# Do robots decrease humans' wages?

**by**

**Thomas Logchies**

**A research thesis forming a part of the requirements for the degree of Master of Commerce**

**at the**

**University of Canterbury**

**August 2024**

**Acknowledgements**

First, I am extremely grateful to my supervisors, Tom Coupé and Bob Reed. I would not have been able to do this without your guidance, enthusiasm, and expertise. Your detailed feedback on my analysis and writing drafts, and your patience with my never-ending questions, have been invaluable throughout my research. Thank you also for your encouragement every step of the way, which gave me the confidence to complete this thesis. I am truly appreciative of your support.

I would also like to acknowledge the research assistants who have contributed to this work. Your effort and collaborative spirit in the coding process have played a significant role in completing this thesis.

A special thanks goes to my mum, dad, and brother. Your continued support gave me the opportunity and motivation to undertake this thesis and complete it to the best of my ability. Your belief in me, coupled with your constant encouragement, kept me in high spirits whenever things got tough. This can also be attributed to my friends and classmates. I am thankful to be surrounded by such a great group of people, even if I sometimes went on about my thesis a bit too much.

I truly appreciate everyone who helped me in any capacity during this journey. Your contributions, whether big or small, were instrumental in the completion of this thesis.

**Abstract**

Robots and how they influence labour market outcomes like wages are a current point of contention within the economic literature. Empirical studies show mixed results, and therefore, using 2,586 estimates from fifty-two studies, I undertake a meta-analysis to study the effects of robots on wages. Overall, while one can find individual studies that show a positive or negative impact of robots on wages, my analysis suggests that when we look at the literature as a whole, there is no clear evidence for a sizeable impact, as most estimates are not statistically significant, and those that are, fail to reach the minimum threshold for a small effect size. I do find evidence that differences in study characteristics lead to differences in estimated effect sizes: how wage is defined and to what extent the data are (dis)aggregated has an influence on the estimates. Finally, I find inconsistent results about whether publication bias exists in this literature.

**Contents**

Acknowledgements ..... ii

Abstract ..... iii

1. Introduction ..... 1

2. Literature review ..... 4

2.1 Literature search and collection ..... 5

2.2 Robots and Wages literature ..... 7

2.2.1 Robot variables ..... 8

2.2.2 Wage variables ..... 11

2.2.3 Data level ..... 17

2.2.4 Data restriction ..... 19

2.2.5 Model specification ..... 26

2.3 Broad automation/technology and wages ..... 31

2.4 Robots and other outcomes ..... 32

2.5 Meta-analyses ..... 34

3. Analysis methods ..... 37

3.1 PCC & Fisher's Z ..... 37

3.2 OLS, fixed effect, random effects & UWLS ..... 39

3.3 Additional non-linear estimators – WAAP, TOP10, Stem & Endogenous Kink 44

3.4 Endogeneity robust estimators – MAIVE, p-uniform* & caliper test ..... 45

3.5 P-hacking robust estimators – RTMA & MAN ..... 46

3.6 Bayesian model averaging ..... 47

4. Results ..... 48

4.1 Data characteristics ..... 49

4.2 Funnel plot ..... 58

4.3 OLS, fixed effect, random effects & UWLS results ..... 59

4.4 Additional non-linear estimators results ..... 64

4.5 Endogeneity robust estimators results ..... 66
4.6 P-hacking robust estimators results ..... 70
4.7 Bayesian model averaging results ..... 71
5. Conclusion ..... 78
References ..... 80

## List of tables

Table 1: Data characteristics descriptions ..... 49
Table 2: Collected estimates summary statistics ..... 54
Table 3: OLS, FE, RE & UWLS results ..... 60
Table 4: Highest weighted papers in FE and RE models ..... 64
Table 5: WAAP, Top10, Stem & Kink results ..... 65
Table 6: MAIVE & P-uniform* results ..... 67
Table 7: Caliper test results ..... 68
Table 8: RTMA & MAN results ..... 70
Table 9: BMA results ..... 72

## List of figures

Figure 1: PRISMA flow diagram ..... 7
Figure 2: Funnel plot of Fisher's Z correlations ..... 59
Figure 3: T-statistic distribution of collected estimates ..... 68

## 1. Introduction

News headlines mentioning that industrial robots “slash wages” (Claburn, 2022), or that there’s “proof robots are taking jobs and cutting wages” (Harthorne, 2017) are becoming more commonplace as robot technology advances and is increasingly being introduced into the workforce. Are these statements just fearmongering, or is there truth behind them? This thesis aims to answer this question by conducting a meta-analysis to summarise the literature looking at the impacts of robots, specifically on wages.

The relationship between automation and labour market outcomes, including wages, has been a topic of discussion for as long as these technologies have existed. John Maynard Keynes popularised the term “technological unemployment” in the 1930s in reference to the way worker efficiency increased at a greater rate than the ability to find new uses for the now excess labour (Keynes, 2008). However, it can be traced back much earlier, with popular examples such as the Luddite movement which began in 1811, that comprised English textile workers opposing the machinery which replaced skilled workers and cut wages (Manuel, 1938). A modern example is the rise of Artificial Intelligence, where Acemoglu (2024) warns of many economic problems, including “inefficiently pushing down wages” if it remains unregulated on its current trajectory.

Interest in the impact of robots has been growing in parallel with the rising use of robots in the workforce. The International Federation of Robotics (2023a) states as part of their 2022 World Robotics Report that within the previous six years, the number of robots installed annually more than doubled, and that 2021 was an all-time high for new industrial robot installations. Determining the true effect of robots on labour markets, including wages, thus has become an important issue. This is especially true because, compared to other forms of automation like basic machinery or computers, robots require little human intervention, possibly amplifying any negative effects on employment.

Theoretically, the direction of the effect of robots on wages is unclear. The relationship between automation like robots and labour market outcomes like wages can be

explained by the degree of substitution or complementarity of robots, and the income effect determined by demand elasticity.

First, the crucial factor that mostly determines the sign of the impact is the degree of substitution or complementarity of an occupation, or the average across all occupations within a region or country. If robots complement an occupation, then they will benefit due to the productivity gains that accompany them. But if robots substitute roles significantly, roles may be diminished or replaced entirely, leading to lower or no wages for some workers. Because, compared to other forms of automation, robots require far less human interaction, substitution is intuitively more likely compared to complementarity, with papers like Graetz and Michaels (2018) going as far as to assume that robots and workers are perfect substitutes in their models.

The second force is the follow-on from the productivity gains through lower costs of production. If demand is sufficiently elastic, firms can pass on the lower costs of production to consumers through lower prices, which can increase demand. This demand is then met by firms through employing additional human and robotic capital, increasing labour demand, and therefore pushing up wages. As Autor (2015) explains, this wage benefit may still be seen even if demand is inelastic. With inelastic income elasticity of demand, as although productivity rises and industries get smaller, the productivity increases lead to surplus income, which consumers then spend elsewhere. This is also a part of the equilibrium model in Acemoglu and Restrepo (2020), where they intuitively explain that “automation lowers the cost of production (thus increasing productivity) and via this channel raises the demand for labour in nonautomated tasks in all industries”.

Overall, this does not clearly determine what the overall outcome would be, as it creates two opposing effects dependent on many distinct factors including income elasticities and the level of substitution between robots and workers within occupations. If the substitution effect is larger than any complementary effects (which is likely), this lowers wages. However, through productivity gains, higher demand for goods and services can increase wages through greater demand for labour. Therefore, it is unclear how these effects will play out across industries and regions showing the need for empirical evidence.

There is already plenty of empirical research done on this topic. However, even with this extensive empirical research, the overall effect across the literature is still unclear. The most prominent paper studying the effects of robots on wages is Acemoglu and Restrepo (2020). They came to the overall conclusion that robots hurt wages, concluding that one more robot per thousand workers reduces wages by 0.42%. This conclusion is shared by many other papers including Dauth et al. (2021) who found a slight reduction in earnings as robots per thousand workers increased. However, these results are contrasted by Graetz and Michaels (2018), another important paper, which found that robot adoption led to an increase in wages for workers, although the gains are only around 10% of the labour productivity gains. Dekle (2020) also supports the idea that an increase in robots leads to higher wages, estimating that a one per cent increase in average robot intensity led to an increase in average industry wages by almost one per cent.

In this thesis, I first focus on estimating the average effect across the existing literature. This is useful in determining the overall impact of the forces that determine the relationship between robots and wages. Then I analyse why studies come to opposing conclusions. Is it because they use data from different countries, because they use different estimation methodologies, or do they use a separate set of control variables? A meta-analysis helps answer these questions. A meta-analysis is a method of analysis that combines results from many different studies to obtain an overall quantitative estimate, and through methods like meta-regression, estimate how factors influence estimated results. The conclusion of a meta-analysis is more reliable and comprehensive than the analysis of a single study. Meta-analysis also can shed light on (and can adjust for) another important consideration, which is the level of publication bias the literature may have when publishing estimates. There may be incentives that lead some results being more likely to be published than others. For example, if there is preference for statistically significant estimates, the published literature might suggest there is a relationship between wages and robots even if no such relationship would be observed if both unpublished and published results were taken into consideration.

In this meta-analysis, I begin by searching for all studies that have estimated the relationship between wages and robots. I obtain 2,586 estimates from fifty-two studies, and then code the different characteristics of these studies. Because these

characteristics are quite different across studies, the estimated coefficient values I collect reflect somewhat different quantities and hence cannot be compared directly, so I use both the Partial Correlation Coefficients and Fisher's Z values of these estimates to measure the strength of the relationship between robots and wages. I then use these values to estimate an average effect size of robots on wages, including estimators that account for endogeneity, p-hacking, and, importantly, publication bias.

My main finding from all the estimation methods I used is that, across the literature, there is little evidence of a consistent relationship between robots and wages. Almost all models fail to achieve statistically significant results, and most coefficients are close to zero. The models that did achieve statistical significance at either the 1%, 5% or 10% level were all far below the thresholds for even a small effect (Doucouliagos, 2011). Additionally, with inconsistent results for the level of publication bias across estimators, the level of publication bias is unclear. Lastly, I find that some study characteristics affect the estimated impact of robots on wages like whether or not the study controls for skill level and import exposure, whether the study uses industry- or country-level data, how wages are defines, what dataset is used (IFR data or not), or whether the dataset only includes developing countries or data from the manufacturing sector.

The rest of this thesis is structured as follows. In section two, after describing the process I undertake to collect the relevant papers, I describe the current literature aiming to determine the effect of robots on wages. Section three introduces the methodology I use in my quantitative analysis of the literature, and section four discusses the results of my coding process and my analysis. Section five then concludes the thesis.

## 2. Literature review

This literature review aims to provide a brief overview of the current literature that explores the relationship between robots and wages. In section 2.1, I start by explaining my literature collection process using the 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses' (PRISMA) flow diagram created by Page et al. (2021). I then summarise the findings of the studies I collected and bring to light

their differences. This includes differences in both the results themselves, and how these results are achieved through the techniques and other decisions implemented in these studies. This may be how they describe their robot variables, their wage variables, and their dataset, such as the data level or whether they restrict their data in some way. I also look at different techniques used, such as adding control variables or panel fixed effects, conducting instrumental variable analysis, and if they weigh their regressions, cluster standard errors or address outliers in some way. Lastly, in section 2.5, I explain why I chose to use a meta-analysis to try and make sense of the current literature and assess the current meta-analysis literature on this topic.

### 2.1 Literature search and collection

The dataset used in this meta-analysis was created following the PRISMA guidelines, which first started with searching for relevant studies across multiple databases (Page et al., 2021). I chose four databases commonly used for economic papers to search for studies related to robots and wages to use in my analysis. This was to ensure my searches would return a wide range of papers from various sources. These four databases were RePEc, Google Scholar, Scopus, and EBSCO[1]. For each of these databases, I searched using “Robots,” and either “Wages” or “Income.” These keywords were the most effective for gathering the most relevant papers, as other synonyms for wages such as “Remuneration” or “Earnings” did not bring any additional relevant papers. For RePEc and EBSCO, I looked at all their results for the searches containing these keywords[2]. However, because of their much larger repository of papers, for Google Scholar and Scopus, I retrieved the first 200 results for each of the keyword combinations[3]. I chose this number as I found there was a drop off in the relevancy of the papers as I approached 200, but I still wanted to review a sufficiently large number of results. To increase the number of papers available, I also included working papers. I felt these papers were necessary to include due to this wave of robot literature still being relatively new, and therefore many papers have not had the chance

[1] For EBSCO I limited my search to their Economics/Business database.

[2] This was because both query results for RePEc had around 200 results, and EBSCO only had 103.

[3] 200 for “Robots” and “Wages”, and 200 for “Robots” and “Income”

to be published, although they still may be intended for eventual publication. I finished my search on August 3, 2023, after which I did not add any additional papers. Overall, these searches left me with 1,320 potentially relevant papers.

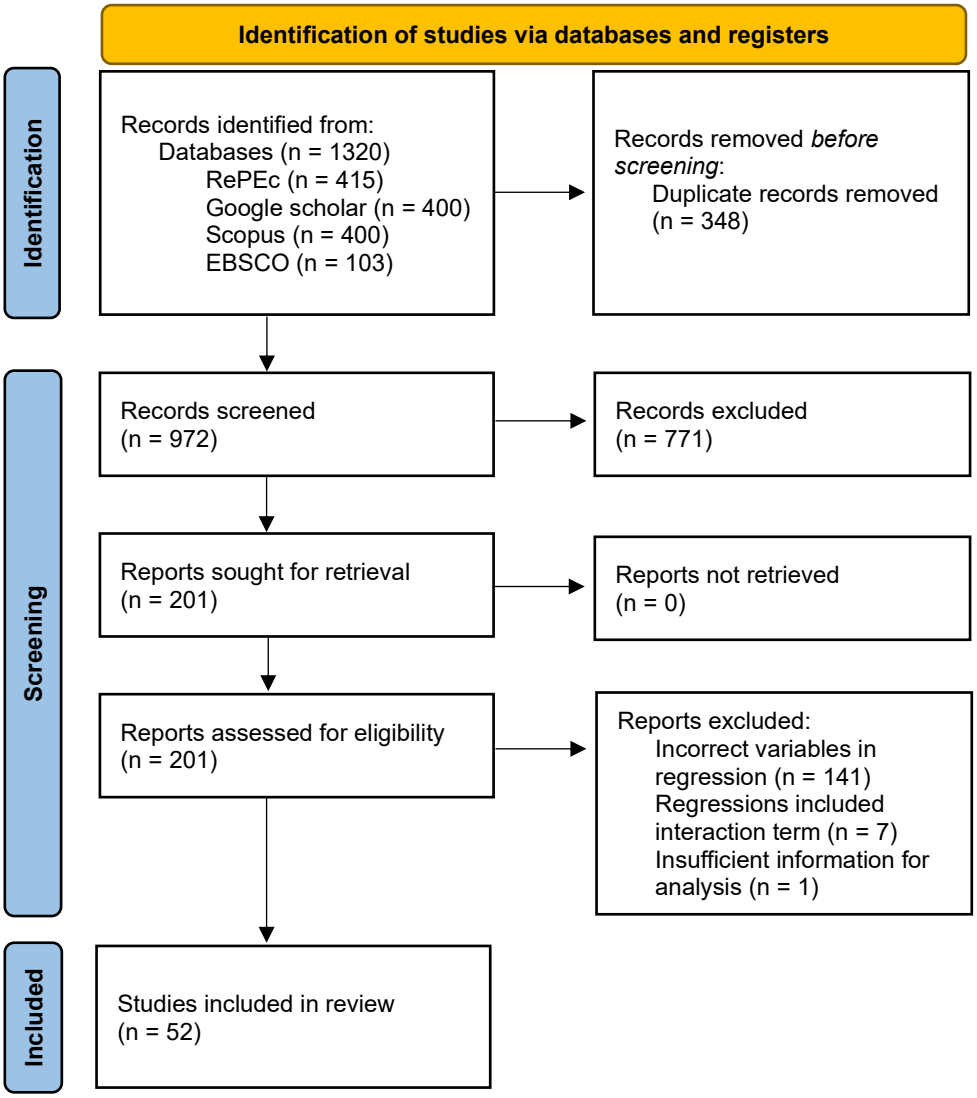
For each of these papers, using scraping tools, I collected the title, authors, abstract and year of publication, along with other identifying features to remove any duplicates from my list of papers. By examining these identifying features, I was able to filter out 348 duplicates, leaving me with 972 unique papers. These 972 papers then underwent the first screening stage, where I further analysed their titles and abstracts. If the abstract suggests that the paper undertakes empirical analysis on the relationship between robots and wages, it moved onto the second round of screening, and if not, I excluded them from further analysis. This process reduced the total number of potential papers from 972 to 201. In the second screening round, I downloaded the paper for full inspection. I was looking for papers that included at least one regression with robots as the independent variable, and wages, or a form of wages like income, earnings, or a firm's wage bill, as the dependent variable. I did not include papers that included variables such as labour shares or wage shares as the dependent variable. Having the share of wage to a country's GDP for example, is very different from simple wage values, and therefore it is much more difficult to compare these estimates, so I chose to exclude them from my analysis. Additionally, these papers had to report sufficient information. This includes the standard error, or a value from which I can calculate the standard error, and the sample size, so I can compute the degrees of freedom. I also excluded papers that only reported regressions with interaction terms, because the interaction with another variable distorts the robot coefficient value by no longer reflecting just robots, but robots when the value for the other variable is zero. This makes the interpretation of estimates with interacted variables difficult to compare to estimates without interaction included in its specification. This second screening stage filtered out a total of 149 studies from the pool of potentially relevant papers, leaving me with fifty-two papers to use in the quantitative analysis of my meta-analysis. Where possible I also collected any working paper versions[4] (or any earlier versions if

[4] If the same estimate was shown in both published and working paper version of a study, I treated it as a published estimate.

the paper is still a working paper) and supplements of these fifty-two papers to expand further the number of estimates I include in my analysis to a total of 2,586.

I show the PRISMA chart explaining the literature collection process in Figure 1 below.

*Figure 1: PRISMA flow diagram*




## 2.2 Robots and Wages literature

Through my literature collection, I collected fifty-two papers that all have at least one estimate measuring the effect of robots on wages. All papers do this in their own way, and therefore, not only are factors such as the methodology and dataset often different for each paper, but the depth of the analysis, and the focus of the robot-wage relationship specifically may also differ. Regardless of the reason for a paper's interest

in this relationship, I included its estimates in my analysis, and in this section, I will describe this literature, giving an insight into key findings and techniques used.

Arguably the most influential paper in the current literature is by Acemoglu and Restrepo (2020). First released as a working paper in 2017, they use commuting-zone data[5] from the US between 1993 and 2007, and find that robots hurt wages, stating that an additional robot per thousand workers reduces wages by 0.42%. They provided a far more comprehensive approach compared to previous papers, along with many robustness checks, making this paper popular across more mainstream media. They were also among the first papers to make use of the robotic dataset created by the International Federation of Robotics (IFR), although notable papers such as Graetz and Michaels (2018), first released as a working paper in 2015, had previously used the IFR data. Graetz and Michaels (2018) used the IFR data to measure the effect of robots on wages across seventeen different countries[6], and instead found that an increase in an industry's robot use resulted in higher hourly wages, although it should be noted that these increases in wages were disproportionately lower than the increases in labour productivity. From the selection of studies I collected, thirty-eight of the fifty-two papers make use of the IFR dataset, not only showing the influence of Acemoglu and Restrepo (2020) but also showing a potential lack of alternative robot data.

### *2.2.1 Robot variables*

However, beyond breathing new life into quantitative analysis on robots' impacts on wages and helping pioneer the use of IFR data, Acemoglu and Restrepo (2020), had arguably the greatest influence on the literature through the new methodology they introduced. This model builds on a previous model of theirs which looks at how technology affects growth, factor shares, and labour outcomes (Acemoglu & Restrepo, 2018). Acemoglu and Restrepo (2020) adapt this model to derive the full equilibrium impact of advances in robotics on wages and employment. The model depends on many different economic variables, but multiple key equations depend on a commuting

[5] They used commuting zones as a proxy for local labour markets.

[6] The countries used in the study were Australia, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, South Korea, Spain, Sweden, United Kingdom, and United States

zone's exposure to robots. They use a Bartik-style instrument to measure robots relative to baseline employment shares. The use of a Bartik-style measure allows for interaction between baseline local industry labour shares and overall industry technological growth rates (Goldsmith-Pinkham et al., 2020), which was necessary when analysing at the commuting zone level, due to IFR only having robot data for each industry at the national level. Additionally, when using regional-level data, often researchers use an industry's employment share within a region such as a commuting zone to weight these values. This approach is similar to the adjusted penetration of robots (APR) used by Acemoglu and Restrepo (2020) in their exposure measure, but as the name suggests, they adjust the value. They do this by removing the relative growth of output in an industry. At least twenty-five out of the fifty-two papers I collected show some level of influence by Acemoglu and Restrepo (2020) on the models they created, especially their robot exposure measures.

Of course, this specific Bartik-style relative measure is not the only measure used in the literature. Graetz and Michaels (2018) produced a "replaceability index" to measure the relative robot density across industries by measuring an industry's fraction of jobs with tasks that can be replaced by robots. Another often-used approach is slightly simpler, using the number of robots per 10,000 workers. Aksoy et al. (2021) and Jung and Lim (2020) both use this approach and find robots increase wages. Aksoy et al. (2021) find that across the twenty European countries they focus on, wages increase, although this increase is uneven, favouring males compared to females. Jung and Lim (2020) conclude that an increase in industrial robot use leads to an increase in hourly compensation within their forty-two-country dataset. This is even though unit labour costs are lower, which they use to determine that increased productivity from robots exceeds any wage-increasing effects, a similar conclusion shared with Graetz and Michaels (2018). Depending on the data used, others instead use robots per 1,000 workers. Shang (2022) focuses on the robot-wage relationship within China at the prefecture level where robot use is common in manufacturing. They find negligible effects on general wages, but focusing on the private sector they find a negative impact.

While most papers use relative measures of robot intensity to try to describe robot use more accurately in the workplace, a simpler approach is sometimes used instead. A minority of papers opt to simply use the absolute values of robots. For example, Dixon

et al. (2021) conducted a firm-level analysis in Canada by adding all robot purchases by a firm recorded each year. Overall, they achieve ambiguous effects, with negative impacts for managerial and medium-skill-level workers, and positive impacts for low and high-skilled workers.

Those that do not use IFR data are usually limited in their data availability. Therefore, often a wide range of sources are utilised, including business survey data (Alguacil et al., 2022; Eggleston et al., 2021; Koch et al., 2021; Ren et al., 2018; Webb, 2020), or customs data (Artuc et al., 2019; Barth et al., 2020; Dixon et al., 2021; Shang, 2022; Wang et al., 2022b). While this sometimes allows for the methods described above, often these do not have the crucial data needed for the previous continuous measures. Instead, these papers opt to use a dummy variable indicating whether a firm uses robots, or if a worker is employed at a robot-using firm, and lead to a similar range of both positive and negative results derived from their analysis. An example is Acemoglu et al. (2020), who use firm-level data to look at the effect of robots on labour market outcomes in France. They utilise a myriad of sources including Ministry of Industry surveys, customs data for robot imports, and even the information on depreciation allowances for industrial robot purchases found in the French fiscal files. Because of their diverse sources, they opt for a simple dummy variable indicating if a firm is a robot adopter. From their analysis, they find some positive, although mostly insignificant results, and therefore cannot come to a concrete conclusion on the effect of robots on wages.

Some papers also choose to lag their robot variable as they may believe that the impact of robots on wages may have a delayed effect. Most offset their robot measures by either one (Compagnucci et al., 2019; Wang et al., 2022b), two (Anelli et al., 2019), or four years (Aksoy et al., 2021). Wang et al. (2022b) use Chinese customs data to create multiple robot measures including dummy variables, the absolute number of robots, and the value of the robots used. They lag these variables by one year to account for robot installation payback periods. Another concern is reverse causation from robots as they use yearly data and therefore argue that the fact that wages increasing could suddenly lead to higher robot use. They find that robots do increase wages, although as a robustness check, they remove the lag from the robot variable and find comparable results. For similar reasons, Aksoy et al. (2021) instead lag by four years as the survey data they use is only captured every four years. Anelli et al.

(2019) measure the relationship between robots and wages but mainly focus on family behaviour within the United States. Because they also look at factors such as childbirth rates, to measure the effect of robots more accurately, they decide to lag their robot measure by two years which of course also affects their effect on wage estimates. Using this lagged analysis, they found that robot penetration significantly lowered income. They also concluded that this negative effect is much more substantial for men, and therefore conclude that the gender income gap also closed.

Lastly, while most papers look at domestic robot use within their country of focus, some instead measure the domestic wage effects of foreign robot use. These mainly focus on how robotic automation in developed countries impact the labour markets in developing economies. Kugler et al. (2020) assess the effect of US robot use on Colombian labour markets. The United States is Colombia's largest trading partner, and therefore they are interested in how changes in robot levels in the US affect labour markets in Colombia. They match industries across the two countries and find that as robot adoption increases in the US, wages reduce in Colombia. They also note that along with women, older workers, and workers in small and medium businesses, the labour markets that predominantly export to the US are among the most affected. This suggests that increased robot use in the US allows for the reshoring of good production back to the US, diminishing the previous advantages countries such as Colombia had due to factors like lower wages. Stemmler (2020) looks at how both domestic and foreign industrial robots affect labour markets in an emerging country like Brazil. They find overall unclear results, and the only statistically significant result is a positive effect for skilled workers. Díaz Pavez and Martinez-Zarzoso (2023) also look at both foreign and domestic robots for ten emerging countries. Again, similar to Stemmler (2020), Díaz Pavez and Martinez-Zarzoso (2023) find that an increase in neither domestic nor foreign robot intensities leads to significant changes in annual wages across the different countries.

### *2.2.2 Wage variables*

As shown above, there is significant heterogeneity in the independent variable (the robot variable) used across papers, and the same can be said about the dependent

variable I am interested in, the wage variable. There are essentially two ways this variable differentiates across studies; how they define the variable used, and the time frame they use for the variable.

While I have been referring to the dependent variable in the regressions I analyse as “wages,” this is not the only definition provided by papers. There are plenty of other terms used by authors, and a lot of the time this originates from the research question they are trying to answer or the dataset they use, and therefore may have an impact on the effect size they are measuring. Of course, the most common definition used by papers is “wages”. This typically refers to employment income, which is the focus of many of the papers, as they aim to estimate the effect of industrial robots in the workplace. Forty-six out of the fifty-two papers I collected measure the effect of robots specifically on wages, showing that much of the literature is interested in employment income and less focused on how robots affect other sources of income.

In a similar vein, some papers choose to instead use the wage bill of a firm. An important difference between wages and the wage bill is that the wage bill takes into account the level of employment, not only the pay rate. Acemoglu and Restrepo (2020) also measure the impact of robots on wage bills, and compared to their wage analysis where they use regional-level data, they instead use industry-level data. This could be due to their focus shifting from workers to firms with the change in variables. Dauth et al. (2021), another influential paper within the literature, looks at how German local labour markets adjust to increases in industrial robot exposure. They also include worker wage bills in their extensive analysis, along with hourly wages and earnings. Their wage bill measure is the individual daily wage multiplied by 365, which they then aggregate to the regional level. For both wages and wage bills, they find insignificant overall results but find manufacturing workers face negative impacts.

Similar to wages, some papers use earnings, which also usually refer to pay gained from working. Giuntella and Wang (2019) look at both wages and earnings, where the difference in definition is the time frame being used. They refer to wages as the hourly rate, whereas earnings use annual data. Another difference is that their earnings measure is for the entire population, not just workers, and therefore they give those who are unemployed zero earnings. This same difference in earnings including non-employed workers is also used by Dauth et al. (2021), who for this reason prefer the

earnings model, as they avoid the selection problem of wages not being observed for unemployed workers. For earnings, Giuntella and Wang (2019) find a negative impact from robots, although this is very small and insignificant. They find slightly more concrete results for wages, as they find that as their robot measure increases by one standard deviation, an individual's hourly wage decreases by 7 per cent. However, the difference between earnings and other definitions like wages is not always so clear, as other papers such as Aksoy et al. (2021) and Borjas and Freeman (2019) use multiple different timeframes when referring to earnings, in a similar context as one would refer to wages.

The last definition found in the literature I collected was income. This mostly includes both employment and non-employment sources. Chugunova et al. (2021) ask how robots affect voter participation and behaviour in the US, and in addressing any potential motivations behind their results, look at household income. They find that robot exposure does lead to a decrease in total household income, and they list it as a potential reason voter turnout decreased with the increase in automation. In further analysis, they then split up their total household income into three categories: wages/salary, self-employment and investment income, and social security and welfare income. They found that wage/salary impacts were similar to total income impacts, business income effects were much less negative and less statistically significant, and welfare impacts were slightly positive. Acemoglu and Restrepo (2020) also separate wage and non-wage income and find that their wage income had large negative effects due to robots, but no significant impacts were found for non-wage income. They use these results to back up their notion that the owners of robot-using firms may not be in commuting zones exposed to robot use.

The other main difference across the dependent variables used is the time frame. Across the papers I collected, almost all papers either use hourly, weekly, monthly, or annual values. Hourly is commonly used because industries where the most robots are found, often use hourly rates. Additionally, hourly rates do not have the same potential issues as other time frames. Unless the annual salary is being used, using larger time frames may be influenced by the number of hours worked, and therefore may not solely reflect the changes in wages. Because of these reasons, hourly is the most common timeframe used across the literature I collected, used by the majority of

papers including the most influential papers such as Acemoglu and Restrepo (2020), Graetz and Michaels (2018), and Dauth et al. (2021).

However, some papers may want to see how robots impact wages without adjusting for other factors like fewer hours worked. As previously discussed, this is what variables such as the wage bill capture, and therefore it is common for these variables to be used with an annual timeframe (Acemoglu & Restrepo, 2020; Dauth et al., 2021). This is further shown by Giuntella et al. (2022) who use both hourly and annual income variables. They find that exposure to robots reduces hourly income by nine per cent on average but find insignificant effects on annual income. They argue that this was because workers who were more exposed to robots, worked more hours to make up for any losses in hourly wages. Additionally, annual time frames are sometimes all that is available in a dataset. For example, Giuntella and Wang (2019) use survey results to calculate their wage and earnings data. This survey did not have data for hours worked for all the periods they used (so couldn't calculate hourly wages) and therefore chose to also include annual data which were available for all years.

Weekly data is another utilised timeframe, although seldomly used in the literature as the use of weekly wages is usually due to the source of the wage data. Chung and Lee (2023) use weekly wages to estimate the effect of robots, as they get their variable data from the American Community Survey which collects weekly data. They focus on a slightly more contemporary dataset, between 2005 and 2016, within US local labour markets. Compared to studies that focus on an earlier period like Acemoglu and Restrepo (2020), they found similar negative results during the early periods (which overlap with the other studies), but this relationship reversed over time and eventually turned positive in more recent years. Kariel (2021) also use weekly wages as they use the Annual Survey of Hours and Earnings (ASHE) to measure the impact of robot use in the United Kingdom, which collects the mean and median weekly earnings across Local Authorities. They follow Acemoglu and Restrepo (2020) to measure the effect of robots on wages at the local labour market level, proxied by Local Authorities. Overall, they find that full-time pay is unaffected by increases in robot use but find slightly positive effects for part-time workers.

The last time frame used in the literature is monthly wage data. Like annual data, monthly data may be influenced by the number of hours worked, although the effects

may not be as severe. Again, the choice of timeframe is usually down to data availability, or convenience as researchers may not feel the need to convert their data to a different timeframe. Brambilla et al. (2023) measured the impact of robots on Latin American labour markets between 2004 and 2016. They find that increased industrial robot use negatively affects the average monthly wage, with formal wage estimates being much larger than estimated effects on informal wages. They find self-employment income is unaffected. Kugler et al. (2020) also use monthly data. They use social security data from the Colombian Social Security Administrative Records to construct their monthly income measure by aggregating each employee's base earnings they report monthly, again showing that the reason for selecting a monthly time frame mostly comes down to data availability.

It is also important to note that some papers do not specify the time frame for their estimates. To include them in my analysis, I did not assign a time frame to them during my coding process. Therefore, in my meta-regressions, the estimates regarding the impact that varying time frames have on effect sizes utilised less information. This omission slightly hinders my analysis, compared to if all estimates reported the time frame of their data.

For both the dependent and independent variables of my regression of interest, some papers transform their variables. Logging the variable is quite common as it not only makes the distributions of these variables more normal (as wages or income are usually right skewed) but also allows for easier interpretation. Famous results, such as the one by Acemoglu and Restrepo (2020), are obtained using a logged wage variable, meaning they can come to their conclusion that as robot exposure increases by an additional robot per one thousand workers, worker wages decrease by 0.42%. It is more common for the dependent variable to be logged, but some papers still opt to log their robot variables, sometimes along with their wage variable. Abeliansky and Beulmann (2021) aim to measure the impact of robots on workers in Germany, focusing on their mental health, and using wages as a potential explanation. They use a logged robot variable for two reasons. First, they say it reduces heteroskedasticity from the substantial differences in robot adoption rates across sectors. The second reason was that the regression coefficient now shows how the impact on the dependent variable due to relative changes in robot intensities. They argue this more closely captures a sector's perceived robotic automation. Overall, they find results

roughly in line with Dauth et al. (2021) (who also focus on Germany) as they find slight negative effects, but these do not pass the robustness checks they conduct. Petit et al. (2023) convert both the wage and robot variables into logarithm form when trying to estimate labour market adjustments to robotic use in twenty-two European countries from 1995 to 2017. They do this to interpret their estimates as elasticities, i.e. the percentage change in wages due to a one per cent increase in robot penetration within a region.

Throughout the papers I collected, researchers often turn their variables into a change over a longer period. Stemmler (2020) uses long-difference models, something that many other papers make use of. They claim that the advantage of using this model is that changes in robot stocks on a year-by-year basis may not be enough for meaningful analysis as there may need to be time for the effects to fully develop. Of course, this comes with the disadvantage that important data points are compared to the yearly variation. Acemoglu et al. (2023) also include long differences in their firm-level analysis of the impact of robots in the Netherlands between 2009 and 2020. They use long differences as part of their analysis due to the advantage of focusing on a time horizon that allows the realisation of any potentially slow-acting effects of robot use. This analysis along with annual panel data, brings them to the conclusion that, at the market level, robots have negative effects on wages, but this relationship is flipped at the firm level. Similar to long-differences, studies also include stacked-differences. This is essentially long-differences but using shorter subperiods. For example, Acemoglu and Restrepo (2020) use stacked-differences for two 7-year subperiods. They use this model because they can focus on two crucial periods of robot adoption: 1990–2000 and 2000–2007. These two time periods represent different periods of robot adoption and the level of robot technology. They claim it also allows them to control for linear commuting zone trends. They also state that although this model exploits a different variation source compared to long differences. The negative effects of wages are similar to what they find in other models.

### *2.2.3 Data level*

While the variables a paper uses are important, another crucial factor in a paper's estimates is the dataset they use. The level the data focuses on is among the most important aspects of dataset selection. As previously mentioned, many papers such as Acemoglu and Restrepo (2020) and Dauth et al. (2021) use regional-level data as part of their analysis whereas Dixon et al. (2021), Acemoglu et al. (2023), and Acemoglu et al. (2020) use data at the firm level. The use of regional-level data aims to focus on entire labour markets, often covering both industries using robots and those that don't. While there are regions that will have higher levels of exposure to robots compared to others, if they do not restrict their dataset to workers within occupations exposed to robots, it will still allow a more overall look at the effect of robots. Firm-level analysis is often (although not always, especially when using dummy variables) performed to look at firms that use robots and therefore measure the impact on workers that are directly exposed to robots. The choice between these two often boils down to the question the researchers are trying to answer, but also the data they have available.

There are of course other levels of aggregation commonly used by researchers. Industry-level analysis is like firm level as the data mainly focuses on industries that are directly exposed to robots. Additionally, the IFR data, which many papers use, aggregates data at the country or industry level (International Federation of Robotics, 2023b). The IFR classify their robots into nineteen different industries, and therefore it is also simpler for studies to keep their analysis at the industry level, within these selected industries. Dekle (2020) uses industry-level data to determine the impact of industrial robots on labour demand in Japan. Using a model derived from Acemoglu and Restrepo (2020), instead of focusing on regional-level analysis, Dekle (2020) does industry-level analysis because they claim that this is an easier way to get an overall effect for the whole country. Additionally, that claim allows them to assess three important effects of robots: the negative substitution effect of robots on workers, the positive cost decline leading to more labour demand, and lastly the positive productivity increases leading to industry expansion. Overall, they find that robots have increased productivity and helped lower costs, leading to a higher average industry wage within Japan.

A paper uses country-level analysis when it covers a large assortment of countries. The data availability for this kind of analysis is often an issue, and therefore it is seldom seen in the literature. Using forty-two countries between 2001 and 2017, Jung and Lim (2020) is one of the only studies able to conduct this type of analysis as a vast amount of data is required to get the number of observations required for meaningful analysis. Because many other studies do not focus on this many countries, they often opt to conduct country-industry-level analysis. This allows them to not only have the data points necessary for their analysis but also use the industry-level analysis across different countries, which is made easier with the IFR data being at the country-industry level.

Lastly, there is also individual-level analysis. This is the most detailed form as it captures additional variation across observations that are lost when they are aggregated to other data levels. However, it comes with its own set of issues. Firstly, data availability is often an issue, so researchers may have to decrease the scope of their analysis. For example, Eggleston et al. (2021) look at how robots impact the service sector but exclusively look at nursing homes which have survey data available. Additionally, other papers like Giuntella and Wang (2019) who also use survey data, are unable to reach their entire population of interest, and therefore their sample only consisted of 15,000 families and 30,000 individuals. However, some studies can obtain a more comprehensive dataset. As I touched on previously, Kugler et al. (2020) use social security data for their analysis of US robots in Colombian labour markets. This allows them to have an extremely large dataset with over 226 million observations from nine million workers. Similarly, Dottori (2021) use data from the Italian Social Security Institute (INPS) for their worker-level analysis of how employment outcomes are affected by increases in the use of robots. Overall, they find that robots have positive impacts on worker wages. They also state that the individual-level analysis allows them to explore the heterogenous effects due to differing worker characteristics more effectively, compared to regional-level analysis who at most may break their datasets into demographic cells within regions. They found that white-collar workers benefited from higher wages compared to blue-collar workers, but the youngest group of workers from their pool of data had the worst outcome in terms of wages. Other potential sources of heterogeneity such as gender or geographical location had little effect.

### *2.2.4 Data restriction*

#### *Restricting by occupation*

Another important aspect of the data a study uses is whether they restrict their data to a particular subgroup of the overall population. Papers often use this as a method to focus on the potential heterogeneity caused by any characteristics represented in these subgroups. Due to the largest variation in robot density mostly coming from differences between industries, many papers choose to focus their wage data on a specific industry. The most common industries used in the literature are manufacturing industries, service industries, and agriculture.

The isolation of manufacturing is common because it is the industry with the most robots. Koch et al. (2021) focus their analysis on Spanish manufacturing firms between 1990 and 2016. They find that the labour cost share decreases by 7 per cent due to robot adoption, but failed to find a statistically significant effect on wages. Ren et al. (2018) also solely look at manufacturing firms in their analysis of the impact of robots in Chinese labour markets. Overall, they conclude that robot investment may lead to higher worker wages but are careful to recognise that their estimates do not have causal meaning. While these two papers focus solely on manufacturing firms, it is also common for researchers to include this in their analysis as a robustness check, or a way to explain potential heterogeneity of results. Popular papers such as Acemoglu and Restrepo (2020) and Dauth et al. (2021) do this and find that manufacturing workers face negative impacts on their wages from robots. Dauth et al. (2021) even compare this to other industries like services which have positive effects. Some estimates do the opposite, they measure estimates on exclusively non-manufacturing firms. Dottori (2021) conducted this type of analysis, and their estimates suggest that non-manufacturing firms experience larger negative effects than manufacturing firms.

Like Dauth et al. (2021), other papers also choose to focus on service industries in their analysis. It is important to note that this sector is not focused on frequently because the robot density is much lower than manufacturing industries. Also, because the IFR dataset commonly used by researchers only looks at industrial robots, and although they do publish separate service robot data, this data is not as extensive as their industrial robot data. Eggleston et al. (2021) focus their entire analysis on the

service industries by looking at Japanese nursing homes. The robots in these care facilities are estimated to reduce the monthly wages of care staff. Cali and Presidente (2021) analysed plant-level data from Indonesia between 2008 and 2015 when robot imports into the country increased. Alongside their overall analysis, they also include estimates looking exclusively at service industry data. They find that robots have increased service wages by five per cent, which is consistent with their overall analysis.

Lastly, a few papers focus some of their estimates solely on the agricultural sector. In their attempt to explain the effects of different types of automation technologies (including robots) across 227 regions from twenty-two European countries, Petit et al. (2023) conducted a subgroup analysis of many different sectors including agriculture. Their estimates indicate a negative impact of robots on average agricultural worker wages. The estimates for the agricultural industry were among the most negative estimates across all industries. On the contrary, Giuntella et al. (2022) look at non-agricultural workers. As previously mentioned, their analysis of robot impacts on Chinese households finds negative effects on wages and inconclusive evidence for annual income. Even though their samples consist only of non-agricultural workers between the ages of sixteen and fifty-nine, they state that if they relax their industry and age restrictions, their main results stay qualitatively the same, with similar levels of statistical significance.

*Restricting by skill*

It is also common for studies to restrict their datasets to certain skill groups. It is commonly theorised that higher-skilled workers are harder to replace with robots, and therefore their wages will be less affected by robots, and some argue that wages should increase due to the productivity benefits from robots. The skill level of a worker is often determined by either the occupation of the worker or their education level. Therefore, it is also common for studies to restrict their datasets to certain education levels. Zhang et al. (2023) aim to answer this question as part of their analysis of the impact of industrial robots on Chinese labour markets. Using similar methods to Acemoglu and Restrepo (2020) but adapted for the Chinese market, their results indicate that, overall, firm adoption of robots can increase worker wages. They further

break their analysis down at the skill level to assess any potential skill bias in their results. They use the education level of a worker as a proxy for their skill level, which they separate into high-skilled workers (college level or higher), medium-skilled workers (high school level), and low-skilled workers (below high school). Overall, they find that this wage benefit from firms adopting robots is concentrated on high and medium-skilled workers. Similarly, Giuntella and Wang (2019) use education as a proxy for skill level and find the negative effects on annual income are concentrated on low-skilled workers who experience a 22.6% decrease in annual income with a one standard deviation increase in robot exposure. Aksoy et al. (2021) instead differentiate by occupation skill levels. Because they focus on the gender pay gap, they focus on occupational skill levels as women may be underrepresented in high and medium-skilled occupations. They find that in high and medium-skilled occupations, while both experience an increase in wages, the benefit for males is greater than for females, and they find little effect for low-skilled occupations. Therefore, while the literature looking at the overall impact is conflicted, it seems that there is more of a consensus that higher-skilled workers are better off when it comes to the impacts of robots on wages.

*Restricting by age*

Like skill level, the literature also explore age as a potential source of heterogeneity across results, and therefore many papers break up the typical working age range and conduct estimations focusing on more specific age brackets. The reason they do this may be because younger workers are able to adapt more easily to any workplace changes due to robots, and therefore may not be subject to as much of the negative effects from robots. In Giuntella and Wang's (2019) analysis, they break up their sample into three age groups: 16-24, 25-44, and 45-65. They refer to the 25-44 group as "prime age" and find that for these workers annual income reduced by 15% with one standard deviation increase in robot exposure. The effect was 14% for over forty-five workers, and there were slightly positive but insignificant results for young (16-24) workers. Similar results are found in Giuntella et al. (2022) where they instead focus more on households. Again, there is no evidence of a statistically significant effect on younger workers, but prime-age (25-44) and older (45-65) workers experience more

of the negative effects of robots on labour markets. Their estimates state that a one standard deviation increase in robot exposure led to an eleven per cent decrease in hourly wages for prime-age workers and an eight per cent decrease for older workers. Wang et al. (2022a) look at workers below forty and above forty when comparing gender differences in the impact of robots in China. They find that for both men and women, wages improve only for the younger generation. Therefore, like restricting an estimate's dataset to a certain skill/education group, there is somewhat of a consensus that older workers are worse off compared to younger workers when it comes to how their wages are impacted by robots.

*Restricting by gender*

The effect across genders is something many papers also focus on, especially as this investigates important topics such as the gender pay gap. I have already mentioned a few papers that look at the effects of robots on the gender pay gap (Aksoy et al., 2021; Anelli et al., 2019; Wang et al., 2022a) and unlike the skill and age analysis in the literature, there are conflicting results. Anelli et al. (2019) find that in the United States, the negative effect of robots on male income is far greater than females. Aksoy et al. (2021) support this conclusion as they find that, through a productivity effect, robotization in medium and high-skilled jobs disproportionately benefits males compared to females. However, Wang et al. (2022a) find in their analysis of Chinese labour markers, that robots help promote equality in households, as they improve female wages relative to males. These differences could be due to cultural differences as these studies mostly focus on different countries. Costanzo (2023) also looks at the gender pay gap and therefore restricts their data to male and female subgroups. They conduct regional-level analysis of US commuting zones, focusing on households. They find that robots reduce hourly wages in the aggregate, but these effects are greater for men compared to women.

There are of course other papers that simply as a robustness check want to study any differences between men and women. In their extensive analysis and robustness checks, Giuntella and Wang (2019) also separate their Chinese dataset by gender. They conclude that men experience far greater negative effects than women. Male income decreases by 16.8% with a 1 standard deviation increase in robot exposure,

which is a greater and more precise estimate compared to females. Borjas and Freeman (2019) expand onto their main estimates by separating men and women. They look at robots as a supply shock and explore these effects at the industry level within the United States. Overall, they find that the robot supply shock reduces wages, and like many papers that separate the sexes, find modest differences, with males bearing most of the negative effects compared to women. Interestingly, when focusing on gender differences, Abeliansky and Beulmann (2021) choose to only focus on the male sample. This could be due to the smaller sample size of females leading to their estimators lacking enough power for them to include in the analysis. Nonetheless, they find that, compared to the overall sample of German households, the difference with the male sample is mostly not statistically significant, and therefore they are unable to come to any conclusions.

*Restricting by country type*

Lastly, arguably the most crucial factor when it comes to the dataset is the country that the data originates from. This factor is also one of the most prevalent differences across papers because often papers extend previous methods of analysis to a new country or group of countries. However, because some countries are only focused on by a single paper, it is difficult to separate studies by country. Therefore, I differentiate by developed and developing countries. To do this I use the list of countries provided by the International Monetary Fund (IMF) who separate countries into two groups: advanced economies and emerging/developing economies. According to the IMF, there are three main criteria used in their classification: per capita income level, export diversification, and the level of integration into the global financial system (International Monetary Fund, 2024). Significant differences in factors such as the per capita income level can lead to a difference in wage variable distributions, and therefore can lead to different results on the impact the robots have on wages. When it comes to robot use, there is also usually a significant difference between the two country types. Because of their more advanced economies and industries, developed countries usually have a greater number of robots, but that does not mean that developing countries are less at risk. Developing countries are home to a far larger proportion of lower-skilled workers who may feel more of the potential negative effects of robots that they import

into their country, indicating that while developing countries might have fewer robots within their economies, the effects of those robots could be more pronounced. Additionally, compared to offshoring production, domestic robots may be cheaper for advanced economies leading to a decline in exports for developing countries, therefore possibly amplifying potential negative impacts on wages (Artuc et al., 2022). Lastly, papers such as Brambilla et al. (2023) note that some of the developing country data is limited. In their analysis of Latin America, they state that the household surveys struggle to capture the top worker incomes, possibly due to underreporting.

Looking at the results from the literature that focuses solely on one country type, the results may not be as clear-cut as expected. The most common country analysed in the literature I collected was the United States of America, which can be considered the most advanced economy in the world. Fifteen of the papers I collected focused solely on robot impacts in the US, which include results showing negative impacts on wages such as Acemoglu and Restrepo (2020), Anelli et al. (2019), and Borjas and Freeman (2019). However positive results are also shown in papers like César et al. (2022), Lyu and Liu (2021) and Chung and Lee (2023) (although they find negative results initially, as time goes on these results become positive). Other developed countries looked at exclusively for analysis are European countries like Germany (Abeliansky & Beulmann, 2021; Dauth et al., 2021; Deng et al., 2020), Spain (Alguacil et al., 2022; Koch et al., 2021), the Netherlands (Acemoglu et al., 2023), France (Acemoglu et al., 2020), Italy (Dottori, 2021) and Norway (Barth et al., 2020) and again a consensus is unclear with results showing both positive and negative impacts of robots on wages. Another advanced economy commonly looked at is Japan. However, Japan is a special case due to the considerable number of robots used compared to other countries. Papers such as Graetz and Michaels (2018) choose to drop Japan from their selection of countries, something that the IFR also recommends, due to reclassifications of what machines are considered robots making Japan hard to compare to other countries. Dekle (2020) chose to solely look at Japan and highlighted the differences between Japan and other countries like the US. They state that robot use in Japan is over ten times as intensive compared to the US, therefore showing that while both are considered developed countries, this classification also has its downfalls as there can still be significant differences between developed nations. Papers like Compagnucci et al. (2019), Aksoy et al. (2021), and Chiacchio et al. (2018)

look at a slew of developed countries, usually focusing on a group of OECD, or European countries. Again, results are similar to studies focusing on a single country, where there is not a clear trend for the effects of robots on wages within developed countries.

You may expect that trends within developing countries to be a bit more clear-cut, but this is not the case. Some of the developing countries exclusively looked at in papers I have previously mentioned include Brazil (Stemmler, 2020), Colombia (Kugler et al., 2020), Mexico (Artuc et al., 2019), and Indonesia (Cali & Presidente, 2021). Overall, a clear effect cannot be seen, with some stating negative results, and others unable to reach a strong conclusion. However, there is a clear outlier in China when it comes to the robot intensity within a developing country. China has the largest manufacturing sector in the world (The World Bank, 2022), and therefore has a much larger number of robots in use compared to other developing countries, making it an interesting case study that many papers look into (Giuntella et al., 2022; Giuntella & Wang, 2019; Ren et al., 2018; Shang, 2022; Wang et al., 2022a; Wang et al., 2022b; Zhang et al., 2023). While the robot use is similar, or even greater, compared to developed countries, it is still categorized as a developing country due to factors like lower wages. This allows China to keep their comparative advantage compared to other developing countries which, because they have fewer robots, are at risk of being replaced by foreign robots in developed nations. However, like other developing countries, there are again mixed, although slightly more positive results (although there are more papers in general). Additionally, due to the expected amplified effect of robots in developing countries, more papers look at a range of countries to measure the effects of robots on wages in developing countries as a whole. Díaz Pavez and Martinez-Zarzoso (2023) is one of these papers, and as previously mentioned, they look at ten emerging economies and across these countries, they fail to find any overall significant impacts of either domestic or foreign robots on wages. However, they did find more negative effects on sectors that are the most exposed to foreign competition, possibly alluding to the replacement of domestic workers by foreign robots in more developed. Brambilla et al. (2023) focus on three emerging Latin American countries (Argentina, Brazil, and Mexico) and share results with other papers covering developing countries, with negative, although statistically insignificant, results.

There are of course papers that have a dataset consisting of both developed and developing countries (Graetz & Michaels, 2018; Jung & Lim, 2020; Kromann et al., 2019; Petit et al., 2023), with similar results. You may therefore think that the country/countries the data is taken from has little importance on the result of the paper, but because each paper uses slightly different techniques for their estimation, simply looking at the results from each of these papers will not tell the whole story. Therefore, I conduct more rigorous testing later in this paper.

### *2.2.5 Model specification*

Another difference across papers is the specification that papers use in their regression analysis. These include (but are not limited to) adding control variables, adding fixed effect variables (panel fixed effects), and conducting instrumental variable analysis.

#### *Fixed effects*

Not to be confused with the fixed effect regression I conduct later in this paper; due to many researchers using panel data they often use panel fixed effects in their specifications. Regional fixed effects are the most common. This is due to regional-level analysis being the most popular in the literature I collected, and therefore regional fixed effects are especially useful in controlling for the differences in wage levels across different regions that remain consistent over time. This is because while commuting zones close to each other may not show significant differences, two commuting zones on opposite sides of a country may be vastly different, requiring these differences to be accounted for when calculating the effect size across all the regions within a country, or countries. Country and US state fixed effects are also used when necessary and capture even larger time-invariant differences across areas. Kromann et al. (2019) use data from nine different countries, both developed and developing and therefore may see significant differences in factors such as average wages or productivity levels across each country. They state that if these differences across countries are systematic and not caused by differences in the level of robot use

(although still may be related), then their OLS estimators will be inconsistent. Therefore, to combat this, they included country fixed effects. They also included industry fixed effects, which are also commonly found in other papers. The reasoning is again similar to other fixed effects, as different industries have various levels of productivity and therefore wages, which are not caused by their robot use and therefore need to be controlled for. Like industry fixed effects, some papers also use firm fixed effects (although not as often as the use firm-level data is uncommon). For example, in their firm-level analysis, Zhang et al. (2023) add firm fixed effects to account for firm characteristics that don't change much over time, like "firm culture." Another type of fixed effects used is individual fixed effects although this is among the least utilised, due to the type of data required. For this, individual-level panel data must be used, which, as I touched on previously, is difficult to obtain. However, papers that do include individual fixed effects in their individual-level regressions such as Giuntella and Wang (2019) do this so they can focus exclusively on within-worker variation in robot exposure. Lastly, while the previous fixed effects have been used to account for time-invariant differences within panel data, one of the most significant differences is the changes over time. Therefore, papers commonly include time fixed effects, usually by year to account for the common trends over time. However, these trends over time may not be common across all observation groups. Therefore, some papers opt to combine time fixed effects with other fixed effect types to control for an additional level of potential variation. For example, Acemoglu et al. (2023) add both industry-year fixed effects, and region (municipality)-year fixed effects in some of their regressions. This allows them to account for both time-invariant and time-variant changes across industries and regions, hopefully leading to higher quality estimates of the effect that robots have on wages.

*Control variables*

There are also further variables that researchers try to control for to try and mitigate omitted variable bias within their specifications. Most control for multiple variables at a time according to what the researchers believe may play an important part in explaining the relationship between robots and wages, and there are some clear popular choices across the literature. One of these is the population, usually of the

region or country depending on the data being used. This is mostly due to the population being vastly different across each region, which may influence factors like the labour demand or supply and therefore wages. Along with population, the gender share is also often controlled for (if individual-level data is used, an indication of gender for the observation is used instead), possibly due to factors such as the gender pay gap, and the job differences across genders which may lead to different impacts on wages aside from robots. Similarly, age, skill level (or education as a proxy), and ethnicity (or foreign worker) shares are also often present in the set of control variables selected by a paper. Lastly, while the previous control variables I mentioned are all related to the underlying working population, industry or firm-specific control variables are also used. The two most common are import and capital exposure. Exposure to imports usually focuses on large manufacturing countries like China and is usually a good indication of the amount of offshoring within a region, which therefore influences wages but is not directly related to the level of robot intensity. For the same reason, researchers also include exposure to other forms of capital other than robots, such as computers or AI. Again, these are all factors that may influence a region or industry's average wage level that are not influenced by robot use, and as papers like Shang (2022) state, these are all included to meet the underlying assumption in their specifications that without robot adoption changes, wages would be at similar levels across observation panels. It is also important to note that papers commonly present both estimates with and without fixed effects and other control variables to evaluate their underlying assumptions and the robustness of their results. Additionally, they may remove some fixed effects or control variables as a basic way to see which variables have significant impacts on results.

*Instrumental variables*

Another quite common approach is instrumental variable analysis or two-stage least squares (2SLS) to address potential endogeneity biases. There may be related reasons why an industry or firm adopts robots, which may also impact their wages. An example Acemoglu and Restrepo (2020) give is unions pushing for higher wages, which may push some firms to raise wages, and others to instead increase robot use. Additionally, regional shocks to wages or labour demand, like the 2008 global financial

crisis or the COVID-19 pandemic, have impacts on business decisions like robot use. Therefore, to account for this, researchers undertake instrumental variable analysis. The most common instrument used is the robot measures from a foreign country. A good instrument is strongly correlated with the variable you are instrumenting while not correlated with the outcome of interest or any omitted variables. Foreign robot use is used as it acts as a proxy for the overall technological improvements in robots. Papers like Acemoglu and Restrepo (2020) instrument their US robot variable with a combination of five European countries[7] which are all high-income, advanced economies although they are all more technologically advanced than the United States. Therefore, US robot data has historically followed a similar path to the robot data from these countries, although at roughly a twenty per cent lower intensity. This is beneficial as the fact that they are ahead of the United States means that the instrument follows global technological advancements, rather than impacts made by the US. Additionally, because they are not one of the top trading countries with the US, it is unlikely that robot use in the selected European countries impacts wages in the US, making it an overall strong and valid instrument. I counted thirty-one other papers that follow Acemoglu and Restrepo (2020) in using foreign robots as an instrument in their 2SLS analysis, and overall most found similar results to their non-instrumental analysis. Acemoglu and Restrepo (2020) themselves still find negative and precise estimates, just like their other analysis. However, foreign robots are not the only variable used. Graetz and Michaels (2018) construct two instrumental variables, and this approach is followed by multiple papers (Aksoy et al., 2021; Bekhtiar et al., 2021; Stemmler, 2020). The first is a "replicability" measure where they take occupational data from 1980 (before robots became popular), and consider the jobs replaceable if, by 2012, robots could've replaced the job. They then use the fraction of the 1980s hours worked that became prone to robot replacement as the instrument. Their second instrument makes use of the advancements in robotic arms. This instrument is a measure of how many occupations required "reaching-and-handling" compared to other physical tasks in 1980. Both of these approaches were found to be strong instruments in Graetz and Michaels' (2018) analysis. Other, less common approaches include Chiacchio et al. (2018) using the intensity of Employment Protection Legislation, and Eggleston et al. (2021) using variation in the planned amount of robot

[7] They call this the EURO5 which comprised of Denmark, Finland, France, Italy, and Sweden.

subsidies across Chinese prefectures, showing the wide range of instrument options. It should be important to note that, while there were varying types of instruments and levels of instrument strength, I found the common trend was that the paper's instrumental variable analysis was not significantly different to the previous, non-instrumental analysis.

*Other specification differences*

The last common specification changes commonly used within an estimate's specification include clustering standard errors, sample weighting, and addressing outliers. Clustering standard errors is usually done at the level of the data that is a paper uses, so if it uses regional-level data, they often cluster by region. However, some papers opt to cluster by slightly larger by a larger scope. For example, Giuntella et al. (2022) use commuting zone-level data from the United States but cluster their standard errors by US state. They do this to account for "spatial correlation" between commuting zones, but because they conduct long difference analysis, they do not have enough observations to cluster by commuting zone and therefore group them by which state they belong to.

Some papers also weight their regressions, often by the sample population. This is because countries, regions, firms, or industries can all have varying population levels, but those that have a higher population likely give more insight into the effects of robots on wages. For example, when controlling for other factors, the effect of an increase in robot intensity within a commuting zone with a population of several million will give more information on the effect on the country, compared to a commuting zone with a population of 100,000. Other weights commonly used are employment level and shares of the population or employment level.

Lastly, a key factor also considered by many papers is the issue of outliers. They often address outliers by simply removing them, usually as a robustness check of previous analysis to see if outliers have a significant impact. For example, in some of their estimates, Acemoglu and Restrepo (2020) remove the commuting zones that fall within the top one per cent of exposure to robots. The robot variable is the factor most often addressed by researchers. While most opt to remove outliers, Gan et al. (2023)

instead address outliers in all their continuous variables by winsorizing at the one and ninety-nine per cent quantile levels.

### 2.3 Broad automation/technology and wages

Throughout my literature collection, there were of course papers I was unable to use because they focused on other forms of automation capital, or general technology measures which encompass many different forms of capital. For example, Dinlersoz and Wolf (2024) look at technology in a more general sense. While this does include robots, it also includes other forms of automation like vehicle guidance systems, sensor-based inspection, and testing equipment, along with non-automated technology like computer-aided-design, -engineering and -manufacturing which may aid automation. They average responses to survey questions[8] regarding technology use to construct an overall technology index. Using their technology index, they look at how United States labour markets are impacted at the manufacturing plant level. They find that the impacts on wages are positive, with a one per cent increase in the technology index leading to a 0.08 to 0.09 per cent increase in wages. There are also papers that I did include in my analysis which also include other regressions I was unable to use as they did not focus on robots. For example, while Webb (2020) analysed the effects of robots on wages, their main focus was on artificial intelligence. They use patent data to collect information on AI use and use this to measure the impact of AI on wages at the industry level within the United States, with a particular focus on wage inequality. They find that AI had a negative relationship with wages overall, but also estimate that AI reduces income inequality between the top ninety and bottom ten per cent of earners, although the top one per cent is left unaffected. While these results may not seem relevant to this meta-analysis study, they can provide useful insights into other aspects of how firms may adapt over time, especially because robots are not the only type of technology used in a business. So, while this study focuses on robots and therefore automation or other forms of capital are often

[8] The survey included questions about the degree of automation, value of investment in automation, future reliance on automation, and any plans for future automation.

controlled for in regressions, their influence on wages, along with many other labour market outcomes should not be ignored.

### 2.4 Robots and other outcomes

Similar to how researchers are often interested in how other automation technologies influence wages, researchers also consider how robots affect factors other than wages. The most prominent outcome of interest within the literature is the impact on employment. Acemoglu and Restrepo (2020) also include employment in their commuting zone-level analysis and find that an extra robot per thousand workers reduces the employment-to-population ratio in the United States by 0.2 percentage points. Dauth et al. (2021) find little evidence of a general robot-employment effect, which can be explained when these results are broken down by different employment types, as manufacturing employment suffers a negative effect, which is then counteracted by positive employment effects in the service sector. Lastly, while Graetz and Michaels (2018) find positive impacts on wages, their analysis shows negative, although insignificant, results on employment, with harsher impacts on low-skilled workers.

Another alternative area of interest in the literature is the wage gap. While my focus has been on wages themselves, wage gaps between genders and skill levels are also a common focus. As previously mentioned, Aksoy et al. (2021) is one of the papers that focus on the gender wage gap. They found that while both male and female wages do increase with robot use, the gap between them also increases, as they estimate that a ten per cent increase in robotization leads to a 1.8 per cent increase in the gender pay gap. They go on to attribute these pay differences to countries with prevalent gender inequality, where robots disproportionately help men in medium and high-skilled occupations. However, these results are not replicated by Ge and Zhou (2020), who focus on US labour markets. They conducted regional-level analysis between 1990 and 2015, and instead found that robots reduce wages, although this decrease is greater for males compared to females, so the gender pay gap decreases. Alternatively, Han (2022) looks at the pay gap between high- and low-skilled workers.

Using data from China, they conducted a regional-level analysis and found that the use of industrial robots results in an expansion of the skill wage gap.

Lastly, some more niche, or indirect topics include the more psychological impacts concerning mental health, but also consider political beliefs and attitudes on gender roles. Similar to the gender wage gap, Wang et al. (2022a) focus on how gender role attitudes are impacted by robot use within China. They find that robot use leads to a more equal viewpoint on gender roles, especially for younger, female, and urban workers. These changes are tied to greater female employment and wage opportunities, improving their relative economic status. Abeliansky and Beulmann (2021) instead look at the mental health of workers and found that, on average, an increase of one standard deviation of robot intensity results in a 0.11 standard deviation decrease in mental health. This mental health measure is an index made up of twelve mental health-related survey questions, and the fear of an individual's financial well-being worsening drove this decrease, with males and those in routine occupations affected the most. Lastly, Chugunova et al. (2021) focus on changes in voter turnout due to robots. Their motivation is that robot labour market impacts (good or bad) may have a psychological effect on an individual, potentially resulting in different political beliefs and therefore may affect overall voter turnout. The paper finds that an increase of one robot per thousand workers leads to a 0.64 percentage point decrease in US presidential election voter turnout. Furthermore, individuals who are directly at risk of automation are fifteen per cent less likely to vote. The researchers' reasoning behind these results is that while potential voters perceived the government as important in addressing any trade shocks (their analysis of Chinese trade shocks on voter turnout was insignificant), the government is not seen as influential in automation trends, so voters may choose to abstain. Again, these results are not relevant for the analysis in this paper and therefore I will not mention them again. However, it is still important to consider the many other impacts that robots have when discussing their impacts on wages, especially when it comes to any potential policy choices.

## 2.5 Meta-analyses

Across the literature only one thing is consistent, and that is inconsistency. Be it the plenty of positive, negative, and inconclusive results on the effect of robots on wages, or the methodology behind these results and how they may affect them. I am unable to come to a clear conclusion on what the overall effect of robots is on wages based on this literature review alone. Therefore, I conduct a meta-analysis to help me achieve an overall conclusion on what the literature believes is the effect that robots have on worker wages. A meta-analysis is the quantitative review of previously reported findings within a subject area of interest. This research type is an essential part of fields such as medicine, education, and psychology, and is growing in popularity within the field of economics. Irsova et al. (2023b) use Google Scholar to estimate that 107,000 meta-analyses were published in 2022 alone. This popularity can be attributed to their ability to, as Stanley and Doucouliagos (2012) put it, “integrate conflicting research findings and to reveal the nuggets of ‘truth’ that have settled to the bottom”. Also, through methods like Meta-Regression Analysis (MRA), it is possible to deconstruct estimates to see how methodology choice may influence results. Stanley and Doucouliagos (2012) go on to further state that “Meta-analysis is the most objective and statistically rigorous approach to systematic reviews, which, in turn, provides the evidence for the evidence-based practice movement.” A systematic review is a more comprehensive literature review, as it involves a detailed search plan that attempts to find and analyse all relevant studies to the chosen subject matter (Uman, 2011).

There have been previous meta-analyses on associated topics to the one I am undertaking. One of the first was by Terzidis et al. (2019). Their study conducts a meta-analysis focusing on the effect that general automation technologies and trade have on multiple aspects of the labour market. Using 623 technology and 1094 trade elasticities from seventy-seven studies, they find that overall “technology and trade benefit both wages and employment in a statistically significant and economically meaningful way”. Like the present study, they also recognise the significant heterogeneity across different estimates due to the different focuses of each study (for example, some focus on employment and others on wages), and the different subsets used as most empirical studies only focus on one. Therefore, they perform multivariate meta-regression analysis and find that their overall conclusions are dependent on

numerous factors. The most crucial factor is the skill level. They suggest a “skill bias” where the gains from technology disproportionately benefit high-skilled workers compared to lower-skilled ones. However, robots are only a subset of their overall definition of technology. Terzidis et al. (2019) count automation, computerization, innovation, research and development, technical change and robotization all as technological progress, and therefore cannot conclusively state the specific impact robots have on wages.

Tier (2022) does look at robots specifically but analyses the impact of robots on employment rather than wages. Overall, he finds little evidence of a clear relationship between changes in robot use and employment. He further examines how the size of the robot-employment connection changes with several factors such as population size and gender shares, but he found these to only have a small effect. These results show multiple reasons behind the power of meta-analyses. First, they provide a valuable quantitative summary of the literature, often in the form of a coefficient estimate, allowing for an estimated effect when the literature may be unclear. Secondly, various inconsistencies may only be revealed through meta-analytic methods like meta-regressions. This is because estimates from a single paper may not show the whole story, as specifications or methods across papers are often different, as shown by my literature review. It is only when all the estimates available are combined that the effects of factors like skill bias or gender shares become clearer, especially because these effects can be quantified through meta-regression coefficients.

Lastly, some papers accomplish a similar goal to mine that were conducted simultaneously with this paper. Schneider et al. (2023) also conduct a meta-analysis focusing on the relationship between robots and wages. They use 2,143 estimates from fifty-three studies[9] and find the overall effect that robots have on wages is insignificant (both statistically and economically) and near zero. Dario et al. (2024) also conducted a meta-analysis, looking at the effect that robots have on both wages and employment. They use 195 estimates from nineteen studies, and like Schneider et al.

[9] Thirty-four of these studies overlap the studies included in this meta-analysis. This is in part due to the slightly different search methods used in gathering the studies, but also due to other factors like my choice to not include estimates with interaction terms.

(2023), Dario et al. (2024) report minimal impacts of robots on wages overall across the current literature. Because robots are a trending topic, it should not be surprising that similar studies have by conducted, however, I conducted this study independently of these two papers. They were not yet made available through pre-print when I began my research, and both are yet to be published. Therefore, unlike the studies by Terzidis et al. (2019) and Tier (2022), I was not influenced by any results or methods conducted by Schneider et al. (2023) or Dario et al. (2024). This paper should be perceived as a completely independent study looking at the same topic, although, compared to these two other papers, this study does make use of more models that provide more estimates and robustness checks on the shared research question. Overall, the combination of results from these three papers should simply lead to more complete and reliable results on the effect that robots have on wages across the current literature.

The typical process of a meta-analysis first involves the extensive literature collection, which I have outlined above in section 2.1. After the studies have been collected, the data collection process begins. Unlike most typical economics or econometrics studies, meta-analyses create their own datasets by collecting the values and characteristics of estimates. Irsova et al. (2023b) recommend at least two researchers independently collect data, to minimise mistakes in the manual “coding” process. For this thesis, I worked on the coding of this meta-analysis with two students working towards their PhD in economics. Independently from myself, they each worked on half of the papers, while I coded all of them. I then compared the results, adding more reliability to my dataset as any mistakes are more likely to be identified and corrected. Once the data collection is completed, the statistical analysis is undertaken. I outline the analysis I conducted in this paper in depth in the following section.

## 3. Analysis methods

### 3.1 PCC & Fisher's Z

Many papers use different measures for both robots and wages. As a result, it is not appropriate to compare these estimates directly. For example, Acemoglu et al. (2020) use hourly wages, whereas César et al. (2022) use annual income as the dependent variable in their regressions. Because of the differences in magnitude between these variables, directly comparing the coefficient sizes will not accurately reflect any differences in the estimates between the two papers. Therefore, it is necessary to transform the estimates into partial correlation coefficients (PCC) to compare them.

PCCs can be interpreted as a correlation between two variables ranging from -1 (perfectly negative) to 1 (perfectly positive). This removes the problem from the previous example, as the use of correlations will no longer be affected by different scales and measures allowing for comparison between studies.

Estimates can be transformed into PCCs using the following formula:

$$PCC_i = \frac{t_i}{\sqrt{t_i^2 + df_i}}$$ [10]

However, according to van Aert (2023), using PCCs violates two of the assumptions required for fixed effect and random effects models. Firstly, many of the meta-analysis models assume that each study's effect size follows normal distributions (Jackson & White, 2018). Therefore, because PCCs range between [-1,1], the normal distribution assumption is violated and gets worse the further the true PCC is from zero. The second violation mentioned by van Aert (2023) lies in the equation for PCC's sampling variance:

$$Var(PCC_i) = \frac{1 - PCC_i^2}{df_i}$$ [11]

Because the PCC estimate is in the variance equation and is therefore dependent on the PCC estimate, the assumption of known sampling variances is violated. There are

[10] Where $t_i$ is the t-statistic for the $i^{th}$ estimate, and $df_i$ is the degrees of freedom for the $i^{th}$ estimate.
[11] Where $PCC_i$ is the partial correlation coefficient for the $i^{th}$ estimate.

three problems with this. The first is that, because the standard error is not independent of the effect size, the effect size to standard error ratio may not follow a t-distribution. Secondly, many of the publication bias tests depend on the correlation between effect size and standard errors. Because PCC standard errors depend on the effect size, these tests will be biased. Lastly, estimation methods such as weighted least squares that weight observations using the standard error (or variance) will be biased.

Therefore, I apply a Fisher's Z transformation to the PCCs which van Aert (2023) claims aligns more with the assumptions of meta-analytic models compared to PCCs.

A PCC can be transformed into Fisher's Z using:

$$\mathcal{Z}_i = \frac{1}{2} \times \log\left(\frac{1 + PCC_i}{1 - PCC_i}\right)$$

And the variance for Fisher's Z values can be computed using:

$$Var(\mathcal{Z}_i) = \frac{1}{N_i - 3 - (M_i - 1)}$$ [12]

The Fisher's Z transformation is variance-stabilizing which means that the above variance formula does not rely on the Fisher's Z/PCC value, reducing some of the previously mentioned problems with PCCs. Additionally, Fisher's Z values are distributed closer to a normal distribution than PCCs as estimates are allowed to range from [-∞, ∞]. This is especially true if the true PCC is non-zero.

However, it is important to note that problems with PCC values are not completely solved with a Fisher's Z transformation, only improved upon. The assumption of known variances is still not solved as the above equation is only a large sample approximation. However, it is a better estimate of the variance, as even with small sample sizes they are still considered accurate (Borenstein & Hedges, 2019). Also, the Fisher's Z values will not truly follow a normal distribution (Fisher, 1950), but they go towards normality much faster than PCCs as the sample size increases.

[12] Where N is the sample size for the $i^{th}$ estimate, and M is the number of independent variables in the linear regression model for the $i^{th}$ estimate.

However, van Aert (2023) also states that it is common practice to revert estimates achieved from meta-analytic models back into a PCC for interpretation. This can be achieved using the following formula:

$$\mathrm{PCC}_i = \frac{e^{2\times Z_i} - 1}{e^{2\times Z_i} + 1}$$

The PCCs I obtained can then be interpreted using the thresholds from Doucouliagos (2011). He used a large sample of over 22,000 partial correlations from literature in a wide range of economic fields. Using these estimates, he developed guidelines for PCC interpretation. He defined a small effect as a PCC above 0.07, and medium effect as above 0.17, and a large effect as 0.33.

### 3.2 OLS, fixed effect, random effects & UWLS

Next, I aim to synthesize the individual estimates from my sample of studies to determine the overall relationship between robots and wages.

The first option is to just take the simple average of the Fisher's Z for all the estimates, which can be done by running an Ordinary Least Squares (OLS) regression with the Fisher's Z values on a constant:

$$Z_i = \mu_i + \varepsilon_i$$ [13]

However, while OLS is consistent and unbiased, it is inefficient, and unreliable in hypothesis testing in this current context. The efficiency can be increased by using weighted least squares (WLS) as some estimates have more information about the true effect than others and should therefore be weighted higher. This can be due to some studies having lower standard errors, so they have less sampling error and will therefore be closer to the true effect they are trying to estimate.

The first weighting method is the Fixed Effect (FE) Model (also called the Common Effect model). The FE model assumes that the effect sizes all come from a single

[13] Where $\mu_i$ is the true effect size for the robot-wages relationship, and $\varepsilon_i$ is the i$^{th}$ estimate's error term.

homogenous population, so that $\mu_i = \mu$ for all i. Thus, all studies try to estimate the same true effect size between robots and wages. Because there is only one true effect size, the explanation that this model gives as to why there are differences in estimates is simply sampling error, which cannot be removed as studies can only use samples from the overall population. Therefore, in the FE model, because the standard error reflects the level of sampling error (a higher standard error implies higher sampling error), those effect sizes with a higher precision[14], should have a higher weight when averaging across all estimates. This is because, with less sampling error, those estimates will be closer to the true effect size. The fixed effect estimator does this by using inverse variance weighting.

The second weighting method I include is the Random Effects (RE) model. While the FE model assumes a single true effect size, the RE model allows the idea that each reported estimate has its own true effect size the study is trying to estimate ($\mu_i$). This approach may be considered much more reasonable compared to the FE model. Different estimation methods, datasets, and research objectives are among the many explanations why there is little reason to expect a treatment to have the same effect on all units. If there are different effects, then the assumptions of the FE model are violated, and the RE model, which assumes there is a distribution of true effect sizes, may be more appropriate. The RE model tries to estimate the overall mean ($\mu$) from this distribution.

However, this now means that an observation's error term is made up of two components (Borenstein et al., 2010). First, the estimates reported in each study differ from their true value due to sampling error (within-study variance, as in the FE model). The second is that even if each effect size perfectly reflects the true value they are trying to estimate, reported effect sizes may be different from the distribution mean (between-study variance), as an important assumption of the RE model is that the distance between the distribution mean and a particular estimate's true effect is random (Thompson et al., 2001).

To account for this second source of heterogeneity, the variance of the true effect sizes distribution is estimated ($\tau^2$ or tau-squared). The model still uses inverse variance

[14] precision is 1/SE, so lower SE means higher precision.

weighting, but under the RE model, the total variance for an observed effect size is now the observation's variance around its true effect size ($Var(Z_i)$), plus the variance of the true effect size around the distribution mean ($\tau^2$).

The RE and FE weighting schemes are shown in the following equation:

$$\frac{Z_i}{\omega_i} = \frac{\mu_i}{\omega_i} + \frac{\varepsilon_i}{\omega_i}, i = 1, 2, \dots, N$$

$$\text{Where: } \omega_i = \begin{cases} \text{Fixed Effect: } \sqrt{Var(Z_i)} \\ \text{Random Effects: } \sqrt{Var(Z_i) + \tau^2} \end{cases}$$

I also include an alternative estimator that gives the same effect estimates as the FE model (it uses FE weights) but builds in more flexibility when estimating heteroskedasticity. This is what Stanley and Doucouliagos (2017) call unrestricted weighted least squares (UWLS).

Additionally, the number of estimates collected from each study can vary greatly. This means that a study with many estimates can take up a large overall weight in the model, possibly skewing the resulting estimate for the mean. This can lead to biased results if a study with imprecise results (and therefore likely further away from true effect size) reports many estimates. Even though each estimate receives little weight, with many estimates the overall weight can be substantial, potentially biasing the estimated overall effect. Therefore, I adjust the previously mentioned weighting models (FE, RE & OLS) to also include weighting by the inverse of the number of estimates per study. This gives each study equal weighting. This new weighting scheme is shown below:

$$\frac{Z_i}{\omega_i} = \frac{\mu_i}{\omega_i} + \frac{\varepsilon_i}{\omega_i}, i = 1, 2, \dots, N$$

$$\text{Where: } \omega_i = \begin{cases} \text{Fixed Effect: } \sqrt{Var(Z_i) \cdot n_{i \in s}} \\ \text{Random Effects: } \sqrt{(Var(Z_i) + \tau^2) \cdot n_{i \in s}} \end{cases}$$ [15]

[15] where $n_{i \in s}$ is the number of estimates from study s where estimate i is taken from.

One of the crucial factors to consider when performing a meta-analysis is publication bias. Publication bias is when estimates reported in the literature are distorted from the population of the estimates. This can happen when researchers selectively report results, emphasizing the statistically significant estimates. In our case, because the effects of robots are a "hot" topic with many news articles mentioning that the increasing prevalence of robots in the workforce is cutting wages (Claburn, 2022; Harthorne, 2017), there is an incentive to produce noteworthy (and potentially negative) results to be referenced in popular articles.

A common way of checking for publication bias is through a funnel plot, which plots effect size estimates on the horizontal axis, with the standard error (descending) on the vertical axis. Assuming there is no publication bias, the most precise estimates should be close to the pooled effect size as they have the least amount of sampling error and will therefore be closer to the true effect. As the standard error increases, estimates will stray further away, forming a funnel shape as the higher sampling error produces a wider range of estimates. In the presence of publication bias (and other small-sample biases) the funnel may be asymmetrical around the mean. This is because estimates with larger standard errors are more likely to suffer from publication bias. If there is publication bias, then smaller studies which show statistically insignificant results or results that fail to fit a certain narrative do not get published, which removes a portion of data points in the funnel plot leading to asymmetry. Additionally, to be published, researchers may re-estimate or adjust their methods to inflate their effect sizes to make up for their high standard errors leading to further asymmetry.

The asymmetry can be assessed using a funnel asymmetry test (FAT) on our previous models. I expand on this by using the FAT-PET-PEESE method, which shows both funnel plot asymmetry, and a bias-corrected estimate (Stanley & Doucouliagos, 2014). The formula below shows the FAT-PET-PEESE method.

$$\text{PET: } \frac{Z_i}{\omega_i} = \frac{\mu_i}{\omega_i} + \beta \frac{SE(Z_i)}{\omega_i} + \frac{\varepsilon_i}{\omega_i}$$

$$\text{PEESE: } \frac{Z_i}{\omega_i} = \frac{\mu_i}{\omega_i} + \beta \frac{Var(Z_i)}{\omega_i} + \frac{\varepsilon_i}{\omega_i}$$ [16]

[16] Where $\omega_i$ is determined whether RE or FE methods are being used.

As shown above, I first regress effect sizes on their standard errors. This is the Funnel Asymmetry and Precision Effect Tests or FAT-PET regression. The reasoning behind regressing on standard errors is that, to get a statistically significant result as the estimate becomes less precise, a more extreme effect size is required. Therefore, if publication bias is present, it is expected that absolute effect sizes and standard errors to have a positive linear correlation. In this case, according to Stanley (2005), the intercept can be interpreted as the publication bias corrected estimated true effect. This is because the intercept shows an estimate with a standard error of zero, so it is the expected effect size with no sampling error. Therefore, with standard error as a control variable, the intercept should show a publication bias-corrected estimate of the true value. Additionally, the level of correlation between the standard error and the effect sizes (or the coefficient for the SE variable), can be interpreted as the size of the publication bias.

Following the FAT-PET regression, if I reject the null hypothesis that the estimated true effect is zero, then the Precision Effect Estimate with Standard Error (PEESE) regression is used to estimate the true effect. PEESE regresses effect sizes on their standard errors squared (their variance). The linear model of the PET regression is known to give biased estimates when the true effect does not equal zero. Therefore, Stanley and Doucouliagos (2014) recommend a quadratic specification.

If the true effect size is non-zero, very precise estimates (those with small standard errors) will produce a significant estimate without any need to selectively report estimates. It is only as the standard error increases that the need for statistical significance dominates, so the relationship between standard errors and absolute effect sizes becomes positive and linear. This change from no relationship (horizontal relationship) to a positive linear relationship more closely resembles a parabola from a quadratic relationship, so PEESE is used (Stanley & Doucouliagos, 2012).

However, it is important to note that while the PEESE is used for estimating effect sizes, it is not used to test for publication bias. Instead the results from the PET analysis are used to determine the presence of publication bias in the data (Stanley & Doucouliagos, 2012).

### 3.3 Additional non-linear estimators – WAAP, TOP10, Stem & Endogenous Kink

The previous methods, apart from PEESE regressions, assume linearity when it comes to the relationship between standard errors and effect sizes. However, Stanley and Doucouliagos (2014) suggest that this relationship in general is instead a complex non-linear function. As briefly explained in the PEESE model, a reason for this may be that once statistical significance is reached, there is little incentive for researchers to adjust specifications to obtain significance, so the relationship between publication bias and standard errors breaks down. Therefore, I include multiple alternative methods that allow for a non-linear relationship.

One non-linear method I use is the weighted average of adequately powered estimates (WAAP) developed by Ioannidis et al. (2017). This is based on the previously mentioned unrestricted weighted least squares model (UWLS), which gives the same estimate as FE, but adds flexibility when estimating the level of heterogeneity (Stanley & Doucouliagos, 2017). Importantly, the WAAP model only uses estimates with a power of at least eighty per cent, which is designed to reduce the impact of selective reporting (publication bias). Simulations by Stanley et al. (2017) found that when some study results have been selected to be statistically significant, WAAP had a smaller bias than other methods such as RE.

Another method I implement is the top10 method by Stanley et al. (2010). This is simply an average of the top ten per cent most precise estimates. Stanley et al. (2010) use Monte Carlo simulations and actual research examples to claim that this method decreases publication bias and improves efficiency compared to traditional summary statistics that use all observations.

I also implement the stem method by Furukawa (2019). This method is similar to the first two techniques in the sense that it only uses a portion of the most precise estimates, but in this case, it is endogenously determined. The name comes from the fact that the most precise estimates come from the “stem” of the funnel plot. The model determines which portion to use by minimising the sum of the bias that is gained when adding more imprecise estimates, and the variance which decreases as more observations are added. Although this estimator is more robust compared to other options, it also comes with the disadvantage of having, on average, larger confidence intervals.

Lastly, I also include the endogenous kink method (Bom & Rachinger, 2019). As previously mentioned, once statistical significance is reached the linear relationship between publication bias and standard errors weakens. Therefore, this method uses a piecewise linear model where estimates are regressed on standard errors. This kink in the model occurs at an endogenous cutoff value of the standard error where publication selection is unlikely to occur. It is determined as a function of a first stage estimate from the FAT-PET-PEESE model and then aims to improve upon it. At the kink (once standard errors are sufficiently small), the linear relationship is replaced by a constant term.

### 3.4 Endogeneity robust estimators – MAIVE, p-uniform* & caliper test

It is important to note that all the models up to this point assume that the correlation between estimated effect sizes and their standard errors is solely due to publication bias. However, this may not always be the case. Because estimated effect sizes and their standard errors in a study are both still estimates, methodology choices can impact both the estimate and the standard errors, which can lead to endogeneity bias for the standard error.

Therefore, I also include instrumental variable analysis as an additional robustness check. I do this using the MAIVE estimator (Irsova et al., 2023a). The MAIVE instruments the standard error with the inverse of the square root of the sample size used in the study. This is a viable instrument as it is correlated with the standard error (as $SE = \sigma/\sqrt{n}$, where n is the sample size), but is not likely to be greatly correlated with the specific estimation techniques a study uses. For the MAIVE, I use the PET-PEESE estimator with no weights (as opposed to inverse-variance weighting) as Irsova et al. (2023a) find that, through simulations, this specification worked the best.

Another estimator I include in my analysis is the p-uniform* technique (van Aert & van Assen, 2018). This is offered as an improvement on their previous p-uniform method (van Assen et al., 2015). This estimator is also robust to endogeneity bias between the effect sizes and standard errors because it does not directly rely on the relationship between the two variables. Instead, p-uniform* relies on the assumption that the distribution of p-values will be uniform when the hypothesized effect size is set equal

to the true value. It then computes the effect size that will make the corresponding p-values uniformly distributed. I chose the p-uniform* over the p-uniform because van Aert and van Assen (2018) claim it is more efficient and takes into account between-study heterogeneity[17]. It does this by allowing for the probability of a significant and non-significant study to be published to be different (although both probabilities still need to be constant) and takes this into account when calculating the conditional probabilities.

The third endogeneity robust test I include in my analysis is the caliper test by Gerber and Malhotra (2008). Like p-uniform*, this test is endogeneity robust because it does not directly use the relationship (linear or non-linear) between effect sizes and standard errors, and instead looks at the distribution of t-statistics. However, this is simply another test for publication bias and does not produce a bias-corrected mean effect. This test assumes that publication bias shows up around specific statistical significance thresholds such as the 1.96 t-statistic value[18]. Therefore, if publication bias exists, then even with a narrow caliper at the threshold there will be an uneven number of results on the significant side instead of a smooth distribution. I use a graphical representation of the distribution of reported t-statistics, along with the number of t-statistics reported above and below each threshold at varying caliper widths to perform a "caliper test." I also include regression outputs where the estimated coefficient represents the imbalance in the per cent of estimates above and below the threshold[19].

### 3.5 P-hacking robust estimators – RTMA & MAN

When defining publication bias, it is common to define it as only statistically significant studies being published compared to non-significant studies. This is referred to as selection across studies (SAS), however, publication bias also encompasses selection within studies (SWS). SWS is when researchers choose which estimates of the they

[17] p-uniform, like FE estimators, assumes a single underlying effect.
[18] The t-statistic value for a 5% significance level.
[19] The value of the reported coefficient shows the percentage of estimates above the expected 50% e.g. a value of 0.15 means that 65% of estimates are above the threshold and 35% of estimates are below.

obtain to submit for publication. This is known as “p-hacking,” and it results in different estimate distributions to what would occur if the estimates reported followed the originally planned analysis.

Therefore, I include an additional method that accounts for both SAS and SWS developed by Mathur (2024). This estimator is called a right-truncated meta-analysis or RTMA. RTMA uses the published non-affirmative[20] estimates to model the underlying distribution of estimated effects following the originally planned analysis, so the meta-analytic mean can be estimated consistently.

Mathur (2024) also introduces another estimator that accounts for p-hacking, which is a meta-analysis of non-affirmative results (MAN). This method assumes a “worst-case” scenario with extreme levels of publication bias and accounts for this by only retaining the non-affirmative studies when conducting a standard meta-analysis (running FE and RE models).

### 3.6 Bayesian model averaging

So far, none of the previously mentioned models incorporate heterogeneity in estimated effects due to study, estimation, and data characteristics. To investigate this type of heterogeneity, a meta-regression can be performed. A meta-regression includes a study’s characteristics as covariates in the previously discussed models such as FE or RE regressions. This not only estimates the effect of robots on wages after correcting biases such as publication bias, but also allows me to see what study characteristics impact estimated effect sizes. To be able to do this, I collected fifty-seven variables that I believe capture the distinguishing factors associated with each estimate and study. In section 2.2, I explain the differences across estimates, which are reflected in the variables I collected. However, the problem is deciding which of these variables to include in the model. You can of course include every variable in the model, but this will be inefficient. Additionally, the selection of which covariates to use is complex and subjective.

[20] Affirmative means statistically significant positive point estimates

Therefore, Steel (2020) recommends conducting Bayesian model averaging (BMA). BMA addresses the problem that the estimated heterogeneity effects will differ depending on model specification. BMA looks at many different combinations of explanatory variables. I added fifty-six of the fifty-seven variables to the BMA model (one was removed to be used as a reference variable). However, this means there will be $2^{56}$ possible combinations of covariates, so a Monte Carlo Markov Chain (MCMC) sampling algorithm by Zeugner and Feldkircher (2015) is used to go through only the most important models.

The model also requires two additional specifications for the prior beliefs used in the Bayesian framework. As suggested by Eicher et al. (2011) the Zellner's g-prior I used is the Unit Information Prior (UIP). The g-prior looks at the prior belief that the coefficients are zero, and the UIP sets the prior so that it contains as much information on coefficient probabilities as is typically found in a single observation. The other prior necessary for BMA is the model prior, which specifies prior beliefs for each model's probability. For this, I use a dilution prior (George, 2010) which accounts for collinearity by adjusting the weight based on the determinant of the correlation matrix[21]. The results from this BMA model then allow me to see how each of the fifty-seven variables effect estimates across studies, and which are the most important in explaining why studies achieve different results.

## 4. Results

In this section, I perform a quantitative review of the literature. I first look at the characteristics of the data I have collected. I then estimate the overall effect of robots on wages as estimated by the literature. Lastly, I use Bayesian Model Averaging to try and understand the factors responsible for the observed heterogeneity in estimates.

[21] The Collinearity matrix is an overall measure of collinearity.

### 4.1 Data characteristics

From fifty-two papers I collected, I obtained 2,586 individual estimates. As explained in Section 2.5, for each estimate, another researcher and I coded fifty-seven variables. A brief description of these variables can be found in Table 1, along with basic summary statistics of the variables. Because all the variables (apart from publication year) are binary variables, the mean simply shows the percentage of estimates that have the characteristic.

*Table 1: Data characteristics descriptions*

| Variable | Description | Mean | SD |
|---|---|---|---|
| Wage variable: | | | |
| Wage Logged | 1 if dependent variable is logged | 0.895 | 0.306 |
| Hourly | 1 if dependent variable is measured hourly | 0.294 | 0.456 |
| Weekly | 1 if dependent variable is measured weekly | 0.115 | 0.319 |
| Monthly | 1 if dependent variable is measured monthly | 0.173 | 0.378 |
| Annually | 1 if dependent variable is measured annually | 0.308 | 0.462 |
| Wages | 1 if dependent variable is defined as wages | 0.689 | 0.463 |
| Income | 1 if dependent variable is defined as income | 0.181 | 0.385 |
| Earnings | 1 if dependent variable is defined as earnings | 0.086 | 0.280 |
| Wage-bill | 1 if dependent variable is defined as a company's wage-bill (reference category) | 0.045 | 0.206 |
| Real wages | 1 if dependent variable is measured in real terms | 0.160 | 0.366 |
| Wage differences | 1 if dependent variable is measured as a change over time | 0.524 | 0.500 |
| Robot variable: | | | |
| Robots logged | 1 if the independent variable is logged | 0.084 | 0.278 |
| Robots Absolute | 1 if the independent variable is measured as an absolute number | 0.026 | 0.160 |
| Robots relative | 1 if the independent variable is measured as a relative number, e.g. the number of robots per capita | 0.923 | 0.266 |

| Variable | Description | Mean | SD |
|---|---|---|---|
| Robots dummy | 1 if the independent variable is measured as a dummy (reference category) | 0.050 | 0.219 |
| Robots differences | 1 if the independent variable is measured as a change over time | 0.600 | 0.490 |
| Robots Contemporaneous | 1 if the independent variable is measured in the same time period as wages | 0.922 | 0.269 |
| Foreign Robots as variable | 1 if the independent variable measures robots outside of the country of the observation | 0.265 | 0.441 |
| Dataset used: | | | |
| Not IFR data | 1 if the regression does not use IFR data | 0.077 | 0.267 |
| Individual-level data | 1 if individual-level data is used | 0.170 | 0.376 |
| Firm-level data | 1 if firm level-data is used | 0.046 | 0.210 |
| Industry-level data | 1 if industry-level data is used | 0.145 | 0.352 |
| Regional-level data | 1 if regional-level data is used | 0.638 | 0.481 |
| Country/US state-level data | 1 if data is country or US state-level data is used | 0.099 | 0.299 |
| Manufacturing | 1 if the data is just covering the manufacturing industry | 0.149 | 0.356 |
| Services | 1 if the data is just covering the services industry | 0.068 | 0.253 |
| Agriculture | 1 if the data is just covering the agricultural industry | 0.028 | 0.165 |
| Young | 1 if the data is just covering young people, cut off is roughly 40 | 0.027 | 0.161 |
| Middle-aged/old | 1 if the data is just covering middle aged/old people, cut off is roughly 40 | 0.026 | 0.160 |
| Male | 1 if the data is just covering males | 0.064 | 0.245 |
| Female | 1 if the data is just covering females | 0.060 | 0.238 |
| Low/medium skilled/educated | 1 if the data is just covering people with not high education and/or not high skills | 0.086 | 0.280 |
| Highly skilled/educated | 1 if the data is just covering people with higher education and/or high skills | 0.021 | 0.143 |
| Developed countries | 1 if the data is just covering developed countries | 0.754 | 0.431 |

| Variable | Description | Mean | SD |
|---|---|---|---|
| Developing countries | 1 if the data is just covering developing countries | 0.175 | 0.380 |
| Fixed effects used: | | | |
| Time fixed effects | 1 if time fixed effects are used in the regression | 0.532 | 0.499 |
| People fixed effects | 1 if people fixed effects are used in the regression | 0.026 | 0.160 |
| Firm fixed effects | 1 if firm fixed effects are used in the regression | 0.039 | 0.195 |
| Industry fixed effects | 1 if industry fixed effects are used in the regression | 0.104 | 0.306 |
| Regional fixed effects | 1 if regional fixed effects are used in the regression | 0.749 | 0.434 |
| Country/US state fixed effects | 1 if country or US state fixed effects are used in the regression | 0.281 | 0.449 |
| Controls used: | | | |
| Population | 1 if size of the population is included as a control variable | 0.315 | 0.465 |
| Gender | 1 if gender composition indicator is included as a control variable | 0.640 | 0.480 |
| Age | 1 if age composition indicator is included as a control variable | 0.715 | 0.452 |
| Education | 1 if education composition indicator is included as a control variable | 0.662 | 0.473 |
| Ethnicity | 1 if ethnicity composition indicator is included as a control variable | 0.463 | 0.499 |
| Occupation | 1 if occupation or industry of occupation composition indicator is included as a control variable | 0.589 | 0.492 |
| Skill | 1 if skills composition indicator is included as a control variable | 0.174 | 0.379 |
| Exposure to capital | 1 if a capital indicator is included as a control variable | 0.253 | 0.435 |
| Exposure to imports | 1 if an import indicator is included as a control variable | 0.505 | 0.500 |
| Foreign Robots | 1 if a foreign robot variable is included as a control variable | 0.025 | 0.155 |
| Further regression specifications: | | | |
| Clustered SE | 1 if standard errors are clustered | 0.854 | 0.353 |
| Sample weighting | 1 if sample weights are used in regression | 0.489 | 0.500 |

| Variable | Description | Mean | SD |
|---|---|---|---|
| IV/2SLS | 1 if IV or 2SLS are used | 0.517 | 0.500 |
| Addresses outliers | 1 if exclusions or changes being made for outliers | 0.105 | 0.306 |
| Publication details: | | | |
| Published | 1 if the paper is published | 0.486 | 0.500 |
| Publication year | Demeaned year of publication of paper | 0 | 1.955 |

Notes: Description and summary statistics of coded variables. SD = standard deviation.

Table 1 shows the wide range of techniques used by studies. Looking at how they measure the wage variable, most studies log the variable (89.5%). Additionally, most studies define their dependent variable as wages (68.9%). Income and earnings were the next two most-used terms and tend to be associated with annual pay rates. Annual and hourly were the most common time frames in the data (30.8% and 29.4%). Hourly rates may have been popular in the literature as they are regularly used in manufacturing and service sectors where robot use is common.

The way robots are defined in the literature is a bit more homogeneous compared to wages. In contrast to wages, only 8.4% use logged robot values. The reason may be that most estimates (92.3%) use relative numbers (mostly per 1,000 or 10,000 workers). A common interpretation for these variables is that an increase in one robot per 10,000 workers leads to a certain percentage difference in wages, which is a simple and easy to understand result for the effect of robots on wages. Additionally, almost all estimates are contemporaneous (92.2%), which may show that many believe that the level of robots used has an immediate impact. However, the other reason could be that most (60%) regress the wage level on a change in robots over time instead.

When looking at the datasets used, regional-level data is the most common (63.8%), then individual (17%), industry (14.5%), country or US state (9.9%), and lastly firm level (4.6%). This could be a data availability issue, as IFR data is regional, and wage data is usually in a regional or industry format. Alternatively, it may give insight into the level at which researchers think the biggest impact is likely to be. They may not think that there will be much variation across individuals, rather, there may be more of an effect across, say, different commuting zones.

There is also an interest across the literature in focusing estimates on a specific subgroup. Around 24.5% of estimates group their data by industry (agriculture, services or manufacturing). 12.4% exclusively look at the impact on either females or males. 10.7% look at skill/education level, and 5.3% look at a specific age group.

Only 7.7% of estimates do not use IFR data, so the accuracy of the literature is dependent on the quality of this data. Lastly, the data mostly focuses on developed countries rather than developing ones (75.4% vs 17.5%, with 7.1% focusing on both developed and developing countries). Methods such as Bayesian model averaging may give an insight into the differences between developed and developing countries.

Many estimates try to control for potential non-robot influences by using fixed effects and control variables. Because a lot of papers use panel data, it makes sense that 53.2% of estimates include time fixed effects. However, the most common fixed effect used is regional fixed effects (74.9%), with many papers using country or US state fixed effects as well (28.1%), which allow for differences in wage levels across geographical units.

Control variables are also something that many estimates use. Age shares are the most popular control variables (71.5%), but more than half the estimates also use gender shares (64%), education level shares (66.2%), occupational shares (58.9%), and exposure to imports (50.5%) as control variables. This indicates that there is a belief in the literature these factors have a large enough influence on the robots-wages relationship that omitted variable biases need to be avoided by controlling for these variables.

Lastly, the results from Table 1 also give an insight into the different methods used by researchers. About half of the reported effect sizes are estimated using instrumental variable analysis (51.7%). While IV analysis is often used as a robustness check, it still shows that many researchers believe there may be some level of endogeneity between robots and wages. Additionally, 85.4% cluster standard errors (usually by regions), 48.9% use sample weighting (mostly by population), and 10.5% address outliers in some way. 48.6% of my estimates also come from unpublished papers, giving me plenty of observations for my BMA analysis to potentially see how publication status impacts effect sizes.

Table 2 shows summary statistics for various subsets. These show PCC values that have been retransformed back from Fisher's Z values. The unweighted section gives each estimate equal weighting, whereas the weighted section weights each observation by the number of estimates from its study. This gives each study equal weight when calculating the summary statistics. I only provide brief overviews of their summary statistics, showing the number of times the binary variable is a one, the mean PCC when the variable is a one, and ninety-five per cent confidence intervals. I look more into the heterogenous effects for each subset when looking at the results of the Bayesian model averaging in Section 4.7.

*Table 2: Collected estimates summary statistics*

| | Unweighted | | | | Weighted | | |
|---|---|---|---|---|---|---|---|
| | N | Mean | 95% CI Lower | 95% CI Upper | Mean | 95% CI Lower | 95% CI Upper |
| All estimates | 2586 | -0.016 | -0.019 | -0.013 | -0.001 | -0.004 | 0.003 |
| Wage variable: | | | | | | | |
| Wage Logged | 2315 | -0.018 | -0.022 | -0.015 | -0.003 | -0.007 | 0.000 |
| Hourly | 759 | -0.008 | -0.010 | -0.005 | 0.001 | -0.002 | 0.003 |
| Weekly | 297 | -0.011 | -0.015 | -0.008 | -0.013 | -0.016 | -0.010 |
| Monthly | 446 | -0.002 | -0.003 | 0.000 | 0.002 | 0.000 | 0.003 |
| Annual | 796 | -0.040 | -0.050 | -0.031 | -0.025 | -0.034 | -0.015 |
| Wages | 1782 | -0.005 | -0.007 | -0.004 | 0.002 | 0.000 | 0.003 |
| Income | 468 | 0.005 | 0.002 | 0.007 | -0.017 | -0.019 | -0.014 |
| Earnings | 221 | 0.005 | 0.004 | 0.006 | 0.001 | -0.001 | 0.002 |
| Wage-bill | 115 | -0.295 | -0.328 | -0.260 | -0.141 | -0.177 | -0.104 |
| Real wages | 413 | 0.005 | 0.001 | 0.008 | 0.008 | 0.004 | 0.012 |
| Wage differences | 1354 | -0.034 | -0.039 | -0.028 | -0.003 | -0.009 | 0.003 |
| Robot variable: | | | | | | | |
| Robots logged | 218 | -0.010 | -0.017 | -0.003 | -0.012 | -0.019 | -0.005 |
| Robots absolute | 68 | -0.024 | -0.041 | -0.007 | -0.002 | -0.019 | 0.015 |
| Robots relative | 2388 | -0.017 | -0.020 | -0.013 | 0.001 | -0.003 | 0.004 |
| Robots dummy | 130 | 0.002 | -0.001 | 0.005 | 0.002 | -0.001 | 0.005 |
| Robots differences | 1552 | -0.027 | -0.032 | -0.022 | 0.009 | 0.004 | 0.014 |

| | | Unweighted | | | Weighted | | |
|---|---|---|---|---|---|---|---|
| | N | Mean | 95% CI Lower | 95% CI Upper | Mean | 95% CI Lower | 95% CI Upper |
| Robots Contemporaneous | 2383 | -0.018 | -0.021 | -0.015 | -0.001 | -0.004 | 0.003 |
| Foreign robots as variable | 684 | -0.060 | -0.070 | -0.049 | -0.018 | -0.029 | -0.008 |
| <u>Dataset used:</u> | | | | | | | |
| NIFR | 199 | -0.006 | -0.012 | -0.001 | -0.007 | -0.012 | -0.001 |
| Individual-level data | 440 | 0.001 | 0.000 | 0.002 | 0.001 | 0.000 | 0.002 |
| Firm-level data | 120 | 0.004 | 0.002 | 0.007 | 0.001 | -0.001 | 0.004 |
| Industry-level data | 375 | -0.091 | -0.110 | -0.072 | -0.028 | -0.047 | -0.009 |
| Regional-level data | 1651 | -0.005 | -0.006 | -0.004 | -0.008 | -0.009 | -0.006 |
| Country/US state-level data | 256 | 0.008 | 0.003 | 0.013 | 0.021 | 0.016 | 0.026 |
| Manufacturing | 385 | -0.060 | -0.076 | -0.044 | -0.019 | -0.036 | -0.003 |
| Services | 177 | 0.002 | -0.002 | 0.007 | 0.012 | 0.008 | 0.016 |
| Agricultural | 72 | 0.007 | 0.001 | 0.014 | 0.008 | 0.002 | 0.015 |
| Young | 69 | 0.003 | 0.001 | 0.005 | 0.006 | 0.004 | 0.008 |
| Middle aged/old | 68 | 0.000 | -0.001 | 0.001 | -0.004 | -0.005 | -0.003 |
| Male | 165 | 0.000 | -0.004 | 0.004 | -0.013 | -0.017 | -0.009 |
| Female | 156 | 0.002 | -0.001 | 0.005 | -0.008 | -0.011 | -0.005 |
| Low/medium skilled/ educated | 222 | 0.001 | -0.001 | 0.004 | -0.005 | -0.007 | -0.002 |
| Highly skilled/educated | 54 | -0.001 | -0.011 | 0.009 | -0.001 | -0.010 | 0.009 |
| Developed countries | 1949 | -0.020 | -0.024 | -0.016 | -0.003 | -0.008 | 0.001 |
| Developing countries | 453 | -0.001 | -0.003 | 0.002 | 0.004 | 0.002 | 0.006 |
| <u>Fixed effects used:</u> | | | | | | | |
| Time fixed effects | 1375 | -0.010 | -0.014 | -0.006 | 0.002 | -0.002 | 0.006 |
| people fixed effects | 68 | -0.006 | -0.008 | -0.004 | -0.005 | -0.007 | -0.003 |
| firm fixed effects | 102 | 0.003 | 0.002 | 0.004 | 0.003 | 0.001 | 0.004 |
| Industry fixed effects | 270 | -0.076 | -0.096 | -0.056 | -0.017 | -0.037 | 0.003 |

| | | Unweighted | | | Weighted | | |
|---|---|---|---|---|---|---|---|
| | N | Mean | 95% CI Lower | 95% CI Upper | Mean | 95% CI Lower | 95% CI Upper |
| Regional fixed effects | 1936 | -0.004 | -0.005 | -0.003 | -0.001 | -0.002 | 0.000 |
| Country/US State fixed effects | 726 | 0.007 | 0.004 | 0.009 | 0.001 | -0.001 | 0.003 |
| Controls used: | | | | | | | |
| Population | 814 | -0.011 | -0.013 | -0.009 | -0.006 | -0.008 | -0.004 |
| Gender | 1656 | -0.005 | -0.006 | -0.004 | -0.006 | -0.007 | -0.004 |
| Age | 1848 | -0.004 | -0.006 | -0.003 | -0.009 | -0.010 | -0.007 |
| Education | 1711 | -0.004 | -0.005 | -0.003 | -0.005 | -0.006 | -0.004 |
| Ethnicity | 1196 | -0.008 | -0.010 | -0.007 | -0.014 | -0.016 | -0.013 |
| Occupation | 1523 | -0.004 | -0.005 | -0.003 | -0.004 | -0.005 | -0.003 |
| Skill | 450 | 0.011 | 0.008 | 0.014 | 0.025 | 0.022 | 0.028 |
| Exposure to capital | 653 | 0.000 | -0.003 | 0.003 | 0.019 | 0.016 | 0.022 |
| Exposure to imports | 1306 | -0.034 | -0.040 | -0.029 | -0.005 | -0.010 | 0.001 |
| Foreign robots | 64 | -0.001 | -0.004 | 0.002 | 0.000 | -0.003 | 0.003 |
| Further regression specifications: | | | | | | | |
| Sample weighted | 1264 | -0.033 | -0.039 | -0.027 | -0.003 | -0.009 | 0.003 |
| IV/2SLS | 1337 | 0.000 | -0.002 | 0.002 | 0.001 | 0.000 | 0.003 |
| Addresses outliers | 271 | -0.003 | -0.006 | -0.001 | -0.011 | -0.013 | -0.008 |
| Publication details: | | | | | | | |
| Published | 1256 | -0.032 | -0.038 | -0.026 | 0.007 | 0.001 | 0.013 |

Notes: Shows the PCC value for the estimated effect of robots on wages. All effect estimates have been transformed from Fisher's Z back into PCC and thus can be interpreted as a PCC. N = number of observations where the binary variable is a one. Table 1 shows definitions of variables. Weighted = each estimate is weighted by the inverse of the number of estimates in that subset reported by that observations study.

Overall, the table suggests that average effect sizes are small, far below the threshold for even a small effect as determined by Doucouliagos (2011). When changing the weighting scheme, differences are minimal and the means across the subgroups all remain low.

How wages are defined can lead to slightly different results as “Wages” and “wage-bill” have negative average PCC values whereas “earnings” and “income” have positive values. This shows that the definition may have an impact on the results of a paper. In contrast, the timeframe results show little variation. All show slightly negative results, with monthly being the closest to positive, and annual being the most negative.

For the robot variable, relative measures are predominantly used compared to other definitions such as absolute values or the use of dummy variables. It is then unsurprising that the average relative robot measure effect size is very close to the overall mean, but interestingly, the effect size for absolute measures is slightly more negative, and for dummy variable measures it is slightly positive. However, there are not many observations for each of these subgroups and the dummy variable confidence interval eclipses both positive and negative values.

When looking at the data level of the datasets being used, industry- and regional-level data PCC values are slightly negative, and country, firm and individual-level values are slightly positive. But again, all are close to zero, with the only exception being the industry-level data which just passes the threshold by Doucouliagos (2011) for a small effect. Surprisingly, when only looking at certain skill subgroups, being low or medium skill seems to have a slight positive effect, whereas being high skilled seems to have a negative effect (although very close to zero). However, this is likely due to the small number of estimates which is reflected in the large confidence interval. Additionally, the weighted average shows effects are negative for both skill levels. When looking at developed vs. developing countries, the negative effects seem to be greater in the developed countries.

Analysing the differences that control variables make, the estimates that had skill-based shares as a control variable were the only ones that averaged a positive PCC value. However, it is harder to make conclusions on the effect that control variables have based solely on these summary statistics, as a lot of the time many of the control variables are used simultaneously. For the fixed effects used, there is little difference, with country or state fixed effects being the only positive average, and industry fixed effects being included led to the most negative estimates.

Lastly, it seems that using instrumental variables led to more positive estimates compared to the overall mean. This may indicate some level of endogeneity. If there

was no endogeneity, then you would expect that with sufficiently strong instruments, IV analysis would lead to similar results. Of course, we cannot be certain of the severity of this endogeneity due to different instruments being used across studies, and varying levels of instrument strength. Published data had more negative results compared to the overall sample, possibly indicating a difference between published and yet to be published results.

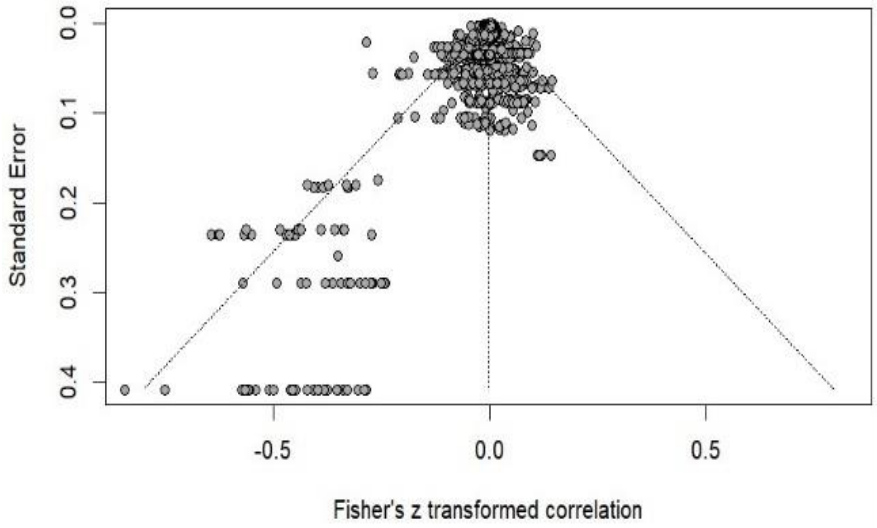
However, it is important to note that one must be wary of coming to conclusions based on these results alone. Getting the average of each subgroup is like running an OLS regression on a constant, and therefore comes with the inefficiencies and biases from the model. With the little estimates that I have in some subgroups, this means I may not have the most accurate results compared to other models. This is especially true because I also have not controlled for publication bias, and therefore the results from the table above may be misleading. In the following analysis, I will try to control for publication bias when estimating the overall mean.

Overall, while each subgroup has its own differences to the overall mean, most of these are too small to be economically meaningful. This is shown by the 95% confidence interval still encompassing the overall mean. Additionally, many of the differences are much smaller than 0.07 which is considered a small effect by Doucouliagos (2011). This means that while each of the subgroups show different effect sizes, these differences are rarely big enough to make any meaningful impacts, and do not change any of the conclusions reached.

### 4.2 Funnel plot

Publication bias can be identified using a funnel plot. As explained in Section 3.2, the funnel plot shows publication bias through its asymmetry. If there is publication bias or other small-sample biases, the funnel will be asymmetrical around the mean as those studies that show statistically insignificant results or results that don't fit a desired narrative may not get published, leading to asymmetry in the plot. Therefore, as a preliminary test for publication bias, Figure 2 shows a funnel plot from the estimates I collected across the literature.

*Figure 2: Funnel plot of Fisher's Z correlations*




*Notes: This funnel plot shows Fisher's Z correlations on the x-axis and standard error in descending order on y-axis.*

The plot shows the heterogeneity of estimates, as there is a quite a large range of estimates across studies. Looking at the observation points from the funnel plot, there is clear asymmetry, with most of the less precise estimates being negative. This indicates that positive estimates may have been discarded (unpublished, unrecorded, or re-estimated). However, it should be noted that these observations are clearly outliers, and the number of these observations in the bottom-left quadrant is few compared to the total number of 2,586 observations.

The vast majority of the observations are a lot more precise than these outliers, and therefore closer to zero and showing a slightly more asymmetrical shape. However, due to the aforementioned outliers, it is hard to visually assess the symmetry of the funnel plot. Therefore, in the following section, I conduct formal funnel plot asymmetry tests (FAT) get a better understanding of the funnel plot and get a more reliable measure of the level of publication bias.

### 4.3 OLS, fixed effect, random effects & UWLS results

As explained in Section 3.2, I use Fixed Effect (FE) and Random Effects (RE) meta-analysis estimators, and Ordinary Least Squares (OLS) and Unrestricted Weighted

Least Squares (UWLS) models to obtain an overall effect size estimate for the literature. The result of this analysis is shown in Table 3 below.

I present estimates that both do and do not correct for publication bias using the FAT-PET-PEESE method. For each of these, I additionally present both unweighted estimates and estimates weighted by the inverse number of observations from each study. It is also important to note that while the FAT-PET-PEESE method was used, the PEESE method was only used twice because estimates of the publication bias adjusted effects were almost always statistically insignificant.

*Table 3: OLS, FE, RE & UWLS results*

| Panel A: Not correcting for Publication bias | OLS | FE | RE | UWLS |
|---|---|---|---|---|
| ***Unweighted:*** | | | | |
| Estimate | -0.016 | -0.000 | -0.003 | -0.000 |
| | (0.014) | (0.000) | (0.003) | (0.002) |
| Observations | 2586 | 2586 | 2586 | 2586 |
| ***Weighted:*** | | | | |
| Estimate | -0.001 | -0.000** | -0.000 | -0.000 |
| | (0.005) | (0.000) | (0.002) | (0.001) |
| Observations | 2586 | 2586 | 2586 | 2586 |
| Panel B: Correcting for Publication bias (PET) | OLS | FE | RE | UWLS |
| ***Unweighted:*** | | | | |
| Effect beyond bias | 0.021** | -0.000 | -0.003 | -0.000 |
| | (0.006) | (0.000) | (0.003) | (0.002) |
| Publication bias | -0.803 | -0.476 | -0.066 | -0.476 |
| | (0.429) | (0.483) | (0.150) | (0.478) |
| Observations | 2586 | 2586 | 2586 | 2586 |
| ***Weighted:*** | | | | |
| Effect beyond bias | -0.005 | -0.000*** | 0.001 | -0.000 |
| | (0.0067) | (0.000) | (0.002) | (0.001) |
| Publication bias | 0.140 | -0.068 | -0.244 | -0.068 |
| | (0.291) | (0.186) | (0.181) | (0.201) |

| | | | | |
|---|---|---|---|---|
| Observations | 2586 | 2586 | 2586 | 2586 |
| Panel C: Correcting for Publication bias (PEESE) | OLS | FE | RE | UWLS |
| ***Unweighted:*** | | | | |
| Effect beyond bias | -0.002 | | | |
| | (0.006) | | | |
| Observations | 2586 | | | |
| ***Weighted:*** | | | | |
| Effect beyond bias | | -0.000** | | |
| | | (0.000) | | |
| Observations | | 2586 | | |

Notes: Effect beyond bias is constant term in regression equation. Publication bias is standard error or variance coefficient (depending on whether PET or PEESE was used). OLS = Ordinary Least Squares, FE = Fixed Effect, RE = Random Effects, UWLS = Unrestricted Weighted Least Squares. Estimates are corrected for publication bias using the FAT-PET-PEESE method. All effect size estimates have been transformed from Fisher's Z back into PCC and thus can be interpreted as a PCC. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level respectively. Standard errors are clustered at study level and given in brackets.

The overall takeaway from these results is that when using the linear models to estimate the overall effect size across the literature, there is little to no evidence for a statistically or economically significant relationship between robots and wages.

The OLS estimates are equivalent to the corresponding simple averages in Table 2. Of course, these may not be reliable due to inefficiencies and/or biases. As a result, the other estimators may give a more accurate insight into the true relationship. When looking at the FE, RE and UWLS estimates, they all indicate that the true effect size of robots on wages is closer to zero compared to a simple average.

All results that are not corrected for publication bias have negative coefficients. However, I cannot use this to determine a negative relationship between robots and wages, as almost all results are statistically insignificant at either the 1%, 5%, or 10% level. The only statistically significant result is from the weighted FE model, which is significant at the 5% level. However, this result is far below the value for what can be

considered a small effect between robots and wages recommended by Doucouliagos (2011).

Of course, these results may not be accurate as they may be plagued with publication bias. Therefore, I also report corrected results and test for the presence of publication bias. Looking at effect sizes, results are again close to zero. The unweighted OLS PET results produced statistically significant positive results (at the 5% level); however, this significance and positive sign is removed once the PEESE estimation is used. This aligns closer to the other estimators, including the weighted OLS PET. The weighted FE model is the only other estimator that requires further estimation using the PEESE model. This remains statistically significant at the 5% level, and the effect size becomes slightly larger. But again, this result is nowhere near the threshold for any form of significant effect (Doucouliagos, 2011), supporting earlier conclusions that robots and wages have little to no relationship on average across the literature.

In addition, there is little evidence of publication bias from the FAT-PET-PEESE tests. All coefficients for the standard error in the PET analysis fail to reach statistical significance at the 1%, 5%, and 10% level. This is further shown by the little difference in effect sizes between estimates that do correct for publication bias, and those that do not. This indicates that, according to the test, the funnel plot from the observations I collected is roughly symmetrical. This also confirms that while the observation points in Figure 2 do seem asymmetrical, these are just outliers, and the rest of the estimates follow enough of a symmetrical shape where the test states publication bias is doubtful.

It is somewhat reassuring that the results from the models are unlikely to be biased from small study effects. Therefore, results that do not correct for publication bias, such as the summary statistics in Table 2, can be seen as more reliable as they are also less likely to be negatively impacted by publication bias. However, it is important to note that while the FAT-PET-PEESE test for publication bias is valid, it is known to have low power (Stanley & Doucouliagos, 2012). This means that the little evidence of publication bias as shown above does not guarantee that the bias is not present, so some caution should still be taken when interpreting results.

As explained in Section 3.2, the weighted least squares methods (FE and RE) weight each observation based on how much information they give about the true effect size.

For the FE model, which assumes that one true effect size exists, estimated effects are weighted by their precision (1/SE). This is because estimates with higher precision are believed to have smaller sampling error, the only reason for any deviation from the true effect size according to the FE model. Therefore, if one study has estimates with very high precision, then those estimates will receive high weighting in the model. This is not problematic if the key assumption of a single true effect size holds.

However, if the assumption fails, then results from weighting solely on precision may be biased, as the estimates from the highly weighted study may not necessarily be close to the mean value of the true effect size distribution[22]. As discussed in Section 2.2, I have collected estimates covering a wide range of countries, wage and robot definitions, and timeframes. Therefore, it may not be reasonable to assume that the FE model assumption holds, leading to the possibility that, if a single study receives a lot of weight in the model, results from the FE model may be biased.

Table 4 shows the five highest weighted studies for the FE and RE models. For the fixed effect model this is Kugler et al. (2020), which makes up almost 75% of the total weight in the model. This is because they use individual-level data with extremely large sample sizes, leading to incredibly precise estimates. This calls into question the validity of the FE and UWLS estimates[23] because I believe heterogeneity is likely across the estimates I collected, indicating that the RE estimates may be more reliable.

For the random effects model, the highest weighted paper is Acemoglu and Restrepo (2020), accounting for over 50% of the weight in the model. This is likely due to the significant portion of observations coming from this paper, as I retrieved 655 out of 2586 from this paper alone.

For both models, after the highest weighted paper, there is a significant drop off in the weight the second most weighted paper takes up, with Aksoy et al. (2021) being the second highest weighted paper for both models.

[22] This is accounted for in the RE model.

[23] UWLS uses the same weights as the Fixed Effect estimator.

*Table 4: Highest weighted papers in FE and RE models*

| **Fixed Effect:** | | **Random Effects:** | |
|---|---|---|---|
| **Study:** | **Weight:** | **Study:** | **Weight:** |
| 1. Kugler et al. (2020) | 74.90% | 1. Acemoglu and Restrepo (2020) | 53.68% |
| 2. Aksoy et al. (2021) | 15.07% | 2. Aksoy et al. (2021) | 9.30% |
| 3. Acemoglu and Restrepo (2020) | 4.50% | 3. Rodgers and Freeman (2019) | 9.22% |
| 4. Dauth et al. (2021) | 3.46% | 4. Dauth et al. (2021) | 8.90% |
| 5. Zhang et al. (2022) | 0.56% | 5. Dottori (2021) | 3.55% |

Notes: The above table shows the top five studies that receive the most weight in estimating the overall effect size for the fixed effect and random effect models. The percentage of the overall weight a study makes-up in the estimation models is shown next to the study name.

It is important to note that having highly weighted papers may not necessarily be a bad thing. This is the case for the random effects model, which accounts for heterogeneity and therefore a highly weighted paper should, in theory, more accurately reflect the true effect of robots on wages. There are also estimators that exclusively look as these very precise and therefore highly weighted papers. These include the WAAP estimator by Ioannidis et al. (2017), the Top10 estimator by Stanley et al. (2010), and the Stem estimator by Furukawa (2019). The results from these estimators are shown in Table 5. Therefore, having highly weighted papers can be a double-edged sword, it can be beneficial as the papers with high weighted likely have very precise estimates, but this can be unfavourable if there is significant heterogeneity and therefore may not accurately show the average effect across the entire literature.

### 4.4 Additional non-linear estimators results

As previously stated in Section 3.3, the linearity assumption of the above models may fail to hold. Other than the few estimates that used the PEESE method, the previous results consisted of linear specifications to adjust for publication bias. Table 5 shows the results from four models that are based on a non-linear regression: the WAAP,

Top10, Stem, and Kink[24]. The results from these methods test the robustness of the linear models in Table 3.

*Table 5: WAAP, Top10, Stem & Kink results*

| Value | WAAP | Top10 | Stem | Kink |
|---|---|---|---|---|
| Effect beyond bias | -0.000*** | -0.002 | -0.000 | -0.000 |
| | (0.000) | (0.002) | (0.000) | (0.000) |
| Publication bias | | | | -0.517** |
| | | | | (0.032) |
| Observations | 5 | 245 | 2586 | 2586 |

Notes: WAAP = weighted average of adequately powered estimates (Ioannidis et al., 2017), Top10 = average of top 10 per cent most precise estimates (Stanley et al., 2010), Stem = stem-based method by Furukawa (2019), Kink = endogenous kink method (Bom & Rachinger, 2019). ∗∗∗, ∗∗, and, ∗ denote statistical significance at the 1%, 5%, and 10% level respectively. Standard errors are given in brackets and are clustered at study level for the TOP10 estimator.

Overall, the non-linear models help confirm what the linear models previously showed – that there is little to no evidence of a relationship between robots and wages. Similar to the linear models, the non-linear results are all statistically insignificant, apart from the WAAP model.

The WAAP model is the only statistically significant result, but like previously obtained statistically significant results, the effect is far below what would be considered a small effect of robots on wages. Additionally, the model only uses five observations, seriously calling into question the validity of these results. The WAAP uses an inverse-variance weighted model, only using estimates with a power of at least eighty per cent. The fact that only five estimates are used shows that many estimates from the literature lack power. All five estimates are from Kugler et al. (2020), which as I showed earlier has the majority of weight in the FE model, showing how much more precise their estimates are compared to other estimates from the literature, and how the WAAP results are likely unreliable.

There is however one important difference between the linear and non-linear models. While the linear models show no indication of publication bias, the endogenous kink

[24] Section 3.3 explains these methods in greater detail.

method (Bom & Rachinger, 2019) shows statistically significant levels of publication bias at the 1% level. This contrasts with the previous results from the funnel asymmetry tests that estimate the level of publication bias. It is therefore hard to come to an overall conclusion for the level of publication bias in the literature as my analysis shows differing results depending on how the issue is analysed. The endogenous kink method uses a piecewise linear model where a kink in the model occurs at an endogenous cutoff value of the standard error where publication selection is unlikely to occur. This is a similar method to the FAT-PET-PEESE model as it assumes a linear function like PET, but similar to PEESE, it assumes that this linear function breaks down for very precise estimates. Therefore, it is surprising that the significance levels of the results are so different.

### 4.5 Endogeneity robust estimators results

As a further robustness check, I relax the assumption that the correlation between estimated effect sizes and their standard errors is solely due to publication bias. I do this by running three additional models that account for any endogeneity between the effect size and the standard error. First is an instrumental variable analysis using the MAIVE estimator, which uses the inverse of the square root of the number of observations as the instrument for the standard error. It then uses the PET-PEESE method to estimate an overall effect size.

Second, I run the P-uniform* model, which doesn't directly rely on the effect size-standard error relationship. Instead, it uses the assumption that p-values should be uniformly distributed at the true effect size. Lastly, I conduct a caliper test to test for potential publication bias. Results from the first two models can be found in Table 6.

*Table 6: MAIVE & P-uniform* results*

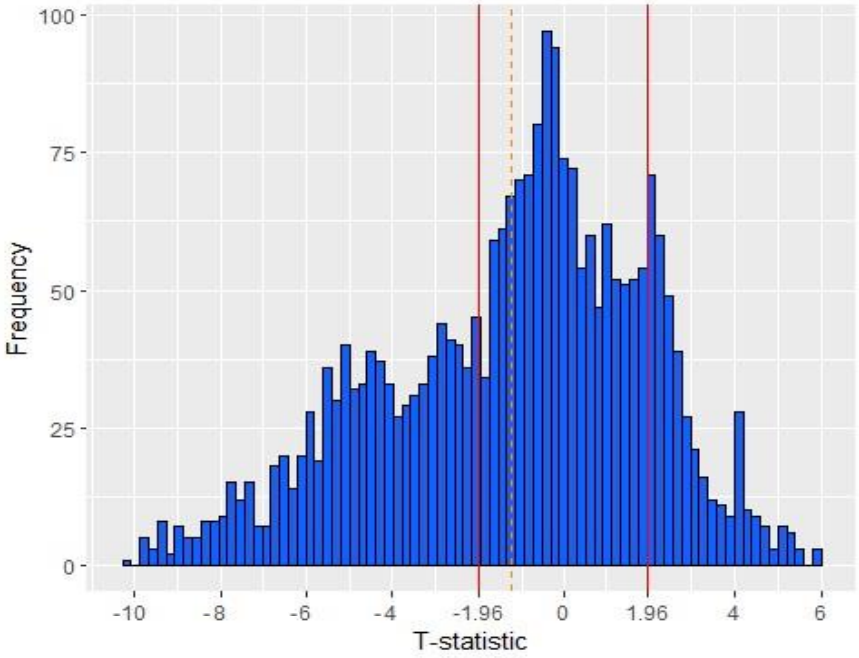
| MAIVE | | P-uniform* | |
|---|---|---|---|
| Effect beyond bias | 0.001 | Effect beyond bias | 0.001 |
| | (0.005) | | (0.005) |
| Robust first-stage F-test | 6828.429 | Observations | 2586 |
| Observations | 2586 | | |

Notes: MAIVE = Meta-Analysis Instrumental Variable Estimator (Irsova et al., 2023a). ***, **, and * denote statistical significance at the 1%, 5%, and 10% level respectively. Standard errors are given in brackets and MAIVE standard errors are clustered at study level.

The results from the MAIVE and P-uniform* models further confirm my previous analysis that the relationship between robots and wages is very small (Doucouliagos, 2011). Like the majority of the linear and non-linear results, both the MAIVE and the P-uniform* estimator show a statistically insignificant estimated effect that is close to zero. Looking at the instrument that the MAIVE uses, the F-stat shows that the instrument is very strong, which adds confidence to my MAIVE results, and further supports the conclusion shared by the previous models that the overall effect of robots on wages across the literature is very close to zero.

However, while the MAIVE and P-uniform* estimators provide an overall estimate that is corrected for publication bias, it does not provide a formal test for the level of the bias. Therefore, I conduct a caliper test which does not provide an estimated effect size but does provide a test for publication bias which is robust to any potential endogeneity bias, as it looks at the distribution of T-statistics. These results can be found in Figure 3 and Table 7.

*Figure 3: T-statistic distribution of collected estimates*




*Notes: The vertical orange dotted line shows the mean across all estimates, and the vertical red lines show the critical values associated with a 5% significance level.*

*Table 7: Caliper test results*

| | Threshold: 1.96 | Threshold: -1.96 |
|---|---|---|
| Caliper width 0.05 | 0.377*** | -0.388*** |
| | (0.041) | (0.047) |
| n1/n2 | 43/15 | 6/19 |
| Caliper width 0.1 | 0.337*** | -0.347*** |
| | (0.031) | (0.044) |
| n1/n2 | 51/27 | 14/29 |
| Caliper width 0.2 | 0.298*** | -0.326*** |
| | (0.026) | (0.026) |
| n1/n2 | 70/54 | 30/50 |

Notes: n1/n2 = number of observations within the caliper above and below the threshold respectively i.e. for n1 for a 1.96 threshold with a caliper width of 0.05 is number of T-statistics whose values are above 1.96 and below 2.01. The value of the reported coefficient shows the percentage of estimates above the expected 50% e.g. a value of 0.15 means that 65% of estimates are above the threshold and 35% of estimates are below. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level respectively. Standard errors are clustered at study level and given in brackets.

Figure 3 shows the distribution of T-statistics from the estimates I collected. The distribution shows that a range of T-statistics are present in the literature. However, it is not an even distribution with the graph showing a clear left skew. The maximum for positive results is around six, whereas for negative results is about ten, showing that the more significant results tend to be negative. Additionally, the mean (shown by the vertical orange dotted line) is only slightly negative, showing that while the literature does show many different, significant results, combining these results to obtain an overall measure across the literature, results in an insignificant combined effect.

Figure 3 can also be used for the caliper test which tests if publication bias exists by checking the smoothness of the distribution around points where statistical significance is determined. If publication bias does exist, then there will be an uneven number of observations on the significant side instead of a smooth distribution.

In Figure 3, the two vertical red lines show the t-statistic values of 1.96 and -1.96, the threshold associated with a 5% significance level. As you can see on the graph, there is a clear spike in the frequency of observations just beyond the threshold. Therefore, the caliper test does indicate significant levels of publication bias as those estimates that are just significant are more likely to be published than those that are just insignificant. Because I collected both published and unpublished estimates, the publication status of estimates on either side of the threshold is not clear. If the published estimates were disproportionately above the threshold, then this would be further evidence of publication bias. However, even if the estimates around the thresholds are all published, this still indicates selection bias. Because of how disproportionate the estimates are, this may indicate not only selection across studies as there is a “gap” in reported estimates, but also selection within studies. The bunching just above the thresholds may be due to researchers using methods such as “p-hacking” to achieve statistically significant results when their originally planned analysis achieved significance levels just below the threshold. This causes the disproportionate shape as shown by the t-statistic distribution. There is also an even larger jump around the zero point, meaning that estimates that are slightly negative are more likely to be reported than those that are slightly positive.

These spikes in frequencies are so obvious that the results of a formal caliper test are not required. However, for completeness the results are shown in Table 7. In all three

caliper widths, t-statistic observations are much more common above the 1.96 threshold than below it, whereas the opposite is true around the -1.96 threshold. This is shown both by the number of observations within the caliper above and below the threshold, and the regression coefficients which are all statistically significant at the 1% level. The coefficients indicate the percentage of estimates that exceed the expected 50% of estimates above the threshold. These coefficients are positive for the 1.96 threshold and negative for the -1.96 threshold, which is what you expect when publication bias exists where estimates that are slightly significant are more likely to be published than slightly insignificant estimates.

Therefore, from the caliper test, I get further evidence that publication bias does exist when trying to estimate the relationship between robots and wages, supporting the findings by the Endogenous Kink method, but contrasting the results from the RE, FE, and UWLS models.

### 4.6 P-hacking robust estimators results

As a last robustness check, I wanted to see if p-hacking robust estimators, which account for both selection across studies and selection within studies, provided significantly different results compared to previous models that mostly account for selection across studies. Therefore, I run right-truncated meta-analysis (RTMA) and meta-analysis of non-affirmative results (MAN) models which account for selection within studies. The results from these regressions can be found in Table 8:

*Table 8: RTMA & MAN results*

| RTMA: | | MAN: | |
|---|---|---|---|
| Estimated mean | -0.003*** | Worst case estimate | -0.004 |
| | (0.000) | | (0.019) |
| Estimated standard deviation | 0.005*** | Publication bias | -0.004 |
| | (0.000) | | (0.019) |

Notes: RTMA = right-truncated meta-analysis, MAN = meta-analysis of non-affirmative results. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level respectively. Standard errors given in brackets.

These results further confirm what the previous models state, that there is little to no overall effect of robots on wages across the literature. The RTMA results are statistically significant at the 1% level, however, the results accounting for potential p-hacking and publication bias are far below the small effect threshold (Doucouliagos, 2011). The heterogeneity (the reported standard deviation) is also quite small, further supporting the conclusion that the overall effect is close to zero. The MAN results are also close to zero but show even less support for an effect of robots on wages as neither the overall estimated mean nor the test for publication bias is statistically significant. This means that even if the level of publication bias is the worst-case scenario, the estimated overall effect cannot be confidently assumed to be different from zero.

The results from the publication bias test also support the results found in the FE, RE, and UWLS models rather than the positive results for publication bias found in the caliper, and endogenous kink models, highlighting that across all models, little to no relationship is found between robots and wages, but the level of publication bias is unclear. While the differences in results from the publication bias tests do not allow me to come to a definite conclusion, the fact that some tests do indicate publication bias should not be overlooked, especially with such strong results as shown in the caliper test. Therefore, I believe that the publication bias correcting estimates have more validity than those that do not take publication bias into account. Overall, while there is little difference between these estimates as neither show any statistically or economically significant results, to be safe, my overall interpretation of results was conducted using the bias correcting estimators.

### 4.7 Bayesian model averaging results

While the overall effect size that robots have on wages across the literature is essentially zero, it can still be a useful exercise to see whether there are factors that do contribute to robots having an appreciable effect on wages. As shown by the t-statistic distributions and summary statistics, while the average effect across studies is insignificant, there is a wide range of estimates across the literature, and it may prove useful to try to explain why. Is it just random variability between studies, or are there factors that influence results in a specific way?

To answer this, I run Bayesian model averaging (BMA). As explained in Section 3.6, this method calls for estimating all possible combinations of the fifty-six control variables I included in the model (one was omitted as a reference variable). Due to the high number of variables, my model includes a Monte Carlo Markov Chain (MCMC) sampling algorithm to only go through the most important combinations, along with a Unit Information Prior (UIP) and a dilution prior. The results from this BMA model are shown in Table 9, displaying their Posterior inclusion probability (PIP), Posterior mean, and Posterior standard error.

Assuming the respective specifications all identify causal effects, a variable's PIP is the probability that a certain variable belongs in the true model (Hinne et al., 2020). The PIP is calculated by the sum of the posterior probabilities for the models that include that variable. It can be used to see how important a variable is across all the models that the BMA performed and can be interpreted as the estimated probability that the variable is included in the actual relationship that occurs. The posterior mean and standard deviation for each variable show the weighted average of a variable's coefficients and standard errors respectively across all the specifications run by the BMA model. The weighting used is the posterior model probabilities (estimated probability a specification is the actual specification that exists) for each specification. The results in Table 9 are sorted by PIP, showing the most important variables for the model at the top, and the least important at the bottom.

*Table 9: BMA results*

| | PIP | Post Mean | Post SD |
|---|---|---|---|
| Standard error | 1 | -0.468 | 0.031 |
| Not IFR data | 1 | 0.034 | 0.005 |
| Wages | 1 | 0.099 | 0.007 |
| Income | 1 | 0.103 | 0.008 |
| Real | 1 | 0.040 | 0.004 |
| Earnings | 1 | 0.068 | 0.012 |
| Industry-level data | 1 | -0.115 | 0.008 |
| Country/US state-level data | 1 | 0.131 | 0.008 |
| Manufacturing | 1 | -0.021 | 0.003 |

| | PIP | Post Mean | Post SD |
|---|---|---|---|
| Developing countries | 1 | 0.042 | 0.009 |
| Control: Skill | 1 | 0.021 | 0.004 |
| Control: Exposure to imports | 1 | -0.016 | 0.003 |
| Robots differences | 0.999 | 0.022 | 0.004 |
| Sample weighting | 0.999 | -0.011 | 0.002 |
| Developed countries | 0.979 | 0.026 | 0.009 |
| Country/US state fixed effects | 0.960 | 0.013 | 0.005 |
| Individual-level data | 0.960 | -0.019 | 0.006 |
| Foreign Robots as variable | 0.948 | -0.011 | 0.004 |
| People fixed effects | 0.933 | 0.025 | 0.009 |
| Robots logged | 0.923 | -0.017 | 0.007 |
| Published | 0.734 | -0.006 | 0.004 |
| Time fixed effects | 0.704 | -0.006 | 0.004 |
| Wage Logged | 0.661 | -0.009 | 0.008 |
| Publication year | 0.385 | 0.001 | 0.001 |
| Robots Absolute | 0.298 | 0.005 | 0.009 |
| Foreign Robots as control | 0.297 | -0.005 | 0.009 |
| IV/2SLS | 0.227 | -0.001 | 0.003 |
| Control: Occupation | 0.170 | -0.001 | 0.003 |
| Control: Population | 0.123 | 0.001 | 0.004 |
| Robots Contemporaneous | 0.096 | -0.001 | 0.003 |
| Firm-level data | 0.056 | 0.001 | 0.005 |
| Regional-level data | 0.055 | 0.001 | 0.004 |
| Wage Weekly | 0.055 | 0.000 | 0.001 |
| Low/medium skilled/educated | 0.052 | 0.000 | 0.002 |
| Control: Ethnicity | 0.042 | -0.000 | 0.001 |
| Wage Hourly | 0.036 | 0.000 | 0.001 |
| Highly skilled/educated | 0.035 | 0.000 | 0.002 |
| Wage differences | 0.035 | -0.000 | 0.002 |
| Control: Exposure to capital | 0.033 | 0.000 | 0.001 |
| Addresses outliers | 0.031 | -0.000 | 0.001 |
| Regional fixed effects | 0.029 | -0.000 | 0.001 |

| | PIP | Post Mean | Post SD |
|---|---|---|---|
| Control: Age | 0.028 | -0.000 | 0.001 |
| Middle-aged/old | 0.028 | -0.000 | 0.001 |
| Industry fixed effects | 0.028 | -0.000 | 0.001 |
| Wage Annual | 0.027 | -0.000 | 0.001 |
| Wage Monthly | 0.026 | 0.000 | 0.001 |
| Control: Gender | 0.023 | 0.000 | 0.001 |
| Clustered SE | 0.023 | 0.000 | 0.001 |
| Male | 0.021 | 0.000 | 0.001 |
| Female | 0.021 | 0.000 | 0.001 |
| Firm fixed effects | 0.020 | -0.000 | 0.002 |
| Services | 0.020 | 0.000 | 0.001 |
| Control: Education | 0.019 | 0.000 | 0.001 |
| Robots relative | 0.018 | 0.000 | 0.001 |
| Agriculture | 0.017 | 0.000 | 0.001 |
| Young | 0.013 | 0.000 | 0.001 |

Notes: The above table shows a BMA analysis where the dependent variable is the Fisher's Z value corresponding to the effect of robots on wages. The covariates are found in the table, see Table 1 for a more detailed description of the variables. Post mean = posterior mean, Post SD = posterior standard deviation, PIP = posterior inclusion probability.

The results of the BMA model show that twenty-three out of the fifty-six variables have a PIP of more than 0.5, where it is more likely than not that the variable belongs in the true model. Adding these variables to an OLS regression gives me an adjusted R-squared of 0.785. Therefore 78.5% of the variation in Fisher's Z values can be explained by these variables, showing that the majority of the variation can be explained by less than half of the variables I collected. This is further shown by an R-squared of 0.783 when only including variables with a PIP greater than 0.9 and an R-squared of 0.769 when including variables with a PIP of one. Twenty variables have a PIP of at least 0.9, followed by a significant drop off to less than 0.75 for the following three variables. Twelve of these twenty-three variables have a PIP of one. Additionally, the best model that the BMA achieved included twenty-two out of the fifty-seven possible variables (all of which have a PIP > 0.5).

Therefore, while many of the factors I collected have little influence on an estimate's value, I have collected a few that are crucial in explaining these differences. I explain these results in more depth below.

One of the notable results of the BMA analysis was for the standard error variable. This was one of the most important variables as it had a PIP of one, along with a negative posterior mean coefficient and a relatively low standard deviation. This has important implications for publication bias as this shows that when controlling for many other aspects of the study, such as study design or publication status, there is still a strong connection between standard error and effect size. This is something that publication bias tests and correctors such as the FAT-PET-PEESE model use to determine if publication bias exists. Therefore, the results from the BMA analysis may imply that publication bias is present in the data as the posterior small standard deviation points towards the variable being statistically significant. This contrasts the linear models such as FAT-PET but in line with other estimators like the caliper, and endogenous kink methods.

The posterior mean of the standard error's coefficient is -0.46, which indicates that estimates with higher standard errors are more negative. This relationship would lead to an asymmetrical funnel plot when graphed. If no publication bias was present, then this coefficient would be zero, and would show a symmetrical funnel plot. This conclusion is further supported by the publication status of an estimate also being an important covariate. The posterior mean for this variable is negative implying that published studies have more negative results than unpublished studies, indicating a potential bias in the literature.

Another noteworthy finding is that the effect on wages is different depending on the level of data. While the effects of regional, firm, and individual-level data are close to zero, industry-level data have larger and more statistically significant negative effects, whereas country/state-level data have larger positive effects. This is likely because country or regional-level data looks at all the jobs in an area, some of which may not have robots in their workplace, and therefore will see less of an effect.

Industry-level data usually looks at industries that are directly impacted by robots and therefore will see a more substantial effect. This may indicate that, within the industries where robots are more common, robots do hurt wages, but overall, in a region or

country, this effect is insubstantial as other industries not directly impacted by robots may see no influence, or even a positive influence when robot use increases.

This may be related to job retention as well. Employment level is also a crucial aspect of the impacts that robots have in the workplace, and when looking at industry or regional-level data, job loss may influence an area or industry's wage levels. Therefore, industries related to robots may see more job loss compared to a region with many industries present which may help explain some of the findings from my analysis.

When looking at specific industries, compared to estimates without a specific industry focus, manufacturing-focused estimates do seem to have different results. The manufacturing variable has a PIP of one, showing that it is an important variable in determining effect size as it likely occurs in the true specification. Manufacturing estimates have a lower effect size as the posterior mean is negative, supporting the claim that robots negatively impact wages in industries where robots are more prevalent, as manufacturing is among the industries with the highest exposure to robots. However, robots in other industries such as agriculture and services, where robots are also common, seem to have little influence. Although, this could be due to their lower number of observations.

Next, other aspects of the datasets other than the data level also seem to matter, with both the variables determining whether the study uses the IFR dataset, and data from developing countries are also among the most important variables in determining the effect of robots on wages. Results show that using the IFR dataset lowers Fisher's Z values, although this effect is only small. Similarly, using data from developing countries led to slightly more positive results compared to developed countries.

The way wages are defined in a study is also important, whereas the timeframe specified for a study's wage variable seems to have little effect. Compared to wage-bill, if a study's dependent variable is defined as wages, income, or earnings, estimates are more positive, with income leading to the most positive estimates. These effects have a noticeable impact on estimates as their coefficient's posterior means are larger compared to other values. However, how they define their robot variable seems to have less impact on Fisher's Z values, as none of the variables can be considered

important based on their PIP. Nonetheless, similar to wages, logging the robot variable and taking the differences were important in explaining variation across studies.

Other factors that seem to be included in the actual robots-wages specification include controlling for skill level and imports, but these are the only important control variables. Additionally, sample weighting is the only important estimation method that seems to have an impact as clustered standard errors, instrumental variables, and addressing outliers all have a low PIP and therefore small posterior means. Lastly, factors that we would expect to have large effects such as skill level and age all have little impact on an estimate's Fisher's Z value. However, this may be due to the low number of observations that only focus on a specific skill level or age.

Overall, the results from the Bayesian Model Averaging show that the variables that were collected during the coding process include some of the factors crucial in explaining the apparent heterogeneity across the studies I have included in my meta-analysis. These variables all have very high Posterior Inclusion Probabilities indicating their estimated importance in the true specification in determining the effect that robots have on wages. Some of the variables include the way the wage definition is defined, the level of data being used, controlling for specific variables or restricting the dataset in some way, such as a specific industry. However, there are also some variables that one may think would have an impact but are insignificant in determining the relationship between robots and wages, such as worker skill level and whether fixed effects were used in an estimate's regression.

The effect that these variables have on an estimates Fisher's Z value of course also vary. The Posterior means for their coefficients show this impact, and while none of the overall effects across the literature from any of the estimators I ran show what can be considered a small effect of robots on wages, some of these BMA coefficients may have large effects on the reported effect. These include the definition of a studies wage variable which have a positive effect compared to the reference variable, and some of the data level options with country-level data having a positive effect, and industry data having a negative effect. The posterior means for the coefficients I have mentioned are also statistically significant, as the variables have high PIP values and low Post SD values, adding validity to these interpretations. Of course, even though the coefficients are significant, simply adding these elements to a specification may not

lead to statistically significant results, and for this reason, I also refrain from using the thresholds by Doucouliagos (2011) to interpret these coefficient values. For example, if the average without a variable is negative, by adding a variable that has a positive effect, the overall effect might still be small. Table 2 indicates that this may be case, as the results show insignificant results almost exclusively.

## 5. Conclusion

In this thesis, I present the results of a meta-analysis of the current literature measuring the impact the robots have on wages. A new wave of literature on the topic has added more fire to the continuing debate on whether robots lower worker wages, but these new studies show conflicting results. In this context, a meta-analysis can be used to obtain an overall estimate of the effect of robots on wages across the entire literature.

By searching a variety of economic research databases, I collect fifty-two papers on the topic, resulting in 2,586 individual estimates. The current literature shows a wide range of findings, and methods used to achieve these findings. Differences include how authors describe their robot or wage variables and their dataset such as the level or scope of the data. Techniques also differ, like the control variables or panel fixed effects used, opting to use instrumental variables, and if they weigh their regressions, address outliers, or cluster standard errors.

Reading through this literature does not suggest there is a clear consensus within the literature. A meta-analysis is thus used as the most effective way of quantitatively summarizing this literature.

My main finding from my analysis is that, across the literature I find on average no significant effect of robots on wages. Almost all ordinary least squares models and weighted least squares fixed effect and random effects models fail to achieve statistically significant results. Further, additional non-linear regression-based estimators like the Top10 and Endogenous Kink models, along with endogenous robust estimators such as the MAIVE and p-uniform* models, and the MAN designed to account for p-hacking, also do not reach statistical significance. The models that do achieve statistical significance at either the 1%, 5% or 10% level include the weighted

fixed effect model, WAAP, and RTMA. However, are all far below the thresholds for even a small effect (Doucouliagos, 2011).

Overall, the tests for publication bias are less clear. While the FAT-PET-PEESE procedure is unable to find any evidence of publication bias, other models such as the caliper test, and Endogenous Kink method do point towards significant levels of bias. Therefore, it is hard to make any conclusions regarding the level of publication bias within the literature.

Even though I do not find a clear relationship between robots and wages, I do find some factors behind the strong heterogeneity of the estimates I collected. I do this using a meta-regression, namely, Bayesian Model Averaging (BMA) with fifty-six variables that reflect different characteristics of studies. I find that the study characteristics that affect most what estimate one gets are controlling for skill level and import exposure, using industry or country-level data, the definition of wages, using the IFR data set, and restricting the dataset to only include developing countries or data from the manufacturing sector. Interestingly, factors that one may think are important such as restricting the dataset to highly skilled or educated workers, or whether researchers use (panel) fixed effects in an estimate's specification are found to have negligible impact on the resulting effect size.

Overall, these results indicate that the question "Do robots reduce worker wages?" has no simple answer. The results suggest overall, there is little to no relationship between robots and wages. However, this does not mean that there is never a relationship between robots and wages, as study characteristics do influence the estimated relationship. If more studies were available, we cannot exclude that a meta-analysis on the subset of studies that focus on a given variable, a given definition of a variable, a specific technique or a subset of countries could potentially find an average positive or negative effect for that subset.

## References


Abeliansky, A. L., & Beulmann, M. (2021). *Are They Coming for Us? Industrial Robots and the Mental Health of Workers* (cege Discussion Papers No. 379). Center for European, Governance and Economic Development Research (cege). https://www.econstor.eu/bitstream/10419/201804/1/1671551540.pdf

Acemoglu, D. (2024). Harms of AI. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford Handbook of AI Governance* (pp. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.65

Acemoglu, D., Koster, H. R. A., & Ozgen, C. (2023). Robots and Workers: Evidence from the Netherlands. *National Bureau of Economic Research Working Paper Series*, *No. 31009*. https://doi.org/10.3386/w31009

Acemoglu, D., Lelarge, C., & Restrepo, P. (2020). Competing with Robots: Firm-Level Evidence from France. *AEA Papers and Proceedings*, *110*, 383-388. https://doi.org/10.1257/pandp.20201003

Acemoglu, D., & Restrepo, P. (2018). The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review*, *108*(6), 1488-1542. https://doi.org/10.1257/aer.20160696

Acemoglu, D., & Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, *128*(6), 2188-2244. https://doi.org/10.1086/705716

Aksoy, C. G., Özcan, B., & Philipp, J. (2021). Robots and the gender pay gap in Europe. *European Economic Review*, *134*, 103693. https://doi.org/https://doi.org/10.1016/j.euroecorev.2021.103693

Alguacil, M., Lo Turco, A., & Martínez-Zarzoso, I. (2022). Robot adoption and export performance: Firm-level evidence from Spain. *Economic Modelling*, *114*, 105912. https://doi.org/https://doi.org/10.1016/j.econmod.2022.105912

Anelli, M., Giuntella, O., & Stella, L. (2019). *Robots, Labor Markets, and Family Behavior* (IZA Discussion Papers No. 12820). Institute of Labor Economics (IZA). https://www.iza.org/publications/dp/12820/robots-labor-markets-and-family-behavior

Artuc, E., Bastos, P., Copestake, A., & Rijkers, B. (2022). Robots and Trade Implications for Developing Countries. In L. Y. Ing & G. M. Grossman (Eds.), *Robots and AI A New Economic Era* (1st ed., pp. 232-274). Routledge. https://doi.org/https://doi.org/10.4324/9781003275534-8

Artuc, E., Christiaensen, L., & Winkler, H. J. (2019). *Does Automation in Rich Countries Hurt Developing Ones?: Evidence from the U.S. and Mexico* (Policy Research Working Paper 8741). World Bank Group. https://doi.org/10.1596/1813-9450-8741

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, *29*(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Barth, E., Roed, M., Schone, P., & Umblijs, J. (2020). *How Robots Change Within-Firm Wage Inequality* (IZA Discussion Papers No. 13605). Institute of Labor Economics (IZA). https://hdl.handle.net/10419/224047

Bekhtiar, K., Bittschi, B., & Sellner, R. (2021). *Robots at Work? Pitfalls of Industry Level Data* (WIFO Working Papers 639/2021). WIFO. https://www.wifo.ac.at/wp-content/uploads/upload-4341/wp_2021_639_.pdf

Bom, P. R. D., & Rachinger, H. (2019). A kinked meta-regression model for publication bias correction. *Research Synthesis Methods*, *10*(4), 497-514. https://doi.org/10.1002/jrsm.1352

Borenstein, M., & Hedges, L. V. (2019). Effect sizes for meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (3 ed., pp. 207-244). Russell Sage Foundation. https://doi.org/10.7758/9781610448864

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97-111. https://doi.org/10.1002/jrsm.12

Borjas, G. J., & Freeman, R. B. (2019). From Immigrants to Robots: The Changing Locus of Substitutes for Workers. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *5*(5), 22-42. https://doi.org/10.7758/rsf.2019.5.5.02

Brambilla, I., César, A., Falcone, G., & Gasparini, L. (2023). The impact of robots in Latin America: Evidence from local labor markets. *World Development*, *170*, 106271. https://doi.org/https://doi.org/10.1016/j.worlddev.2023.106271

Cali, M., & Presidente, G. (2021). *Automation and Manufacturing Performance in a Developing Country*. World Bank Group. https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-9653

César, A., Falcone, G., & Garriga, P. (2022). *Robots, Exports and Top Income Inequality: Evidence for the U.S* (307). CEDLAS, Working Papers. https://ideas.repec.org/p/dls/wpaper/0307.html

Chiacchio, F., Petropoulos, G., & Pichler, D. (2018). *The impact of industrial robots on EU employment and wages: A local labour market approach* (Bruegel Working Paper No. 2018/02). Bruegel.

Chugunova, M., Keller, K., & Samila, S. (2021). *Robots, China and Polls: Structural Shocks and Political Participation in the US* (Research Paper No. 21-22). Max Planck Institute for Innovation and Competition. https://dx.doi.org/10.2139/ssrn.3929377

Chung, J., & Lee, Y. S. (2023). The Evolving Impact of Robots on Jobs. *ILR Review*, *76*(2), 290-319. https://doi.org/10.1177/00197939221137822

Claburn, T. (2022, December 6). Industrial robots in China push people out of jobs, slash wages. *The Register*. https://www.theregister.com/2022/12/06/robots_china_employment/

Compagnucci, F., Gentili, A., Valentini, E., & Gallegati, M. (2019). Robotization and labour dislocation in the manufacturing sectors of OECD countries: a panel VAR approach. *Applied Economics*, *51*(57), 6127-6138. https://doi.org/10.1080/00036846.2019.1659499

Costanzo, C. (2023). *Automation, Relative Wages, and Time Use Distortions within Households*. SSRN. http://dx.doi.org/10.2139/ssrn.4350864

Dario, G., Alessandro, P., & Jelena, R. (2024). *Will robot replace workers? Assessing the impact of robots on employment and wages with meta-analysis* (GLO Discussion Paper No. 1395). https://hdl.handle.net/10419/282318

Dauth, W., Findeisen, S., Suedekum, J., & Woessner, N. (2021). The Adjustment of Labor Markets to Robots. *Journal of the European Economic Association*, *19*(6), 3104-3153. https://doi.org/10.1093/jeea/jvab012

Dekle, R. (2020). Robots and industrial labor: Evidence from Japan. *Journal of the Japanese and International Economies*, *58*, 101108. https://doi.org/https://doi.org/10.1016/j.jjie.2020.101108

Deng, L., Plümpe, V., & Stegmaier, J. (2020). *Robot adoption at German plants* (IWH Discussion Papers, No. 19/2020). Leibniz-Institut für Wirtschaftsforschung Halle (IWH). https://nbn-resolving.de/urn:nbn:de:101:1-202101201115386267810 3

Díaz Pavez, L. R., & Martinez-Zarzoso, I. (2023). *The impact of local and foreign automation on labor market outcomes in emerging countries* (INFER Working Paper Series No. 1). Internation Network for Economic Research (INFER). https://infer-research.eu/publication/the-impact-of-local-and-foreign-automation-on-labor-market-outcomes-in-emerging-countries/

Dinlersoz, E., & Wolf, Z. (2024). Automation, labor share, and productivity: plant-level evidence from U.S. manufacturing. *Economics of Innovation and New Technology*, *33*(4), 604-626. https://doi.org/10.1080/10438599.2023.2233081

Dixon, J., Hong, B., & Wu, L. (2021). The Robot Revolution: Managerial and Employment Consequences for Firms. *Management Science*, *67*(9), 5586-5605. https://doi.org/10.1287/mnsc.2020.3812

Dottori, D. (2021). Robots and employment: evidence from Italy. *Economia Politica*, *38*(2), 739-795. https://doi.org/10.1007/s40888-021-00223-x

Doucouliagos, C. (2011). *How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics*. Deakin University. https://hdl.handle.net/10536/DRO/DU:30106880

Eggleston, K., Lee, Y. S., & Iizuka, T. (2021). Robots and Labor in the Service Sector: Evidence from Nursing Homes. *National Bureau of Economic Research Working Paper Series*, *No. 28322*. https://doi.org/10.3386/w28322

Eicher, T. S., Papageorgiou, C., & Raftery, A. E. (2011). Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics*, *26*(1), 30-55. https://doi.org/10.1002/jae.1112

Fisher, R. A. (1950). *Statistical methods for research workers* (11 ed.). Oliver and Boyd, Edinburgh.

Furukawa, C. (2019). *Publication Bias under Aggregation Frictions: Theory, Evidence, and a New Correction Method*. SSRN. http://dx.doi.org/10.2139/ssrn.3362053

Gan, J., Liu, L., Qiao, G., & Zhang, Q. (2023). The role of robot adoption in green innovation: Evidence from China. *Economic Modelling*, *119*, 106128. https://doi.org/https://doi.org/10.1016/j.econmod.2022.106128

Ge, S., & Zhou, Y. (2020). Robots, computers, and the gender wage gap. *Journal of Economic Behavior & Organization*, *178*, 194-222. https://doi.org/https://doi.org/10.1016/j.jebo.2020.07.014

George, E. I. (2010). Dilution priors: Compensating for model space redundancy. In *Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown* (Vol. 6, pp. 158-166). Institute of Mathematical Statistics. https://doi.org/10.1214/10-IMSCOLL611

Gerber, A., & Malhotra, N. (2008). Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals. *Quarterly Journal of Political Science*, *3*(3), 313-326. https://doi.org/10.1561/100.00008024

Giuntella, O., Lu, Y., & Wang, T. (2022). How do Workers and Households Adjust to Robots? Evidence from China. *National Bureau of Economic Research Working Paper Series*, *No. 30707*. https://doi.org/10.3386/w30707

Giuntella, O., & Wang, T. (2019). *Is an Army of Robots Marching on Chinese Jobs?* (IZA Discussion Paper No. 12281). Institute of Labor Economics (IZA). https://www.iza.org/publications/dp/12281/is-an-army-of-robots-marching-on-chinese-jobs

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, *110*(8), 2586-2624. https://doi.org/10.1257/aer.20181047

Graetz, G., & Michaels, G. (2018). Robots at Work. *The Review of Economics and Statistics*, *100*(5), 753-768. https://doi.org/10.1162/rest_a_00754

Han, Y. (2022). The Impact of Industrial Robots on the Skill-Based Wage Gap. *American Journal of Industrial and Business Management*, *12*, 571- 602. https://doi.org/10.4236/ajibm.2022.124031

Harthorne, M. (2017, March 29). Here's proof robots are taking jobs and cutting wages. *Fox News*. https://www.foxnews.com/tech/heres-proof-robots-are-taking-jobs-and-cutting-wages

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2020). A Conceptual Introduction to Bayesian Model Averaging. *Advances in Methods and Practices in Psychological Science*, *3*(2), 200-215. https://doi.org/10.1177/2515245919898657

International Federation of Robotics. (2023a, October 13). *World Robotics Report: "All-Time High" with Half a Million Robots Installed in one Year*. IFR Press Room. https://ifr.org/ifr-press-releases/news/wr-report-all-time-high-with-half-a-million-robots-installed

International Federation of Robotics. (2023b). *WR Industrial Robots 2023 - Sources and Methods*. https://ifr.org/img/worldrobotics/WR_Industrial_Robots_2023_Sources_and_Methods.pdf

International Monetary Fund. (2024, April 16). *WORLD ECONOMIC OUTLOOK (WEO) Frequently Asked Questions*. https://www.imf.org/en/Publications/WEO/frequently-asked-questions

Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The Power of Bias in Economics Research. *The Economic Journal*, *127*(605), F236-F265. https://doi.org/10.1111/ecoj.12461

Irsova, Z., Bom, P. R. D., Havranek, T., & Rachinger, H. (2023a). *Spurious Precision in Meta-Analysis of Observational Research*. Charles University, Prague. https://meta-analysis.cz/maive/

Irsova, Z., Doucouliagos, H., Havranek, T., & Stanley, T. D. (2023b). Meta-analysis of social science research: A practitioner's guide. *Journal of Economic Surveys*, *00*, 1-20. https://doi.org/https://doi.org/10.1111/joes.12595

Jackson, D., & White, I. R. (2018). When should meta-analysis avoid making hidden normality assumptions? *Biometrical Journal*, *60*(6), 1040-1058. https://doi.org/10.1002/bimj.201800071

Jung, J. H., & Lim, D.-G. (2020). Industrial robots, employment growth, and labor cost: A simultaneous equation analysis. *Technological Forecasting and Social Change*, *159*, 120202. https://doi.org/https://doi.org/10.1016/j.techfore.2020.120202

Kariel, J. (2021). Job Creators or Job Killers? Heterogeneous Effects of Industrial Robots on UK Employment. *LABOUR*, *35*(1), 52-78. https://doi.org/https://doi.org/10.1111/labr.12192

Keynes, J. M. (2008). Economic Possibilities for our Grandchildren (1930). In L. Pecchi & G. Piga (Eds.), *Revisiting Keynes* (pp. 17-26). The MIT Press. http://www.jstor.org/stable/j.ctt5hhftz.6

Koch, M., Manuylov, I., & Smolka, M. (2021). Robots and Firms. *The Economic Journal*, *131*(638), 2553-2584. https://doi.org/10.1093/ej/ueab009

Kromann, L., Malchow-Møller, N., Skaksen, J. R., & Sørensen, A. (2019). Automation and productivity—a cross-country, cross-industry comparison. *Industrial and Corporate Change*, *29*(2), 265-287. https://doi.org/10.1093/icc/dtz039

Kugler, A. D., Kugler, M., Ripani, L., & Rodrigo, R. (2020). U.S. Robots and their Impacts in the Tropics: Evidence from Colombian Labor Markets. *National Bureau of Economic Research Working Paper Series*, *No. 28034*. https://doi.org/10.3386/w28034

Lyu, W., & Liu, J. (2021). Artificial Intelligence and emerging digital technologies in the energy sector. *Applied Energy*, *303*, 117615. https://doi.org/https://doi.org/10.1016/j.apenergy.2021.117615

Manuel, F. E. (1938). The Luddite Movement in France. *The Journal of Modern History*, *10*(2), 180-211. https://doi.org/10.1086/243523

Mathur, M. B. (2024). P-hacking in meta-analyses: A formalization and new meta-analytic methods. *Research Synthesis Methods*, *1*(17). https://doi.org/10.1002/jrsm.1701

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Petit, F., Jaccoud, F., & Ciarli, T. (2023). *Heterogeneous Adjustments of Labor Markets to Automation Technologies* (CESifo Working Paper No. 10237). Center for Economic Studies and ifo Institute (CESifo). https://www.cesifo.org/en/publications/2023/working-paper/heterogeneous-adjustments-labor-markets-automation-technologies

Ren, H., Zhou, S., Hu, A., & Cheng, H. (2018). *Industrial Robots and Jobs Turnover: Evidence from Chinese Firm Level Data*. SSRN. http://dx.doi.org/10.2139/ssrn.3213959

Schneider, F., Jurkat, A., & Klump, R. (2023). *Robots and Wages: A Meta-Analysis*. ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg. doi.org/10.2139/ssrn.4666798

Shang, Q. (2022). *The Effects of Robots on Labor Markets and Political Attitudes: The Case of China*. Center for Monetary and Financial Studies (CEMFI). https://qianqian-shang.com/Documents/Paper2_Robots%20in%20China_QianqianShang.pdf

Stanley, T. D. (2005). Beyond Publication Bias. *Journal of Economic Surveys*, *19*(3), 309-345. https://doi.org/10.1111/j.0950-0804.2005.00250.x

Stanley, T. D., & Doucouliagos, H. (2012). *Meta-Regression Analysis in Economics and Business*. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/canterbury/detail.action?docID=1016122

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60-78. https://doi.org/10.1002/jrsm.1095

Stanley, T. D., & Doucouliagos, H. (2017). Neither fixed nor random: weighted least squares meta-regression. *Research Synthesis Methods*, *8*(1), 19-42. https://doi.org/10.1002/jrsm.1211

Stanley, T. D., Doucouliagos, H., & Ioannidis, J. P. A. (2017). Finding the power to reduce publication bias. *Statistics in Medicine*, *36*(10), 1580-1598. https://doi.org/10.1002/sim.7228

Stanley, T. D., Jarrell, S. B., & Doucouliagos, H. (2010). Could It Be Better to Discard 90% of the Data? A Statistical Paradox. *The American Statistician*, *64*(1), 70-77. https://doi.org/10.1198/tast.2009.08205

Steel, M. F. J. (2020). Model Averaging and Its Use in Economics. *Journal of Economic Literature*, *58*(3), 644-719. https://doi.org/10.1257/jel.20191385

Stemmler, H. (2020). *Does automation lead to de-industrialization in emerging economies? Evidence from Brazil* (cege Discussion Papers No. 382). Center for European, Governance and Economic Development Research (cege). https://wwwuser.gwdg.de/~cege/Diskussionspapiere/DP382.pdf

Terzidis, N., Brakman, S., & Ortega-Argiles, R. (2019). *Labour Markets, Trade and Technological Progress. A Meta-Study* (CESifo Working Paper No. 7719). CESifo. https://www.cesifo.org/node/43851

The World Bank. (2022). *Manufacturing, value added (current US$)*. World Bank Group. https://data.worldbank.org/indicator/NV.IND.MANF.CD?most_recent_value_desc=true

Thompson, S. G., Turner, R. M., & Warn, D. E. (2001). Multilevel models for meta-analysis, and their application to absolute risk differences. *Statistical Methods in Medical Research*, *10*(6), 375-392. https://doi.org/10.1177/096228020101000602

Tier, M. (2022). *A meta-analysis of the effect of robotics upon labour employment* [Master's thesis, University of Canterbury]. UC Research Repository. https://dx.doi.org/10.26021/12648

Uman, L. S. (2011). Systematic reviews and meta-analyses. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *20*(1), 57-59. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3024725/

van Aert, R. C. M. (2023). Meta-analyzing partial correlation coefficients using Fisher's z transformation. *Research Synthesis Methods*, *14*(5), 768-773. https://doi.org/10.1002/jrsm.1654

van Aert, R. C. M., & van Assen, M. A. L. M. (2018). *Correcting for Publication Bias in a Meta-Analysis with the P-uniform* Method*. MetaArXiv Preprints. https://doi.org/10.31222/osf.io/zqjr9

van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, *20*(3), 293-309. https://doi.org/10.1037/met0000025

Wang, R., Shi, J., & Ye, B. (2022a). Can robots reshape gender role attitudes? *China Economic Review*, *75*, 101852. https://doi.org/https://doi.org/10.1016/j.chieco.2022.101852

Wang, T., Zhang, Y., & Liu, C. (2022b). *Robot Adoption and Employment Adjustment: Firm-Level Evidence From China*. SSRN. https://dx.doi.org/10.2139/ssrn.4210055

Webb, M. (2020). *The Impact of Artificial Intelligence on the Labor Market*. SSRN. https://dx.doi.org/10.2139/ssrn.3482150

Zeugner, S., & Feldkircher, M. (2015). Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R. *Journal of Statistical Software*, *68*(4), 1 - 37. https://doi.org/10.18637/jss.v068.i04

Zhang, L., Gan, T., & Fan, J. (2023). Do industrial robots affect the labour market? Evidence from China. *Economics of Transition and Institutional Change*, *31*(3), 787-817. https://doi.org/https://doi.org/10.1111/ecot.12356