that the publication bias persists among male respondents even after introducing 28 variables to capture heterogeneity.

The thesis is organized as follows. Chapter 2 delves into the theoretical background, discusses the main determinants of the relationship between family size and labour market outcomes, unravels the temporal patterns of labour supply and fertility, and evaluates distinct approaches to estimating the impact. Chapter 3 outlines the data collection procedure and highlights the key high-level trends observed in the female and male samples. We proceed with the tests of publication bias in Chapter 4, presenting the results from the visual analysis, linear and non-linear approaches. Chapter 5 expands the analysis by examining heterogeneity in our dataset to determine whether publication selectivity persists after accounting for other study characteristics. Finally, we present the subjective best practice estimate in Chapter 6 and summarize the results in Chapter 7.

# Chapter 2

# Theoretical background

## 2.1 Family size and parental employment: An introductory survey

In the recent past, a growing body of research has delved into the nuanced and multi-dimensional relationship between family size and parents' labour market outcomes (Angrist & Evans 1996a;a; Brewster & Rindfuss 2000; Cools *et al.* 2017). This paper focuses on heterosexual couples to study the work-family balance and differences in effects on parents of different genders. Although the effect on female labour supply is generally observed to be negative (Jacobsen *et al.* 1999; Agüero & Marks 2008), the impact on male labour supply is usually reported to be positive (Kaufman & Uhlenberg 2000; Baranowska-Rataj & Matysiak 2022), but further research in this domain is needed as the existing body of literature on this topic is limited in scope.

Before delving into the understanding of these gender-specific effects, it is essential to clarify the key definitions used throughout the paper. In this context, we employ the term "family size" as a synonym for fertility. There is a great degree of diversity in the definition of fertility across studies: while some scholars define fertility as the total number of children in the household (Agüero & Marks 2011), others define it as the number of children born during a specified period (Chevalier & Viitanen 2003). Besides, there are different proxies for the labour market outcome: hours worked (Lundberg 2005), wages (Waldfogel 1997), and labour force participation (Cáceres-Delpiano 2006). Since we are interested in the overall effect, we consider all definitions of family size and labour outcome to be of equal importance. Therefore, within the theoretical sections of this thesis, the reader may observe the application of diverse terminology in relation to labour force participation. Likewise, when analysing the male effect we consider the effect on mother's partner living in the same household as child or children, regardless of the genetic link with children (i.e., spouse, child's father).

The literature on the impact of fertility on female labour supply reports divergent results, with the majority of researchers presenting negative impact (Angrist & Evans 1996a;b; Jacobsen *et al.* 1999) while only some suggest a positive impact (Angrist & Schlosser 2010). These results support the maternal role incompatibility hypothesis (Stycos & Weller 1967), which in essence suggests that the negative effect exists only when women face conflicts between their roles as workers and mothers (Mason & Palan 1981). This idea brings about a situation in which women are forced to make tradeoffs between family and work. The work-family dilemmas arise mostly in the context of industrial organisation of production, where home-offices are rather unwelcome, if at all feasible (Mason & Palan 1981). Due to the insufficient supportive environment for new parents, women may have to suspend their career until childcare needs less assistance. Apart from unwanted career suspension, women may suffer from wage penalties (Waldfogel 1997; Avellar & Smock 2003; Budig & Hodges 2010), passing up career promotions (Aisenbrey *et al.* 2009), unwanted occupation changes to more family-friendly ones and switching to part-time employment (Gangl & Ziefle 2009). Moreover, some researchers state that women with children may be discriminated against in their workplaces and perceived as less committed to job and less qualified than childless colleagues (Correll *et al.* 2007).

The impact of children on men's labour supply is more heterogeneous, as some studies report positive effect (Kaufman & Uhlenberg 2000; Baranowska-Rataj & Matysiak 2022), while others find no (Carlin & Flood 1997) or negative effect (Lundberg & Rose 2000). Whatever the findings, parenthood facilitates significant changes in men's lives (Kaufman & Uhlenberg 2000). Bielby & Bielby (1989) propose that identification with work and family roles is inherent for each person, regardless of gender, but the identity formation process is gender dependent. Assuming scarcity in personal resources, the authors postulate that women may experience social pressure concerning the family-work tradeoff, whereas men may not be subject to equivalent expectations regarding their "husband" and "father" roles but are instead perceived as "providers". Thus, in traditional families commitment to work or family for men is not a zero-sum process, while the same may not be true for women. Kaufman & Uhlenberg

(2000) argue that from the "good provider" perspective, men tend to work more as they become fathers. However, if men are equally involved in childrearing, they suffer from the same family-work trade-off as women do. In general, literature about the effect of parenting on male labour market outcomes is quite limited as at the date of this thesis.

## Individual-level determinants

The relationship between family size and parents' labour market outcome can be examined from both microeconomic and macroeconomic perspectives. Studies following the microeconomic approach examine interrelation at the individual level. Considerable attention has been paid to how fertility decisions affect mothers' labour supply (Fleisher & Rhodes 1979; Angrist & Evans 1996a; Aaronson *et al.* 2021; Lundborg *et al.* 2021). Angrist & Evans (1996a) state that having children has a negative impact on mothers' labour supply, but the magnitude of the causal effect is not uniform across women with different levels of schooling. They claim that low-skilled women are exposed to the largest effects of childbearing on labour supply and therefore are more likely to leave the labour market. Desai & Waite (1991) argue that although education increases the likelihood of women working, it dwindles the probability of employment at the early stages of parenthood. However, in the long run, as kids mature, the impact of higher education on labour market return appears to reassert itself. Lundberg (2005) reports that the effect of children under the age of 3 on men's work hours varies substantially by education level: male respondents with lower level of education suffer a larger negative impact compared to highly educated fathers. Men with lower education levels may have less family-friendly or flexible jobs, resulting in less time left for parenting. They may also have lower-paying jobs and thus be unable to afford childcare without working longer hours, let alone time off work (Levy & Murnane 1992).

From the individual-level perspective, marital status or cohabitation are essential contributors to the employment decisions of new parents. Wenk & Garrett (1992) conclude that having a spouse in the household helps reduce female labour market exits and facilitates faster returns to work after childbirth by providing additional childcare, transportation, and necessary housekeeping activities. Having other household members who can provide childcare is positively associated with mothers' workforce participation (Floge 1989). Altintas & Sullivan (2017) claim that men contribute more into housework and childcare, although the intensity of results differ by country regimes (Nordic, Southern, Liberal and Corporatist). In countries where the burden of parenthood falls primarily on women and fathers are only weakly involved, men are more likely to be promoted (Baranowska-Rataj & Matysiak 2022). Thus, marital status or cohabitation may indirectly influence men's labour participation after childbirth if they are involved in parenthood, as time is a scarce resource. "Out-of-wedlock" childbirth elevates the importance of the mother's nonmarket time and is presumed to reduce her involvement in labour force participation and academic endeavours (Bronars & Grogger 1994). Since women with children born outside of marriage may already have a lower wealth position in some more conservative countries, the impact of marital status is likely affected by endogeneity. Bronars & Grogger (1994) apply "twins at first birth" as an exogenous fertility event on the sample of unwed women and confirm that mothers indeed experience sizable and negative short-run effects on labour force participation, although the impact varies by race.

The presence of a new-born seems to inhibit the labour market behaviour of white mothers more than their black peers (Shapiro & Mott 1979). Shapiro & Mott (1979) explain a major portion of the difference by the higher level of unemployment among blacks. Lehrer (1992) proposes that racial differences play a vital role among low-educated women. Based on her findings, childbearing has an insignificant effect on low-education blacks, while the labour supply of high-education women (both races) and low-education whites is negative and substantial. However, lower depressing impact on labour force participation of black mothers does not imply they care less about their descendants. Bell (1974) suggests that black husbands' sustained inadequacy in earning ability compels black wives to constantly contribute to the household's financial stability. Moreover, black women are more likely to work part-time evening jobs and benefit from enhanced opportunities for childcare (Sweet 1973). The fatherhood premium, which stands for an increase in hourly wages and annual earnings, appears to vary by race as well (Glauber 2008). The findings suggest that white men and latinos tend to benefit from a larger fatherhood premium compared to black men. Van Winkle & Fasang (2020) confirm the racial differences among fathers: the largest fatherhood premiums are observed for white men, followed by hispanic men, with the smallest premiums for black fathers.

The timing of the first child, and early childbearing, in particular, is recognised as a significant factor influencing labour force participation among both men and women (Trussell 1976; Hofferth & Moore 1979). Adolescent pregnancy exerts overt and covert negative influence on early child-bearers. Since parenthood requires a significant investment of time, evidence suggests that teenage mothers complete fewer years of school, have more children over their lifespan, accumulate less work experience, and consequently earn less than their child-free peers (Hofferth & Moore 1979). Thus, postponing a first birth until adulthood augments the likelihood of achieving a robust employment background and yields higher income for both parents. Delaying the first birth has also an indirect impact: it reduces the total family size. Among others, it leads to an increase in accumulated work experience, working hours and earnings of both parents (Hofferth & Moore 1979).

Child gender appears to have a noteworthy impact on parents' labour decisions (Lundberg & Rose 2002; Choi *et al.* 2008; Agüero *et al.* 2020). Having son as a first child increases the father's working hours by 100 hours per year (Choi *et al.* 2008) and earnings by 5.3% (Lundberg & Rose 2002), compared to having a firstborn daughter. In general, the birth of a male child is associated with higher marital stability and greater likelihood of marriage in case of "out-of-wedlock" birth (Choi *et al.* 2008). However, the impact on mothers may vary depending on regional and cultural differences. In low-income countries adolescent daughters are assuming command over household responsibilities with age and incrementally substitute mothers' tasks, which facilitate better work-life balance and imply significant wage premium (Agüero *et al.* 2020).

Angrist & Evans (1996a) claim that parents of same-sex siblings tend to have a higher likelihood of subsequent childbearing. However, while American and European parents may strive for a balanced sex composition of their offspring (Angrist & Evans 1996a; Andersson *et al.* 2007), the preferences of, for instance, Korean households are different. Despite incurring relatively higher bearing costs, boys are still preferred over girls due to the expectation of greater material support in old ages, which appears to be the most crucial factor in parental decisions to have sons (Chun & Oh 2002).

## Aggregate-level perspective

Another way to study the relationship between family size and labour supply is to approach it from the aggregate (or macroeconomic) perspective. Cross-country variations in women's work patterns depend largely on leave policies, availability of childcare and work-family supporting programs (Brewster & Rindfuss 2000). Hence, the effect of a new born on labour supply can be moderated by proper governmental support (Meyers & Gornick 2005). For instance, leaving the labour force for lengthy intervals in Germany, Japan and Ireland is primarily dictated by either shortage of childcare or conventional beliefs in intensive maternal participation. Conversely, new mothers in the United States and Scandinavian countries tend to take shorter maternity leaves due to the widespread availability of childcare services (Brewster & Rindfuss 2000).

Likewise, labour market stability, as indicated by the unemployment rate, significantly influences the fertility decisions of young individuals (Del Boca *et al.* 2003). Countries with high unemployment rate report postponed household formation and fertility, since young couples tend to seek confidence in their financial affluence before parenthood. Moreover, high unemployment and economic uncertainty imply challenging re-entering the labour market for women after childbirths. In Italy, Greece and Spain, where the unemployment among youth is high, women participation rate is lower than in other European countries. Consequently, fewer females take maternity leaves because of the difficulties in re-entering (Del Boca *et al.* 2003).

Rights to paid parental leave are frequently claimed to increase female labour force participation rate and minimize the need for women to change jobs (Ruhm 1998). Even though the EU legislation established the minimum period of maternity leave of 14 weeks (European Parliamentary Research Service 2022), the related policies across the EU differ. While the base period benefits are low for Greece, Spain, Portugal, the Netherlands and the UK, the Nordic countries employ more generous policies (Del Boca *et al.* 2003; European Parliamentary Research Service 2022). From one perspective, longer parental leaves may induce skill deterioration and postpone job promotions. Employers may prefer to rather hire male applicants to avoid the risk of having absent workers for 14 weeks. From another perspective it gives job security, enhances overall labour participation and in some cases even positively contributes to fertility (Del Boca *et al.* 2003).

The evidence from the growing body of panel research about female labour market participation rates suggests that declining fertility rates may arguably be the reason for higher women's labour force participation over the last centuries, although the correlation is reported to weaken compared to previous generations (Bernhardt 1993; Engelhardt *et al.* 2004) or even turn positive (Ahn & Mira 2002).

## 2.2 Evolution of patterns in family size

Over the last decades, the temporal pattern of female labour supply and fertility undergone significant changes, leading to greater variations across countries (Del Boca *et al.* 2003). Del Boca *et al.* (2003) characterizes the trend in fertility by the composition of two effects: tempo and quantum. The tempo effect stands for postponing first birth to later ages, while the quantum effect portrays total decline in fertility.

In the United States, the total fertility rate (TFR) dropped by 4% compared to 2019, reaching 1.6 children per woman in 2020 (Hamilton *et al.* 2024). In Germany, the TFR declined to 1.53 children per woman in 2020 (Bundesamt 2021). Since the 1970s, the researchers observe a notable decline in fertility rates among the OECD countries, accompanied by a general shift towards deferred parenting (Sleebos 2003). Robey *et al.* (1993) summarise the four main direct contributors to fertility: effective use of conception control (i.e., contraception, family planning), age at which woman first marry, length of postpartum infertility and pregnancy termination.

The so-called "postponement transition" has resulted in the mean age at first birth of around 26-30 years (Sobotka & Berghammer 2021). Arguably, the primary driver behind this trend is the expansion of academic and professional avenues for women, leading to a delay in marriage and parenthood. Hence, with the deferral in childbirth, family sizes have generally decreased, but the extent of the effect varies across countries. The employment rate among mothers has largely increased in Northern Europe, whereas in Southern European countries, the increase is comparatively lower, symbolizing the obstacles faced by women in these regions (Del Boca *et al.* 2003).

Kessler-Harris (2003) stated that whilst childbearing years were reduced nearly by half, women's life expectancy lengthened. Thus, women's extended post-childbearing years are contributing to a positive trend in their workforce participation.

## 2.3 Approaches to Estimation of Family Size Impact

The effect of family size on parent's labour market outcome can be represented using the following econometric model:

$$\text{Labour market outcome}_i = \beta_0 + \beta_1 \text{Family size}_i + \beta_2 X_i + \epsilon_i \tag{2.1}$$

where Labour market outcome$_i$ denotes labour market outcome of an $i$-th individual, Family size$_i$ denotes his/her family size, $X_i$ stands for the set of included control variables and $\epsilon_i$ is the error term.

However, to ensure reliability of this model, endogeneity is a significant challenge that must be addressed. The possible endogeneity captures one of the key challenges in research about family size and labour supply, as it inhibits scholars from estimating the true effect (Clarke 2018). The concept of endogeneity implies that the explanatory variable might be correlated with the error term (Wooldridge 2013), which suggests that unobserved factors affecting both family size and labour market outcome may exist. These factors, such as female autonomy, can lead to systematic variations between parents and non-parents and result in biased outcomes if not appropriately addressed (Cáceres-Delpiano 2008). For example, if we assume that autonomy correlates positively with the labour market outcome variables and negatively with family size, omitting this variable from the model biases the OLS estimates upward (Agüero & Marks 2011). Additionally, there may be unobserved characteristics that affect both family size and parents' labour force participation, such as individual preferences or societal conventions described in the previous sections.

Beyond the concerns of endogeneity, reverse causation is another critical aspect to consider. Reverse causation proposes that parents' labour force participation may not only be the outcome of family size but also its determinant, implying that the relationship is rather bidirectional. As per traditional economic theory, children are perceived as a normal good, and the demand for them within a family are determined by both income and substitution effects (Becker 1992). More independent and career-oriented women may choose to have fewer offspring and be disproportionately represented in the workforce relative to their share of the female population (Agüero & Marks 2008; 2011).

Consequently, the negative relationship between family size and parents' labour market outcomes, as established in a plethora of research employing simple OLS models, may be biased and exaggerate the true effect. It is necessary to recognise the complexity and dynamic nature of these variables, whihc require more sophisticated econometric strategies. One possible approach is to incorporate an exogenous source of variation in family size in the model. The implementation of Instrumental Variables (IVs) and Two-Stage Least Squares

(TSLS) has become prevalent in the literature, owing to their relative ease of use and intuitive interpretation.

What would have been the parents' labour market outcome if they had not given birth to their child/children? Due to ethical constraints, the experimental method of randomly assigning (or withholding) children is not feasible.

The first approach is the use of **natural experiments** that exploit external shocks and policy changes that may lead to variations in family size. For instance, the introduction of child subsidies, parental leave or family planning policies can affect a household's fertility decisions. Angrist & Evans (1996a) used the 1970 US state abortion reforms to investigate the effects of teen and out-of-wedlock childbearing on employment rates. The authors concluded that these reforms were associated with higher employment rates. Although the credibility and precision of results obtained from natural variation are challenging to establish, this approach highlights the potential of natural experiments in further research.

Another strategy involves implementation of non-experimental or quasi-experimental methods to examine the effect of family size on parents' labour supply. For instance, instruments can be employed to determine the sources of heterogeneity in fertility across households. However, this approach does not guarantee the elimination of all sources of endogeneity and requires precise selection and validation of the instruments used (i.e. addressing the weak instruments problem).

**Two-Stage Least Squares** (**TSLS** or **2SLS**) is a popular approach used in econometrics to further refine the analysis and address the endogeneity problem using instruments (Wooldridge 2013). The TSLS method breaks the model into two stages. In the first stage, a set of instruments is employed to predict the endogenous variable. These predicted values are then incorporated into the second stage regression as independent variables, enabling the estimation of the causal effect of the endogenous variable on the dependent variable.

$$\text{Family size}_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 X_i + v_i \tag{2.2}$$

$$\text{Labour market outcome}_i = \beta_0 + \beta_1 \text{Family size}_i + \beta_2 X_i + \epsilon_i \tag{2.3}$$

The first stage involves estimating Equation 2.2, where we regress family size on the set of chosen Instrumental Variables and controls. This yields the fitted values for family size. In the second stage, as per Equation 2.3, these fitted values are employed to explain the variation in the labour market outcome. Since the fitted values of family size are utilised instead of the actual family size, the results of TSLS can differ significantly from those obtained from OLS estimation.

For the Instrumental Variable to be considered valid, it must satisfy the conditions of relevance and exogeneity (Wooldridge 2013). The relevance condition stipulates that the IV must have a strong relationship with the variable it aims to predict, thereby enabling the identification of the causal effect (i.e., $\text{Cov}(Z_i, \text{Family size}_i) \neq 0$). The exogeneity condition requires that the IV must be uncorrelated with the error term, implying that it is not influenced by the same unobserved factors that affect the outcome variable (i.e., $\text{Cov}(Z_i, \epsilon_i) = 0$). Amongst the most commonly used IVs in our topic under study are sex composition of the first two children, miscarriage, twins at first birth, and infertility shocks.

In the context of analysing family size and labour market outcomes, another powerful and widely employed econometric technique is the **Fixed Effects** (**FE**) methodology. This approach complements TSLS by providing a comprehensive understanding of the dynamic relationship under investigation. The FE models are particularly helpful for controlling for unobserved individual-specific factors that might influence both family size and their labour market outcomes and do not vary over time (i.e., personal preferences, innate abilities). For instance, more career-oriented individuals tend to have higher earnings, reflecting a significant individual effect in the wage equation, and also are less engaged in housework. Consequently, this individual effect will be negatively correlated with housework, leading to a negatively biased OLS estimate. Taking the issue from the employer perspective, employees with greater housework responsibilities may be perceived as less motivated to work due to their time constraints and therefore may face discrimination by their employer (Bryan & Sevilla-Sanz 2011).

The FE model mitigates the bias from these factors by differencing out the individual-specific effect. Thus, Equation 2.1 takes the following form:

$$(\text{Labour market outcome}_{it} - \overline{\text{Labour market outcome}}_i) = \beta_1(\text{Family size}_{it} - \overline{\text{Family size}}_i) \quad + \beta_2(X_{it} - \overline{X}_i) + (\epsilon_{it} - \overline{\epsilon}_i) \tag{2.4}$$

where the bar represents the individual-specific mean over time. By subtracting the average of each variable over time for each respondent from the observed values, individual-specific effects are differenced out ensuring that the estimates are no longer biased by these effects.

In essence, the complexity of the relationships under scrutiny is driven by multiple factors, both observable and latent. Rigorous econometric models aim to reveal the true impact and offer deeper insights, thereby guiding future research and policy formulation.

# Chapter 3

# Data Overview

To conduct a thorough investigation of how family size influences parents' labour market outcomes, we employ the quantitative research design of meta-analysis. This approach involves compiling a dataset composed of various relevant variables extracted from primary studies. Our research draws inspiration from the literature review by Clarke (2018) and extends the focus of previous meta-analyses conducted by Matysiak & Vignoli (2008), Cukrowska-Torzewska & Matysiak (2020) and de Linde Leonard & Stanley (2020). This chapter aims to describe the dataset collection procedure, followed by the necessary data transformations. Subsequently, we will conclude with the presentation of descriptive evidence pertaining to the dataset.

## 3.1 Data Compilation

Following the guidelines proposed by Stanley & Doucouliagos (2012), we start by compiling a list of primary studies using Google Scholar's advanced search capabilities. It is a common practice to prefer Google Scholar over other databases since it includes all research papers ever appeared on the Internet and performs search through the full text of the papers, not just keywords and abstracts (Irsova *et al.* 2023). The search was done via applying specific key words as search queries. We decided to use various combinations including "motherhood wage penalty", "fatherhood wage penalty", "family wage gap" and "parenthood and labour force participation".

Our search was delimited to articles written in English to ensure an accurate interpretation of the included estimates, as translations might introduce distortions to the meaning of the results (Stanley & Doucouliagos 2012). Adhering to standard practice, we identify those papers containing empirical estimates and falling into the subject of inquiry. Moreover, we ensure the studies referenced in previous meta-analyses by Cukrowska-Torzewska & Matysiak (2020) and de Linde Leonard & Stanley (2020) are included in the dataset, altogether with those mentioned in the literature review by Clarke (2018).

It is worth highlighting that we encountered numerous challenges while replicating datasets from previous meta-analyses. One such challenge was the absence of a list of papers included in the dataset by Matysiak & Vignoli (2008). Following our email correspondence with the authors, they declined to provide access to the list of studies. Additionally, the authors of the other two meta-analyses Cukrowska-Torzewska & Matysiak (2020) and de Linde Leonard & Stanley (2020) were reluctant to provide access to their respective datasets to facilitate more precise replication. Thus, several studies included in the previous meta-analyses were not included in our dataset due to their inaccessibility in open sources.

Besides the mentioned challenges, we were able to successfully compile the dataset completely anew. Moreover, we have expanded the scope of our dataset beyond the restriction of estimates based exclusively on the logarithm of mothers' wages, as observed in the previous meta-analyses by Cukrowska-Torzewska & Matysiak (2020) and de Linde Leonard & Stanley (2020). Instead, we have collected all available estimates of parents' labour market outcomes that offer insights into the effect of interest. Consequently, the dependent variable in our analysis may manifest in various forms (i.e., dummy indicating the labour force participation status, hours worked, the logarithm of wages, etc.). Furthermore, to the best of our knowledge, this is the first meta-analysis to present results for the male sample.

The search started in May 2023 and terminated in December 2023. We identify the following criteria for the papers to be included in the dataset:

- Study investigates the impact of family size on parents' labour market outcome.
- Study uses quantitative methods.
- Study provides a clear coefficient of family size.
- Study either reports standard errors of the coefficients or presents other statistics allowing to calculate standard errors (i.e., p-values, t-statistics) and clearly states the number of observations.

Following the approach proposed by Irsova *et al.* (2023) we refrained from excluding any articles solely based on the perceived quality inferred from its publication prestige. Thus, we included working papers and studies from local journals alongside those published in peer-reviewed journals. Subsequently, we performed a “snowballing” to ensure that the most important studies were not unintentionally omitted from our analysis.

Based on the criteria mentioned above, we identified 107 papers as fitting. After a careful examination of each study, the final list of studies included 89 papers. The studies were mainly excluded due to the missing number of observations and the irrelevance of the dependent variable. The list of papers included in the final dataset is presented in Table 3.1. The lists of studies for the female and male samples separately are presented in Table A.1 and Table A.2

Table 3.1: Primary studies used in the meta-analysis

| | |
|---|---|
| 1. Green, C.A. & Ferber, M.A. (2008) | 46. Hotz, V.J., Mullin, C.H. & Sanders, S.G. (1997) |
| 2. Aguero, J.M. & Marks, M.S. (2008) | 47. Jacobsen, J.P., Pearce III, J.W. & Rosenbloom, J.L. (1999) |
| 3. Aguero, J.M. & Marks, M.S. (2011) | 48. Jia, N., & Dong, X.Y. (2013) |
| 4. Andersen, S.H. (2018) | 49. Albrecht, J.W., Edin, P.A., Sundström, M. & Vroman, S.B. (1999) |
| 5. Angrist, J., Lavy, V. & Schlosser, A. (2010) | 50. Bollen, K.A. & Brand, J.E. (2010) |
| 6. Angrist, J.D. & Evans, W.N. (1996) | 51. Kalist, D.E. (2008) |
| 7. Angrist, J.D. & Evans, W.N. (1996) | 52. Killewald, A. & Gough, M. (2013) |
| 8. Bari, L. (2023) | 53. Kim & Assve (2006) |
| 9. Bronars, S.G. & Grogger, J. (1994) | 54. Klesment, M. & Van Bavel, J. (2017) |
| 10. Bryan, M.L. & Sevilla-Sanz, A. (2011) | 55. Korenman, S. & Neumark, D. (1992) |
| 11. Buchmann, C. & McDaniel, A. (2016) | 56. Kühhirt, M. & Ludwig, V. (2012) |
| 12. Budig, M.J. & England, P. (2001) | 57. Livermore, T., Rodgers, J. & Siminski, P. (2011) |
| 13. Budig, M.J. & Hodges, M.J. (2010) | 58. Loughran, D.S. & Zissimopoulos, J.M. (2009) |
| 14. Buligescu, B., Crombrugghe, D.D., Menteşoğlu, G. & Montizaan, R. (2008) | 59. Lowen, A., & Sicilian, P. (2009) |
| 15. Amuedo-Dorantes, C. & Kimmel, J. (2005) | 60. Lundberg, S. & Rose, E. (2000) |
| 16. Caceres-Delpiano, J. (2006) | 61. Lundborg, P., Plug, E. & Rasmussen, A.W. (2014) |
| 17. Caceres-Delpiano, J. (2008) | 62. Marshall, M.I. & Flaig, A. (2014) |
| 18. Caceres-Delpiano, J. (2012) | 63. Meurs, D., Pailhé, A. & Ponthieux, S. (2010) |
| 19. Casal, M.D.P. & Barham, B.L. (2013) | 64. Miller, A.R. (2011) |
| 20. Chevalier, A. & Viitanen, T.K. (2003) | 65. Mincer, J. & Polachek, S. (1974) |
| 21. Choi, S. (2011) | 66. Molina, J.A. & Montuenga, V.M. (2009) |
| 22. Chun, H. & Oh, J (2002) | 67. Moore, W. J. & Wilson, R. M. (1982) |
| 23. Cools, S., Markussen, S. & Strøm, M. (2017) | 68. Nielsen, H. S., Simonsen, M. & Verner, M. (2004) |
| 24. Cristia, J.P (2008) | 69. Noonan, M. C. (2001) |
| 25. Cukrowska-Torzewska, E. (2015) | 70. Nsiah, C., DeBeaumont, R. & Ryerson, A (2013) |
| 26. Cukrowska-Torzewska, E. & Lovasz, A. (2016) | 71. Pacelli, L., Pasqua, S. & Villosio, C. (2013) |
| 27. Daniel, F.K., Lacuesta, A. & Rodríguez-Planas, N. (2013) | 72. Phipps, S., Burton, P. & Lethbridge, L. (2001) |
| 28. Davies, R. & Pierre, G. (2005) | 73. Piras, C. & Ripani, L. (2005) |
| 29. Anderson, D.J., Binder, M. & Krause, K. (2003) | 74. Ribar, D.C. (1999) |
| 30. Doren, C. (2019) | 75. Rosenzweig, M.R. & Schultz, T.P. (1985) |

31. Duvivier, C. & Narcy, M. (2015)
32. Felfe, C. (2012)
33. Fletcher, J.M. & Wolfe, B.L. (2009)
34. Gangl, M. & Ziefle, A. (2009)
35. Gash, V. (2009)
36. Geronimus, A.T. & Korenman, S. (1992)
37. Giffin, S. & White, Q. (2008)
38. Glauber, R. (2012)
39. Glauber, R. (2007)
40. Gupta, N. D. & Smith, N. (2002)
41. Hardoy, I. & Schøne, P. (2008)
42. Hersch, J. (1991)
43. Hill, M.S. (1979)
44. Hirvonen, L. (2009)
45. Hotz, V.J., McElroy, S.W. & Sanders, S.G. (2005)
76. Rosenzweig, M.R. & Wolpin, K.I (1980)
77. Rosholm, M., & Smith, N. (1996)
78. Avellar, S. & Smock, P.J. (2003)
79. Simonsen, M. & Skipper, L. (2012)
80. Staff, J. & Mortimer, J.T. (2012)
81. Taniguchi, H. (1999)
82. Waldfogel, J. (1998)
83. Waldfogel, J. (2008)
84. Waldfogel, J. (1995)
85. Waldfogel, J. (1998)
86. Weeden, K. A. (2005)
87. Wilner, L. (2016)
88. Yu, W.H., & Hara, Y. (2021)
89. Yu, W.H., & Kuo, J.C.L. (2017)

Given the prevalent practice among researchers in this field to conduct extensive robustness checks and hypothesis testing on various subsamples, we adopt a similar approach as articulated by Lang (2023a). Specifically, we exclude robustness checks or heterogeneity analyses unless explicitly emphasized by the authors as pivotal. Thus, if authors specify the primary regression outcomes, but highlight the significance of the estimates derived from a particular subsample, we denote it using a designated dummy variable *Main* in our dataset.

In total, we collected 1542 estimates from 89 papers (1323 with female-related effects, 209 with male-related effects, and 10 including both). The estimates including effects on both parents irrespective of their gender were subsequently excluded from the dataset due to their insufficiency, with only 10 estimates collected from 2 studies. Moreover, grounded in the theoretical framework presented in Chapter 2, it is essential to distinguish between females and males since the effect might differ in both size and direction. While we report the results for both samples, we center our analysis primarily on the female sample, as a significant portion of the discourse pertains specifically to women. Moreover, the existing literature on the effect of fatherhood on labour market outcomes is relatively sparse, not allowing us to conduct robustness checks on specific subsamples within the male sample.

Aside from the estimates of the effect of interest, we also collected information delineating the regressions employed, including estimation techniques and major sources of heterogeneity identified across the body of papers.

## 3.2 Data Processing

After concluding the data collection phase, we undertook essential data adjustments to ensure the inclusion of comparable effect estimates.

In our dataset, several studies did not present standard errors, but instead reported t-statistics (Mincer & Polachek 1974; Rosenzweig & Wolpin 1985; Albrecht *et al.* 1999; Lundberg & Rose 2000; Hotz *et al.* 2005; Kim 2006; Lowen & Sicilian 2009; Molina & Montuenga-Gomez 2009; Bryan & Sevilla-Sanz 2011). To adhere to the specified restrictions, we controlled for the number of observations in these cases and transformed t-statistics into standard errors. Additionally, in the dataset by Buchmann & McDaniel (2016), zero standard errors were reported. To address this issue, we replaced these standard errors with a low value of 0.001. Furthermore, in the study by Pacelli *et al.* (2013), the dependent variable was *exiting employment after t=0*. For consistency, we adjusted the results to reflect the dependent variable as *not exiting employment after t=0* by changing the sign of the coefficients; the standard errors remained unchanged. We performed the same transformation for the estimate of *no job* from the dataset by Cristia (2008).

Although we endeavored to standardise both dependent and independent variables within our dataset, direct comparison of their coefficients still remains infeasible. Previous meta-analyses by Cukrowska-Torzewska & Matysiak (2020) and de Linde Leonard & Stanley (2020) attempted to address this issue by exclusively collecting coefficients for the dependent variable *log(wages)* and independent variables such as *number of children* or *motherhood status*, following the Mincerian wage equation. In contrast, we opted not to confine the dataset to these specifications. Instead, we recalculated the effect sizes using the standardization method called partial correlation coefficients (PCCs), as advocated by Doucouliagos (2005), Zigraiova & Havranek (2016) and Cazachevici & Horvath (2020). This approach allows for the assessment of the relationship between two variables while keeping all other variables constant, thereby segregating the impact of other factors. The closer a partial correlation to the absolute value of 1, the larger the estimated effect (Doucouliagos & Stanley 2013). Following the approach employed in recent meta-analyses (Zigraiova & Havranek (2016); Havranek *et al.* (2016); Kroupova *et al.* (2024)), the PCCs are calculated as follows:

$$PCC_{is} = \frac{t_{is}}{\sqrt{(t^2)_{is} + df_{is}}}, \tag{3.1}$$

Where $t$ represents the t-statistic of the reported coefficient, $df$ indicates the number of degrees of freedom used in the analysis, $s$ indexes the individual study and $i$ specifies the estimate within that study. The corresponding standard errors are calculated as:

$$SE(PCC)_{is} = \sqrt{\frac{1-(PCC^2)_{is}}{df_{is}}}, \tag{3.2}$$

As a final stage of our data transformation process, we address outliers present in our dataset by applying the winsorization technique, a method recently adopted in meta-analyses (see, for instance Bajzik *et al.* (2020); Zigraiova *et al.* (2021)). Winsorization involves the replacement of outliers with less extreme values to mitigate the impact of a few highly influential data points unless the source study is substantial and trustworthy (Irsova *et al.* 2023). In our case, we opt for a 1% level winsorization approach. Please note that empirical tests are performed on winsorized PCCs, while descriptive evidence and funnel plots are presented on unwinsorized data.

## 3.3 Descriptive Analysis

As stated earlier, we collected 1542 estimates, from which we distinguished 1323 estimates of the relationship for mothers, 209 cases for fathers, and 10 estimates for parents irrespective of their gender from 89 primary studies. The oldest study we use was published by Mincer & Polachek (1974). The most recent study is Bari (2023). Based on the information provided by Google Scholar, the most cited study is Mincer & Polachek (1974) with 3882 citations as at the time of the dataset collection. The most cited study per year after publication after adjustments for citations is Budig & England (2001).

Before proceeding with the analysis of publication bias and heterogeneity, we provide a detailed overview of the constructed dataset by presenting summary statistics for the female and male samples separately.

### Sample of Female Respondents

Figure 3.1 presents the distribution of PCCs estimating the relationship between family size and mothers' labour market outcomes. The solid vertical line signifies the mean value, while the dashed vertical line represents the median. As depicted in Table 3.2, the minimum value of the PCC is -0.954 and the maximum is 0.241. The distribution demonstrates negative skewness with the mean (-0.033) being lower than the median (-0.016) and a higher frequency of values appearing to the left of the peak. We reveal an imbalance in our dataset in the number of reported estimates per study, with 10 studies providing only two estimates, while other studies such as those by Cáceres-Delpiano (2012) and Angrist & Evans (1996a), presenting 96 and 66 estimates, respectively. To mitigate the uneven distribution of estimates in our primary studies with the higher number of reported estimates, we calculated a weighted mean of PCC (-0.008), using the inverse of the number of estimates per study as weights. As per the guidelines provided by Doucouliagos (2011), weighted and unweighted means are regarded as small.

Table 3.2: Descriptive statistics for PCCs. Female sample

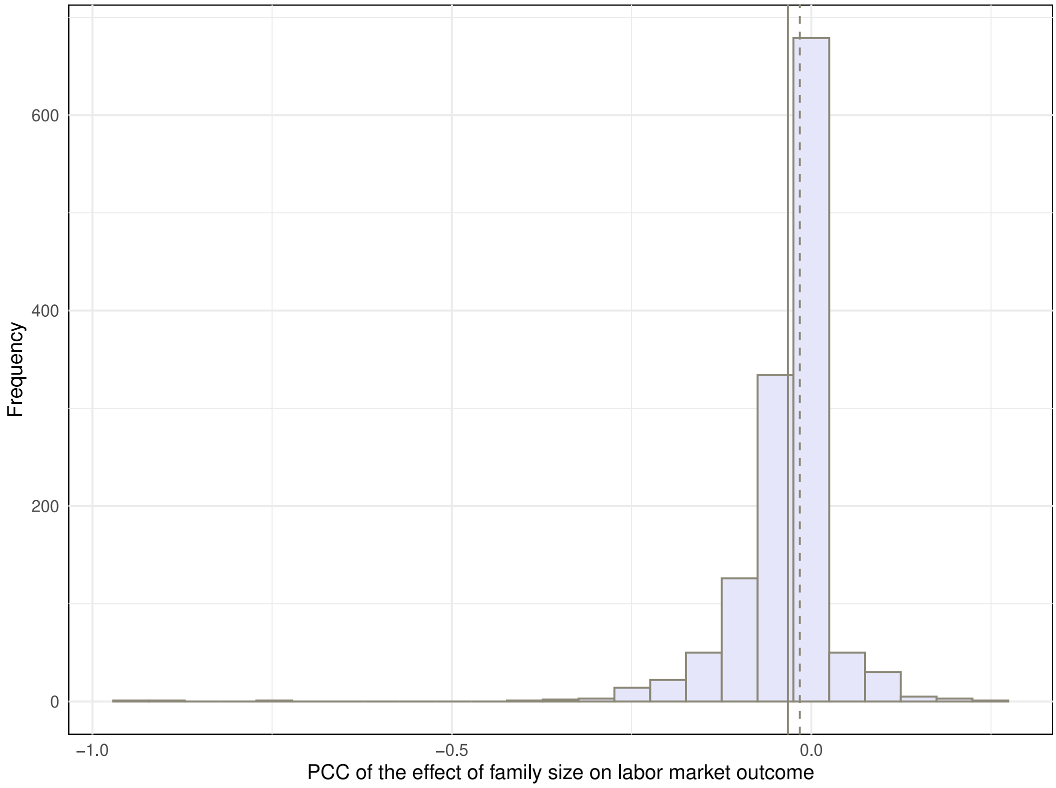
| | Unweighted mean | Weighted mean | Median | Standard deviation | Min | Max |
|---|---|---|---|---|---|---|
| Female sample | -0.033 | -0.008 | -0.016 | 0.073 | -0.954 | 0.241 |

The country-specific estimates of the relationship can be found in Figure 3.2. The boxes show the interquartile range, representing the data spread from the 25th to 75th percentiles. The solid line inside the box represents the median. The lower and upper whiskers display the bottom and top 25% of the data, respectively. The outliers are shown as individual points outside the whiskers. Thus, the boxplot in Figure 3.2 demonstrates that the estimates from the academic literature exhibit greater variability within geographical locations such as *North America*, *Europe* and *Oceania* and include a portion of outliers. In contrast, geographical regions like *Asia*, *South America* and *Other* show more consistent estimates with fewer outliers present.

The graphical representation of the variability within and across studies is presented in Figure A.1 in Appendix. To investigate the variability within our dataset, we begin by evaluating the average partial correlation coefficients across different groups, as shown in Table 3.4. As per the discussion above, we prioritize the more reliable weighted mean values in our subsequent analysis.

From the information obtained from Table 3.4 we can conclude that there is no substantial difference in the mean values across publication status category. Results presented in the main body of the papers, as well as those from panel datasets tend to contain less extreme mean PCC values than ancillary results and those from cross-sectional datasets, respectively. Not surprisingly, controlling for endogeneity leads to less pronounced mean value of PCC. The

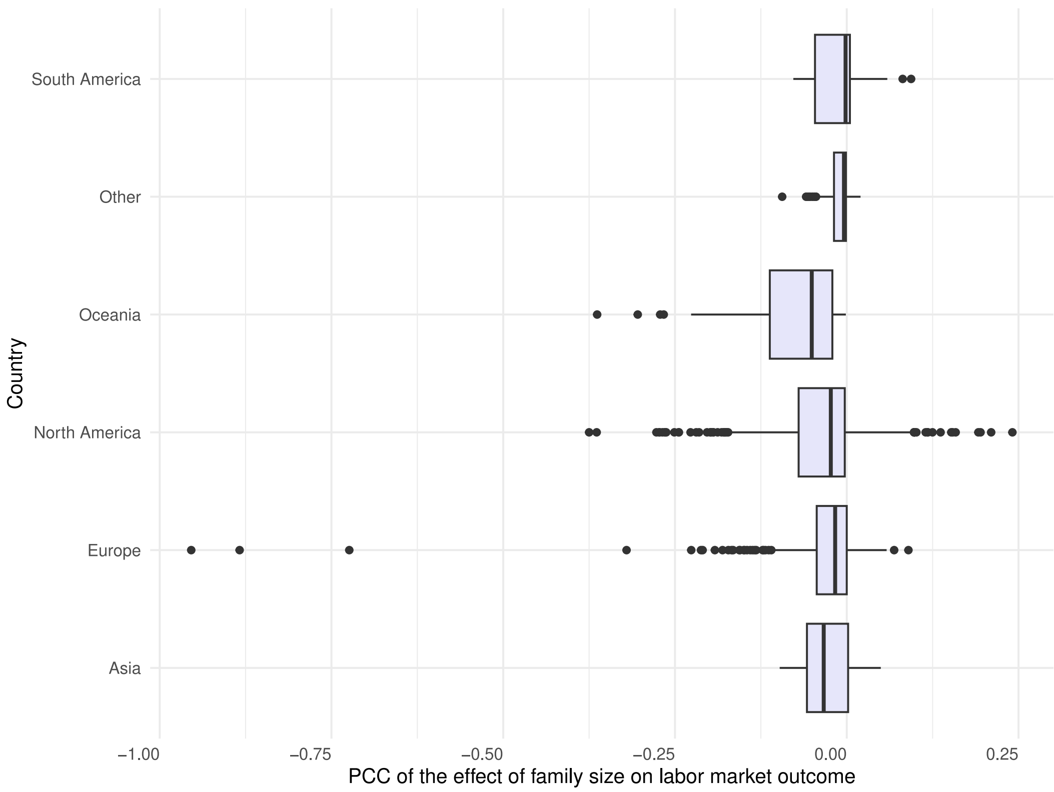
Figure 3.1: Distribution of PCCs for the female sample




Note: The graph illustrates the distribution of PCCs corresponding to the estimated effects of family size on labor market outcomes reported in individual studies for the female sample. The solid vertical line indicates the mean, whereas the dashed line marks the median. The figure includes unwinsorized PCCs.

variability in estimation methods employed results in variations in mean values ranging from -0.0119 to -0.0002. Studies from Europe and Oceania present substantially lower mean values of PCCs compared to those from North and South America, Asia and the Other category (includes studies analyzing data from the list of undefined regions, primarily developing countries). When comparing labour characteristics, earnings exhibit the highest mean value in absolute terms, whereas experience yield the least substantial mean PCC results within this category. Various family characteristics imply high variations in the mean values of PCCs. Years of motherhood used as a proxy for family size is the only category specified for the female sample that lead to a positive mean of PCCs, primarily because motherhood years were employed as a family variable in only one study in our dataset.

Figure 3.2: Forest plot of PCCs for the female sample across countries




Note: The figure illustrates the forest plot of PCCs across countries for the female sample. The PCCs correspond to the estimated effects of family size on labor market outcomes reported in individual studies for the female sample. The solid vertical line marks the median, boxes present the interquartile range encompassing the 25th to 75th percentiles. Data points that lie beyond the whiskers are considered outliers. The figure includes unwinsorized PCCs.

## Sample of Male Respondents

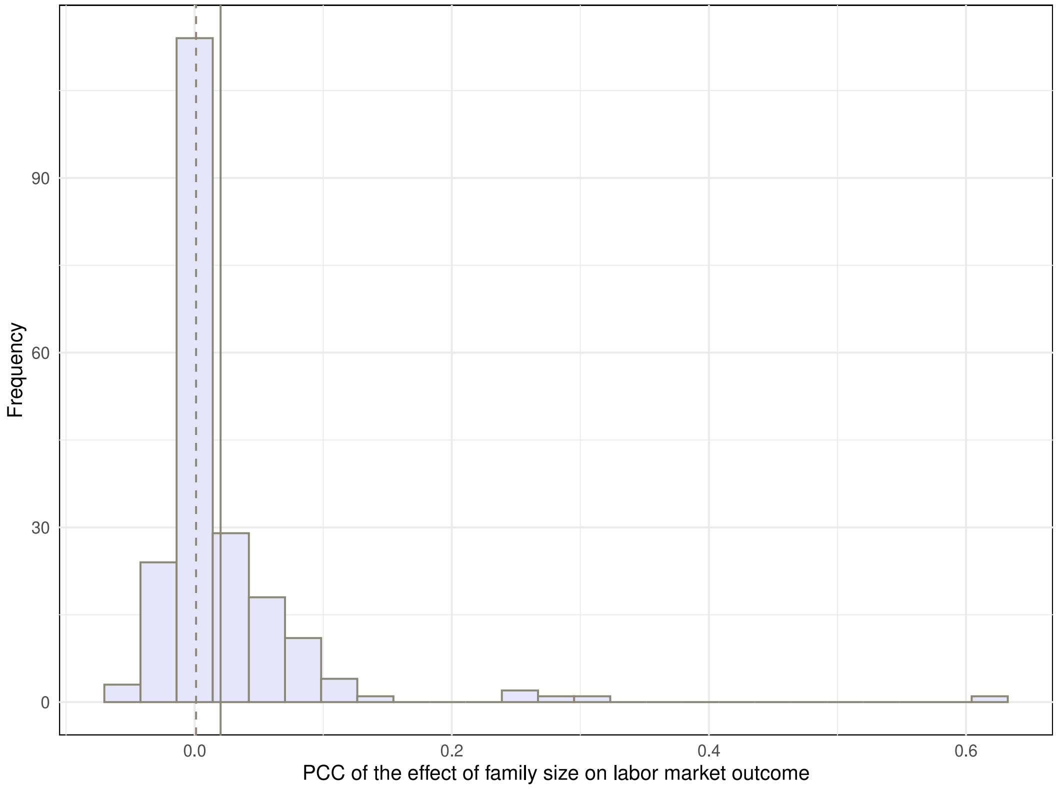
The histogram of PCCs for the male respondents Figure 3.3 demonstrates a positive skewness in this subsample, with the mean (0.020) exceeding the median (0.001). The dataset also exhibits an imbalance, with seven studies reporting two estimates, while Hirvonen (2010) reported 24 estimates. This issue is addressed using the same approach as applied for the female sample. Both weighted (0.008) and unweighted (0.020) mean values are classified as representing a small effect (Doucouliagos 2011). Figure A.2 provides a graphical illustration of the forest plot across countries, while Figure A.3 represents the variations within and between studies. To obtain an initial understanding of the heterogeneity within our dataset for the male subsample, please refer to Table 3.4.

From the descriptive statistics for the male sample, we can conclude that the variance in data dimension, publication status, and result type yield signif-

Table 3.3: Descriptive statistics for PCCs. Male sample

| | Unweighted mean | Weighted mean | Median | Standard deviation | Min | Max |
|---|---|---|---|---|---|---|
| Male sample | 0.020 | 0.008 | 0.001 | 0.001 | -0.048 | 0.627 |

Figure 3.3: Distribution of PCCs for the male sample




Note: The graph illustrates the distribution of PCCs corresponding to the estimated effects of family size on labor market outcomes reported in individual studies for the male sample. The solid vertical line indicates the mean, whereas the dashed line marks the median. The figure includes unwinsorized PCCs.

icant differences in the mean values of PCCs. Winthin the estimation methods, the most pronounced absolute values of mean PCC are presented for advanced regression and panel data techniques, whereas IV and 2SLS are associated with a mean value close to zero. There are no results for South America, Oceania, and the "Other" location categories, but the variation within and between the remaining categories appear to be substantial. Additionally, studies examining labour force participation present negative mean estimates, while those for earnings, hours worked and experience yield positive results. Finally, primary studies employing the number and presence of children as the variable indicating family size yield more notable positive results than studies employing other female characteristics.

Table 3.4: Descriptive statistics for weighted PCCs

| **Variable** | **Female sample** | | | **Male sample** | | |
|---|---|---|---|---|---|---|
| | **Mean** | **Lower CI** | **Upper CI** | **Mean** | **Lower CI** | **Upper CI** |
| *Data type* | | | | | | |
| Cross-sectional | -0.0084 | -0.0105 | -0.0062 | 0.005 | 0.0023 | 0.0076 |
| Longitudinal | -0.007 | -0.0083 | -0.0058 | 0.0135 | 0.0059 | 0.0021 |
| *Publication status* | | | | | | |
| Published | -0.0077 | -0.0089 | -0.0064 | 0.0116 | 0.0071 | 0.0161 |
| Unpublished | -0.0073 | -0.0095 | -0.005 | -0.0013 | -0.002 | -0.0006 |
| *Result type* | | | | | | |
| Main | -0.0075 | -0.0088 | -0.0062 | 0.0089 | 0.0051 | 0.0128 |
| Not main | -0.0085 | -0.0108 | -0.0062 | 0.0042 | -0.0007 | 0.009 |
| *Endogeneity control* | | | | | | |
| Controlled | -0.0045 | -0.0057 | -0.0033 | 0.0099 | 0.0038 | 0.016 |
| Not controlled | -0.0118 | -0.014 | -0.0096 | 0.0069 | 0.0036 | 0.0101 |
| *Estimation method* | | | | | | |
| OLS | -0.0119 | -0.0141 | -0.0097 | 0.007 | 0.0031 | 0.0108 |
| IV and 2SLS | -0.0002 | -0.001 | 0.0006 | 0.0006 | -0.0002 | 0.0015 |
| Panel Data Techniques | -0.009 | -0.0115 | -0.0064 | 0.0162 | 0.0063 | 0.0261 |
| Binary Outcome Models | -0.0083 | -0.0151 | -0.0014 | 0.0092 | -0.0005 | 0.0189 |
| Advanced Regression Techniques | -0.0042 | -0.0082 | -0.0001 | 0.0121 | -0.1277 | 0.1519 |
| Other | -0.0095 | -0.0239 | 0.0049 | 0.0029 | -0.0098 | -0.0156 |
| *Location* | | | | | | |
| Asia | -0.0066 | -0.0092 | -0.0041 | 0.0121 | -0.1277 | 0.1519 |
| Europe | -0.0105 | -0.014 | -0.0069 | 0.005 | 0.0031 | 0.0069 |
| North America | -0.0084 | -0.0101 | -0.0068 | 0.0123 | 0.0051 | 0.0195 |
| South America | -0.0036 | -0.0062 | -0.001 | - | - | - |
| Oceania | -0.0133 | -0.017 | -0.0096 | - | - | - |
| Other | -0.0013 | -0.0016 | -0.001 | - | - | - |
| *Labour characteristics* | | | | | | |
| Earnings | -0.0092 | -0.0108 | -0.0075 | 0.0112 | 0.0066 | 0.0158 |
| Experience | -0.0016 | -0.018 | 0.0148 | 0.0023 | -0.0017 | 0.0062 |
| Hours worked | -0.0026 | -0.0068 | 0.0016 | 0.0017 | -0.001 | 0.0043 |
| Participation | -0.0059 | -0.0073 | -0.0044 | -0.0007 | -0.0012 | -0.0002 |
| *Family characteristics* | | | | | | |
| One child | -0.0091 | -0.0139 | -0.0044 | 0.0042 | -0.0001 | 0.0085 |
| 2+ children | -0.0105 | -0.0133 | -0.0078 | 0.0028 | 0.0006 | 0.005 |
| Number of children | -0.0055 | -0.0067 | -0.0042 | 0.0112 | 0.003 | 0.0193 |
| Presence of children | -0.0071 | -0.0093 | -0.0049 | 0.0138 | 0.0058 | 0.0218 |
| Motherhood years | 0.0016 | -0.0008 | 0.0041 | -0.0007 | -0.0012 | -0.0001 |

Note: The table presents the descriptive statistics for weighted PCCs for the female and male samples across various data categories.

# Chapter 4

# Publication Bias

In the previous chapters, we have thoroughly examined the theoretical framework and methodological approaches relevant to studying the impact of family size on parents' labour market outcomes. We have also summarised the collected estimates to identify the average effect and eminent empirical trends in this area. However, the credibility of the presented results remains a concern.

This chapter aims to tackle the concerns related to the reliability of existing findings by investigating potential publication bias in the literature. Publication bias is a key issue of primary studies that is almost impossible to counter for individual primary studies but can be addressed in meta-analyses (Irsova *et al.* 2023). There are multiple definitions of publication bias (or a "file drawer problem"). In general, these two issues are observationally equivalent, though represent distinct terms (Havranek *et al.* 2022). They refer to the phenomenon describing the tendency of the statistically significant, favorable and/or widely expected results to be more likely published over those non-significant or unfavorable (Rosenthal 1979).

This issue may sometimes arise from the conscious preferences of journal editors and reviewers for significant findings, as well as authors' tendency to produce and submit studies that are more likely to be accepted. Thus, the decision-making process might distort the overall body of empirical evidence. However, the underrepresentation of insignificant findings is not always attributable to conscious intentions. It can also stem from unintentional biases in research design and reporting practices.

In either case, as mentioned by Irsova *et al.* (2023), meta-analysts should not overlook publication bias, as doing so compromises the validity and informativeness of their conclusions. As observed in more than half of the meta-analyses published in 2022, ignoring these issues fails to provide a thorough and accurate picture of the empirical evidence.

The authors of previous meta-analyses employed several statistical approaches. Cukrowska-Torzewska & Matysiak (2020) utilised Funnel Asymmetry Tests (FAT) and Precision Estimate Tests (PET) and concluded that they obtained no evidence for the presence of the publication bias in their analysis. Similarly, de Linde Leonard & Stanley (2020) adopted FAT-PET, as well as the Precision-Effect Estimate with Standard Error (PEESE) method, revealing robust evidence of selective reporting, but confirming a negative effect despite examined biases.

In our paper we start with the visual analysis by conducting a funnel plot. Then we employ more formal linear and non-linear tests to inspect the dataset. At the end of this section, we present several robustness checks to support our findings. The results are split by subsamples since the publication bias is expected to affect the results in distinct directions for females and males.

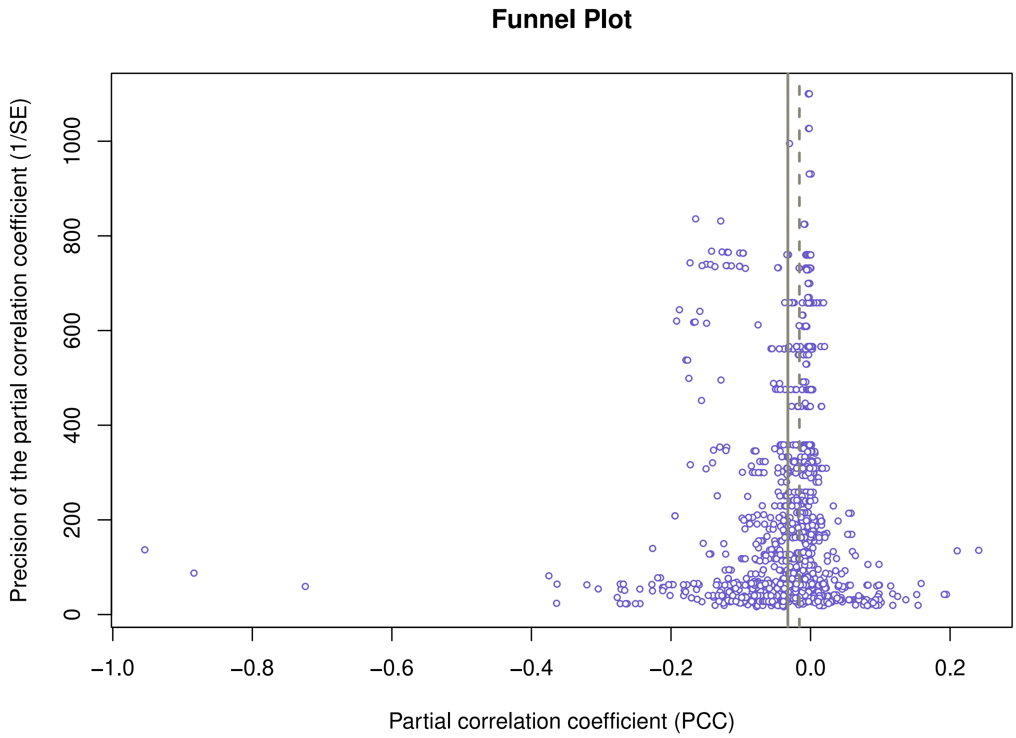
## 4.1 Visual Test: Funnel Plot

The funnel plot serves as an initial visual instrument for the assessment of publication bias in meta-analysis (Sterne *et al.* 2005). This scatter plot straightforwardly represents the treatment effect from individual studies along the horizontal axis, while the vertical axis depicts precision, quantified as the inverse of the standard error. When bias is absent, the plot resembles a symmetrical inverted funnel. Conversely, the plot appears asymmetrical. The level of asymmetry reflects the magnitude of the bias, with greater asymmetry indicating a higher degree of bias. In case a funnel plot appears hollow and extensively wide, Stanley (2005) suggests that publication selection favours statistically significant results.

### Sample of Female Respondents

The funnel plot presented in Figure 4.1 illustrates the relationship between the values of PCCs (on the horizontal axis) and their precision (on the vertical axis) across distinct studies for the female sample. The funnel plot appears to be roughly symmetric, with the majority of the data points concentrated around the center and some outliers to the left that are addressed through winsorization. Although the mean is located to the left of the median, the funnel plot does not provide clear evidence of skewness.

Figure 4.1: Funnel plot of partial correlation coefficients for the female sample.




Note: The graph illustrates the funnel plot of PCCs corresponding to the estimated effects of family size on labor market outcomes reported in individual studies for the female sample. The solid vertical line indicates the mean, whereas the dashed line marks the median. The figure includes unwinsorized PCCs.

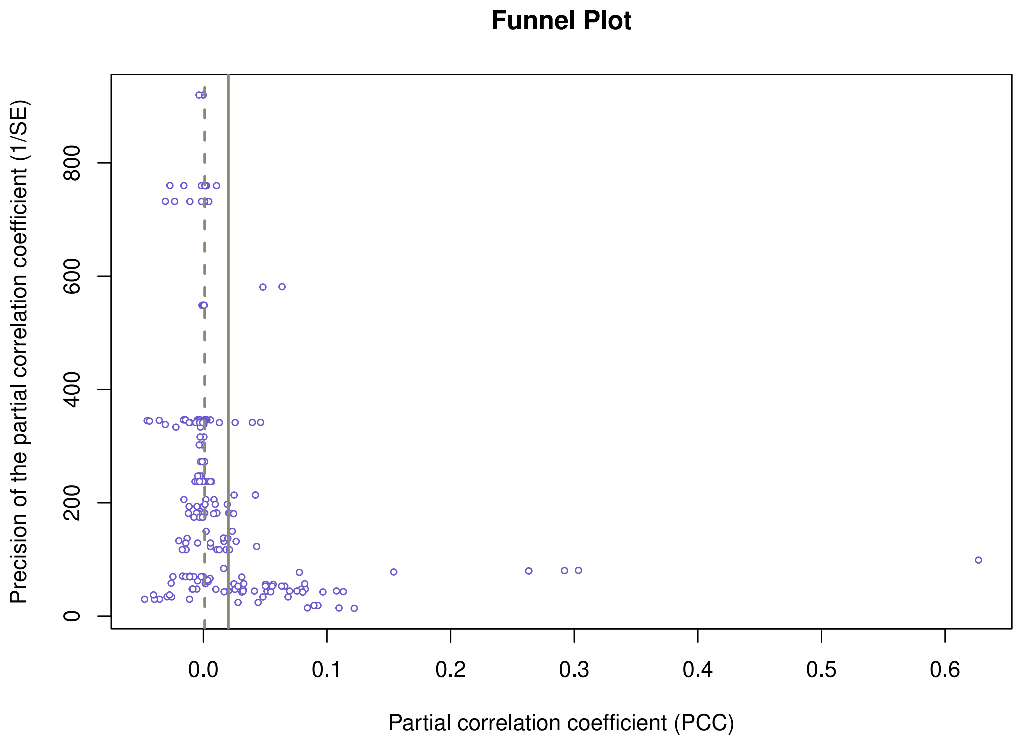
Additionally, upon closer examination of Figure 4.1, there is a distinct cluster of observations to the left of the mean. Nearly half of the data points stem from the studies Angrist & Evans (1996b), Angrist & Evans (1996a), Cáceres-Delpiano (2006) and Angrist & Schlosser (2010). All data points constructing this hump resulted from the application of the OLS method, potentially introducing bias into the reported results.

Nevertheless, the outcome of this visual test might not necessarily indicate the absence of publication bias due to its graphical nature. Therefore, the results represent a highly subjective assessment, underscoring the need for further, more sophisticated analysis.

### Sample of male respondents

Figure 4.2 displays the relationship between the partial correlation coefficients (PCCs) and their precision for the male sample, analogous to the analysis presented for the female sample. As anticipated, the results suggest the presence of positive publication selectivity. The paucity of data points to the left of the primary cluster suggests that studies reporting non-significant negative results might be underreported.

Figure 4.2: Funnel plot of partial correlation coefficients for the male sample.




Note: The graph illustrates the funnel plot of PCCs corresponding to the estimated effects of family size on labor market outcomes reported in individual studies for the male sample. The solid vertical line indicates the mean, whereas the dashed line marks the median. The figure includes unwinsorized PCCs.

However, the observed asymmetry could stem from variations in estimation methods and other heterogeneity-related issues (Stanley 2005). This issue is explored in greater detail in Chapter 5. The graphical nature of this test does not allow for an extensive examination of the potential sources of skewness beyond publication selectivity and relies heavily on the researcher's visual evaluation.

Further investigation reveals a distinct horizontal line of points with nearly identical precision (around 750), but different PCCs. This pattern indicates that these data points have similar standard errors but varying partial correlation coefficients. After examining the dataset, we found that these results originate exclusively from two studies already mentioned in the description of the clustered points for the female sample (Angrist & Evans 1996b;a). This observation suggests that the consistency across studies might contribute to the uniformity in precision. Despite the consistency, the varying PCCs reflect different methodologies applied within these studies, with OLS primarily associated with smaller PCCs.

Additionally, there is a notable horizontal clustering of data points within the precision range between 330 and 400. These points stem also from only two studies by Hirvonen (2010) and Wilner (2016). We do not observe any other patterns that could indicate a particular subset of the male sample is responsible for these results. The absence of additional clustering within the sample suggests that the observed consistencies are likely attributable to the methodological approaches rather than demographic or other sample-specific characteristics.

## 4.2 Linear Tests

While investigating publication bias, we utilize a linear regression technique inspired by Egger *et al.* (1997). These linear tests for publication bias operate under the assumption that the likelihood of publication is a linear function of the standard error.

Thus, we estimate the following model:

$$PCC_{is} = \beta_0 + \beta_1 SE(PCC_{is}) + \epsilon_{is}, \tag{4.1}$$

Where $PCC_{is}$ denotes the partial correlation coefficients, $SE(PCC_{is})$ represents their standard errors, and $\epsilon_{is}$ stands for the error term. The constant term $\beta_0$ represents the bias-corrected true effect size, while the slope $\beta_1$ indicates the magnitude and direction of bias, which is evaluated statistically to assess its significance.

The regression above is commonly referred to as the funnel asymmetry test (FAT), as it quantitatively estimates the symmetry of the funnel plot (Egger *et al.* 1997). In the absence of publication bias, the correlation between the estimates and their standard errors would be zero, whereas a non-zero correlation suggests its presence (Stanley 2005).

We estimate the model using several specifications. We begin with ordinary least squares (OLS). Next, we account for between-study variance. Although our dataset is imbalanced, each study includes at least two estimates for both female and male samples, allowing us to explore the variability within individual studies, but to a limited extend. Additionally, we implement a weighting scheme that assigns weights to estimates calculated as the inverse of the number of observations per study, ensuring the influence of studies with different sample sizes is balanced (Havranek *et al.* 2018; Gechert *et al.* 2019).

Acknowledging the possibility of heteroscedasticity in our model, we give greater weight to more precise estimates, in line with standard meta-analytical procedures (Stanley 2005; Zigraiova & Havranek 2016), resulting in Equation 4.2:

$$PCC_{is} \cdot \frac{1}{SE(PCC_{is})} = \beta_0 \cdot \frac{1}{SE(PCC_{is})} + \beta_1 + \nu_{is}, \tag{4.2}$$

where $\nu_{is} = \frac{\epsilon_{is}}{SE(PCC_{is})}, \quad \nu_{is} \sim N(0, \sigma^2)$.

In Equation 4.2, $\beta_0$ stands for the extent and direction of publication bias and $\beta_1$ denotes the mean effect after the correction for publication bias. As per Stanley (2005), this equation is called a precision asymmetry test (PAT).

Furthermore, to account for the heteroscedasticity in the FAT-PAT model, we cluster standard errors at the study level. This approach accounts for potential breaches of the independent and identically distributed (iid) standard errors assumption by incorporating possible within-study correlations. Moreover, it maintains the assumption of independence across distinct studies. Additionally, we employ the wild bootstrap method to provide robust confidence intervals, thereby addressing potential imbalances in cluster sizes (Gechert *et al.* 2019).

As was introduced in Chapter 2, family-labour relationships commonly suffer from endogeneity. Thus, we relax the exogeneity assumption and present the findings from the instrumental variable model as per the approach proposed by Stanley (2005). We employ the inverse of the number of observations, the inverse of the number of observations squared, the inverse of the square root of the number of observations, and the inverse of the logarithm of observations as instruments for the standard errors. Relevance tests and the exogeneity assumption confirm the validity of all instruments used.

### Sample of Female Respondents

The results for the sample of female respondents are presented in Table 4.1. According to Doucouliagos & Stanley (2013), the publication selectivity ranges from little to modest across all model specifications described above. We observe a negative publication bias that is significant at the 5% and higher significance levels for three out of five specifications.When adjusting for the number of observations per study, we found a positive yet insignificant publication bias, indicating that certain studies might be driving this bias. Conversely, the mean beyond bias appears to be negative and highly significant across all model specifications, with values ranging from -0.030 to -0.023, while the weighted mean for female sample, as described in the previous section, amounted to -0.008. These results suggest that the true effect is constantly negative and higher in absolute values after correcting for the publication bias.

Table 4.1: Linear Tests' results: female sample

| | **OLS** | **Between** | **Within** | **Weighted by study** | **Weighted by precision** |
|---|---|---|---|---|---|
| Standard error | -0.514** | -0.624*** | 0.055 | 0.012 | -0.676*** |
| *Publication bias* | (0.166) | (0.139) | (0.039) | (0.271) | (0.191) |
| | [-1.459, 0.281] | | | [-0.898, 0.917] | [-1.870, 0.450] |
| Constant | -0.025*** | -0.024*** | -0.026*** | -0.030*** | -0.023*** |
| *Mean beyond bias* | (0.002) | (0.002) | (0.000) | (0.004) | (0.002) |
| | [-0.038, -0.014] | | | [-0.007, 0.032] | [-0.015, 0.012] |
| Observations | 1323 | 1323 | 1323 | 1323 | 1323 |
| Studies | 88 | 88 | 88 | 88 | 88 |

Note: The table presents the results the five linear tests for publication bias detection for the female sample. Standard errors are presented in round brackets, and confidence intervals are reported in square brackets. The asterisks indicate significance levels: * - at 10%, ** - at 5%, *** - at 1%.

Based on the high F-statistics and low p-values in the Weak Instruments Test, we can conclude that all four instruments are robust. When the exogeneity assumption is relaxed by introducing the instrumental variables, the publication bias is observed to be statistically significant and remains negative across three specifications. Additionally, the mean beyond bias is negative and highly significant for all instrumental variables employed.

Table 4.2: IV's results: female sample

| | $\frac{1}{\text{obs}}$ | $\frac{1}{\text{obs}^2}$ | $\frac{1}{\sqrt{\text{obs}}}$ | $\frac{1}{\log(\text{obs})}$ |
|---|---|---|---|---|
| Standard error | -0.410* | -0.195 | -0.059*** | -0.683*** |
| *Publication bias* | (0.201) | (0.259) | (0.171) | (0.163) |
| Constant | -0.026*** | -0.029*** | -0.024*** | -0.023*** |
| *Mean beyond bias* | (0.002) | (0.003) | (0.002) | (0.002) |
| Observations | 1323 | 1323 | 1323 | 1323 |
| Studies | 88 | 88 | 88 | 88 |

Note: The table presents the results of the IV tests for publication bias detection for the female sample. Standard errors are presented in round brackets. The asterisks indicate significance levels: * - at 10%, ** - at 5%, *** - at 1%.

### Sample of Male Respondents

The results for the male sample are detailed in Table 4.3. Four model specifications indicate a positive and significant publication bias, which Doucouliagos & Stanley (2013) classify as substantial to severe. The mean beyond bias is negative and significant at the 10% level or higher for the within-effect specification and the model weighted by precision indicating that after adjusting for publication bias, the true effect of family size on fathers' labour market outcomes is likely to be detrimental. This mean is lower than the weighted mean of 0.008 presented in Chapter 3, suggesting that after correcting for publication bias, the true effect might indeed shift its direction.

Table 4.3: Linear Tests' results: male sample

| | **OLS** | **Between** | **Within** | **Weighted by study** | **Weighted by precision** |
|---|---|---|---|---|---|
| Standard error | 1.435*** | 1.565*** | 2.304*** | 1.192 | 2.054*** |
| *Publication bias* | (0.194) | (0.347) | (0.102) | (0.218) | (0.363) |
| | [0.356, 2.383] | | | [0.025, 2.015] | [0.771, 4.137] |
| Constant | 0.003 | 0.002 | -0.005*** | 0.009 | -0.004* |
| *Mean beyond bias* | (0.003) | (0.005) | (0.000) | (0.005) | (0.002) |
| | [-0.009, 0.024] | | | [-0.007, 0.032] | [-0.015, 0.012] |
| Observations | 209 | 209 | 209 | 209 | 209 |
| Studies | 27 | 27 | 27 | 27 | 27 |

Note: The table presents the results of the five linear tests for publication bias detection for the male sample. Standard errors are presented in round brackets, and confidence intervals are reported in square brackets. The asterisks indicate significance levels: * - at 10%, ** - at 5%, *** - at 1%.

The publication selectivity results from instrumental variables tests for the male sample are positive and highly statistically significant. However, we cannot reject the hypothesis that the underlying effect on male sample is non-existent based on the insignificance of all mean beyond bias results with instrumental variables.

Table 4.4: IV's results: male sample

| | $\frac{1}{\text{obs}}$ | $\frac{1}{\text{obs}^2}$ | $\frac{1}{\sqrt{\text{obs}}}$ | $\frac{1}{\log(\text{obs})}$ |
|---|---|---|---|---|
| Standard error | 1.332*** | 1.422*** | 1.487*** | 1.669*** |
| *Publication bias* | (0.186) | (0.200) | (0.201) | (0.243) |
| Constant | 0.004 | 0.003 | 0.002 | 0.000 |
| *Mean beyond bias* | (0.004) | (0.004) | (0.003) | (0.002) |
| Observations | 209 | 209 | 209 | 209 |
| Studies | 27 | 27 | 27 | 27 |

Note: The table presents the results of the IV tests for publication bias detection for the male sample. Standard errors are presented in round brackets. The asterisks indicate significance levels: * - at 10%, ** - at 5%, *** - at 1%.

## 4.3 Non-Linear Tests

Moving beyond the strong linear assumption and leveraging the independence between estimates and standard errors, we employ various non-linear tests to detect publication bias. Stanley *et al.* (2010) contend that linear tests tend to amplify publication bias, particularly for highly precise estimates, which results in an underestimated true effect size. Therefore, we employ various non-linear techniques to obtain a more reliable estimation of the true effect.

We begin with the Top 10 method, which focuses on the top 10% of the most precise estimates, assuming these estimates are less likely to be affected by publication bias. The Top 10 method presents an average effect size derived from the most precise subset of studies, which is further considered the true effect (Stanley *et al.* 2010).

We then utilize a stem-based method proposed by Furukawa (2019), which enhances the idea of selecting an arbitrary number of most precise primary studies by optimising the bias-variance tradeoff, rather than applying a fixed cutoff as used in the Top 10 model. Analogous to the stem-based approach, the Endogenous Kink model introduced by Bom & Rachinger (2019), is a novel approach that also incorporates the principle of the selection of primary studies based on the estimates' precision. The Endogenous kink model identifies the most precise estimates through a linear meta-regression of estimates on their standard errors, with a kink at the cutoff value of the standard error, below which publication selectivity is unlikely. The kink, or intersection of two linear segments (one flat for precise estimates with no publication selectivity based on the significance of results and one sloped for biased ones), acts as a threshold that isolates the most accurate estimates.

We proceed with the Weighted Average of Adequately Powered (WAAP) method proposed by Ioannidis *et al.* (2017). The WAAP method includes only those estimates that are "adequately" powered (with statistical power exceeding 80%), weighting them by the inverse of their standard errors squared ($1/SE^2$). This approach differs from the conventional meta-analysis that uses all estimates. Instead, the WAAP aims to reduce the reporting and small sample biases. The main weakness of this approach is that it is ineffective when applied to samples without sufficiently powered studies (Stanley *et al.* 2017). To mitigate this constraint, we utilize unwinsorised data for the WAAP methodology.

Finally, we employ a Selection Model following the approach by Andrews

& Kasy (2019). The authors propose two approaches to accounting for selective publication. The first pertains to replication studies, while the second is tailored to meta-analysis. The meta-analysis approach corrects for publication bias by estimating the probability of publication as a function of statistical significance through the maximum likelihood equation. The authors propose that selection bias may arise due to the tendency of researchers to select results for publication based on their statistical significance. Consequently, the Selection Model recalibrates the weights of the estimates to account for selective reporting, thereby assigning greater weights to underrepresented estimates.

### Sample of Female Respondents

The results for non-linear tests Table 4.5 for the sample of female respondents are consistent with those based on linear specifications. We observe that the effect beyond bias is negative, with significant values of -0.033 and -0.026. On average, the values of the significant true effect outcomes are slightly more negative compared to those derived from linear tests. This observation suggests that linear tests might inherently overestimate publication bias and underestimate the actual effect.

Table 4.5: Non-linear Tests' results: female sample

| | **Top 10** | **Stem** | **Endogenous kink** | **WAAP** | **Selection model** |
|---|---|---|---|---|---|
| Effect Beyond Bias | -0.033*** | -0.015 | -0.026*** | -0.024 | 0.011 |
| | (0.005) | (0.013) | (0.002) | (0.002) | (0.007) |
| Observations | 1323 | 1323 | 1323 | 1323 | 1323 |
| Studies | 88 | 88 | 88 | 88 | 88 |

Note: The table presents the results of the non-linear tests for publication bias detection for the female sample. Standard errors are presented in round brackets. The asterisks indicate significance levels: * - at 10%, ** - at 5%, *** - at 1%.

### Sample of Male Respondents

The results for the male sample in Table 4.6 show that the effect beyond bias is highly statistically significant for Endogenous kink and Selection model. The direction of effect is however different. The variation may arise from different methodologies: the positive effect observed in the selection model adjusts for publication bias by estimating the probability of study publication, whereas the Endogenous Kink model addresses endogeneity and structural breaks.

Table 4.6: Non-linear Tests' results: male sample

| | **Top 10** | **Stem** | **Endogenous kink** | **WAAP** | **Selection model** |
|---|---|---|---|---|---|
| Effect Beyond Bias | -0.004 | 0.001 | -0.005** | 0.001 | 0.046*** |
| | (0.002) | (0.004) | (0.002) | (0.003) | (0.010) |
| Observations | 209 | 209 | 209 | 209 | 209 |
| Studies | 27 | 27 | 27 | 27 | 27 |

Note: The table presents the results of the non-linear tests for publication bias detection for the male sample. Standard errors are presented in round brackets. The asterisks indicate significance levels: * - at 10%, ** - at 5%, *** - at 1%.

As per the results from the linear and non-linear tests and IV regression, there is evidence of negative publication bias in the female sample and positive publication bias in the male sample. These findings align with our initial expectations of negative publication bias for mothers and positive for fathers. However, we further enhance our analysis with the results from the Caliper test presented overleaf and robustness checks on two female subsamples.

## 4.4 Caliper Test

To expand our analysis of publication bias while relaxing the linearity assumption between effect estimates and their respective standard errors inherent in FAT-PET tests, we employ a method proposed by Gerber & Malhotra (2008) and used in recent meta-analyses (e.g., Kroupová (2021); Bajzik *et al.* (2023)). This test evaluates the distribution of effect estimates within designated ranges around specific t-statistic thresholds. In our analysis, we set the thresholds at -1.96 and 1.96.

The caliper test operates on the assumption that, in the absence of publication bias, estimates should be evenly distributed around these established thresholds. Deviations from this symmetry indicate the existence of publication bias, implying selective reporting, with estimates surpassing the specified thresholds being more likely to be published.

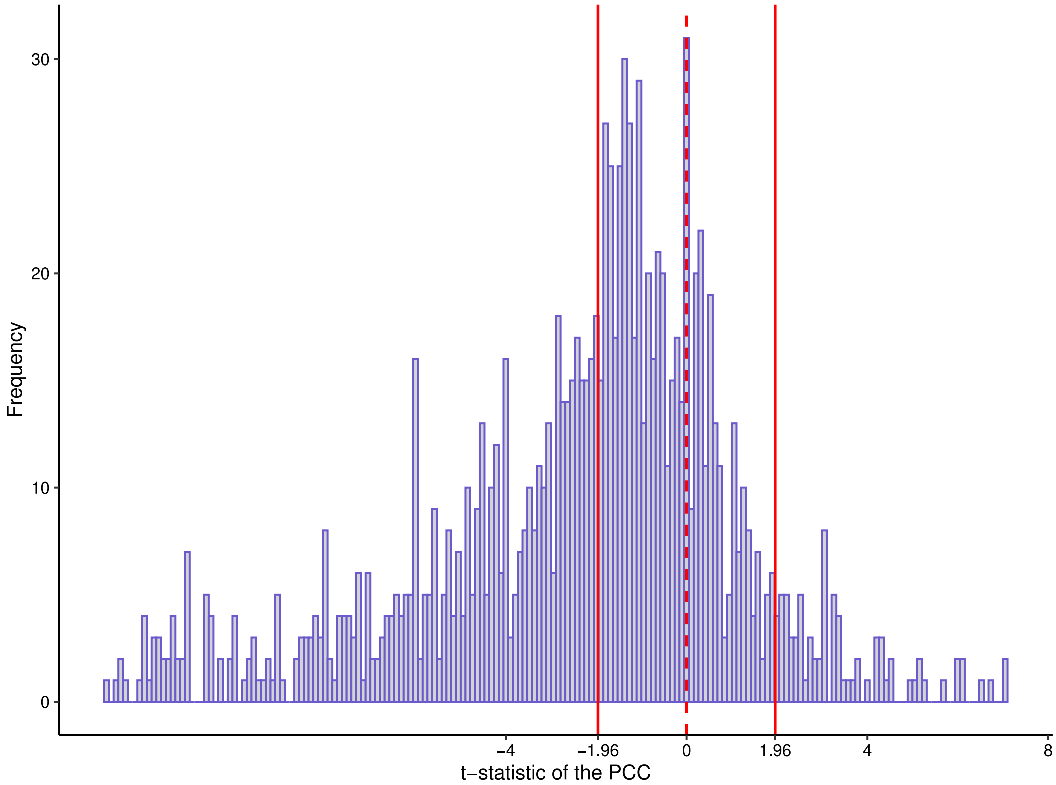
Figure 4.3 presents the distribution of t-statistics of partial correlation coefficients for the female sample. Red vertical lines illustrate the critical values of t=-1.96; 0 and 1.96.

As per the results reported in Table 4.7, we can conclude that there is strong evidence of publication selection bias at the lower bound, but not at the upper bound for the female sample. At the lower bound, the results for caliper sizes 0.05 and 0.1 are 0.727 and 0.694, respectively, indicating that authors exhibit a marked preference for publishing significant results over non-significant ones (e.g., the ratio of significant results to non-significant ones is 73% to 27% for t=-1.96 and caliper size of 0.05). However, at the upper bound, we do not observe any evidence of publication selection bias; the results are steadily distributed.

The results of the Caliper test for the male sample are unrepresentative. For t=1.96, there are only 2 and 6 observations for caliper sizes 0.05 and 0.1, respectively. For t=-1.96, there is only 1 observation. Therefore, we conclude that it is not feasible to proceed with the analysis on the male sample. The distribution of t-statistics of PCCs for the male sample is presented in Figure 4.4.

### Sample of Female Respondents

Figure 4.3: Distribution of t-statistics of PCCs: female sample




Note: The figure presents the distribution of t-statistics of PCCs for the female sample. The solid vertical lines illustrate the critical values of t=-1.96 and t=1.96, while the dashed line represents the value of t=0.

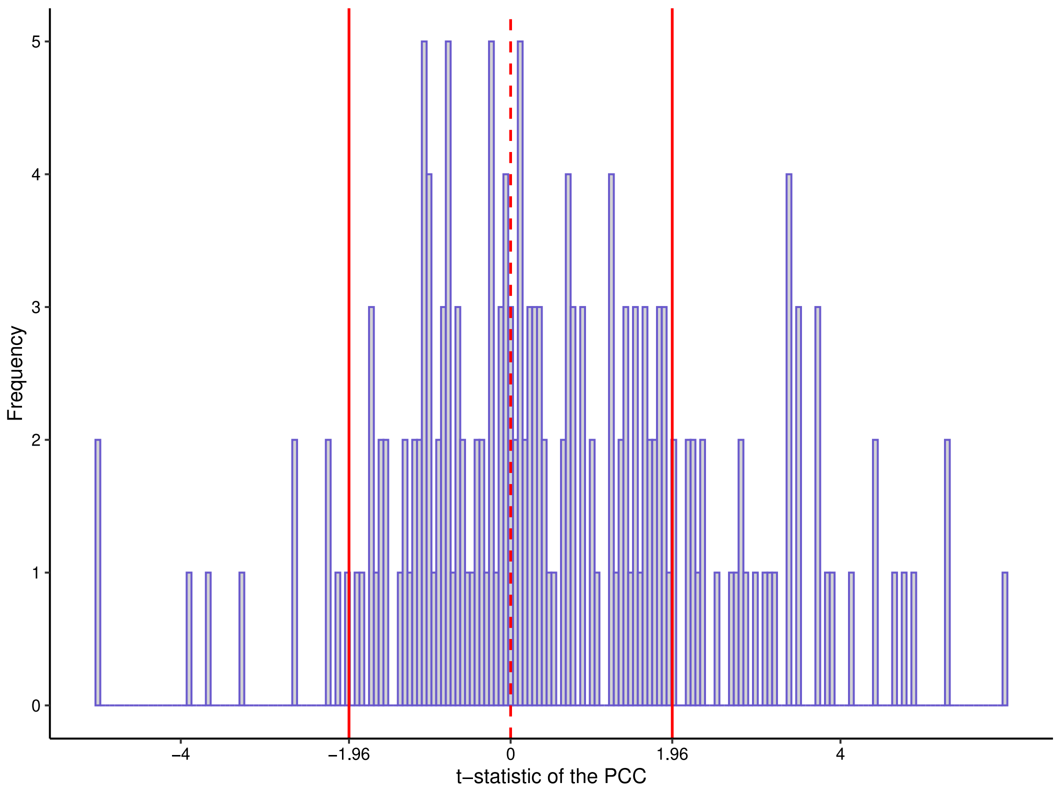
Table 4.7: Caliper Test results: female sample

| | **-1.96** | **1.96** |
|---|---|---|
| Caliper size: 0.05 | 0.727*** | 0.500* |
| | (0.097) | (0.189) |
| Caliper size: 0.1 | 0.694*** | 0.500** |
| | (0.078) | (0.151) |
| Observations | 1323 | 1323 |
| Studies | 88 | 88 |

Note: The table presents the results of the Caliper Test for the female sample for t=-1.96 and t=1.96. Standard errors are presented in round brackets. The asterisks indicate significance levels: * - at 10%, ** - at 5%, *** - at 1%.

## Sample of Male Respondents

Figure 4.4: Distribution of t-statistics of PCCs: male sample




Note: The figure presents the distribution of t-statistics of PCCs for the male sample. The solid vertical lines illustrate the critical values of t=-1.96 and t=1.96, while the dashed line represents the value of t=0.

## 4.5 Robustness check

To enhance our results, we perform robustness checks on two distinct subsamples of the female sample since motherhood penalty is a primary focus of our analysis. We do not perform robustness checks on the male sample due to the smaller number of reported estimates in our dataset.

The first subsample includes only studies employing *Earnings* as a variable measuring labour force participation. This dataset consists of 823 observations collected from 81 studies, with weighted and unweighted means of PCCs equal to -0.01 and -0.036, respectively. The publication bias results obtained from the linear tests (Table B.1) are statistically insignificant for all specifications except the specification weighted by precision, where the result of the standard error is negative (-0.72) and highly statistically significant. The mean beyond bias is negative and statistically significant for all specifications, ranging from -0.03 to -0.022. The instrumental variable $\frac{1}{\log(\text{obs})}$ confirms our assumption of negative publication bias, statistically different from zero (Table B.3). As per the non-linear tests presented in Table B.5, results obtained from the WAAP and Top 10 methods yield negative and statistically significant effects beyond bias (-0.024 and -0.023, respectively).

Next, we perform the robustness check on the subsample of studies controlling for *endogeneity* (63 studies with 757 estimates in total). In this subsample, the weighted mean of PCC is equal to -0.005, and the unweighted mean is equal to -0.02. Linear tests reveal negative and significant evidence of publication bias in methods such as OLS, Between, and Weighted by precision (Table B.2). However, OLS reveals a statistically significant mean beyond bias of 0.015, while other models report negative means between -0.015 and -0.012. IV and non-linear tests show results consistent with the previous subsample (see Table B.4 and Table B.6).

To sum up, the outputs from the robustness checks confirm the findings in the female sample from the main body of the research and those presented in the previous meta-analyses by Cukrowska-Torzewska & Matysiak (2020) and de Linde Leonard & Stanley (2020). As noted by Cukrowska-Torzewska & Matysiak (2020), there is significant evidence of negative publication bias even after the model was extended with explanatory variables. de Linde Leonard & Stanley (2020) also report the presence of publication selectivity and uncover that the motherhood wage penalty persists even after correcting for publication bias. The authors of the least recent meta-analysis, Matysiak & Vignoli (2008), primarily focus on the discussion around the heterogeneity of the results and do not conclude on publication selectivity. These observations set the stage for the examination of heterogeneity in Chapter 5.

# Chapter 5

# Heterogeneity

In the theoretical sections, we have already discussed the various factors that potentially influence the relationship between family size and parents' labour market outcomes. As presented, the reported coefficients exhibit notable variability across different subsamples (see Table 3.4). In the previous section, we introduced the concept of publication bias and its potential impact on study results and now we delve deeper to explore the sources of heterogeneity among studies and their impact on the reported effects.

According to Clarke (2018), decisions regarding childbirth are not made in isolation. The heterogeneity in these decisions and their outcomes is influenced by various factors, including the methodology and data of the primary study, as well as the detailed characteristics of the sample. Thus, our investigation aims to determine whether publication selectivity persists after accounting for diverse study characteristics and seeks to uncover factors that contribute most significantly to the observed variations.

To perform this analysis, we have compiled an extensive set of potential attributes related to heterogeneity, and we will describe these collected variables in detail further in this section. Next, we will employ the Bayesian Model Averaging (BMA) on both female and male samples and present the results. Finally, we will provide robustness checks by presenting our findings from the Frequentist Model Averaging (FMA) and Frequentist Check.

## 5.1 Explanatory variables

To thoroughly investigate the heterogeneity in the literature examining the relationship between family size and parents' labour market outcomes, it is necessary to analyse the underlying data characteristics beyond simple effect estimates. The following subsections will classify distinct groups of variables and discuss their relevance within our context. The descriptive statistics of variables used in the analysis is presented in Table 5.1. We denote the reference variables as *ref.var*.

### 5.1.1 Data parameters

First, we consider the *sample size*. The number of observations in a sample affects both credibility and precision of the estimates, as larger sample sizes generally reduce the likelihood of random errors. In our datasets, sample sizes differ significantly. For example, Jacobsen *et al.* (1999) employs the data sample of 1,210,215 individual observations, while Hersch (1991) analyses the sample of only 217 respondents.

The second characteristic is the *average data year*, which indicates the period during which the data were collected. This variable is crucial as it reflects the changes in economic conditions, labour policies, and social norms over time within the context of the study, making it a critical factor to consider. We assume that the effect varies over time. To standardize this variable, we transform it to the logarithm of the average year of study in our analysis.

Lastly, we examine whether the authors employ cross-sectional or panel data. The variable *panel* takes values of 1 if the dataset is of a panel nature and 0 if it is cross-sectional. Panel data allows researchers to control for unobserved heterogeneity and analyse changes over time, whereas cross-sectional data provides a snapshot at a single point in time and may lack depth and control.

### 5.1.2 Publication specifics

The dummy variable *published* takes the value of 1 if a primary study was published in a peer-reviewed journal. This variable aims to reflect the credibility and rigorous evaluation associated with peer-reviewed publications.

The *publication year* of a primary study is another explanatory variable that captures the time trend. We denote this variable as the logarithm of the year of publication of the primary study minus the year in which the first study in our datatset was published. The adjustment allows for a standardized comparison across papers published in different years.

To assess the scholarly impact of primary studies, we introduce the variable *adjusted citations*. We normalise this metric to ensure that recent studies are not disadvantaged compared to older ones. The number of citations was collected from Google Scholar in March 2024.

The *impact factor* of the journal is another significant publication characteristic. The variable serves as a proxy for the journals' quality, with an impact factor sourced from RePEc (Research Papers in Economics). Since our dataset includes several primary studies that are either unpublished (i.e., working papers) or published in journals not included in the RePEc list, these studies are assigned an impact factor of zero.

### 5.1.3 Methodology

We code six distinct analytical methods employed in the analysis of the relationship between family size and parents' labour market outcomes. *OLS* is a widely used method for estimating linear relationships between variables by minimizing the sum of squared differences between observed and predicted values. *Panel Data Techniques* encompass fixed and random effects models. *IV and 2SLS* methods aim to address endogeneity issues and isolate causality by utilizing instrumental variables. *Binary Outcome Models* include logit and probit estimation techniques suitable for models with binary dependent variables. *Advanced Regression Techniques* reflect methodologies as GLS, GMM, quantile regressions and other advanced methodologies employed to handle specific model issues. We also use the *Other* category to include estimates from three primary studies with undefined methodologies.

### 5.1.4 Design of the analysis

To capture the analytical design of the dataset, we include variables as *Endogeneity*, *Location*, *Number of explanatory variables* and *Main*.

Addressing endogeneity issues such as omitted variable bias, is paramount, as they can distort the results by introducing systematic errors that bias the estimated relationships. Therefore, we anticipate significant heterogeneity between the samples that tackle endogeneity and those that do not. In the context of our analysis, unobserved heterogeneity is a significant concern primarily due to the potential correlation between the error term and housework. Career-oriented individuals tend to spend less time on housework, resulting in a negative corelation between the individual effect in wage equation and housework, which in turn leads to negatively biased OLS estimates (Bryan & Sevilla-Sanz 2011).

The location of the studies is another crucial factor in shaping the results. We split our sample into six geographical regions: *Asia*, *Europe*, *North America*, *South America*, *Oceania*, and the *Other* (i.e., category mainly including clusters of developing countries). The value of comparative analysis by location of the respondents lies in its ability to assess whether parents with the freedom to choose their market outcomes can leverage these opportunities and face fewer penalties and address cultural differences within the dataset (Gash 2009).

In our dataset, the number of explanatory variables ranges from one to 49 estimates. It is essential to maintain a balance in the number of independent variables in the model to avoid multicollinearity. Consequently, we introduce the variable *Number of explanatory variables*. On average, the authors of our primary studies employed nine covariates for the female sample, and eleven for the male.

Finally, we introduce the *Main* variable specific to our topic. Since it is common in our analysis to report distinct results for different sets of covariates and subsets of data, we employ the approach introduced by Lang (2023b). If multiple specifications tested the same hypothesis, we used the one favored by the authors. Robustness checks and heterogeneity analyses were included only if they were specifically mentioned as a focus of the article. The variable *Main* takes a value of 1 if the result is principal, and 0 if the result is not central to the article.

### 5.1.5 Variable specifications

Researchers employ various specifications to define both family size and labour market outcome. As mentioned earlier, we do not restrict the results to only those representing *log(wages)*, but aim to cover all possible effect proxies instead. Therefore, we code four variables to control for the variation in labour market outcomes: *Earnings* (mainly includes logarithm of wages and hourly pay), *Experience* (defines the work experience in months or years), *Hours worked* (covers solely the duration of work) and *Participation* (includes variables defining whether the respondent participates or not in the labour market). The following categories cover the variability in the specification of the main independent variables: *One child* (presents studies using the presence of one child as an independent variable), *2+ children* (denotes studies measuring the impact of the second and further children on the parents' labour market outcomes), *Motherhood years* (captures the studies using years of motherhood as an independent variable), *Number of children* (includes number of birth or number of children) and *Presence of children* (includes all binary variables related to the presence of children).

### 5.1.6 Set of controls

As we proceed with including crucial variables, we account for a set of control variables, such as *Age*, *Education*, *Marriage*, *Race*, *Sector*, *Schedule*, and *Location*. We decided to include these variables into the dataset during the data collection phase based on the theoretical background described in Chapter 2 and the level of details available in the primary studies. These variables are coded as 1 if the authors of the primary studies accounted for them, and 0 otherwise.

Table 5.1: Descriptive statistics of variables used for heterogeneity

| | | Female sample | | Male sample | |
|---|---|---|---|---|---|
| **Variable** | **Description** | **Mean** | **SD** | **Mean** | **SD** |
| PCC | partial correlation coefficient | -0.0314 | 0.0590 | 0.0185 | 0.0525 |
| Standard error | standard error of PCC | 0.0126 | 0.0122 | 0.0110 | 0.0120 |
| *Data parameters* | | | | | |
| Panel | = 1 if primary study uses panel data | 0.5593 | 0.4967 | 0.4067 | 0.4924 |
| Sample size | logarithm of the sample size | 9.7837 | 2.1653 | 10.0603 | 2.0778 |
| Average data year | logarithm of the average data year | 7.5967 | 0.0049 | 7.5969 | 0.0059 |
| *Publication specifics* | | | | | |
| Published | = 1 if primary study was published in peer-reviewed journal | 0.9033 | 0.2957 | 0.7512 | 0.4334 |
| Publication year | logarithm of the publication year minus the base year | 3.3706 | 0.9981 | 2.6724 | 3.3121 |
| Impact factor | RePEc impact factor | 0.6758 | 0.9884 | 0.3341 | 0.5151 |
| Adjusted citations | logarithm of the total number of citations divided by the years since publication | 2.3029 | 1.2158 | 2.2337 | 1.2341 |
| *Methodology* | | | | | |
| OLS | = 1 if primary study uses OLS | 0.4104 | 0.4921 | 0.3732 | 0.4848 |
| Panel Data Techniques *(ref.var)* | = 1 if primary study uses FE and RE models | 0.2101 | 0.4076 | 0.3062 | 0.4620 |
| IV and 2SLS | = 1 if primary study uses IV or 2SLS | 0.2472 | 0.4315 | 0.2057 | 0.4052 |
| Bianry Outcome Models | = 1 if primary study uses logit or probit | -0.0026 | -0.0018 | 0.0046 | 0.0025 |
| Advanced Regression Techniques | = 1 if primary study uses GLS, GMM or quantile regression | 0.0741 | 0.2620 | 0.0096 | 0.0976 |
| Other | = 1 if primary study does not mention the methodology | 0.1678 | 0.3738 | 0.0046 | 0.0025 |
| *Design of the analysis* | | | | | |
| Endogeneity | = 1 if estimation method accounts for endogeneity | 0.5722 | 0.4949 | 0.5120 | 0.5011 |
| Number of explanatory variables | logarithm of explanatory variables | 9.0786 | 6.0014 | 10.6986 | 6.7447 |
| Main | = 1 if the model is favoured by authors | 0.9002 | 0.2998 | 0.8947 | 0.3076 |
| Asia | = 1 if primary study uses data from Asia | 0.0469 | 0.2114 | 0.0096 | 0.0976 |
| Europe *(ref.var)* | = 1 if primary study uses data from Europe | 0.2419 | 0.4284 | 0.5311 | 0.5002 |
| North America | = 1 if primary study uses data from North America | 0.4618 | 0.4987 | 0.4593 | 0.4995 |
| South America | = 1 if primary study uses data from South America | 0.0438 | 0.2048 | - | - |
| Oceania | = 1 if primary study uses data from Oceania | 0.0378 | 0.1908 | - | - |
| Other | = 1 if primary study uses data from a set of countries (i.e. developing countries, middle-income countries,etc.) | -0.0026 | -0.0018 | - | - |
| *Variable specifications* | | | | | |
| Earnings | = 1 if the dependent variable is specified as earnings | 0.6221 | 0.4851 | 0.7416 | 0.4388 |
| Hours worked *(ref.var)* | = 1 if the dependent variable is specified as hours worked | 0.0748 | 0.2632 | 0.0766 | 0.2665 |
| Participation | = 1 if the dependent variable is specified as labour force participation | 0.2918 | 0.4547 | 0.1340 | 0.3414 |
| Experience | = 1 if the dependent variable is specified as experience | 0.0113 | 0.1059 | 0.0478 | 0.2140 |
| One child *(ref.var)* | = 1 if the independent variable is specified as one child | 0.1081 | 0.3106 | 0.0622 | 0.2421 |
| 2+ children | = 1 if the independent variable is specified as 2 children and more | 0.2525 | 0.4346 | 0.3445 | 0.4763 |
| Number of children | = 1 if the independent variable is specified as number of children | -0.0026 | -0.0018 | 0.0046 | 0.0025 |
| Presence of children | = 1 if the independent variable is specified as presence of children | 0.3190 | 0.4663 | 0.3828 | 0.4872 |
| Motherhood years | = 1 if the independent variable is specified as years of motherhood | 0.0060 | 0.0776 | 0.0383 | 0.1923 |
| *Set of controls* | | | | | |
| Age | = 1 if primary study controls for age | 0.7067 | 0.4554 | 0.3971 | 0.4905 |
| Education | = 1 if primary study controls for education | 0.6712 | 0.4700 | 0.4545 | 0.4991 |
| Marriage | = 1 if primary study controls for marriage | 0.4339 | 0.4958 | 0.4306 | 0.4964 |
| Race | = 1 if primary study controls for race | 0.2389 | 0.4265 | 0.2153 | 0.4120 |
| Sector | = 1 if primary study controls for job sector | 0.0665 | 0.2493 | 0.0526 | 0.2238 |
| Schedule | = 1 if primary study controls for work schedule | 0.1194 | 0.3244 | 0.1005 | 0.3014 |
| Location | = 1 if primary study controls for work location | 0.1920 | 0.3940 | 0.0574 | 0.2332 |

Note: The table presents the descriptive statistics for the female and male samples. We denote the reference variables as *ref.var*.

## 5.2 Estimation methods

### 5.2.1 Overview of Bayesian Model Averaging

In the preceding section, we identified 37 variables for the female sample and 34 variables for the male sample that possibly reflect the heterogeneity across the partial correlation coefficients derived from the reported estimates. In this section, we investigate whether a relationship exists between the partial correlation coefficient (dependent variable) and these variables. Specifically, we aim to determine whether these distinct study characteristics can account for the variability observed in the primary studies. To achieve this objective, we extend the model used for publication bias detection by introducing additional control variables representing study heterogeneity. The newly specified model is as follows:

$$PCC_{is} = \beta_0 + \beta_1 X_{is} + \beta_2 SE(PCC_{is}) + \epsilon_{is} \tag{5.1}$$

where $PCC_{is}$ stands for the $i$-th partial correlation coefficient from the $s$-th primary study, $\beta_0$ represents the constant term, $X_{is}$ corresponds to the vector of control variables introduced earlier, $SE(PCC_{is})$ is a standard error, $\beta_2$ identifies the direction and size of publication bias, and $\epsilon_{is}$ is the error term.

Nonetheless, due to the uncertainty regarding the exact set of variables, regressing partial correlation coefficients on all possible controls can lead to inflated standard errors and imprecise results. To build an optimal regression model, one could use a stepwise selection method, which involves iteratively adding and removing predictors based on t-test outcomes. However, the statistical validity of this approach is debatable. In our case, this method would yield $2^{37}$ (for the female sample) and $2^{34}$ (for the male sample) different models, given that we have 37 and 34 variables, resulting in $2^{37}$ and $2^{34}$ possible combinations, respectively. This makes the analysis complex and time-consuming.

To mitigate this problem and account for model uncertainty, we apply the Bayesian Model Averaging (BMA) method. BMA generates a weighted average of multiple models by considering various combinations of the independent variables. This approach is widely utilized in contemporary meta-analyses (see, for instance Havranek *et al.* (2015); Elminejad *et al.* (2022)). The application of BMA needs an understanding of several fundamental concepts, which we will elaborate on in this section.

## Posterior Model Probability (PMP)

The Posterior Model Probability (PMP) is a concept derived from the Bayesian theorem and used in BMA to calculate the probability of a specific model being the true model, given the observed data. The PMP is calculated using the model's likelihood and its prior probability, as follows:

$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum_{j=1}^{K} P(D|M_j)P(M_j)} \tag{5.2}$$

where $P(D|M_k)$ corresponds to the likelihood of the data given the model $M_k$, $P(M_k)$ represents the prior probability of model $M_k$. The denominator sums the numerator across $K$ models.

## Posterior Mean

In the context of BMA, the posterior mean denotes the weighted average of parameter estimates across all models with weights represented by PMPs. This approach acknowledges model uncertainty by averaging estimates, accounting for the relative support each model receives from the data. The posterior mean of $\beta$ is given by:

$$E(\beta|D) = \sum_{k=1}^{K} E(\beta|D, M_k)P(M_k|D) \tag{5.3}$$

where $E(\beta|D)$ is the expectation of $\beta$ given the data $D$ and model $M_k$. $P(M_k|D)$ is the PMP for model $M_k$. This approach prevents the risk of over-reliance on a single model (i.e., mitigates the probability of parameter estimates being influenced by any single model).

## Posterior Variance

The posterior variance in BMA measures the uncertainty of the parameter estimates by applying model uncertainty. Posterior variance is computed as:

$$\mathrm{Var}(\beta|D) = \sum_{k=1}^{K} \left[\mathrm{Var}(\beta|D, M_k) + (E(\beta|D, M_k) - E(\beta|D))^2\right] P(M_k|D) \tag{5.4}$$

where $\mathrm{Var}(\beta|D, M_k)$ is the variance of $\beta$ given data $D$ and model $M_k$,

$E(\beta|D, M_k)$ is the posterior mean within model $M_k$. The model captures both within and between model variations.

## Posterior Inclusion Probability (PIP)

Posterior Inclusion Probability (PIP) is a measure denoting the probability that a specific variable is included in the true model, conditional on the observed data. PIP for the variable $X_i$ is calculated as:

$$PIP(X_i) = \sum_{k:X_i \in M_k} P(M_k|D) \tag{5.5}$$

where $P(M_k|D)$ is the PMP for model $M_k$. The closer the PIP value to one, the stronger the evidence for the inclusion of the variable in the true model.

Based on the rule of thumb introduced by Jeffreys (1998):

- The evidence of a regressor having an effect is *weak* if PIP lies between 0.5-0.75.
- The evidence of a regressor having an effect is *positive* if PIP lies between 0.75-0.95.
- The evidence of a regressor having an effect is *strong* if PIP lies between 0.95-0.99.
- The evidence of a regressor having an effect is *decisive* if PIP lies between 0.99-1.

## Distribution Priors

There are two types of priors introduced in BMA: prior on the parameter space $g$ and prior on the model space $p(M_s)$ (Hasan *et al.* (2018)). When a researcher possesses knowledge regarding the parameters, it should be further incorporated in priors and informative priors should be employed (Eicher *et al.* (2011)). However, the amount of prior knowledge is often limited. Based on the paper by Eicher *et al.* (2011), the best priors identified via examination of predictive performance of 12 candidates are unit information prior (UIP) and uniform model prior (UMP). The former is a parameter prior that is designed to provide the same amount of information as a single observation from the data. The latter is a model prior that assigns equal probability to all candidate models under consideration.

### 5.2.2 Application of Bayesian Model Averaging

Before delving into the description of results for different specifications of heterogeneity, we should carefully consider potential distortions of the results. Despite our intentions to include all pre-introduced variables in the model, some issues may arise in estimating the regression model.

First of all, the inclusion of all sets of dummy variables for every category will lead to a dummy variable trap. Therefore, we denote one variable as a reference variable (*ref.var.*) in Table 5.1 for every set of dummy variables. Consequently, we exclude reference variables from our model.

After excluding of reference variables from our model, some variables with high collinearity persist. According to the guidelines proposed by Ratner (2009), correlation coefficients up to 0.7 are considered weak (0 - 0.3) and moderate (0.3 - 0.7). Consequently, we exclude one variable from each pair of variables with correlation coefficients higher than 0.7 in absolute value. Thus, we exclude variable *Size* (correlated with the standard error of PCC) for both samples. For the female sample, we also exclude *Endogeneity* (correlated with *OLS*) and *Participation* (correlated with *Earnings*). For the male sample, we additionally exclude variables *Asia* (correlated with *Advanced Regression Techniques* as all estimates from *Asia* apply *Advanced Regression Techniques*), *Other methodologies* (due to the correlation with the variable *Publication year*), and *Published* (excluded to obtain precise results due to its correlation of 0.7 with *2+ children* variable) for the male sample only. Based on the variance inflation factors (VIFs), the remaining variables do not suffer from collinearity. The correlation matrices for both samples are available in the Appendix (Figure C.4 for the female sample and Figure C.5 for the male sample). As an additional step to address multicollinearity, we relax the zero correlation assumption by applying the collinearity-adjusted dilution model prior suggested by George (2010). We estimate the BMA model in R, using the package provided by Zeugner & Feldkircher (2015).

Furthermore, we address the imbalance in the dataset by weighting the model by the inverse number of estimates for each study.

### 5.2.3 Frequentist Model Averaging

As part of the robustness check for the results obtained from BMA and following the methodological pattern of modern meta-analyses (see, for example, Havranek *et al.* (2018); Bajzik *et al.* (2020); Matousek *et al.* (2022)), we apply Frequentist Model Averaging (FMA) technique. FMA is a statistical technique similar to BMA, aiming to improve the reliability and accuracy of statistical predictions by averaging several models. However, FMA attributes weights based on performance metrics such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or modern Mallows Model Average (MMA), rather than posterior model probabilities derived from prior distributions. We employ Mallow's weights, which Hansen (2007) has demonstrated to be optimal for Frequentist Model Averaging (FMA) due to their ability to minimize squared error among discrete model averaging estimators. Furthermore, we apply orthogonalization to the entire set of covariates, thereby reducing the number of models to $K$, instead of $2^K$, as recommended by Amini & Parmeter (2012). We employ the FMA technique in R, using the code provided in the online appendix by Havranek *et al.* (2021).

### 5.2.4 Frequentist Check

As the final step of our robustness check, we employ a technique based on OLS estimation of the initial Equation 5.1, with standard errors clustered at the study level. This procedure is known as the Frequentist check. In this segment of our analysis, we limit our consideration to variables from the BMA approach that have a PIP of at least 0.5, in line with the lower threshold for weak evidence as defined by Jeffreys (1998).

## 5.3 Results

In this section, we aim to provide a comprehensive summary of the results obtained from the heterogeneity analysis of female and male samples. Following the exclusion of reference variables and variables suffering from multicollinearity, the final number of variables with standard error is 32 for the female sample and 28 for the male sample.

The results of the baseline Bayesian Model Averaging (BMA) model with unit information prior (UIP) and dilution model prior, based on the best 5000 models, are resented in Figure 5.1 (female sample) and Figure 5.2 (male sample). In the figures, blue (darker in grayscale) indicates positive coefficients, while red (lighter in greyscale) denotes negative coefficients. A white color indicates the exclusion of a variable from the model. In both figures, we present the variables in the order of their Posterior Inclusion Probability (PIP) values.

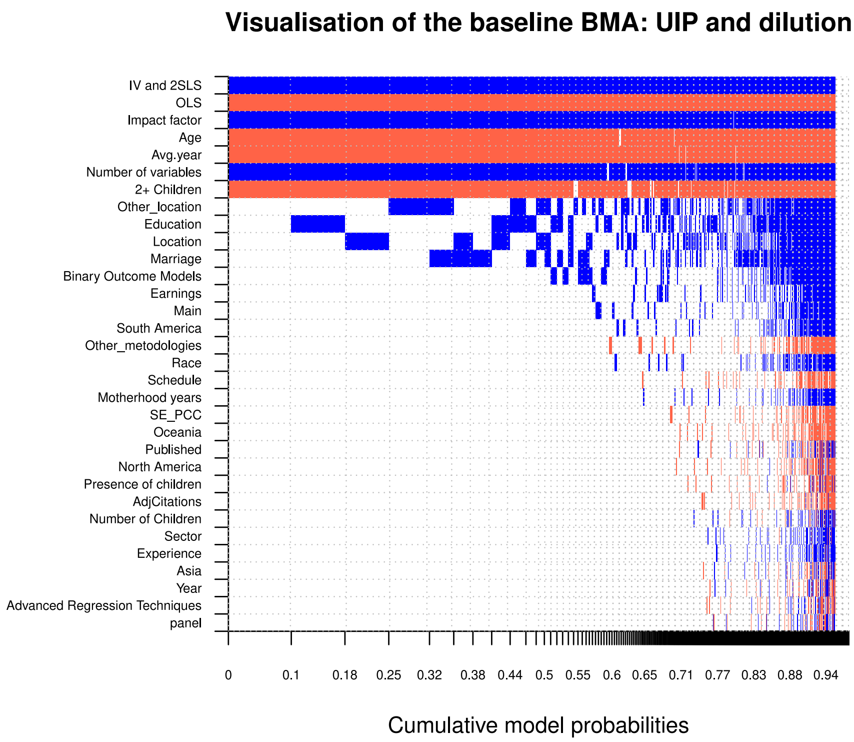
Thus, those variables with the highest PIP are positioned at the top. Additionally, we present the results from BMA estimations across different priors: UIP and dilution, BRIC and random and HQ and random in Figure C.6 for the female sample and Figure C.7 for the male sample.

The graphical output of BMA analysis is followed by quantitative results from BMA, FMA, and Frequentist Check presented in Table 5.2 for the female sample and in Table 5.3 for the male sample.

### 5.3.1 Results: female sample

The highest posterior inclusion probability in the female sample pertains to the *IV and 2SLS* variable, with a positive coefficient. As discussed earlier, relevant and exogenous instrumental variables can address endogeneity by isolating exogenous variation in family size, leading to more reliable estimates of the causal effects.

Figure 5.1: Model inclusion of the BMA estimation: female sample




Note: The variables included in the model are listed on the y-axis based on their PIP results, and the posterior model probabilities are on the x-axis. Blue indicates positive coefficients, while red signifies negative coefficients. White represents the exclusion of the variable from the model.

We can see that the *OLS* variable also belongs among those with the highest PIP value and contributes to the variations in the PCCs, but with a negative sign. This suggests that if the authors apply basic OLS to estimate the relationship between family size and labour market outcome, the results would be more negative. This might be attributable to the endogeneity bias related to OLS estimates and the simplistic assumptions introduced by the OLS models that might not hold in the context of family size and labour market outcomes (i.e., absence of reverse causality, perfect exogeneity of the independent variables).

The *impact factor* also has a positive and decisive effect on the variations in the PCCs. This indicates that studies published in higher-quality journals, typically having higher impact factors, tend to present more positive effects of family size on female labour market outcomes. Several factors may cause the introduced heterogeneity. Firstly, higher-quality journals may attract more established researchers with access to better-quality data and advanced analytical tools, leading to more precise results of the effect under study. Secondly, as the submission process includes rigorous peer reviews, it may contribute to the identification and correction of methodological flaws resulting in biased results, ensuring that only the most reliable results are published.

The negative coefficient for the *age* variable indicates that the impact of the family size increase tends to be more negative as the age is included in the model as a control. Older women may experience more career interruptions due to the challenges in their work-life balance and longer gaps in employment history, having a cumulative impact on their labour market outcomes. Moreover, older women with larger families may experience age-related discrimination in the labour market. Interestingly, in the previous meta-analysis by Cukrowska-Torzewska & Matysiak (2020) the authors report the negative effect of omitting age in the model (meaning that if the model controls for age, the results are more positive), which is not consistent with our findings.

The variable denoting the *average data year* shows a negative coefficient with a decisive effect (PIP higher than 0.99), indicating a significant relationship where studies using more recent datasets report more negative effects. This result was unexpected, as a plethora of studies (Brewster & Rindfuss 2000; Meyers & Gornick 2005; Brewster & Rindfuss 2000) discuss the implementation of family-friendly policies and gender-neutralization among the workforce. However, as can be seen from the results, the effectiveness of the policy changes may be hard to achieve. Another explanation may lie in recent economic uncertainties disproportionately affecting women with larger families.

The positive coefficient for the *number of variables* included in the model may stem from the better model specifications, as models with more control variables are likely to be better-specified and reduce omitted variable bias. Thus, the picture of the true relationship is in general clearer.

As per the results, the models using the *presence of two or more children* as a family size variable tend to have more pronounced negative PCCs. The possible explanations may be from both employer-related (as employers might perceive women with two or more children as less available and committed, potentially affecting career opportunities) and family-related (as more children demand more time from their parents) perspectives (Bryan & Sevilla-Sanz 2011).

The results across different priors are presented in Figure C.6. As per the figure, there are no significant differences among the model specifications for the female sample.

Regarding the results of the *standard error* of the PCC, the BMA results suggest that the standard error variable is not significant and does not contribute to explaining the variability in PCCs. The low PIP value indicates that across all models considered, only a small fraction finds evidence of publication bias. The output obtained from the FMA model suggest the presence of the negative publication bias even after the inclusion of other variables in the model. The reason behind the discrepancy between the significance of the results may be attributed to the distinct approaches to handling model uncertainty and variable inclusion. BMA, taking a more holistic approach, averages over all possible models. In contrast, FMA may highlight variables that seem important in specific models that fit the data well.

Moreover, FMA model suggests that such variables as *Publication year*, *Binary Outcome Models*, *Main*, all variables of geographical locations except *North America*, all specifications of the dependent variable (*Earnings* and *Experience*), and controlling for *marriage* may introduce a significant heterogeneity in the data.

The results of the Frequentist Check validate the significance and direction of all variables with PIP higher than 0.5 for the female sample, except for the *average data year* and controlling for *age*, that appear insignificant in the Frequentist Check.

Table 5.2: Summary of results for BMA, FMA, and Frequentist Check: female sample

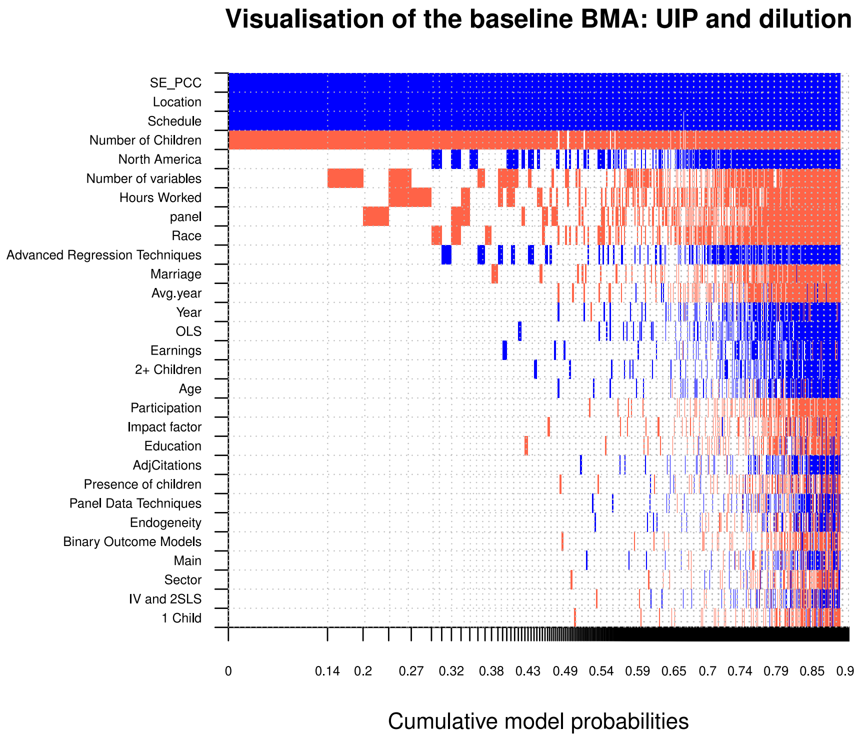
| | BMA | | | FMA | | | Frequentist Check | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **Post.Mean** | **Post.SD** | **PIP** | **Coef.** | **SD** | **p-value** | **Coef.** | **SD** | **p-value** |
| PCC | | | | | | | | | |
| Standard error | -0.0031 | 0.0290 | 0.0212 | -0.4803 | 0.1593 | 0.0026 | | | |
| *Data parameters* | | | | | | | | | |
| Panel | -0.0000 | 0.0004 | 0.0117 | 0.0008 | 0.0048 | 0.8676 | | | |
| Sample size | | | | | | | | | |
| Average data year | -0.0041 | 0.0011 | 0.9848 | -0.5183 | 0.4502 | 0.2496 | 0.6051 | 0.3754 | 0.1072 |
| *Publication specifics* | | | | | | | | | |
| Published | -0.0001 | 0.0018 | 0.0201 | -0.0066 | 0.0074 | 0.3725 | | | |
| Publication year | -0.0000 | 0.0001 | 0.0126 | -0.0038 | 0.0017 | 0.0254 | | | |
| Impact factor | 0.0094 | 0.0021 | 0.9975 | 0.0083 | 0.0023 | 0.0003 | 0.0081 | 0.0015 | 0.0000 |
| Adjusted citations | -0.0000 | 0.0002 | 0.0149 | -0.0022 | 0.0017 | 0.1956 | | | |
| *Methodology* | | | | | | | | | |
| OLS | -0.0174 | 0.0037 | 0.9984 | -0.0180 | 0.0049 | 0.0002 | -0.0159 | 0.0040 | 0.0001 |
| Panel Data Techniques *(ref.var)* | | | | | | | | | |
| IV and 2SLS | 0.0352 | 0.0054 | 1.0000 | 0.0366 | 0.0060 | 0.0000 | 0.0365 | 0.0038 | 0.0000 |
| Binary Outcome Models | 0.0017 | 0.0050 | 0.1175 | 0.0169 | 0.0085 | 0.0468 | | | |
| Advanced Regression Techniques | -0.0000 | 0.0008 | 0.0113 | 0.0144 | 0.0084 | 0.0865 | | | |
| Other | -0.0005 | 0.0031 | 0.0393 | 0.0131 | 0.0149 | 0.3793 | | | |
| *Design of the analysis* | | | | | | | | | |
| Endogeneity | | | | | | | | | |
| Number of explanatory variables | 0.0011 | 0.0003 | 0.9713 | 0.0002 | 0.0003 | 0.5050 | 0.0010 | 0.0002 | 0.0000 |
| Main | 0.0005 | 0.0025 | 0.0490 | 0.0112 | 0.0056 | 0.0455 | | | |
| Asia | -0.0000 | 0.0009 | 0.0125 | -0.0178 | 0.0087 | 0.0408 | | | |
| Europe *(ref.var)* | | | | | | | | | |
| North America | -0.0000 | 0.0006 | 0.0155 | -0.0087 | 0.0051 | 0.0880 | | | |
| South America | 0.0006 | 0.0035 | 0.0411 | 0.0200 | 0.0093 | 0.0315 | | | |
| Oceania | -0.0002 | 0.0023 | 0.0195 | -0.0328 | 0.0107 | 0.0022 | | | |
| Other countries | 0.0114 | 0.0164 | 0.3662 | 0.0396 | 0.0079 | 0.0000 | | | |
| *Variable specifications* | | | | | | | | | |
| Earnings | 0.0005 | 0.0024 | 0.0606 | 0.0217 | 0.0043 | 0.0000 | | | |
| Hours worked *(ref.var)* | | | | | | | | | |
| Participation | | | | | | | | | |
| Experience | 0.0000 | 0.0015 | 0.0130 | 0.0502 | 0.0152 | 0.0010 | | | |
| One child *(ref.var)* | | | | | | | | | |
| 2+ children | -0.0165 | 0.0054 | 0.9622 | -0.0185 | 0.0055 | 0.0008 | -0.0161 | 0.0036 | 0.0000 |
| Number of children | 0.0000 | 0.0006 | 0.0146 | -0.0128 | 0.0067 | 0.0561 | | | |
| Presence of children | -0.0000 | 0.0006 | 0.0161 | 0.0003 | 0.0059 | 0.9594 | | | |
| Motherhood years | 0.0005 | 0.0049 | 0.0227 | 0.0284 | 0.0208 | 0.1721 | | | |
| *Set of controls* | | | | | | | | | |
| Age | -0.0152 | 0.0039 | 0.9906 | -0.0050 | 0.0043 | 0.2449 | -0.0058 | 0.0038 | 0.1342 |
| Education | 0.0037 | 0.0055 | 0.3515 | -0.0010 | 0.0045 | 0.8241 | | | |
| Marriage | 0.0029 | 0.0048 | 0.3065 | 0.0140 | 0.0041 | 0.0006 | | | |
| Race | 0.0002 | 0.0015 | 0.0373 | 0.0092 | 0.0048 | 0.0553 | | | |
| Sector | 0.0000 | 0.0007 | 0.0133 | 0.0132 | 0.0070 | 0.0593 | | | |
| Schedule | -0.0001 | 0.0010 | 0.0231 | 0.0046 | 0.0056 | 0.4114 | | | |
| Location | 0.0045 | 0.0073 | 0.3109 | 0.0056 | 0.0062 | 0.3664 | | | |

Note: The table presents the results of BMA, FMA, and Frequentist Check for the female sample, containing 1323 observations from 88 individual studies.

### 5.3.2 Results: male sample

As for the male sample, the results of the *standard error* of PCCs are consistent across all models, indicating a strongly positive publication bias in the relationship under study even after the model was populated with additional variables. The results confirm those obtained in Chapter 4.

Figure 5.2: Model inclusion of the BMA estimation: male sample




Note: The variables included in the model are listed on the y-axis based on their PIP results, and the posterior model probabilities are on the x-axis. Positive coefficients are presented in blue, and negative coefficients in red. White denotes the exclusion of the variable from the model.

The positive and decisive coefficient of the *location* variable suggests that controlling for geographical factors influences the PCCs and results in more positive outcomes. This effect could stem from variations in economic, cultural, and policy environments between urban and rural regions. By accounting for these differences, a more accurate picture of the relationship between family size and male labour market outcomes can be achieved .

Similarly, controlling for work *schedules* results in more positive outcomes. This indicates that differences between part-time and full-time employees might influence the relationship, and accounting for these variations implies a more positive outcome.

The negative coefficient for the *number of children* suggests that studies using number of children as a proxy for the family size are associated with more negative PCCs.

The results obtained from the FMA suggest that the following variables might be similarly significant in explaining the heterogeneity across the male sample: *Average data year*, *IV and 2SLS*, *North America* and *Race*.

Frequentist Check confirms the direction and significance of all variables with PIP higher than 0.5 for the male sample.

Table 5.3: Summary of results for BMA, FMA, and Frequentist Check: male sample

| | BMA | | | FMA | | | Frequentist Check | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **Post.Mean** | **Post.SD** | **PIP** | **Coef.** | **SD** | **p-value** | **Coef.** | **SD** | **p-value** |
| PCC | | | | | | | | | |
| Standard error | 2.0304 | 0.2731 | 1.0000 | 2.1481 | 0.3755 | 0.0000 | 2.1072 | 0.2254 | 0.0000 |
| *Data parameters* | | | | | | | | | |
| Panel | -0.0040 | 0.0078 | 0.2492 | 0.0003 | 0.0138 | 0.9827 | | | |
| Sample size | | | | | | | | | |
| Average data year | -0.0013 | 0.0035 | 0.1603 | 3.1570 | 0.8332 | 0.0002 | | | |
| *Publication specifics* | | | | | | | | | |
| Published | | | | | | | | | |
| Publication year | 0.0005 | 0.0014 | 0.1415 | 0.0010 | 0.0016 | 0.5320 | | | |
| Impact factor | -0.0001 | 0.0021 | 0.0498 | -0.0024 | 0.0086 | 0.7802 | | | |
| Adjusted citations | 0.0002 | 0.0012 | 0.0495 | 0.0041 | 0.0045 | 0.3622 | | | |
| *Methodology* | | | | | | | | | |
| OLS | 0.0014 | 0.0046 | 0.1074 | 0.0092 | 0.0115 | 0.4237 | | | |
| Panel Data Techniques *(ref.var)* | | | | | | | | | |
| IV and 2SLS | 0.0001 | 0.0028 | 0.0305 | 0.0302 | 0.0142 | 0.0334 | | | |
| Binary Outcome Models | -0.0001 | 0.0028 | 0.0312 | -0.0225 | 0.0254 | 0.3757 | | | |
| Advanced Regression Techniques | 0.0072 | 0.0176 | 0.1806 | 0.0674 | 0.0390 | 0.0840 | | | |
| Other | | | | | | | | | |
| *Design of the analysis* | | | | | | | | | |
| Endogeneity | -0.0001 | 0.0022 | 0.0368 | -0.0194 | 0.0125 | 0.1207 | | | |
| Number of explanatory variables | -0.0002 | 0.0005 | 0.2187 | 0.0007 | 0.0008 | 0.3816 | | | |
| Main | -0.0000 | 0.0023 | 0.0325 | 0.0163 | 0.0153 | 0.2867 | | | |
| Asia | | | | | | | | | |
| Europe *(ref.var)* | | | | | | | | | |
| North America | 0.0100 | 0.0145 | 0.4010 | 0.0599 | 0.0110 | 0.0000 | | | |
| South America | | | | | | | | | |
| Oceania | | | | | | | | | |
| Other countries | | | | | | | | | |
| *Variable specifications* | | | | | | | | | |
| Earnings | 0.0015 | 0.0058 | 0.0918 | 0.0121 | 0.0128 | 0.3445 | | | |
| Hours worked *(ref.var)* | | | | | | | | | |
| Participation | -0.0008 | 0.0051 | 0.0505 | 0.0113 | 0.0139 | 0.4162 | | | |
| Experience | 0.0012 | 0.0078 | 0.0484 | 0.0437 | 0.0264 | 0.0979 | | | |
| One child *(ref.var)* | | | | | | | | | |
| 2+ children | 0.0010 | 0.0047 | 0.0685 | 0.0193 | 0.0129 | 0.1346 | | | |
| Number of children | -0.0377 | 0.0157 | 0.9081 | -0.0474 | 0.0199 | 0.0172 | -0.0443 | 0.0109 | 0.0001 |
| Presence of children | -0.0003 | 0.0026 | 0.0424 | 0.0165 | 0.0145 | 0.2551 | | | |
| Motherhood years | -0.0015 | 0.0090 | 0.0502 | 0.0116 | 0.0209 | 0.5789 | | | |
| *Set of controls* | | | | | | | | | |
| Age | 0.0012 | 0.0052 | 0.0827 | -0.0001 | 0.0123 | 0.9935 | | | |
| Education | -0.0001 | 0.0024 | 0.0424 | 0.0157 | 0.0148 | 0.2888 | | | |
| Marriage | -0.0011 | 0.0041 | 0.0991 | -0.0165 | 0.0100 | 0.0989 | | | |
| Race | -0.0123 | 0.0205 | 0.3372 | -0.0404 | 0.0146 | 0.0057 | | | |
| Sector | -0.0001 | 0.0025 | 0.0270 | -0.0037 | 0.0199 | 0.8525 | | | |
| Schedule | 0.0607 | 0.0159 | 0.9921 | 0.0736 | 0.0173 | 0.0000 | 0.0522 | 0.0145 | 0.0004 |
| Location | 0.0597 | 0.0144 | 0.9973 | 0.0552 | 0.0183 | 0.0026 | 0.0788 | 0.0255 | 0.0023 |

Note: The table presents the results of BMA, FMA, and Frequentist Check for the male sample, containing 209 observations from 27 individual studies.

## 5.4 Robustness check

We perform a robustness check on two distinct subsets from the dataset defined in Chapter 4. For the sake of consistency, we employ the same methodological approach used for the entire sample. Thus, we present the results of the baseline model (UIP and the dilution prior on the weighted coefficients) along with the results from the model extensions - BRIC and Random, HQ and Random. Due to the lack of estimates in the male sample, we present the robustness check for the female sample only. Due to collinearity issues, we excluded the variable *Earnings*, in addition to the reference variables in both subsamples.

The first subsample includes studies with *Earnings* as the labour market outcome variable. The results of the baseline model, presented in Figure C.8, confirm the findings obtained for the full female sample. Similar to the main results, studies employing *more than two children* as a family variable and studies applying *OLS* methodology report more negative results, while studies published in peer-reviewed journals (studies with higher *impact factor*) and those applying *IV and 2SLS* report more positive results. The findings regarding publication bias are also consistent with the results for the full female sample - the *standard error* of PCCs has a low PIP value, indicating no clear evidence of publication bias. The results of BMA estimation across different priors in Figure C.9 do not present significant differences across model specifications.

Subsequently, we assess the robustness of our findings using the second subsample, which consists exclusively of results that account for *endogeneity*. The final baseline model results are presented in Figure C.10, and estimations across different priors are reported in Figure C.11. We can conclude that the outcomes are largely consistent across different priors, showing no significant variation. Nonetheless, the HQ and Random model specifications yield marginally higher PIP values. The findings for *Impact factor* and *IV and 2SLS* variables correspond to the results of the previous subset and those obtained in the main body of the research. The elevated PIP values associated with these variables suggest they significantly contribute to explaining the variability within the subsample and are linked to a more positive effect. Moreover, studies using *Participation* to determine mothers' labour market outcomes and studies with higher number of *adjusted citations* produce more negative estimates. On the other hand, controlling for *Race* and implementing *Binary Outcome Models* result in more positive results. In line with the results obtained from FMA, the visualization presents a negative *standard error* with high PIP values, implying negative publication bias.

In conclusion, our analysis of the entire female sample indicates an absence of publication bias, based on the results from the main model specification of BMA. When examining the sample of primary studies focusing on earnings as a labour market outcome, the standard error of PCCs remains its low PIP value, corroborating the initial findings. However, the robustness check based on the subsample of studies that control for endogeneity reveals a high PIP value of standard error (which goes in line with the presence of publication selectivity reported in the previous meta-analysis by de Linde Leonard & Stanley (2020) after controlling for the heterogeneity-related variables). These varying outcomes suggest that the detection of publication bias might be sensitive to specific methodological approaches within the literature.

The results for the male sample remain consistent across our analysis, indicating a positive publication bias in primary studies even after controlling for heterogeneity.

# Chapter 6

# Best practice estimate

In this section, we will follow the approach proposed by Havranek *et al.* (2020) to construct the *best practice estimate*. This entails examining the mean effect of family size on parents' labour market outcomes while controlling for study characteristics. Specifically, our objective is to generate the estimates adjusted for (i) the underrepresentation of positive estimates for the female sample and negative estimates for the male sample, and (ii) various study determinants.

However, it is necessary to highlight that this exercise is experimental and the results inherently contain a degree of subjectivity, as the assessment of the best practice estimate relies on the author's judgment. It is also pertinent to acknowledge that the application of PCCs lacks explicit economic meaning.

In our view, the best practice estimate should adhere to the following conditions. First, to mitigate publication selection bias, We assign the coefficient of the *Standard error* to its minimum observed value within the sample. Furthermore, we prioritize the utilization of panel data over cross-sectional data due to its enhanced capability in capturing dynamic temporal changes. Moreover, we use the sample maximum for the variables *Published* and *Impact factors*, as we believe these measures enhance the credibility of the estimates and reflect the higher standards of academic rigor. Regarding the control variables, we set their values to 1, regardless of the PIP results in the main body of the research. Finally, we favour advanced econometric techniques over simplistic OLS and undefined methodologies. Consequently, we set the coefficients of *OLS* and *Other methodologies* at their sample minima. We plug the sample means for the remaining variables.

Upon the determination of preferred values for each variable within the dataset, we performed a linear regression analysis on both the female and male samples. The best practice estimates, given the specified preferences, are detailed in Table 6.1 along with their respective 95% confidence intervals. The best practice estimate of the partial correlation coefficient is 0.0513 for the female sample, and 0.0305 for the male sample. Despite the positive estimate for the female sample, it is close to 0 and thus deemed small. These findings indicate that the impact of family size on labour market outcomes is relatively small for both genders within the context of our specified model parameters.

Table 6.1: Best practice estimates for female and male samples

| | Predicted Estimate | 95% Confidence Interval |
|---|---|---|
| Female sample | 0.0513 | [0.0253, 0.0772] |
| Male sample | 0.0305 | [-0.0376, 0.0986] |

Note: The table presents the results of the subjective best practice estimate of the PCC corresponding to the estimated effects of family size on labor market outcomes. The results are clustered at the study level.

# Chapter 7

# Conclusion

This thesis aimed to conduct a quantitative review of studies investigating the impact of family size on parents' labour market outcomes. The contribution of our thesis is twofold. First, although three meta-analyses on a similar topic existed as at the date of this study, this is the first to conduct a quantitative review specifically on the male sample. Furthermore, since none of the previous studies published their datasets for replication, we collected an original dataset with no restrictions on the effect estimates. Moreover, we enhanced our results by modern techniques used in meta-analyses and several robustness checks.

Our analysis was based on 1323 observations from 88 individual studies for the female sample, and 209 observations from 27 individual studies for the male sample. To avoid confining the dataset to specific definitions of family size and labour market outcome variables, we recalculated the effect sizes into Partial Correlation Coefficients (PCCs). Additionally, we performed the robustness checks on two subsamples of the female sample: one focusing on earnings as the labour outcome variable and the other containing only estimates from models that controlled for endogeneity.

In the first part of our analysis, we assessed publication bias through several methods, including the visual test (Funnel plot), Funnel and Precision Asymmetry tests (FAT-PAT), tests relaxing the exogeneity assumption (Instrumental Variables) and non-linear tests (such as Top10, Stem, Endogenous kink, Weighted Average of Adequately Powered and Selection model). Moreover, we conducted the Caliper test using the collected t-statistics. The results of this analysis, along with robustness checks, indicate the presence of negative publication bias in the female sample. It is worth mentioning that only a portion of the tests yielded statistically significant results, with estimates ranging from -0.683 to -0.41. Based on the statistically significant results from the said models, the mean beyond bias for the female sample ranges from -0.033 to -0.023, indicating a small effect.

For the male sample, the results were contrary: publication bias estimates ranged from 1.332 to 2.304, and the mean beyond bias ranged from -0.005 to -0.004. This suggests that after correcting for publication bias, the direction of the effect may change, although the effect size remains negligible.

To examine the variations in PCCs driven by study characteristics, we extended the model by incorporating 32 variables for the female sample and 28 variables for the male sample that potentially contribute to heterogeneity. Subsequently, we employed Bayesian Model Averaging (BMA), Frequentist Model Averaging (FMA), and Frequentist Check to investigate the impact of individual study characteristics. For the female sample, the BMA results yielded a low Posterior Inclusion Probability (PIP) for the standard error of PCC, indicating that after controlling for additional study characteristics, there is no clear evidence of publication bias. The BMA model further highlighted the significance of the average year of the dataset, impact factor of the study, certain methodologies (such as OLS and IV and 2SLS), the number of explanatory variables, the specification of having more than two children, and controlling for age in explaining the variation of PCCs in the female sample. The robustness check results from the first subsample, which focuses on earnings as a labour market outcome variable, corroborated the initial findings of diminishing publication bias with the inclusion of additional variables. However, the results from the second subcsample, which controls for Endogeneity, revealed the presence of negative publication bias.

The results of the male sample remained consistent throughout the analysis, with the standard error of PCCs showing a PIP value of 1 in the BMA model. This outcome indicates that, after controlling for characteristics of the primary studies, publication selectivity persisted. Besides the standard error, only three additional variables achieved PIP values higher than 0.5: the number of children (as the variable specification for family size), and controlling for schedule and location.

Finally, we constructed the best practice estimate by subjectively assigning preferred values to the variables of interest. The resulting mean effects for both samples are positive, but near zero.

While the analysis of the female sample was supported by several robustness checks, the results for the male sample lacks similar robustness. Consequently, we suggest that future research should focus on the male sample separately, even though this may be challenging due to the limited number of primary studies examining the effect on males. Additionally, we recommend analyzing the reported estimates directly, rather than relying on PCCs. In this thesis, we utilised PCCs to achieve comparability among different effect estimates, which imposes constraints on the further interpretation of our results. However, given the diversity of measures used for both dependent and independent variables, we opted not to restrict the analysis to certain interpretations and continue with PCCs.

# Bibliography


Aaronson, D., R. Dehejia, A. Jordan, C. Pop-Eleches, C. Samii, K. Schulze, & B. A. Weinberg (2021): "The effect of fertility on mothers' labor supply over the last two centuries." *Economic Journal* **131(633)**: pp. 1–32.

Agüero, J. M., M. Marks, & N. Raykar (2020): "Economic development and the motherhood wage penalty." *University of Connecticut, Department of Economics, Working Paper 2020-06* .

Agüero, J. M. & M. S. Marks (2008): "Motherhood and female labor force participation: Evidence from infertility shocks." *American Economic Review* **98(2)**: pp. 500–504.

Agüero, J. M. & M. S. Marks (2011): "Motherhood and female labor supply in the developing world: Evidence from infertility shocks." *Journal of Human Resources* **46(4)**: pp. 800–826.

Ahn, N. & P. Mira (2002): "A note on the changing relationship between fertility and female employment rates in developed countries." *Journal of Population Economics* **15(4)**: pp. 667–682.

Aisenbrey, S., M. Evertsson, & D. Grunow (2009): "Is there a career penalty for mothers' time out? a comparison of germany, sweden, and the united states." *Social Forces* **88(2)**: pp. 573–605.

Albrecht, J. W., P. A. Edin, M. Sundström, & S. B. Vroman (1999): "Career interruptions and subsequent earnings: A reexamination using Swedish data." *Journal of Human Resources* **34**: pp. 294–311.

Altintas, E. & O. Sullivan (2017): "Trends in fathers' contribution to housework and childcare under different welfare policy regimes." *Social Politics* **24(1)**: pp. 81–108.

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

Andersson, G., K. Hank, & A. Vikat (2007): "Understanding parental gender preferences in advanced societies: Lessons from sweden and finland." *Demographic Research* **17**: pp. 135–156.

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

Angrist, J. & W. N. Evans (1996a): "Children and their parents' labor supply: Evidence from exogenous variation in family size."

Angrist, J. & W. N. Evans (1996b): "Schooling and labor market consequences of the 1970 state abortion reforms." *Technical report*, National Bureau of Economic Research Cambridge, Mass., USA.

Angrist, J. & A. Schlosser (2010): "Multiple experiments for the causal link between the quantity and quality of children." *Journal of Labor Economics* **28(4)**.

Avellar, S. & P. J. Smock (2003): "Has the price of motherhood declined over time? a cross-cohort comparison of the motherhood wage penalty." *Journal of Marriage and Family* **65(3)**: pp. 597–607.

Bajzik, J., T. Havranek, Z. Irsova, & J. Novak (2023): "Does shareholder activism create value? a meta-analysis." .

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

Baranowska-Rataj, A. & A. Matysiak (2022): "Family size and men's labor market outcomes: Do social beliefs about men's roles in the family matter?" *Feminist Economics* **28(2)**: pp. 93–118.

Bari, L. (2023): "Gendered divergence in the impact of parenthood on wages: The role of family size, human capital and working time." *Journal of Family and Economic Issues* pp. 1–16.

Becker, G. S. (1992): "A treatise on the family." *Population and Development Review* **18(3)**: pp. 53–84.

Bell, C. S. (1974): "Why participation rates of black and white wives differ." *The Journal of Human Resources* **9(4)**: pp. 465–479.

Bernhardt, E. M. (1993): "Fertility and employment." *European Sociological Review* **9(1)**: pp. 25–42.

Bielby, W. T. & D. D. Bielby (1989): "Family ties: Balancing commitments to work and family in dual earner households." *American Sociological Review* **54(5)**: pp. 776–789.

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

Brewster, K. L. & R. R. Rindfuss (2000): "Fertility and women's employment in industrialized nations." *Annual Review of Sociology* **26(1)**: pp. 271–296.

Bronars, S. G. & J. Grogger (1994): "The economic consequences of unwed motherhood: Using twin births as a natural experiment." *American Economic Review* **84(5)**: pp. 1141–1156.

Browning, M. (1992): "Children and household economic behavior." *Journal of Economic Literature* **30(3)**: pp. 1434–1475.

Bryan, M. L. & A. Sevilla-Sanz (2011): "Does housework lower wages? Evidence for Britain." *Oxford Economic Papers* **63**: pp. 187–210.

Buchmann, C. & A. McDaniel (2016): "Motherhood and earnings in African countries." *Demographic Research* **34**: pp. 591–620.

Budig, M. J. & P. England (2001): "The wage penalty for motherhood." *American sociological review* **66(2)**: pp. 204–225.

Budig, M. J. & M. J. Hodges (2010): "Differences in disadvantage: Variation in the motherhood penalty across white women's earnings distribution." *American Sociological Review* **75(5)**: pp. 705–728.

Bundesamt, S. (2021): "Geburten." *Available at: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Geburten/$_i$nhalt.html*.

Cáceres-Delpiano, J. (2012): "Can we still learn something from the relationship between fertility and mother's employment? evidence from developing countries." *Demography* **49(1)**: pp. 151–174.

Carlin, P. S. & L. Flood (1997): "Do children affect the labor supply of swedish men? time diary vs. survey data." *Labour Economics* **4(2)**: pp. 167–183.

Cazachevici, A., H. T. & R. Horvath (2020): "Remittances and economic growth: A meta-analysis." *World Development* **134**: p. 105021.

Chevalier, A. & T. K. Viitanen (2003): "The long-run labour market consequences of teenage motherhood in britain." *Journal of Population Economics* **16(2)**: pp. 323–343.

Choi, H., J. M. Joesch, & S. Lundberg (2008): "Sons, daughters, wives, and the labour market outcomes of west german men." *Labour Economics* **15(5)**: pp. 795–811.

Chun, H. & J. Oh (2002): "An instrumental variable estimate of the effect of fertility on the labour force participation of married women." *Applied Economics Letters* **9(10)**: pp. 631–634.

Clarke, D. (2018): "Children and parents' labor market outcomes." *Demography* **55**: pp. 1009–1032.

Cools, S., S. Markussen, & M. Strøm (2017): "Children and careers: How family size affects parents' labor market outcomes in the long run." *Demography* **54(5)**: pp. 1501–1523.

Correll, S. J., S. Benard, & I. Paik (2007): "Getting a job: Is there a motherhood penalty?" *American Journal of Sociology* **112(5)**: pp. 1297–1338.

Cristia, J. P. (2008): "The effect of a first child on female labor supply: Evidence from Latin America." *Journal of Human Resources* **43**: pp. 166–196.

Cukrowska-Torzewska, E. (2015): "She cares and he earns? the family gap in poland." *Ekonomia. Rynek, Gospodarka, Społeczeństwo* **(42)**: pp. 43–77.

CUKROWSKA-TORZEWSKA, E. & A. MATYSIAK (2020): "The motherhood wage penalty: A meta-analysis." *Social Science Research* **87**: p. 102377.

CÁCERES-DELPIANO, J. (2006): "The impacts of family size on investment in child quality." *Journal of Human Resources* **41(4)**.

CÁCERES-DELPIANO, J. (2008): "Keeping the best for last. impact of fertility on mother's employment. evidence from developing countries." *Economic Series Working Papers* **32(68)**: pp. 23–47.

DEL BOCA, D., S. PASQUA, & C. PRONZATO (2003): "Labour market participation of women and fertility: the effect of social policies." *FRDB Child Conference* pp. 1–37.

DESAI, S. & L. J. WAITE (1991): "Women's employment during pregnancy and after the first birth: occupational characteristics and work commitment." *American Sociological Review* **56(4)**: pp. 551–566.

DOUCOULIAGOS, C. (2011): "How large is large? preliminary and relative guidelines for interpreting partial correlations in economics." .

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

DOUCOULIAGOS, H. (2005): "Publication bias in the economic freedom and economic growth literature." *Journal of Economic Surveys* **19**: pp. 367–387.

EGGER, M. *et al.* (1997): "Bias in meta-analysis detected by a simple, graphical test." *Bmj* **315(7109)**: pp. 629–634.

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

ELMINEJAD, A., T. HAVRANEK, & Z. IRSOVA (2022): "Relative risk aversion: a meta-analysis." *Kiel, Hamburg: ZBW-Leibniz Information Centre for Economics, Working Paper* .

ENGELHARDT, H., T. KÖGEL, & A. PRSKAWETZ (2004): "Fertility and women's employment reconsidered: A macro-level time-series analysis for developed countries, 1960-2000." *Population Studies* **58(1)**: pp. 109–120.

European Parliamentary Research Service (2022): "Maternity and paternity leave in the eu." `https://www.europarl.europa.eu/thinktank`. Authors: Ülla Jurviste, Marie Lecerf; Graphics: Samy Chahri.

Fleisher, B. M. & G. F. Rhodes (1979): "Fertility, women's wage rates, and labor supply." *American Economic Review* **69(1)**: pp. 14–24.

Floge, L. (1989): "Changing household structure, child-care availability, and employment among mothers of preschool children." *Journal of Marriage and the Family* **51(1)**: pp. 177–190.

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

Gangl, M. & A. Ziefle (2009): "Motherhood, labor force behavior, and women's careers: An empirical assessment of the wage penalty for motherhood in britain, germany, and the united states." *Demography* **46(2)**: pp. 341–369.

Gash, V. (2009): "Sacrificing their careers for their families? an analysis of the penalty to motherhood in europe." *Social indicators research* **93**: pp. 569–586.

Gechert, S. *et al.* (2019): "Death to the cobb-douglas production function." *Technical report*, IMK working paper.

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

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.

Glauber, R. (2008): "Race and gender in families and at work: The fatherhood wage premium." *Gender & Society* **22(1)**: pp. 8–30.

Hamilton, B. E., J. A. Martin, & M. J. Osterman (2024): "Births: Provisional data for 2023." .

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

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

Havranek, T., R. Horvath, & A. Zeynalov (2016): "Natural resources and economic growth: A meta-analysis." *World Development* **88**: pp. 134–151.

Havranek, T., Z. Irsova, L. Laslopova, & O. Zeynalova (2020): "The elasticity of substitution between skilled and unskilled labor: A meta-analysis." .

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

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

Havranek, T. *et al.* (2015): "Cross-country heterogeneity in intertemporal substitution." *Journal of International Economics* **96(1)**: pp. 100–118.

Havranek, T. *et al.* (2022): "Publication and attenuation biases in measuring skill substitution." *Review of Economics and Statistics* pp. 1–37.

Hersch, J. (1991): "Male-female differences in hourly wages: The role of human capital, working conditions, and housework." *ILR Review* **44(4)**: pp. 746–759.

Hirvonen, L. (2010): "The effect of children on earnings using exogenous variation in family size: Swedish evidence." *Technical Report 2*, Swedish Institute for Social Science Research.

Hofferth, S. L. & K. A. Moore (1979): "Early childbearing and later economic well-being." *American Sociological Review* **44(5)**: pp. 784–815.

Hotz, V. J., S. W. McElroy, & S. G. Sanders (2005): "Teenage childbearing and its life cycle consequences: Exploiting a natural experiment." *Journal of Human Resources* **40**: pp. 683–715.

Ioannidis, J. P. A., T. D. Stanley, & C. Doucouliagos (2017): *The power of bias in economics research.* Oxford University Press Oxford, UK.

Irsova, Z., H. Doucouliagos, T. Havranek, & T. Stanley (2023): "Meta-analysis of social science research: A practitioner's guide." *Journal of Economic Surveys* .

Jacobsen, J. P., J. W. Pearce III, & J. L. Rosenbloom (1999): "The effects of childbearing on married women's labor supply and earnings: using twin births as a natural experiment." *Journal of Human Resources* pp. 449–474.

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

Kaufman, G. & P. Uhlenberg (2000): "The influence of parenthood on the work effort of married men and women." *Social Forces* **78(3)**: pp. 931–947.

Kessler-Harris, A. (2003): "Out to work: A history of wage-earning women in the united states." *Oxford University Press* .

Kim, J. (2006): "Women's wages in the Korean labor market: The role of family structure and lifecycle events." *Review of Economics of the Household* **4**: pp. 357–377.

Kroupova, K., T. Havranek, & Z. Irsova (2024): "Student employment and education: A meta-analysis." *Economics of Education Review* **100**: p. 102539.

Kroupová, K. (2021): "The impact of student employment on educational outcomes: A meta-analysis." *Technical report.*

Lang, K. (2023a): "Estimating the effect of family size on parental labor market outcomes." *Journal of Labor Economics* **41**: pp. 421–450.

Lang, K. (2023b): "How credible is the credibility revolution?" *Technical report*, National Bureau of Economic Research.

Lehrer, E. L. (1992): "The impact of children on married women's labor supply: Black-white differentials revisited." *Journal of Human Resources* **27(3)**: pp. 422–444.

Levy, F. & R. J. Murnane (1992): "U.S. earnings levels and earnings inequality: A review of recent trends and proposed explanations." *Journal of economic literature* **30(3)**: pp. 1333–1381.

de Linde Leonard, M. & T. D. Stanley (2020): "The productivity-wage premium: A meta-analysis." *Journal of Economic Surveys* **34**: pp. 24–52.

Lowen, A. & P. Sicilian (2009): ""Family-friendly" fringe benefits and the gender wage gap." *Journal of Labor Research* **30**: pp. 101–119.

Lundberg, S. (2005): "Sons, daughters, and parental behaviour." *Oxford Review of Economic Policy* **21(3)**: pp. 340–356.

Lundberg, S. & E. Rose (2000): "Parenthood and the earnings of married men and women." *Labour Economics* **7(6)**: pp. 689–710.

Lundberg, S. & E. Rose (2002): "The effects of sons and daughters on men's labor supply and wages." *Review of Economics and Statistics* **84(2)**: pp. 251–268.

Lundborg, P., E. Plug, & A. W. Rasmussen (2021): "Fertility effects on female labor supply: Iv evidence from ivf treatments." *American Economic Review* **111(6)**: pp. 2023–2063.

Mason, K. O. & V. T. Palan (1981): "Female employment and fertility in peninsular malaysia: The maternal role incompatibility hypothesis reconsidered." *Demography* **18(4)**: pp. 549–575.

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

Matysiak, A. & D. Vignoli (2008): "Fertility and women's employment: A meta-analysis." *European Journal of Population* **24**: pp. 363–384.

Meyers, M. K. & J. C. Gornick (2005): "Policies for reconciling parenthood and employment." *Challenge (05775132)* **48(5)**: pp. 39–61.

Mincer, J. & S. Polachek (1974): "Family investments in human capital: Earnings of women." *Journal of Political Economy* **82**: pp. S76–S108.

Molina, J. A. & V. M. Montuenga-Gomez (2009): "The motherhood wage penalty in Spain." *Journal of Family and Economic Issues* **30**: pp. 237–251.

Pacelli, L., S. Pasqua, & C. Villosio (2013): "Employment decisions of women and the part-time wage differential in Europe." *Economic Policy* **28**: pp. 123–169.

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

Robey, B., S. O. Rutstein, & L. Morris (1993): "The fertility decline in developing countries." *Scientific American* **269(6)**: pp. 60–67.

Rosenthal, R. (1979): "The file drawer problem and tolerance for null results." *Psychological bulletin* **86(3)**: p. 638.

Rosenzweig, M. R. & K. I. Wolpin (1985): "Specific experience, household structure, and intergenerational transfers: Farm family land and labor arrangements in developing countries." *Quarterly Journal of Economics* **100**: pp. 961–987.

Ruhm, C. J. (1998): "The economic consequences of parental leave mandates: Lessons from europe." *Quarterly Journal of Economics* **113(1)**: pp. 285–317.

Shapiro, D. & F. L. Mott (1979): "Labor supply behavior of prospective and new mothers." *Demography* **16(2)**: pp. 199–210.

Sleebos, J. (2003): "Low fertility rates in oecd countries: facts and policy responses." *OECD Labour Market and Social Policy Occasional Papers* **15**: pp. 1–53.

Sobotka, T. & C. Berghammer (2021): "Demography of family change in europe." *Research Handbook on the Sociology of the Family* **18(3)**: pp. 17–38.

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

Stanley, T. D., C. Doucouliagos, & J. P. A. Ioannidis (2017): "Finding the power to reduce publication bias." *Statistics in medicine* **36(10)**: pp. 1580–1598.

Stanley, T. D. & H. Doucouliagos (2012): "Meta-regression analysis in economics and business." *Journal of Economic Surveys* **26**: pp. 77–102.

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

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

Stycos, J. M. & R. H. Weller (1967): "Female working roles and fertility." *Demography* **4(1)**: pp. 210–217.

Sweet, J. A. (1973): "Women in the labor force. studies in population series." *ERIC* .

Trussell, J. (1976): "Economic consequences of teenage childbearing." *Family Planning Perspectives* **8(4)**: pp. 184–190.

Van Winkle, Z. & A. E. Fasang (2020): "Parenthood wage gaps across the life course: A comparison by gender and race." *Journal of Marriage and Family* **82(5)**: pp. 1284–1299.

Waldfogel, J. (1997): "The effect of children on women's wages." *American Sociological Review* **62(2)**: pp. 209–217.

Wenk, D. A. & P. Garrett (1992): "Having a baby: Some predictions of maternal employment around childbirth." *Gender & Society* **6(1)**: pp. 49–65.

Wilner, L. (2016): "Worker-firm matching and the parenthood pay gap: Evidence from linked employer-employee data." *Journal of Population Economics* **29(4)**: pp. 991–1023.

Wooldridge, J. (2013): "Introductory econometrics: a modern approach." *Cengage Learning* .

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

Zigraiova, D. & T. Havranek (2016): "Bank competition and financial stability: Much ado about nothing?" *Journal of Economic Surveys* **30(5)**: pp. 944–981.

Zigraiova, D., T. Havranek, Z. Irsova, & J. Novak (2021): "How puzzling is the forward premium puzzle? a meta-analysis." *European Economic Review* **134**: p. 103714.

# Appendix A

## Additional information for Data description

Table A.1: Primary studies used in the meta-analysis: sample of female respondents

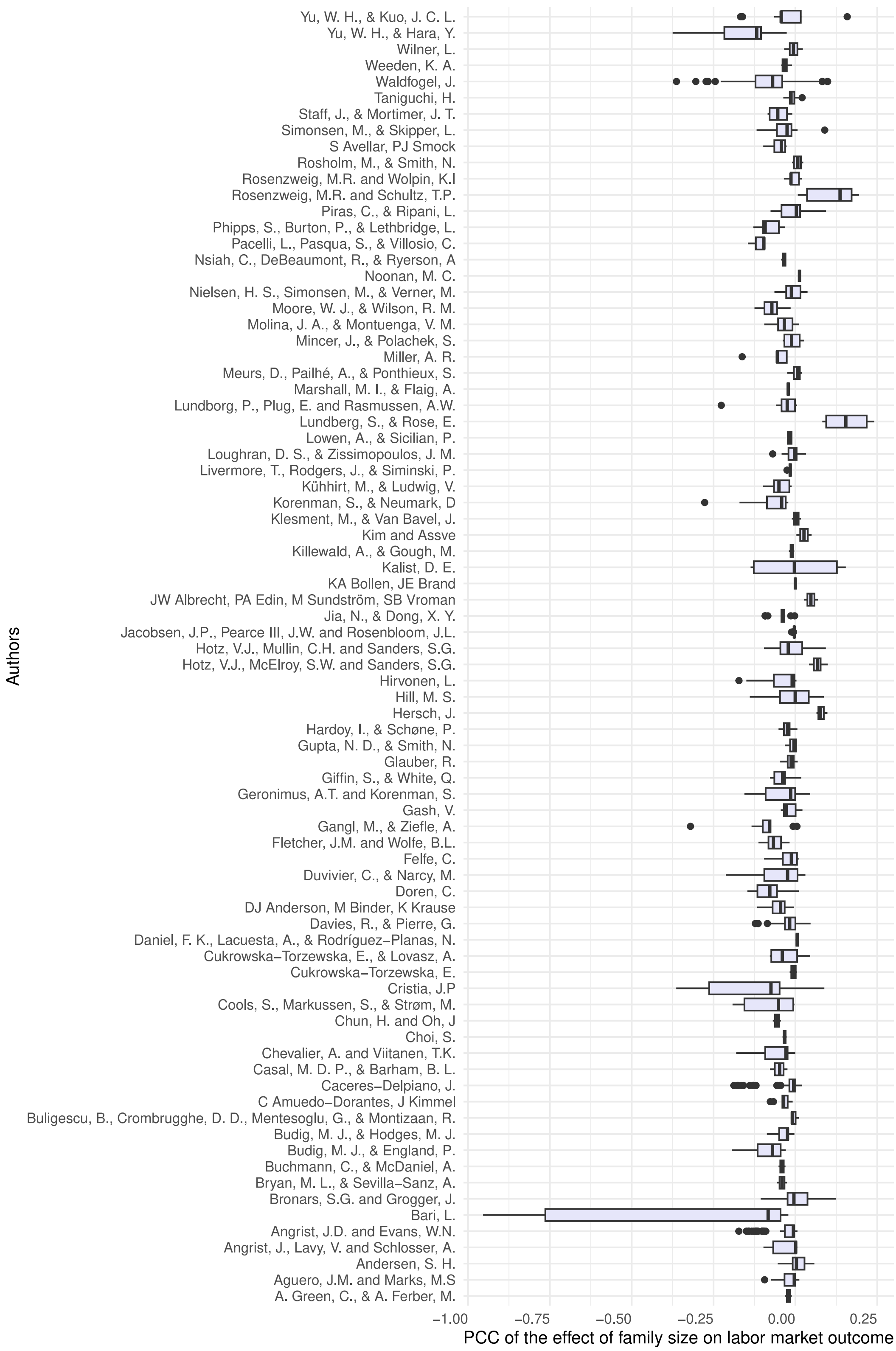
| | |
|---|---|
| 1. Green, C.A. & Ferber, M.A. (2008) | 46. Hotz, V.J., Mullin, C.H. & Sanders, S.G. (1997) |
| 2. Aguero, J.M. & Marks, M.S. (2008) | 47. Jacobsen, J.P., Pearce III, J.W. & Rosenbloom, J.L. (1999) |
| 3. Aguero, J.M. & Marks, M.S. (2011) | 48. Jia, N., & Dong, X.Y. (2013) |
| 4. Andersen, S.H. (2018) | 49. Albrecht, J.W., Edin, P.A., Sundström, M. & Vroman, S.B. (1999) |
| 5. Angrist, J., Lavy, V. & Schlosser, A. (2010) | 50. Bollen, K.A. & Brand, J.E. (2010) |
| 6. Angrist, J.D. & Evans, W.N. (1996) | 51. Kalist, D.E. (2008) |
| 7. Angrist, J.D. & Evans, W.N. (1996) | 52. Killewald, A. & Gough, M. (2013) |
| 8. Bari, L. (2023) | 53. Kim & Assve (2006) |
| 9. Bronars, S.G. & Grogger, J. (1994) | 54. Klesment, M. & Van Bavel, J. (2017) |
| 10. Bryan, M.L. & Sevilla-Sanz, A. (2011) | 55. Korenman, S. & Neumark, D. (1992) |
| 11. Buchmann, C. & McDaniel, A. (2016) | 56. Kühhirt, M. & Ludwig, V. (2012) |
| 12. Budig, M.J. & England, P. (2001) | 57. Livermore, T., Rodgers, J. & Siminski, P. (2011) |
| 13. Budig, M.J. & Hodges, M.J. (2010) | 58. Loughran, D.S. & Zissimopoulos, J.M. (2009) |
| 14. Buligescu, B., Crombrugghe, D.D., Menteşoğlu, G. & Montizaan, R. (2008) | 59. Lowen, A., & Sicilian, P. (2009) |
| 15. Amuedo-Dorantes, C. & Kimmel, J. (2005) | 60. Lundberg, S. & Rose, E. (2000) |
| 16. Caceres-Delpiano, J. (2006) | 61. Lundborg, P., Plug, E. & Rasmussen, A.W. (2014) |
| 17. Caceres-Delpiano, J. (2008) | 62. Marshall, M.I. & Flaig, A. (2014) |
| 18. Caceres-Delpiano, J. (2012) | 63. Meurs, D., Pailhé, A. & Ponthieux, S. (2010) |
| 19. Casal, M.D.P. & Barham, B.L. (2013) | 64. Miller, A.R. (2011) |
| 20. Chevalier, A. & Viitanen, T.K. (2003) | 65. Mincer, J. & Polachek, S. (1974) |
| 21. Choi, S. (2011) | 66. Molina, J.A. & Montuenga, V.M. (2009) |
| 22. Chun, H. & Oh, J (2002) | 67. Moore, W. J. & Wilson, R. M. (1982) |
| 23. Cools, S., Markussen, S. & Strøm, M. (2017) | 68. Nielsen, H. S., Simonsen, M. & Verner, M. (2004) |
| 24. Cristia, J.P (2008) | 69. Noonan, M. C. (2001) |
| 25. Cukrowska-Torzewska, E. (2015) | 70. Nsiah, C., DeBeaumont, R. & Ryerson, A (2013) |
| 26. Cukrowska-Torzewska, E. & Lovasz, A. (2016) | 71. Pacelli, L., Pasqua, S. & Villosio, C. (2013) |
| 27. Daniel, F.K., Lacuesta, A. & Rodríguez-Planas, N. (2013) | 72. Phipps, S., Burton, P. & Lethbridge, L. (2001) |
| 28. Davies, R. & Pierre, G. (2005) | 73. Piras, C. & Ripani, L. (2005) |
| 29. Anderson, D.J., Binder, M. & Krause, K. (2003) | 74. Rosenzweig, M.R. & Schultz, T.P. (1985) |
| 30. Doren, C. (2019) | 75. Rosenzweig, M.R. & Wolpin, K.I (1980) |
| 31. Duvivier, C. & Narcy, M. (2015) | 76. Rosholm, M., & Smith, N. (1996) |
| 32. Felfe, C. (2012) | 77. Avellar, S. & Smock, P.J. (2003) |

| | |
|---|---|
| 33. Fletcher, J.M. & Wolfe, B.L. (2009) | 78. Simonsen, M. & Skipper, L. (2012) |
| 34. Gangl, M. & Ziefle, A. (2009) | 79. Staff, J. & Mortimer, J.T. (2012) |
| 35. Gash, V. (2009) | 80. Taniguchi, H. (1999) |
| 36. Geronimus, A.T. & Korenman, S. (1992) | 81. Waldfogel, J. (1998) |
| 37. Giffin, S. & White, Q. (2008) | 82. Waldfogel, J. (2008) |
| 38. Glauber, R. (2012) | 83. Waldfogel, J. (1995) |
| 39. Glauber, R. (2007) | 84. Waldfogel, J. (1998) |
| 40. Gupta, N. D. & Smith, N. (2002) | 85. Weeden, K. A. (2005) |
| 41. Hardoy, I. & Schøne, P. (2008) | 86. Wilner, L. (2016) |
| 42. Hersch, J. (1991) | 87. Yu, W.H., & Hara, Y. (2021) |
| 43. Hill, M.S. (1979) | 88. Yu, W.H., & Kuo, J.C.L. (2017) |
| 44. Hirvonen, L. (2009) | |
| 45. Hotz, V.J., McElroy, S.W. & Sanders, S.G. (2005) | |

Table A.2: Primary studies used in the meta-analysis: sample of male respondents

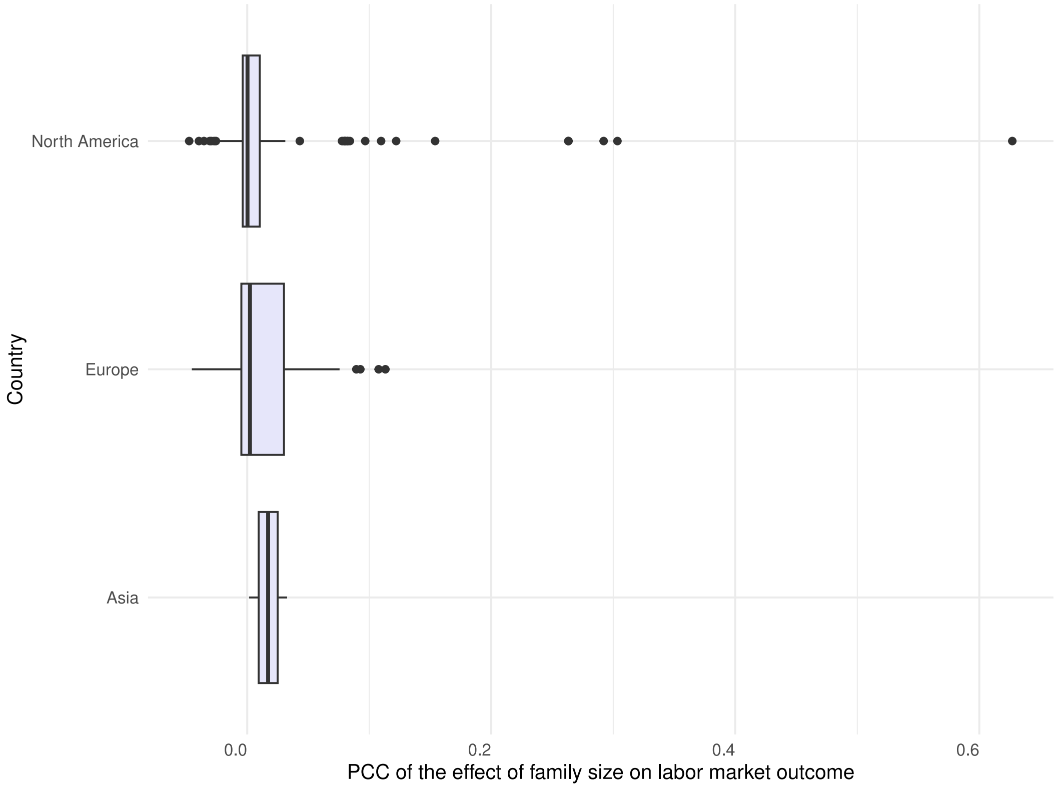
| | |
|---|---|
| 1. Green, C.A. & Ferber, M.A. (2008) | 15. Kim & Assve (2006) |
| 2. Andersen, S.H. (2018) | 16. Loughran, D.S., & Zissimopoulos, J.M. (2009) |
| 3. Angrist, J.D. & Evans, W.N. (1996) | 17. Lowen, A. & Sicilian, P. (2009) |
| 4. Angrist, J.D. & Evans, W.N. (1998) | 18. Lundberg, S. & Rose, E. (2000) |
| 5. Bari, L. (2023) | 19. Meurs, D., Pailhé, A. & Ponthieux, S. (2010) |
| 6. Bryan, M.L., & Sevilla-Sanz, A. (2011) | 20. Noonan, M.C. (2001) |
| 7. Cukrowska-Torzewska, E. (2015) | 21. Phipps, S., Burton, P. & Lethbridge, L. (2001) |
| 8. Cukrowska-Torzewska, E., & Lovasz, A. (2016) | 22. Rosholm, M., & Smith, N. (1996) |
| 9. Gupta, N.D. & Smith, N. (2002) | 23. Simonsen, M. & Skipper, L. (2012) |
| 10. Hersch, J. (1991) | 24. Waldfogel, J. (1998) |
| 11. Hill, M.S. (1979) | 25. Weeden, K.A. (2005) |
| 12. Hirvonen, L. (2009) | 26. Wilner, L. (2016) |
| 13. Albrecht, J.W., Edin, P.A., Sundström, M. & Vroman, S.B. (1999) | 27. Yu, W.H., & Hara, Y. (2021) |
| 14. Killewald, A., & Gough, M. (2013) | |

Figure A.1: Forest plot of PCCs for the female sample across studies



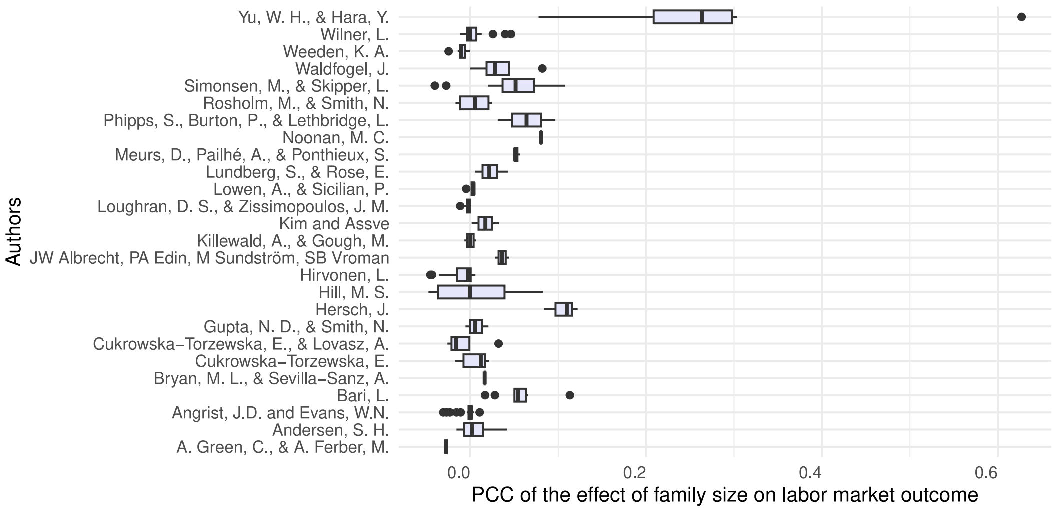

Note: The figure illustrates the forest plot of PCCs across studies for the female sample. The PCCs correspond to the estimated effects of family size on labor market outcomes reported in individual studies for the female sample. The solid vertical line marks the median, boxes present the interquartile range encompassing the 25th to 75th percentiles. Data points that lie beyond the whiskers are considered outliers. The figure includes unwinsorized PCCs.

Figure A.2: Forest plot of PCCs for the male sample across countries




Note: The figure illustrates the forest plot of PCCs across countries for the male sample. The PCCs correspond to the estimated effects of family size on labor market outcomes reported in individual studies for the female sample. The solid vertical line marks the median, boxes present the interquartile range encompassing the 25th to 75th percentiles. Data points that lie beyond the whiskers are considered outliers. The figure includes unwinsorized PCCs.

Figure A.3: Forest plot of PCCs for the male sample across studies




Note: The figure illustrates the forest plot of PCCs across studies for the male sample. The PCCs correspond to the estimated effects of family size on labor market outcomes reported in individual studies for the female sample. The solid vertical line marks the median, boxes present the interquartile range encompassing the 25th to 75th percentiles. Data points that lie beyond the whiskers are considered outliers. The figure includes unwinsorized PCCs.

# Appendix B

## Additional information for Publication bias

Table B.1: Linear Tests' results: female sample, subsample Earnings

| | **OLS** | **Between** | **Weighted by study** | **Weighted by precision** |
|---|---|---|---|---|
| Standard error | -0.171 | -0.350 | 0.189 | -0.720*** |
| *Publication bias* | (0.181) | (0.182) | (0.321) | (0.214) |
| | [-0.944, 0.463] | | [-0.705, 1.100] | [-0.944, 0.463] |
| | | | | |
| Constant | -0.030*** | -0.028*** | -0.030*** | -0.022*** |
| *Mean beyond bias* | (0.003) | (0.003) | (0.004) | (0.003) |
| | [-0.041, -0.020] | | [-0.043, -0.017] | [-0.041, -0.020] |
| Observations | 823 | 823 | 823 | 823 |
| Studies | 81 | 81 | 81 | 81 |

Table B.2: Linear Tests' results: female sample, subsample Endogeneity

| | **OLS** | **Between** | **Weighted by study** | **Weighted by precision** |
|---|---|---|---|---|
| Standard error | -0.453* | -0.490** | -0.564 | -0.439* |
| *Publication bias* | (0.212) | (0.163) | (0.360) | (0.221) |
| | [-1.830, 0.552] | | [-1.840, 0.447] | [-2.080, 0.978] |
| | | | | |
| Constant | 0.015*** | -0.014*** | -0.012** | -0.015*** |
| *Mean beyond bias* | (0.002) | (0.002) | (0.004) | (0.002) |
| | [-0.026, -0.003] | | [-0.025, 0.001] | [-0.033, -0.001] |
| Observations | 757 | 757 | 757 | 757 |
| Studies | 63 | 63 | 63 | 63 |

Table B.3: IV's results: female sample, subsample Earnings

| | $\frac{1}{\text{obs}}$ | $\frac{1}{\text{obs}^2}$ | $\frac{1}{\sqrt{\text{obs}}}$ | $\frac{1}{\log(\text{obs})}$ |
|---|---|---|---|---|
| Standard error | 0.055 | 0.489 | -0.272 | -0.449* |
| *Publication bias* | (0.213) | (0.281) | (0.186) | (0.179) |
| Constant | -0.034*** | -0.041*** | -0.029*** | -0.026*** |
| *Mean beyond bias* | (0.003) | (0.005) | (0.003) | (0.003) |
| Observations | 823 | 823 | 823 | 823 |
| Studies | 81 | 81 | 81 | 81 |

Table B.4: IV's results: female sample, subsample Endogeneity

| | $\frac{1}{\text{obs}}$ | $\frac{1}{\text{obs}^2}$ | $\frac{1}{\sqrt{\text{obs}}}$ | $\frac{1}{\log(\text{obs})}$ |
|---|---|---|---|---|
| Standard error | -0.432 | -0.458 | -0.472* | -0.502* |
| *Publication bias* | (0.258) | (0.338) | (0.215) | (0.201) |
| Constant | -0.015*** | -0.014*** | -0.014*** | -0.014*** |
| *Mean beyond bias* | (0.003) | (0.004) | (0.002) | (0.002) |
| Observations | 757 | 757 | 757 | 757 |
| Studies | 63 | 63 | 63 | 63 |

Table B.5: Non-linear Tests' results: female sample, subsample Earnings

| | **Top 10** | **Stem** | **Endogenous kink** | **WAAP** | **Selection model** |
|---|---|---|---|---|---|
| Effect Beyond Bias | -0.023*** | -0.017 | -0.021 | -0.024*** | -0.002 |
| | (0.004) | (0.009) | (0.002) | (0.004) | (0.005) |
| Observations | 823 | 823 | 823 | 823 | 823 |
| Studies | 81 | 81 | 81 | 81 | 81 |

Table B.6: Non-linear Tests' results: female sample, subsample Endogeneity

| | **Top 10** | **Stem** | **Endogenous kink** | **WAAP** | **Selection model** |
|---|---|---|---|---|---|
| Effect Beyond Bias | -0.031*** | -0.003 | -0.018 | -0.013*** | 0.026 |
| | (0.006) | (0.011) | (0.002) | (0.002) | (0.005) |
| Observations | 757 | 757 | 757 | 757 | 757 |
| Studies | 63 | 63 | 63 | 63 | 63 |

# Appendix C

## Additional information for Heterogeneity

Figure C.4: Correlation matrix between variables of variables capturing heterogeneity: female sample

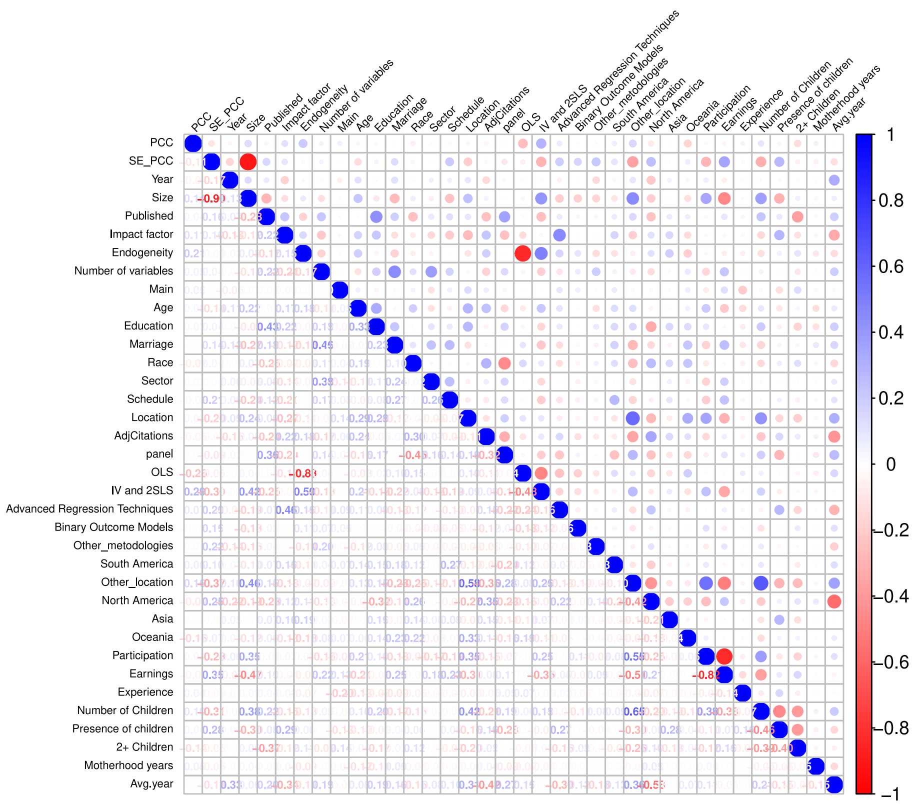


Figure C.5: Correlation matrix between variables of variables capturing heterogeneity: male sample

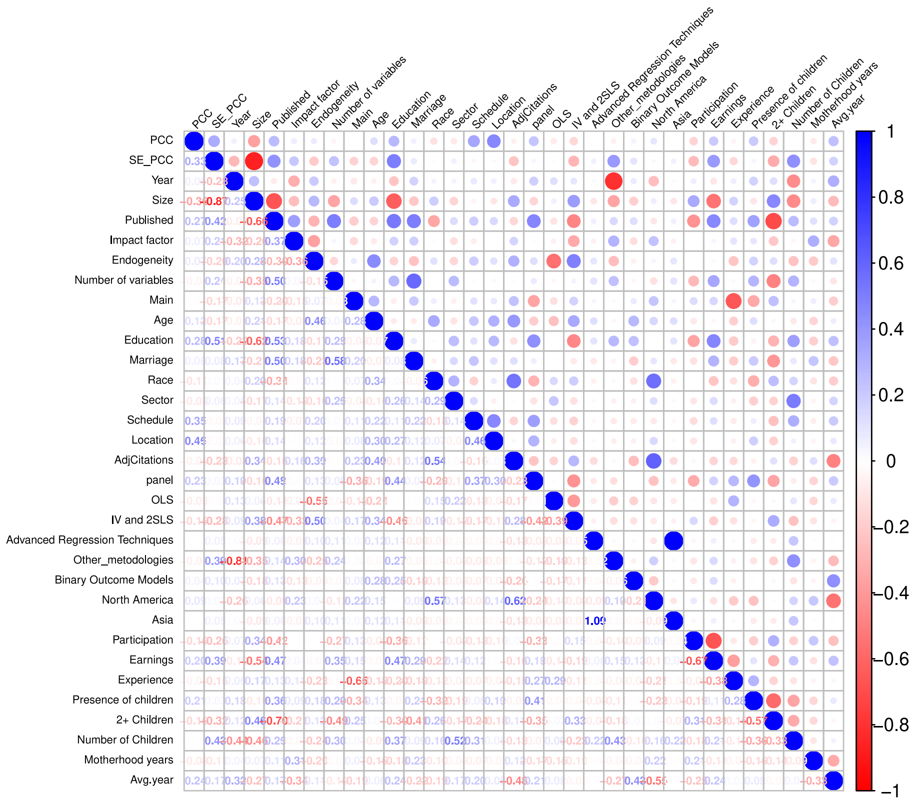


Figure C.6: BMA estimation across different priors: female sample

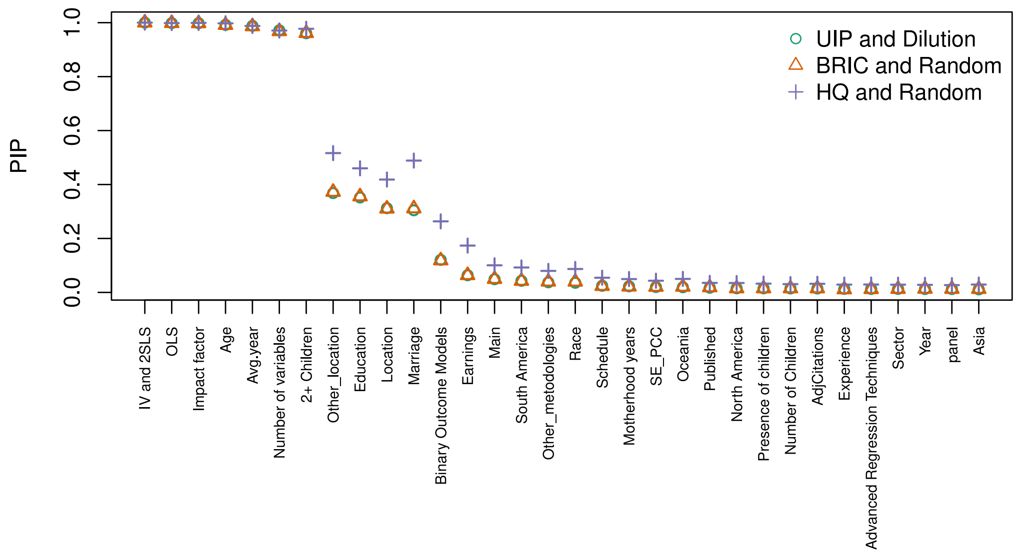



Figure C.7: BMA estimation across different priors: male sample

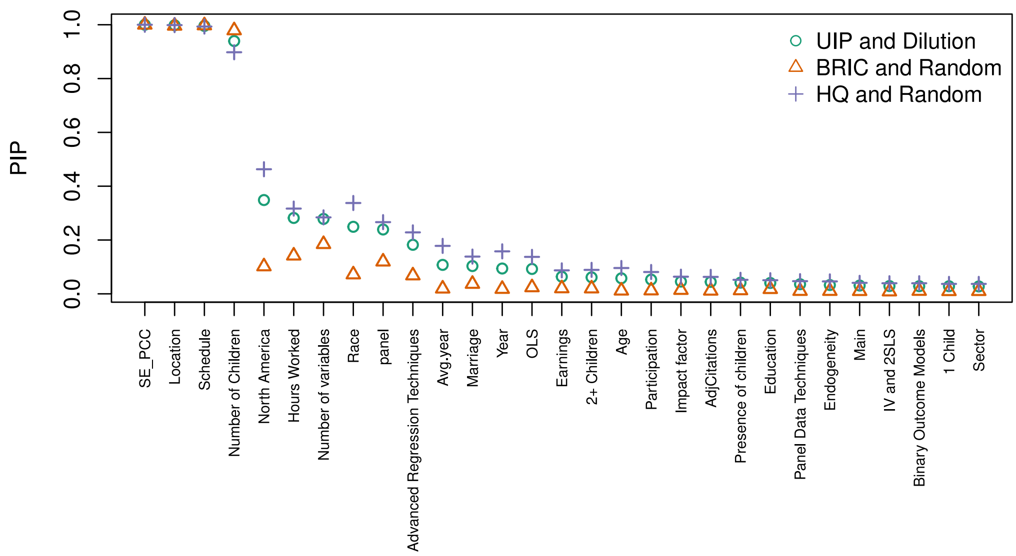


Figure C.8: Model inclusion of the BMA estimation: female sample, subsample Earnings

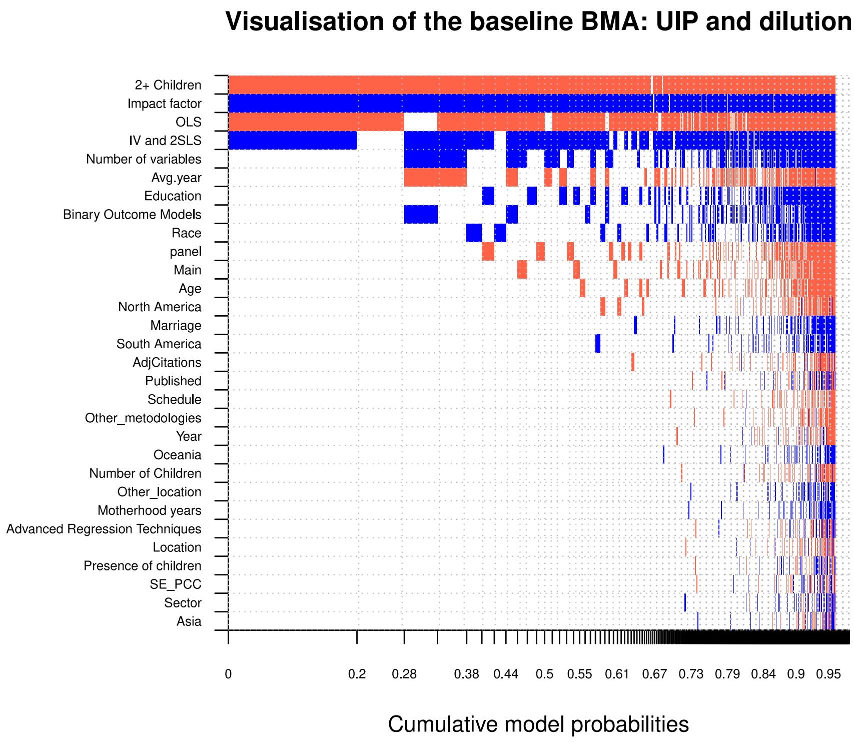



Figure C.9: BMA estimation across different priors: female sample, subsample Earnings

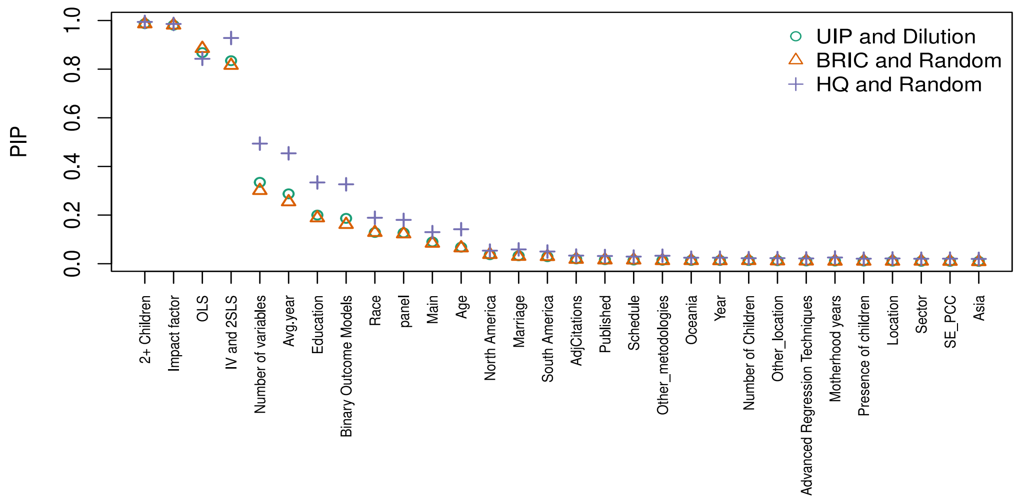


Figure C.10: Model inclusion of the BMA estimation: female sample, subsample Endogeneity

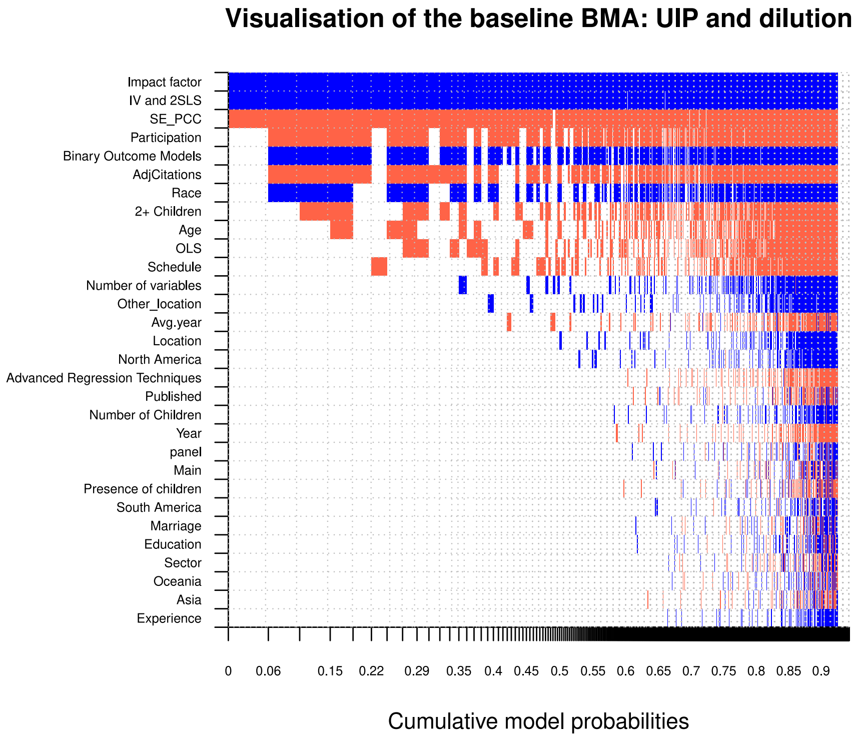



Figure C.11: BMA estimation across different priors: female sample, subsample Endogeneity