CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Effect of Temperature on Suicide - Meta Analysis

Bachelor's thesis

Author: Daniel Bartušek
Study program: Economics and Finance
Supervisor: doc. PhDr. Zuzana Havránková Ph.D.
Year of defense: 2021

Declaration of Authorship

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

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

Prague, May 2, 2021

Daniel Bartusek

Abstract

In this thesis I investigate the empirical evidence on the relationship between temperature and suicide rates. The previous survey on the relationship suggested that a 1°C increase in temperature is associated with a 1% increase in suicide risk. An estimate of this magnitude would play a significant role in the computation of the social cost of carbon, a concept used to set climate-related regulations by policymakers around the world. I challenge this conclusion using novel, state-of-the-art meta-analysis methods. By expanding the dataset and correcting for publication bias I show that the effect of temperature on suicide rates is three times less than previously suggested, deeming the relationship economically insignificant. Moreover, I did not find any robust evidence for a specific study design that would systematically influence the magnitude of the estimated effect quantifying this relationship.

Keywords temperature, weather, climate, suicide, suicidality

Title Effect of Temperature on Suicide - Meta Analysis

Abstrakt

V této práci zkoumám empirické důkazy o vztahu mezi teplotou a počtem sebevražd. Předchozí průzkum tohoto vztahu ukazuje, že zvýšení teploty o 1°C je spjato se zvýšením rizika sebevraždy o 1%. Odhad této velikosti by hrál významnou roli při výpočtu sociálních nákladů uhlíku, což je koncept, který používají regulátoři po světě ke stanovení zákonů souvisejících s klimatem. Tento závěr zpochybňuji pomocí nejmodernějších metod meta-analýzy. Rozšířením datasetu a ošetřením odhadů o publikační selektivitu ukazuji, že vliv teploty na míru sebevražd je třikrát menší, než hodnota prezentovaná v předchozí meta-analýze, což považujeme za ekonomicky nevýznamné. Dále nebyl nalezen důkaz pro konkrétní charakteristiky studie, které by systematicky ovlivňovaly velikost odhadovaného efektu pro tento vztah.

Klíčová slova teplota, počasí, klima, sebevraždy

Název práce Vliv teploty na míru sebevražd: Meta-Analýza

Acknowledgments

I would like to thank my supervisor, doc. PhDr. Zuzana Havránková Ph.D., who introduced me to the topic of meta-analysis, provided detailed commentary during both the technical and the writing part of this thesis, and always responded to my questions in a timely manner. Furthermore, I am grateful to professor Bob Reed and Sanghuyun Hong Ph.D. at the University of Canterbury, who developed an R-code featuring several meta-analytic methods used in my work (Hong & Reed 2021). Their paper served as an inspiration during the coding part of my thesis. Lastly, I thank my good friend Petr Čala who provided valuable advice during the writing of my thesis.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record
Bartusek, Daniel: *Effect of Temperature on Suicide - Meta Analysis*. Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2021, pages 60. Advisor: doc. PhDr. Zuzana Havránková Ph.D.

Contents

List of Tables

List of Figures

Acronyms

PIP Posterior inclusion probability

IRR Incidence Rate Ratio

FAT Funnel-asymmetry test

PET Precision-effect test

FE Fixed-effect

RE Random-effect

BMA Bayesian model averaging

FMA Frequentist model averaging

VIF Variance inflation factor

ICD International Classification of Diseases

BAT brown adipose tissue

WHO World Health Organization

VSL Value of statistical life

SCC social cost of carbon

Chapter 1

Introduction

Globally, suicide is the leading cause of violent death (World Health Organisation 2019). Although the number of suicide deaths has been in decline, WHO estimates about 800 thousand people kill themselves each year, which is more than those murdered in homicides and killed in wars combined. Especially now, in the midst of global pandemics, the suicide rates are increasing due to COVID-19 (Sher 2020). Why do suicides happen? Researchers, such as Deisenhammer (2003), argue that no individual suicide can be causally related to a single event, but one can show that the risk of suicide increases with some risk factors. Such risk factors identify with social, psychological, and demographic influences; the two echoed by many include weather conditions and economic status of an individual (Schneider *et al.* 2020). Fountoulakis *et al.* (2016), for example, indicate that climate variables can explain more than a third of the variation in suicide rates while economic variables explain no more than a quarter of the variation.

Most of the climate variables have a clearly documented contribution to the risk of suicide: for example, less sunlight and higher air pollution significantly increase the risk of suicide (Fountoulakis *et al.* 2016; Kim *et al.* 2016). However, the message about the effect and its significance, becomes less clear when one searches through the literature examining the impact of ambient temperature. Some studies suggest that colder temperatures are associated with a higher rate of suicide (Kim *et al.* 2016; Page *et al.* 2007), while some studies suggest the exact opposite (Deisenhammer 2003; Preti & Miotto 2000). Several others, including Dixon & Kalkstein (2009) and Tsai (2010), propose that there is no link between the temperature and suicide rates. To make sense of the diverse study results, Gao *et al.* (2019) constructed a quantitative literature review, so-called meta-analysis. The authors impose strict limitations for their quantitative analysis. Based on the subsample of 23 observations of the effect in question, they conclude that the relationship is positive and postulate that a 1°C increase in temperature is significantly associated with a 1% increase in the incidence of suicide.

In my thesis, I build on the work of Gao *et al.* (2019) and examine the relationship between the temperature and suicide rates more thoroughly. I ask three main questions: first, how large is the effect beyond biases; second, what drives the magnitude of the effect; and third, what are the economic implications stemming from the risk associated with this effect. To accomplish this goal, I enlarge the sample by Gao *et al.* (2019) to 186 estimates gathered from 31 studies, use state-of-the-art meta-analysis tools to address the issue of publication bias and heterogeneity in the literature, and exploit the Doucouliagos *et al.* (2012) estimate of the statistical value of life and the Havranek *et al.* (2015) estimate of the social cost of carbon (SCC) to tackle the economics behind the lost lives.

The Value of statistical life (VSL) is the economic benefit of avoiding the death of a person. Since negative VSL implies that the society would save money for every non-living person, which is a controversial statement, it is also subject to publication bias. We use the estimate corrected for bias by Doucouliagos *et al.* (2012). The authors report a value of 1.66 million dollars. Using the suicide report from World Health Organisation (2019), a 1% rise in suicides would increase the minimum estimate of annual lives lost due to temperature by 8 000. In that case, the effect of temperature on suicide should be a significant factor in the computation of the SCC.

The SCC metric is defined as the monetary value of the damage done by emitting an additional tone of carbon (Pearce 2003). Carbon produces carbon dioxide, which has been widely recognized as the main driving factor behind global warming (Nordhaus 2017). If release of carbon increases the temperature and if increase of temperature is associated with higher suicide rates, than setting the carbon tax right could be of incremental influence to the suicide rates.

The majority of the methods used, including the precision-weighted Funnel-asymmetry test (FAT) test utilized by Gao *et al.* (2019), identified a presence of publication bias. Contrary to the findings of Gao *et al.* (2019), our results suggest that the temperature association with suicide rates is subject to selection bias among the researchers. Out of the eight methods deployed to identify the mean estimate beyond this bias, only the Random-effect (RE) and simple OLS identified a statistically significant effect. Interestingly, RE was also used as the primary method for Gao *et al.* (2019) with comparable results. Nevertheless, we rely on the more sophisticated non-linear techniques, which suggest that there is no statistically significant effect of temperature on suicide. If we simply average the estimated mean beyond bias from our methods, of the publication bias methods, we obtain a value of 0.003. Thus, according to our results, the relationship of suicide increase due to temperature changes is overvalued threefold.

To study the heterogeneity behind the effect of temperature changes, Gao *et al.* (2019) use the measure of I^2 on subsets ranging from 18 to 31 estimates. According to the authors, variables such as gender, type of temperature used for the analysis, or climate, show higher heterogeneity. Nevertheless, I^2 is imprecise and likely biased when used with a small dataset (von Hippel 2015). We deploy Bayesian model averaging to identify variables driving the heterogeneity behind the effect. In our main specification, as well as the robustness check, the standard error stood out as the main variable driving the heterogeneity. This contradicts the findings of Gao *et al.* (2019), who did not identify the presence of publication bias. Controlling for seasonality and humidity levels also affects the estimate, although only in the weighted specification of Bayesian model averaging (BMA). Seasonality control is crucial for establishing the effect of temperature on suicide rates, as most studies report the peak in suicides during spring. This finding is consistent with the systematic review by Deisenhammer (2003). The author puts forward humidity and temperature as the main meteorological factors associated with suicide rates. Lastly, the weighted specification supports the notion, that males are more affected by temperature changes than females.

The rest of the thesis is structured as follows: Chapter 2 elaborates on the methods used to measure attempted and completed suicides, and further describes the recent meta-analysis by Gao *et al.* (2019). Chapter 3 describes our inclusion criteria, the methods used for standardizing the effect, and the summary statistics of the effect. Chapter 4 includes comments on the importance of correcting estimates for publication bias and presents our results of the utilized methods. Chapter 5 describes the methodology behind Bayesian model averaging and presents the explanatory variables. Furthermore, this chapter provides the results of our heterogeneity test, along with possible reasoning for the effect of temperature on suicide. Lastly, Chapter 6 summarizes the findings and describes the limitations of this meta-analysis.

Chapter 2

Motivation for the effect and its estimation

2.1 Motivation for the effect

It is possible to find support for the association of temperature and suicide in several scientific fields. From a biological point of view, there are two possible explanations at hand. Helama *et al.* (2013) attribute the relationship to brown adipose tissue (BAT), which reacts to temperature by creating tolerance to one temperature extreme at expense of the other. Active BAT is linked to suicides associated with depression. Over-activation of BAT due to outdoor temperature swings or lifestyle choices leads to increased suicide risk. Another explanation points to the effect of L-tryptophan, which changes in the brain into serotonin. Maes *et al.* (1995) show that high ambient temperature and humidity levels are related to lower L-tryptophan availability, which leads to depression and suicidality. These factors are likely more relevant to men, as they traditionally work outdoors more than women.

Furthermore, women regulate body temperature better than men (Barker *et al.* 1994). From a psychological standpoint, Maes *et al.* (1995) also argue that dysfunctional serotonin activity due to L-tryptophan increases impulsivity and aggression, leading to reckless decisions. People with a mental illness treated by drugs could also be at increased risk during higher temperatures due to dehydration from antidepressant intake (Lin *et al.* 2008). The association can also be attributed directly to socioeconomic reasons, namely the agricultural income. A study by Carleton (2017) compared suicides in growing seasons and discovered that low crop yields due to droughts increase the risk of suicide among farmers during the growing season. They also estimate that heat and drought have been responsible for 59 000 suicides in India over 30 years. This effect was not apparent during the non-growing season.

2.2 Suicide data

In 2019, World Health Organisation (2019) published a brochure with statistics of suicide deaths in 2016. Some of the main findings are that men have a 1.8 times higher suicide rate on average than women and almost three times more in high-income countries. Low and middle-income countries report 79% of total suicides. Regarding age-specific suicide rates, 52.1% of suicides occur before the age of 45 years. Moreover, it is the second leading cause of death for the age group 15-29 years. The mean suicide rate around the world is approximately 12 per 100 000, although this statistic varies substantially.

Tracking and preventing suicide attempts is, in many cases, challenging to manage. Although there have been some attempts to create scales for estimating the risk of suicide, there are some issues. Firstly, the description of steps people take while developing suicidal thoughts is not standardized. Studies use different methods to estimate the thinking process behind developing suicide ideation. Suicide itself is often impulsive. It has been stated that the majority of steps preceding suicide, such as thinking of the place or method, occur within 12 hours before the act, or even simultaneously (Mann 2002; Millner *et al.* 2017).

Nevertheless, even data regarding completed suicide are often misreported. World Health Organisation (2019) states that more than 50% out of the 180 countries used for their dataset have unreliable data, which undermines the data usability in the studies. In addition to that, using datasets of self-reported suicides is questionable in the least, as the data are likely to be underestimated.

A recent study by Snowdon & Choi (2020) provides a review of the main reasons for suicide data under-reporting. When reporting deaths by means of the ICD-10 classification, for example, there is evidence that categories ‘ill-defined or unknown cause’, ‘undetermined intent’, or ‘accidental deaths’, are sometimes used in case of completed suicide. Drowning, accidental suffocation, falls from a height, drug overdose, and road accidents are all examples of deaths where suicide could not be recognized.

Furthermore, in Islamic states or South Korea, there is stigma and shame associated with suicide. Family members of the deceased may report the death to be due to natural causes or an accident (Snowdon & Choi 2020), (Karamouzian & Rostami 2019). There are also expenses related to suicide. In some countries, the cost of medical care associated with suicide is not covered by insurance companies (Karamouzian & Rostami 2019). Moreover, the autopsy is usually requested by the family. Thus, in case of the death of an elderly relative, the family might not request an autopsy. Even if requested, the autopsy might not be thorough, resulting in an 'unnatural' cause of death reported (Snowdon & Choi 2020). Lastly, taking account of death causes in developing countries and rural areas can be troublesome, introducing another factor in suicide under-reporting. To sum up, under-reporting is a major limitation of any study that uses suicide data.

2.3 Measuring suicide

To quantify the effect of temperature changes on suicide, it is necessary to standardize the effect to one common measure. Out of the different measures further presented in Chapter 3, the most optimal strategy appears to be standardizing to the risk of suicide, defined as the change in suicide rate divided by the reported suicide rate. By simply multiplying the value by 100, we obtain the percentage change in the suicide rate, which is also a convenient interpretation.

The meta-analysis from Gao *et al.* (2019) uses the Incidence Rate Ratio (IRR) to specify the relationship. Since we want to compare our results with those of Gao *et al.* (2019), it is necessary to find means of recalculating our results to this measure. The incidence rate is simply the frequency of some event. Rate ratio utilizes the incidence rate in groups exposed and unexposed to the effect in question. In our case, the effect is a marginal move in temperature, and by incidence, we mean the suicide rate. The suicide rate is usually presented as the number of suicides per 100 000. Under the presumption that there is no effect of temperature increase on rates of suicide, the incidence rate of the group exposed to an increase in temperature will be similar to the unexposed group. Thus the IRR will be equal to 1.

The formula for incidence ratio is:

$$IRR = \frac{Rate_{exposed}}{Rate_{base}} = \frac{SR_{base} + SR_{change}}{SR_{base}} = 1 + \frac{SR_{change}}{SR_{base}} = 1 + risk$$

Where 'exposed' symbolizes the suicide rate under the effect of 1°C rise in temperature, and 'base' stands for the baseline suicide rate in the study. A clear conclusion from the formula is that we derive IRR from the risk of suicide by simply adding 1 to our results. The effect distribution is therefore expected to be around 0. Conventionally, studies attempt to show that the researched elasticity is statistically different from 0, and methods used in meta-analyzes are tailored for the 0 value. If we were to perform our analysis using the IRR values, we would also have to alter these statistical methods.

2.4 Previous findings

So far, there has been only one meta-analysis conducted on this topic (Gao *et al.* 2019). The authors estimate the IRR associated with suicide to be 1.01. In other words, an increase in temperature by 1°C increases the risk of suicide by 1%. There are several limitations to this finding. First, authors impose strict inclusion criteria: suicides needed to be reported via the International Classification of Diseases (ICD) and effects already needed to be reported in the form of IRR . Furthermore, the authors only select estimates presented as the final results, or estimates with most covariates, and exclude results of simple OLS. For that reason, only 22 relevant estimates from just 13 studies were used in the main model. Moreover, Gao *et al.* (2019) did not perform robustness checks for their publication bias test. Heterogeneity was assessed by doing subgroup analyzes of gender, climate zone, latitude, and income, dividing the sample into even smaller pieces. To measure the heterogeneity, the authors use I^2, which can be biased to either side for small sample sizes (von Hippel 2015). Lastly, the results were rounded to two decimals, so it is impossible to observe smaller differences in the estimates. For example, the IRR associated with temperature for the middle-income level is 1.02, whereas the high-income estimate is only 1.01. The results can hardly be plausible. This study aims to fix these flaws by expanding the sample of studies and conducting methods, which correct the estimates for heterogeneity and publication bias.

Chapter 3

Data

3.1 Inclusion criteria

Studies were gathered using Google Scholar. The algorithm used in its search engine matches queried words in full texts of studies, rather than simply matching the title, keywords, or abstract. The coverage is thus more precise (Gechert *et al.* 2020). The query used was adjusted to feature most studies included in the quantitative analysis of the meta-study by Gao *et al.* (2019). The final form of the query is ('temperature' OR 'temperatures' OR 'climate' OR 'climatic' OR 'weather')AND('suicide' OR 'suicidal' OR 'suicidality').

The query returned approximately 850 000 results, out of which the first 200 were examined. Additional sources include studies provided by the supervisor and studies identified with snowballing. Snowballing is the process of examining references in studies, which could provide additional estimates. The query was also applied to a restricted time span from 2020 in order to capture recently published studies.

In total, 75 studies with usable effects were identified. The cumulative number of effects in those studies was 743. The following inclusion criteria have been set for the quantitative analysis:

- The effect presented in the study can be recalculated to the risk of suicide and IRR.

- The studies have to present the sample size or a feasible way to estimate it. In the end, only a measure of the mean number of suicides over some time span was approved, as it was most likely calculated from the sample size.

- The final estimate has to be calculated using minimum, mean, or maximum temperature measures. This condition was used to restrict studies that used maximum or minimum suicide-temperature (MaxST or MinST). Several studies calculated temperature associated with the lowest and highest risk of suicide and used this measure in the analysis. These measures produced systematically different effects compared to the normal temperature measures. Therefore, they were also left out.

After this process, our dataset was restricted to 31 studies and 186 data points. For comparison, Gao *et al.* (2019) only used 16 studies. An overview of the studies is available in the Appendix.

3.2 Recalculating effects and standard errors

Three types of estimates other than the risk of suicide were identified. Their distribution is presented in Figure 3.2. Methods of standardizing each of these methods are described below.

Incidence rate ratio As stated in Chapter 2, the only standardization necessary is to subtract one from the estimates.

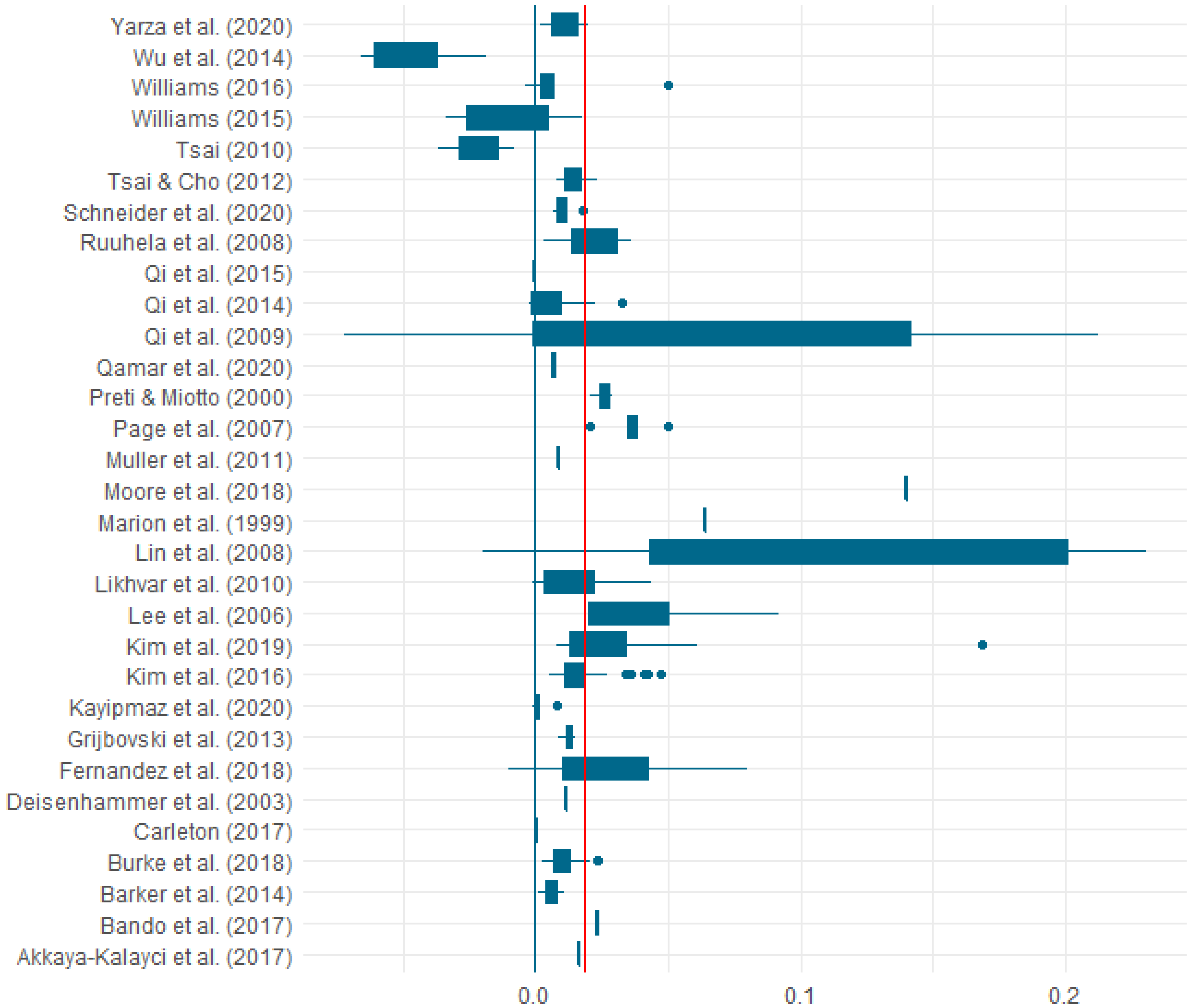
Relative Risk There are inconsistencies in the use of relative risk and rate ratio. Both measures compare rates of incidence. The only difference is that the rate ratio differentiates the group by using two time spans, while relative risk measurement divides groups by the exposure to the observed variable. Nevertheless, the time span in rate ratio is utilized for quantifying the temperature difference. Thus, the rate ratio can be treated similarly to relative risk.

Percentage increase When study reports their estimate in terms of percentage increase in the suicide rate or the number of suicides associated with the change in temperature, the calculation is trivial as well:

$$risk = \frac{percent\_increase}{100}$$

Regression coefficient Some studies reported their result as an absolute change to suicide rate (SR) or the number of daily suicides. Fortunately, most studies also reported the suicide rate or daily suicide count in the area.

Figure 3.1: Distribution of estimates for every study



Notes: Every row represents the distribution of effects in the respective study. Interquantile range is denoted by the box length. Blue vertical line is the 0 intercept, while red line represents simple mean of all estimates. The plot is cutoff at estimate value of 0.2 to better show the between-study variation.

To obtain the risk of suicide, we simply divide the absolute change by the reported suicide rate.

$$risk = \frac{SR_{change}}{SR_{base}}$$

There were 36 estimates, which included suicide attempts in their analysis, 19 of which require the reported suicide rate for the recalculation to IRR. Naturally, attempted suicide has a substantially higher rate than completed suicide. Should we use this value in our calculations, the risk of suicide would always be lower for these estimates. According to McIntosh & Drapeau (2012), for every complete suicide, there are 25 suicide attempts. Therefore, attempted suicide rates were scaled by 25.

Given the heterogeneity of the studies, computing the standard error of suicide risk could not be performed using only one method. We identified three types of cases, which required different computation of the standard error.

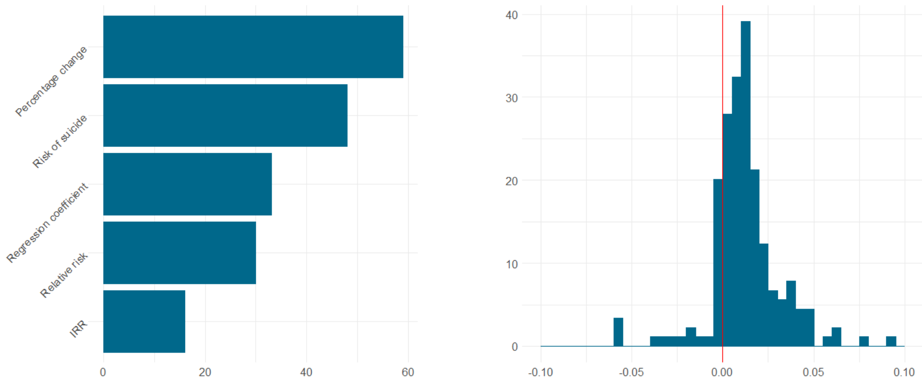
- Ideally, the study would present the standard error along with the effect. In that case, we apply the Delta Method to transform the original standard error. The Delta Method form depends on the method used to calculate the suicide risk. Namely, for the regression coefficient transformation representing an absolute change in suicide rates, we use:

$$se(risk) = var\left(\frac{SR_{change}}{SR_{base}}\right)^{1/2} = \left(\frac{1}{SR_{base}}^{2} * var(SR_{change})\right)^{1/2} = \frac{se(SR_{change})}{SR_{base}}$$

- When there is no standard error, the confidence interval of the reported effect is a sufficient alternative to the Delta Method since it produces comparable results. For the confidence interval of 95%, we use the following formula:

$$se(risk) = \frac{CI_{upper} - CI_{lower}}{3.92}$$

- For some estimates, only the p-value has been reported. In that case, we determined the t-statistics from the p-value and calculated the standard error using the relationship between t-statistics and the effect.



(a) Distribution by effect type (b) Distribution by effect magnitude

Figure 3.2: Cumulative estimate distribution

Notes: Left plot shows the distribution of effect types in our dataset, which have been described in Chapter 3 section. Plot on the right shows the distribution of estimates. For convenience, estimates outside the -0.1 to 0.1 range are not included. There are 7 such estimates. Red vertical line represents the 0 threshold.

3.3 Summary statistics

The effects are dispersed between the studies, as well as within (Figure 3.1). Some studies report only a single value, such as Akkaya-Kalayci *et al.* (2017) with a suicide risk estimate of 0.0016. On the other hand, some produce estimates spanning over the whole range of other estimates (Kim *et al.* 2016). In addition, several studies produce a range of effects significantly further from the mean threshold of 0 (Wu *et al.* 2014).

A simple average of all the effects yields a value of 0.0188, which translates to IRR of 1.019. This value is substantially higher than the results of Gao *et al.* (2019). Clearly, a simple average of the effects is not representative enough to draw any conclusions. By computing the weighted average, where weights are equal to the inverse of the number of estimates in each study, we obtain a significantly lower IRR value equal to 1.0033. Both the minimum and maximum estimate in our dataset are not credible. Should the maximum estimate be the true underlying effect, a 1 °C would raise the risk of suicide by 20%. Summary statistics are presented in Table 3.1. The simple average effect for attempted suicides is 0.007, which is three times less than for the completed suicides. The ratio supports the notion of Gao *et al.* (2019) that completed suicides are more affected by the temperature changes than suicide attempts. Studies using only elderly suicides have, on average, 40% higher estimate than for the unrestricted sample. Nevertheless, there are only 11 estimates for the elderly suicides, so the difference could be due to sampling error. Similarly, studies deploying non-linear methods for their estimation produce, on average, 40%

higher estimates than studies using linear methods. Approximately 22% studies did not control for any meteorological variables or seasonality. These studies produce an average estimate of 0.008, which is 3 times less than the average estimate of studies controlling for both seasonality and meteorological variables. Our measures of income and latitude are in form of continuous variables, so subsetting is not possible. Nevertheless, simple regression shows that the effect size decreases with lower latitude, which contradicts the findings of Gao *et al.* (2019). Each of these observations will be tested for statistical significance in Chapter 5. The data is available upon request.

Due to the high variance in our data, winsorization was considered. Winsorization is a method for treating outliers in data by making them less extreme. Since there are not enough estimates gathered, this method is preferred to simply trimming the outliers. However, leaving outliers untreated could distort the analysis. Moreover, due to our extremely low values of standard errors for some estimates, several methods dealing with publication bias could not have been used. Therefore, we winsorized our data at the 1% level.

Table 3.1: Summary statistics of the estimate and standard error

	Mean	**Standard Error**	**Median**	**Weighted Mean**
Effect	0.0188	0.03755	0.0112	0.0033
Standard error	0.0108	0.0166	0.0053	0.0024

Notes: The summary statistics have been calculated from the full sample. The dataset is consisted of 186 studies from 31 studies. The weighted mean has been computed by dividing every observation by the number of estimates reported by the respective study.

Chapter 4

Publication Bias

4.1 About publication bias

Without any data-based knowledge, the common guess would be that consecutive days of cold temperatures must increase the suicide rate. Early researchers of the effect could be conducting their research with this notion in mind and possibly alter their methods to fit the general view. For example, a model which gives results in accordance with their view could be favored. Furthermore, the journal might be keener to publish statistically significant results that support the current notion (Card & Krueger 1995). Nevertheless, with better methodologies being developed and systematical reviews being published, the conventional view started to be questioned. Review by Deisenhammer (2003) was perhaps the most influential in challenging the classical view:

> *"This finding is a further confirmation of the fact that the emergence of suicidality in a particular person is a phenomenon profoundly distinct from the so-called normal, generally understandable reactions to environmental influences but is the consequence of an individual psychopathological process that is subject to an interaction of exogenous and endogenous factors" (Deisenhammer 2003, p. 403)*

Deisenhammer comments further on the heterogeneity of methodology, and study-specific characteristics, such as seasonality and data granularity, which could have impacted the results of reviewed studies. This review gave rise to more studies analyzing this relationship, the majority of which reported a positive link. However, the motive to publish an effect of a selected sign likely did not disappear. It could have merely changed direction.

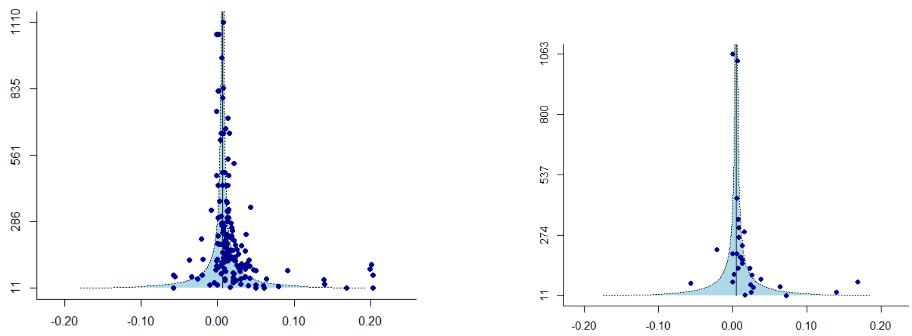
This issue is called publication bias. One reason for it is the need to select results and robustness checks that will be included in the work to keep it concise. Another reason is that the researchers are motivated to produce statistically significant results of a certain sign. When a study fails to reject the null hypothesis due to large standard errors, the conductors will attempt to collect more data or restrict their sample size in order to make their results significant. On the contrary, effects with a lower magnitude that are statistically significant do not meet this constraint and will be published (Card & Krueger 1995; Brodeur *et al.* 2020). Another case could be that effects failing to reject the null hypothesis will stay unpublished, called the file-drawer problem (Stanley 2005; Brodeur *et al.* 2016). Both of these customs might overstate the actual effect.

We can take the work of Blanco-Perez & Brodeur (2020) as an example. In 2015, an editorial statement for health economic journals had been published, urging researchers not to omit findings that do not reject the null hypothesis. Using the difference-in-differences method, they found that the proportion of statistically significant results decreased by 18%. Moreover, Brodeur *et al.* (2020) compared results published in top economic journals and showed that certain methods produce systematically larger estimates than others.

Perhaps the most convenient method for detecting publication bias is simply plotting the data using a funnel plot. When estimates are plotted against the inverse of their standard errors, we can expect the most precise values to be densely distributed around the true effect. In contrast, less precise estimates are scattered on both sides of this effect (Stanley *et al.* 2010). Without publication bias, the plot should be symmetrical around the true effect, forming the shape of an inverted funnel, hence the name ‘funnel plot’. The left plot in Figure 4.1 shows minor right skewness of the estimates, as well as the cut-off at 0. The bias is even less apparent in the right plot, where only median points from each study are plotted. In our case, making a decision solely by a subjective judgement of the funnel plot is insufficient. Thus, it is needed to compute the bias numerically using state-of-the-art baseline methods.

4.2 Baseline methods

As previously mentioned, publication bias happens when researchers lower their standard errors to make their estimates statistically significant. Consequently, we can expect the standard error to be correlated with the effect itself. When we regress the estimate on the standard error, the intercept represents the



(a) Precision (b) Median Precision per Study

Figure 4.1: Funnel plot

Notes: The funnel plot with all estimates is on the left. Estimate values are plotted against the inverse of their standard errors. Precise estimates will be distributed around the true vale, while less precise estimates are dispersed at lower levels of the plot. Funnel plot with only median effect from each study is on the right. To better observe the funnel shape, we plot only estimates with precision less than 1200, which leaves out 12 data points from the left plot.

mean effect corrected for the influence of standard errors. This intercept can therefore be viewed as the 'true effect' (Stanley 2005).

Thus, we are interested in the intercept β_0 in the following model:

$$risk_{ij} = \beta_0 + \beta_1 * SE(\hat{risk_{ij}}) + u_{ij}$$

Where $risk_{ij}$ is the i-th estimate from j-th study. We can then perform a FAT test to identify the publication bias by rejecting the null hypothesis that the beta coefficient of SE is 0. Moreover, Precision-effect test (PET) provides further confirmation of a genuine effect beyond this bias by testing the null hypothesis that the intercept is not 0 (Stanley 2005; Stanley *et al.* 2010).

Since a large sample size also decreases the variance, the standard error is likely heteroskedastic. Therefore, it is recommended to apply the inverse of the standard error as weights for the regression (Ioannidis *et al.* 2017). Another commonly used weighing scheme is the inverse of the number estimates produced by a single study since heterogeneity of studies could also affect the estimates. The range of estimates from a single study in this meta-analysis spans from 1 to 33. By applying weights to the regression, we assure that every study impacts the result in the same way.

Another way of dealing with unexplained heterogeneity in studies is to use the study-level Fixed-effect (FE) and the Random-effect (RE) methods. In FE, it is assumed that studies come from one sample and have one common true effect. Thus, sampling error can only arise within each study. RE recognizes that the effect can not be similar due to heterogeneity between the studies. Therefore, RE uses a weighing matrix of both the within and between-study variance (Bom & Rachinger 2019).

To summarize, we are using five variations of the primary regression, four of which use convenient weighing schemes. The results are presented in Table 4.1, along with their standard errors. The standard errors are clustered on the study level in order to account for the within-study correlation since non-clustered errors could introduce false precision levels of models.

Considering the FAT test, every method we used identified a presence of publication bias in our data, which was not previously apparent from the funnel plots. The tests support our initial notion that studies analyzing the effect of temperature on suicide are indeed subject to publishing results based on their significance. On the other hand, the mean estimates corrected for the relationship with standard errors are not uniform. The reason behind the statistical significance of RE method is that it balances the dataset by putting more weight on smaller studies. The main benefit is that unique environments and methods used in studies are valued, which arguably provides more credible results. In comparison, FE would have been more plausible if studies came from an identical environment, were performed by the same researchers, and used common methods, which is not the case.

Each estimate is below 0.01, which translates to the estimate of 1.01 reported by Gao *et al.* (2019). Due to the correction for publication bias, our results are lower than 0.01 and insignificant in the majority. Since only two out of our five methods identified statistically significant presence of non-zero mean estimate, we can not claim that temperature does have an effect on the suicide rate based on these linear tests.

The methods mentioned above assume a linear relationship between the estimate and its standard error. At some values, however, we could expect to find non-linear jumps or kinks in these variables. The true effect could then be underestimated by the FAT-PET tests, assuming that the mean beyond bias is bigger than 0 (Bom & Rachinger 2019). There could also be bias in the standard error due to random sampling error introduced by the researchers (Stanley 2005). For this reason, we apply several new methods, which account for publication bias, while assuming non-linearity. The general idea behind the majority of the so-called selection models is to impose restrictions on the dataset by either treating subsamples of estimates or removing them completely.

Table 4.1: Linear tests of publication bias

	OLS	**FE**	**RE**	**Precision**	**Study**
SE	0.854***	2.7676***	1.0914*	2.7155***	0.8214*
Publication bias	(0.218)	(0.5189)	(0.4156)	(0.5449)	(0.3335)
Constant	0.0095*	-0.0001	0.0079**	-0.00002	0.0013
Mean beyond bias	(0.0038)	(0.0005)	(0.0025)	(0.0009)	(0.0009)
Studies	31	31	31	31	31
Observations	186	186	186	186	186

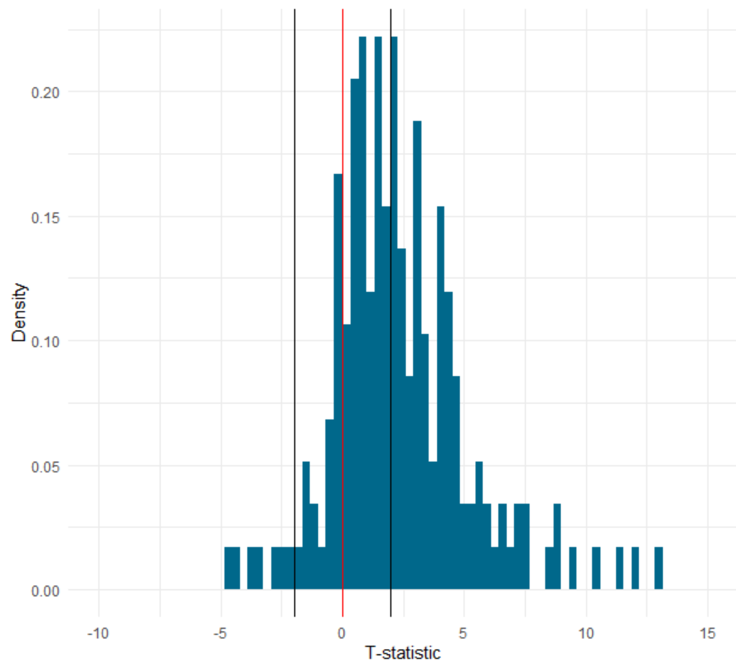
Notes: This table provides results for the linear techniques estimating publication bias. Upper row represents the FAT test of publication bias. Lower row tests for the mean estimate beyond bias. Study-clustered standard errors are provided below each coefficient. All tests are based on the regression $risk_{ij} = \beta_0 + \beta_1 \cdot SE(risk_{ij}) + u_{ij}$. First column provides results of this simple regression. FE accounts only for within-study variation, while RE considers the between, as well as within, study variation. Precision column applies weight proportional to the standard error of each estimates, while the study column weights the effect by the number of estimates reported in every study.
. $p < 0.10$, * $p < 0.05$,* $p < 0.01$, *** $p < 0.001$.

Perhaps the simplest method of non-linear publication bias estimation is the Top10 developed by Stanley *et al.* (2010), whose work shows that if levels of statistical significance affect the chance of a paper being published, the sample of effects is indeed not representative and any statistical computation method will be biased. Therefore, meta-analysts might be better of by simply discarding 90% of data having lower levels of precision and leaving the rest with higher precision. However, Stanley *et al.* (2010) recognize that this method contradicts the traditional Central Limit Theorem and rather presents the work with the intention to highlight this issue. When applied to our dataset, the final estimate is equal to 1.03, which is significantly higher than in other methods, both linear and non-linear. This is due to the fact, that the subset with the highest precision also reports some of the highest estimates. Thus, it is necessary to employ more sophisticated methods.

Another method used to detect publication bias is called the Weighted average of adequately powered by Ioannidis *et al.* (2017). This method also recognizes the tendency to publish estimates simply by gaining statistical significance, or in other words, passing the 1.96 t-statistic threshold. Eligible estimates in this method should have adequate power and their standard error should be smaller than the absolute value of the effect divided by 1.96 + 0.84, where the former comes from the statistical significance and the latter from the definition of adequate power (Ioannidis *et al.* 2017). Our dataset contains eight adequately powered estimates, all of which come from one study by Carleton (2017). High precision of these estimates is likely due to the statistical methods used rather than the structure of the underlying dataset. Therefore, the results of this method should not be valued too highly.

Under the notion that insignificant results are under-reported, Andrews & Kasy (2019) apply weighted distribution theory to the insignificant estimates rather than removing data based on precision. Instead of choosing one weighing scheme for the whole range of values, a step function will be used to apply different weights to every interval of the reported p-values. The cut-offs in these intervals are set to the conventionally reported values of 0.001, 0.01, 0.05. The correct functional form of these weights is determined with the maximum likelihood function (Hedges 1992). Jumps at the conventional values, such as 0 and 1.96, are not visible in our data (Figure 4.2). This method also provides a mean estimate calculated using the weighted data presented in Table 4.2.

Figure 4.2: T-statistics distribution of the effects



Notes: The figure shows the distribution of t-statistics in our dataset. Black vertical lines represent the 1.96 threshold, which translates to the commonly used p-value of 0.05. Red vertical line denotes the 0 threshold. Suspected jumps at these values are not apparent judging by the plot.

The last method we chose is The Stem Based Method (Furukawa 2020), which uses the logic of Stanley *et al.* (2010), but makes the threshold relative to the sample. The optimal number of studies is determined by minimizing the mean squared error in equation $minMSE(n) = Bias^2(n) + Var(n)$. As the number of studies increases, bias increases due to the inclusion of less precise studies. On the other hand, the variance decreases due to more information. In our sample, only 12 out of 186 meets the criteria of this method (Figure 4.2).

Unfortunately, our dataset was not fit to utilize the Endogenous kink method by Bom & Rachinger (2019). Similarly to Hedges (1992), the authors identify important p-values resembling significance thresholds and alter the standard errors at these cut-offs with a piecewise linear function. Nevertheless, the standard errors in our sample are too small, making a key value, calculated as $(SSR/SE)^2$, too high to identify spurious standard errors. In our case, the Endogenous kink only differs from linear estimation by applying $1/SE^2$ weights to the regression, which is known as the PEESE method (Stanley 2005).

Table 4.2: Non-linear tests of publication bias

	Andrews & Kasy (2019)	**Ioannidis et al (2017)**	**Furukawa (2019)**
Mean beyond bias	0.003	0.00015	0.0008
	(0.002)	(0.00018)	(0.0014)

Notes: This table provides results of our 3 main non-linear techniques for publication bias determination. These methods only provide estimation of the mean beyond bias. Clustered standard errors are presented in the parenthesis. None of the techniques identified the estimate to be statistically different from zero.

4.3 Extensions

As an extension to our baseline methods, we use the caliper test, developed by Gerber *et al.* (2008). While the caliper test does not establish an estimate of the underlying effect, it provides a test of publication bias using different rules in comparison with the baseline methods. The caliper test does not assume a relationship between the main effect and the standard error. Instead, it compares frequencies around key values in the distribution of the t-statistics to identify sudden jumps, which would indicate the presence of publication bias. In our case, these values will be set to 0 and 1.96, as we have right-skewed data. It is not feasible to make similar tests for the -1.96 value due to a low number of data points around that value.

Since we have a low number of observations, it is necessary to set the calipers large enough to secure statistical significance. On the other hand, setting the caliper too wide would not capture the possibility of jumps around the threshold well, as values further from the thresholds are less likely to be biased. The lowest caliper we set comprised only 28 variables. All six of our tests failed to identify a publication bias since the frequencies on both sides of the thresholds have similar distribution (Figure 4.2).

Table 4.3: Caliper tests for selected thresholds

Caliper size	**0.4**	**0.6**	**0.8**
Threshold: 0	0.036 (0.096)	0.059 (0.086)	0.059 (0.086)
Observations	28	34	34
Threshold: 1.96	0.081 (0.09)	0 (0.08)	0.012 (0.079)
Observations	31	38	41

Notes: In total, 6 caliper tests have been performed. For the two significant thresholds, 3 calipers of different sizes were used. None of the tests identified a presence of publication bias. Caliper around the 0 threshold did not comprise more values upon expanding from the band 0.6 to 0.8.

To conclude, most of the methods applied in this meta-analysis identified a presence of publication bias. Results concerning the mean corrected for the publication bias are inconsistent. A majority of the methods used estimated the effect to be less than the IRR equal to 1.01, reported by Gao *et al.* (2019). Out of the eight methods utilized to estimate the mean beyond bias, only two found a statistically significant value. One of these methods, the RE, was also used by Gao *et al.* (2019) with comparable results. According to our results, the claim that temperature changes affect suicide rates can not be made. Perhaps the only certain observation is that the standard error affects the magnitude of the estimate. The next chapter of this meta-analysis will put other explanatory variables to use, rather than modelling the regression using only the standard error. Using a broader context of the studies, we should be able to identify the study characteristics, which also affect the estimate.

Chapter 5

Why estimates vary?

5.1 Heterogeneity

Although publication bias affects the underlying mean estimate, there are still other study characteristics that could systematically move the effect in either direction. We divide potential sources of heterogeneity into categories such as the data origin, methods used, publication characteristics, or factors that the original study controls for. Every factor from these categories could be of a significant effect on the final mean estimate. For instance, families in countries with low socioeconomic status might not own an AC unit, which would expose them further to the effect of heat waves (Burke *et al.* 2018). Another example is that studies might restrict their sample to younger or elderly aged people. Elderly people might deal with extreme temperatures worse, resulting in higher suicide rates. Lastly, suicide is a highly seasonal matter. Out of the 31 studies in our dataset, 17 reported a peak in suicide frequency for one of the seasons. Only two studies reported the peak in fall or winter, while the rest finds it in spring or summer. Therefore, studies that do not control for seasonality will likely find a positive association of temperature with suicide since temperature also rises with the spring season. Contrary to Gao *et al.* (2019), we include results of simple OLS since they also carry information of the place and time during which the study was conducted. To individualize the effect of temperature, it is recommended to deploy methods, which identify these variables in the context of the estimates by simple regression. Our data consists of 37 variables which could be of importance for the mean estimate. Thus, the regression will have this form:

$$\hat{risk}_{ij} = \beta_0 + \sum_{l=1}^{37} \beta_l X_{l,ij} + \gamma SE(\hat{\sigma}_{ij}) + u_{ij}$$

However, not every variable affects the estimate. Should every study characteristic be included, we risk overfitting our model and introducing collinearity, which reduces the precision of the model. Using only variables, which we deem logical to use with respect to previous literature, is not ideal as well because we might miss some relationship, which is not apparent at first sight.

The first step is to treat collinearity. Since we have a high number of variables but not too many observations, multicollinearity is likely. The convention in meta-analysis is to reduce the number of variables until the maximum value of Variance inflation factor (VIF) is under 10. By removing variables with high VIF and ambiguous relationship to suicide risk, we cut down the number of variables to 25. Using this procedure, we obtain the maximum VIF value of 9.09.

It is not feasible to manually select the correct variables from the rest. Given our 25 independent variables, we would have to run $2^{25} \approx 33000000$ different combinations, which would take an immense amount of time. This issue, called model uncertainty, can be addressed using BMA (Eicher *et al.* 2011).

BMA does not require a concrete set of independent variables to be chosen in advance. Instead, it runs a set number of models and assigns to each model its posterior model probability, which increases with the model fit, but decreases with the number of variables in the model (Havranek 2019). We can then specify for each variable its Posterior inclusion probability (PIP), which is calculated as the sum of all posterior model probabilities in which the variable has been included (Gechert *et al.* 2020). Markov chain Monte Carlo algorithm is then used to traverse only models with high PIP (Madigan *et al.* 1995). Afterwards, variables with PIP higher than an artificially set threshold are included in the final set of statistically significant variables. The coefficients for these variables is calculated as the weighted average of the coefficients in previously run models, using the posterior model probabilities as weights (Havranek *et al.* 2018).

To use BMA, it is necessary to choose the weight of the prior probability of each coefficient, called the g-prior. Priors are usually set to zero unless there is a strong conviction for some of the variables affecting the main estimate. For the g-prior, we will be using the unit information prior, which assigns weights equal to one individual observation (Havranek *et al.* 2018). Moreover, prior model probabilities also have to be set in advance. Regarding model priors, the dilution prior will be used as our primary choice. Dilution prior treats multicollinearity between the variables by weighing the models based on the correlation matrix of included variables. This is our preferred choice due to a high number of explanatory variables with respect to a relatively small dataset.

To observe the gender differences, we will restrict the subset to studies reporting the proportion of males and females in their data. In this specification, we will only comment on the effect of gender since the full sample is more representative for describing the rest of the variables.

It is appropriate to compute BMA using other modifications. Apart from increasing the number of iterations, the choice of priors in BMA also matters. As reported by Havranek *et al.* (2018), the choice of g-priors rarely produces significantly different results. On the other hand, the choice of model prior affects them considerably. Typically, the uniform prior would be the preferred option. Uniform model prior assigns a similar probability to all models. Therefore, models with a mean number of variables will be more heavily represented, while models with very little or close to all variables will not be valued well. Another option would be to deploy Frequentist model averaging (FMA) with robust standard errors (Hansen 2007). FMA includes all explanatory variables and weights them. Unfortunately, neither of these methods treats the presence of multicollinearity in the data. For that reason, we are not able to use these methods as a valid robustness check.

Apart from focusing on the priors in BMA parameters, it is also possible to perform robustness checks with weighted data. For similar reasons to the ones mentioned in Chapter 4, we will apply weights proportional to the number of estimates per study. Weighing by the standard error was not possible. Small standard errors in our dataset introduce VIF values over 6000, and we would have to remove a majority of explanatory variables to treat it. The last convenient robustness check we use is an ordinary least squares regression with variables from our BMA specifications, unweighted or weighted, which passed the significance PIP threshold of 0.75.

5.2 Variables

Data characteristics This category includes information related to the sample of the respective studies. As we have established in Chapter 4, the standard error affects the effect magnitude, and the number of observations in original studies is by definition correlated with the standard error. Thus, this variable will not be included in our BMA specification.

Regarding data granularity, more than half of the studies use daily suicide data. At first thought, this is preferred to broader time spans. However, temperature changes might affect people with days of delay, which would not have been captured by daily data (Deisenhammer 2003). For that reason, we use a dummy with a baseline category for studies with daily data and with a reference category of weekly, monthly, or even annual data granularity. Some estimates only used data for the elderly population, which is also accounted for using a dummy variable, for reasons stated at the beginning of this chapter.

Some studies, such as Barker *et al.* (1994), used suicide data from multiple countries to determine the effect of temperature on suicide. In comparison with the panel data, analysis of suicide using simple time-series from only one source could be lead by local variables, not necessarily the temperature itself (Dixon & Kalkstein 2009; Fernández-Niño *et al.* 2018). Therefore, this will be included in our BMA model.

Another dummy variable will denote whether the study used only completed suicides or included attempts as well. According to Preti & Miotto (2000), attempted suicide is more sensitive to seasonality and also under-reported in comparison with complete cases. Moreover, it is necessary to account for our scaling of the attempted suicide rate by 25. We presume that attempted suicide will have a negative effect on the overall estimate. As previously stated in Chapter 2, suicide is often misclassified. Therefore, we include a dummy variable controlling for the use of ICD codes to identify suicides. ICD is a credible international classification of diseases.

Specification This category comprises the study setting and the controlling factors in each study. Regarding socioeconomic variables, few studies accounted for them in their analysis. Deisenhammer (2003) points out that controlling for non-climatic variables is needed to identify suicide-temperature association correctly. For that reason, we gathered state-level median expenditures and standard suicide rates in each country, as these variables play an important part in explaining the effect of temperature. Unfortunately, few studies controlled for psychological state or different socioeconomic factors in their dataset. The effect of temperature could be the catalyst for people already having psychosocial risk factors related to suicide (Deisenhammer 2003; Schneider *et al.* 2020).

Data of individuals is difficult to obtain. Thus, we will only make use of the aforementioned macro-level indicators.

Since temperature changes are likely experienced more severely in some climatic zones than others, we include the absolute value of latitude in each country (Dixon & Kalkstein 2009). For example, Kim *et al.* (2019) conducted their results on several areas worldwide, and attributed the non-linearity of temperature-suicide relationship to areas having a broader range of temperature extremes. Moreover, since most studies identified a peak in suicide rates in spring, the non-linearity of the effect is more probable (Schneider *et al.* 2020). The latitude variable reflects an overall climate state in every country. Lastly, a variable for the gender proportion of the sample was also gathered, although only for approximately two thirds of studies. We also attempted to categorize the types of suicide as violent or non-violent. Unfortunately, only 7 out of our studies reported this measure, so a subset analysis was not possible.

Studies use minimum, mean, as well as maximum temperature values to model the association with suicide rates, which will also be represented by dummy variables. This variable captures the possibility that people could be sensitive to extremes rather than responding to deviations from the mean. Dummy variables will also be deployed to denote whether a study controls for seasonality or time trend. Undoubtedly, temperature is closely correlated with the seasons. Day in the week, weekends, and holidays will also be coded in our BMA model. Moreover, temperature is correlated with other meteorological variables, which alter depending on the seasons. These variables, such as sunshine, rainfall, atmospheric pressure, or humidity, could also create biological impulses inducing suicide.

Statistical approach Studies which reported their effect in an absolute change to suicide rate or count could be less biased compared to those with effect measured in IRR. For that reason, we divided the variable definition into categories and included them in our model. As stated earlier, periods with temperatures far from the average might be of greater effect than a simple daily association (Dixon & Kalkstein 2009). Therefore, we deploy a dummy variable equal to 1, if studies allow lagged forms of temperature in their model or compute cumulative effects of temperature on suicide. Surely, the relationship between temperature and suicide could prove to be non-linear. Thus, we control for whether the model of choice in a study allows non-linear relationship. This also includes studies that only allow temperature above a certain level in their analysis, as they also impose non-linear restrictions on their data.

Publication characteristics We include the number of citations of each paper since more cited papers are often conducted using correct statistical procedures and published in trusted journals. Furthermore, suicide data in our dataset ranged from 1967 up to 2019. Over the years, researchers likely developed better techniques, which yield more robust results. Moreover, the effect of temperature on suicide could have changed due to socioeconomic status improvement or global warming. For these reasons, we deploy a variable representing the midpoint year of every study to capture the change over time.

It is important to pronounce one distinguishing feature of our data. Researchers often analyzed suicide in association with other meteorological variables along with temperature. It would not always be clear whether there is one model containing all the variables or whether we are dealing with multiple simple regressions. Moreover, researchers often acknowledged the importance of lagged temperature inputs, non-linear specifications, and seasonal trend but did not include them in their final model after discovering negligible effects of these variables, for example, using the hiearchical regression model (Barker *et al.* 1994; Deisenhammer 2003; Fernández-Niño *et al.* 2018; Grjibovski *et al.* 2013; Likhvar *et al.* 2011; Williams *et al.* 2015; Schneider *et al.* 2020; Preti & Miotto 2000). This issue required increased attention during the coding of the study characteristics.

Table 5.1: Definitions and summary statistics of explanatory variables

Variable	Description	Mean	SD	WM
Variable definition				
Risk Estimate	Estimated risk associated with 1°C increase in temperature	0.019	0.003	0.003
Standard error	Standard error of the risk estimate	0.011	0.001	0.002
Data characteristics				
Panel data	= 1 if panel data are used in study (reference category: time-series)	0.301	0.034	0.04
Daily data	= 1 if study used daily data of suicide and temperature (reference category: weekly, monthly, annually)	0.688	0.034	0.091
Complete cases	=1 if study uses only completed suicides in analysis (reference category: pooled complete suicides and attempts)	0.796	0.03	0.118
Elderly sample	=1 if if study used elderly sample in analysis (reference category: no age restrictions)	0.059	0.017	0.015
Specification				
Day of the week control	=1 if study controls for day of the week, or weekends compared to workdays	0.392	0.036	0.029
Holidays control	=1 if study controls for holidays	0.14	0.025	0.022
Seasonality control	=1 if study controls for seasonality or time trend	0.769	0.031	0.105

Continued on next page

Table 5.1: Definitions and summary statistics of explanatory variables (continued)

Variable	Description	Mean	SD	WM
Daylight control	= 1 if study controls for sunlight, sunshine or radiation	0.403	0.036	0.08
Rainfall control	= 1 if study controls for amount of precipitation or rainy days	0.43	0.036	0.097
Humidity control	= 1 if study controls for humidity	0.43	0.036	0.4
Atmospheric pressure control	= 1 if study controls for atmospheric pressure	0.371	0.036	0.042
Lagged temperature analysis	= 1 if study allows lagged forms of temperature in analysis (reference category: only direct association)	0.296	0.034	0.048
Minimum temperature analysis	= 1 if study uses minimum temperature in analysis of the relationship	0.07	0.019	0.021
Maximum temperature analysis	= 1 if study uses minimum temperature in analysis of the relationship	0.134	0.025	0.029
ICD Coding	= if suicides in study sample have been labeled with ICD codes	0.672	0.035	0.092
Log median expenditures	The logarithm of median expenditures in country, where the study was conducted	6.513	0.055	1.102
Log latitude	The absolute value of latitude in country, where the study was conducted	3.368	0.045	0.579
Log suicide rate	Logarithm of rate of suicide per 100 000	2.102	0.072	0.341
Statistical approach				
Non-linear model	= 1 if study used model, which allows non-linear relationship between temperature and suicide rate, or if input temperature values are restricted	0.280	0.033	0.054
IRR variable measure	= 1 if study reports the effect as rate ratio or relative risk (reference category: association with suicide rate)	0.247	0.032	0.043
Absolute variable measure	= 1 if study reports the effect in absolute change to suicide rate or count (reference category: association with suicide rate)	0.25	0.03	0.26
Publication characteristics				
Log citations	Logarithm of the number of times the study has been cited (Google Scholar citations)	3.238	0.110	0.534
Midpoint	Mean year of the data used minus the earliest mean year in the data	17.422	0.688	2.813

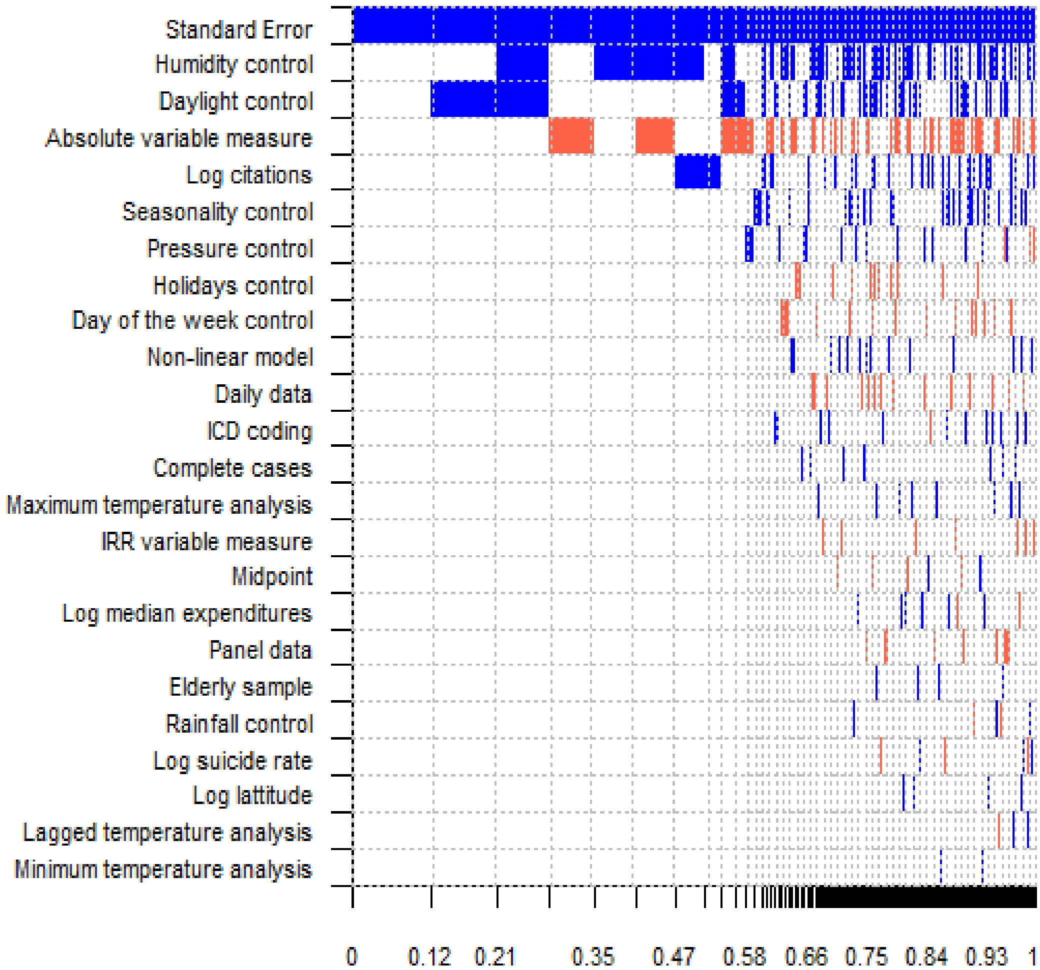
Notes: This table contains the selected summary statistics for the variables included in the BMA model. SD = in sample standard deviation, WM = weighted mean proportionally to the number of estimates in each study, ICD = international classification of diseases, IRR = incidence rate ratio associated with the temperature change

5.3 Results and Robustness Check

Results of BMA are illustrated in Figure 5.1. Variables are presented on the vertical axis, sorted by their PIP. The horizontal axis represents regression models, which are sorted by their posterior model probabilities. The most successful models are on the left. For each model, the colored variables in a column have been included. Red color means that variable has a negative effect on the estimate, while blue color symbolizes a positive sign. Considering the top model in our primary specification of BMA, only one variable is included.

To our surprise, only the standard error has been evaluated as statistically significant for the estimate, with PIP ≈ 1. Other variables have PIP less than

Figure 5.1: Model inclusion in Bayesian model averaging



Notes: This figure illustrates the results of our baseline Bayesian model averaging specification. Every row symbolizes a single explanatory variable used in the models, sorted by their posterior inclusion probability. The columns represent individual models. Blue color means that the variables increases the effect, while red color decreases the effect. If the cell is white, the variable was not used in the model. Models are sorted by their posterior model probabilities. The best models are on the left. In the best model, only the standard error was included.

0.5, which is by common standards viewed as non-significant since there is not enough evidence to show the presence of the underlying relationship (Havránek *et al.* 2021). Thus, the standard error is still the primary variable affecting the overall estimate, and higher values of standard error are associated with higher estimates. This observation is sufficient to claim that the correlation between estimates and their standard errors was not driven by omitted variable bias hidden in the context of studies. Therefore, BMA also serves as a robustness check for the findings presented in Chapter 4.

None of the other variables passed the minimal PIP threshold of 0.75. Since the variables are not statistically significant, there is no reason to comment on the sign of their posterior mean coefficients. The main finding from this specification is that the setting and methods chosen by the researchers have little importance when performing an analysis on the effect of temperature on suicide. PIP and mean values of the explanatory variables are presented in Table 5.2.

Nonetheless, the absence of statistical significance is a valuable observation on its own for several of our inputted variables, specifically the ones denoting controlling for seasonality, the use of non-linear techniques, and inclusion of suicide attempts along with completed suicides. Apart from the explanation associated with the coding of the explanatory variables, we can find further reasoning in the literature.

A potential explanation for the lack of seasonality could be drawn from studies, which reported the peak in suicide rates during winter months. A study by Ajdacic-Gross *et al.* (2007) reported the highest positive correlation of temperature with suicide rates in the winter months. The authors suggest that the lack of cold in winter, rather than periods of hot temperatures in summer, triggers suicidality. Therefore, the positive relationship of temperature and suicide could be observed in any season, which mitigates the effect of seasonality.

Our analysis reports no significant change in using non-linear models, although some authors provided valid reasoning for a possible non-linear relationship (Kim *et al.* 2016). Newer studies by Kim *et al.* (2019) and Sim *et al.* (2020) use non-linear methods and identify an inverted J-shaped curve for the association. The findings of Kim *et al.* (2019) apply only on Asian countries with warmer climates since they have more days with higher temperatures, allowing for precise results. Sim *et al.* (2020) report overall results consistent with Kim *et al.* (2019), whose research was set in Japan. Nonetheless, the colder areas were more subject to the non-linear trend in their regional analysis. A clear argument would be that people living in colder areas adapt worse to sudden temperature jumps. Unfortunately, these studies firstly estimate temperature levels, for which the highest suicide rates are reported, and use that as a measure for their analysis. This measure is called the maximum suicide-temperature (MaxST), and estimates using it are in some cases 20 to 30 times higher than our reported mean. Therefore, using these studies would greatly affect our mean estimate.

No differences were found between studies using only completed suicides in comparison with pooled suicides and attempts. This contradicts the results of Gao *et al.* (2019), who find the association to be higher for completed suicides. To better differentiate between suicide attempts and completed suicide, more studies using only one of the options, such as Müller *et al.* (2011); Preti & Miotto (2000), would be needed.

When using our proposed weighing scheme, it is necessary to apply it to discrete variables and the dummy variables. Characteristics of studies with few estimates will then be more heavily valued since their dummy variables remain equal to one. On the other hand, the same variable in a study with 33 estimates reduces the value of the same dummy variable to 1/33. As a result, using this measure as weights effectively treats the within-study correlation. The presence of a within-study correlation can be expected for studies in which the researchers alter few study aspects between their reported estimates. In our case, several studies report a half of their estimates with control for some meteorological factor and a half without it (Kim *et al.* 2016).

Moreover, since gender and type of suicide are not included in our baseline BMA specification, studies distinguishing their estimates solely by these factors would have a 100% within-study correlation for the characteristics of their estimates (Burke *et al.* 2018; Lee *et al.* 2006; Ruuhela *et al.* 2009; Carleton 2017). By weighing the data, we also increased the maximum variance inflation factor. In order to use weighted BMA as a valid robustness check, we had to reduce VIF from 33 to 15.5 by omitting median expenditures and latitude from the model. Since these variables were not significant in the main specification, we are not omitting any crucial information.

Applying weights equal to the inverse of the number of estimates reported in each study produces several estimates with PIP higher than 0.5 (Figure B.3). The standard error remains the best variable for explaining the variance in estimates, although its PIP has decreased to 0.87. The variable representing control for seasonality now has almost identical PIP, while its statistical significance in the baseline model was negligible. This further highlights the effect of within-study correlation in our data. If certain studies with homogeneous model specification dominate the effect of explanatory variables due to high number of estimates, the within-study differences are not apparent, which is the case in our baseline specification.

There were only four studies, which reported estimates treated, as well as untreated, for seasonality. Seasonality likely became significant due to those studies. The importance of controlling for seasonality has been argued for at the beginning of this chapter. Most researchers likely acknowledge the importance of controlling for time trend or seasonality as a crucial variable and considered untreated potential untreated estimates as not precise enough to be published. Furthermore, seasonality and temperature are hardly distinguishable in studies using monthly or even annual data granularity. In both our model specifications, controlling for seasonality increases the effect size on average. In other words, failing to control for the time trend or seasonality reduces the estimate size, as seasonality itself is likely partly responsible for suicide cases. Frequentist OLS analysis also confirms the positive effect of seasonality control on the 10% significance level.

Although the proportion of males in the sample has not been found as significant with PIP equal to 0.2 for our baseline specification, applying our weighing scheme increased the PIP of this variable to 0.746, making it the third most statistically significant variable in the restricted dataset. Similarly to seasonality, the insignificance in the baseline model could be driven by studies having an above average number of estimates with the same proportion of males to females. By weighing the data, smaller studies that report estimates for subsamples consisting of strictly men or women reveal the likely significance. For both the baseline and the weighted model, the coefficient for the proportion of males was positive. This serves as a confirmation that males are more affected by temperature changes than women, which we briefly argued for in Chapter 1. To an extent, this could be attributed to a higher overall proportion of males in the included studies (Preti & Miotto 2000; Wu *et al.* 2014). Men likely experience temperature changes in a worse manner due to lower ability of temperature regulation. Moreover, men traditionally work more outdoors than women, which makes them exposed to weather effects.

Nonetheless, the seasonality of male suicide could be driven by some other unobserved factor. The difference in suicide between men and women changed drastically with the transformation of traditions in a postmodern age. Among the main reasons for gender discrepancy in suicide rates, Möller-Leimkühler (2003) puts forward masculine inexpressivness, increased alcohol and drug consumption, and strengthening male gender-role as a competitive and self-actualized human being, which can create overload-induced stress. Although none of these reasons is clearly seasonal, some unmentioned factors might better explain the difference in suicide.

In the weighted BMA version, humidity almost passed the PIP significance threshold with a value of 0.749. Similarly to controlling for seasonality and time trend, it holds that when humidity levels are not controlled for in the model, the effect of temperature on suicide rates is underestimated. Therefore, humidity also affects the estimate. In our baseline BMA specification and the OLS model, the positive sign of this relationship is confirmed. The coefficient in our OLS model is significant on the 5% significance level. Temperature and humidity were identified as main meteorological factors by Deisenhammer (2003), whose study represented the turning point with regards to studying the relationship between meteorological variables and suicide.

The last variable with a relatively higher PIP of 0.73 in the weighted model is atmospheric pressure control. While the sign of this variable is positive in the baseline specification, which would imply similar reasoning as above, weighted BMA and OLS report a negative sign. Nonetheless, this variable is not statistically significant in the OLS model, and the PIP threshold was not passed.

Table 5.2: Why estimated elasticities vary

Response variable:	Bayesian model averaging (base)			Bayesian model averaging (weighted)			OLS		
Risk of Suicide	Post. mean	Post. SD	PIP	Post. mean	Post. SD	PIP	Coef.	SE	p-value
Constant	0.00140	NA	1.00	-0.00182	NA	1.00	-0.00538	0.00577	0.35
Standard error (SE)	0.88001	0.16	1.00	0.51288	0.24584	0.88	0.93988	0.15346	>0.001
Data characteristics									
Panel data	-0.00011	0.00116	0.02	0.00031	0.00283	0.04			
Daily data	-0.00027	0.00184	0.03	0.00011	0.00149	0.04			
Complete cases	0.00026	0.00198	0.03	-0.00156	0.00516	0.12			
Elderly sample	0.00014	0.00178	0.02	0.01212	0.01549	0.44			
Specification									
Day of week control	-0.00044	0.00253	0.04	-0.00004	0.00218	0.03			
Holidays control	-0.00058	0.00321	0.04	-0.00305	0.00871	0.14			
Seasonality control	0.00112	0.00420	0.08	0.02500	0.01163	0.88	0.01149	0.00629	0.07
Daylight control	0.00509	0.00772	0.34	0.00291	0.00722	0.18			
Rainfall control	0.00005	0.00092	0.02	-0.00118	0.00522	0.09			
Humidity control	0.00722	0.00857	0.46	0.03299	0.02239	0.75	0.01812	0.00791	0.02
Pressure control	0.00058	0.00304	0.06	-0.02519	0.01789	0.73	-0.00711	0.00835	0.4
Lag temp. analysis	0.00001	0.00066	0.01	-0.00147	0.00518	0.11			
Min. temp. analysis	0.00004	0.00112	0.01	0.00032	0.00232	0.04			
Max. temp.analysis	0.00018	0.00172	0.02	0.00980	0.01422	0.39			
ICD Coding	0.00026	0.00181	0.03	0.00013	0.00137	0.03			
Log median exp	0.00007	0.00076	0.02						
Log latitude	0.00003	0.00063	0.01						
Log suicide rate	-0.00001	0.00034	0.01	0.00135	0.00262	0.26			
Statistical approach									
Non-linear model	0.00039	0.00221	0.04	-0.00086	0.00372	0.08			
IRR measure	-0.00014	0.00141	0.02	0.00002	0.00132	0.03			
Absolute measure	-0.00605	0.00971	0.32	-0.00013	0.00167	0.04			
Publication characteristics									
Log citations	0.00067	0.00176	0.15	0.00097	0.00183	0.27			
Midpoint	>0.00001	0.00009	0.02	-0.00008	0.00023	0.15			
Studies		31			31			31	
Observations		186			186			186	

Notes: First specification reports results of our baseline specification of BMA using dilution prior. Next are results of BMA using weights proportional to the number of estimates reported by each study. Finally, coefficients of simple OLS using variables with PIP higher than 0.70 in the weighted BMA specification. In weighted BMA, the importance of controlling for other meteorological variables and time trend. BMA uses the PIP as a measure of statistical significance, whereas OLS uses common p-value measure.

Chapter 6

Conclusion

The purpose of this work was to shed some light on the topic of suicide rates in relation to temperature changes. A recent meta-analysis by Gao *et al.* (2019) reported that an increase of 1°C resulted in a 1% increase in suicide risk. This association would play a significant role in the computation of the social cost of carbon, which is used by policy makers to set the value of carbon tax (Nordhaus 2017). By correcting the estimate for publication bias, we did not identify a statistically significant effect of temperature changes on suicide rates. Furthermore, a simple estimate average of our methods used to identify the bias produces a value three times lower than that of Gao *et al.* (2019).

The authors limited the dataset with strict criteria and used only 23 estimates from 13 studies to determine their pooled estimate. Moreover, the authors only used one technique for the publication bias determination and did not identify an effect of the standard error on the estimates. Heterogeneity tests were performed using the measure of I^2 on subsets ranging from 18 to 31 estimates, which likely makes the measure biased (von Hippel 2015).

This thesis expands on the temperature-suicide association topic in three important domains: bigger dataset, sophisticated treatment of publication bias, and heterogeneity analysis using detailed study characteristics. We create methodology for standardizing various measures of this relationship, and increase the dataset used in the quantitative analysis to 186 estimates from 31 studies.

Apart from Gao *et al.* (2019), the majority of our tests identified the presence of publication bias in our data, including the precision weighted FAT test used by the authors of the aforementioned meta-analysis. In other words, researchers of the temperature-suicide relationship likely under-report negative or non-significant estimates. A possible explanation for this under-reporting is that researchers select their robustness checks in accordance with the primary results of their research since they can not include every specification at hand.

Analysis of the variance between estimates further strengthened the notion that bigger estimates are accompanied by inflated standard errors. In every robustness check we present, the standard error explains the most variance in the estimates. Our baseline specification states that the mean estimate is inflated by 0.165, on average. In our baseline model averaging specification, no other variable was statistically significant, meaning that the study characteristics have little importance when measuring the effect of temperature on suicide. By applying weights proportional to the number of estimates in each study, we discover the importance of controlling for seasonality and humidity levels. The weighted model also showed that males are affected by temperature changes more than females, although this finding needs to be supported by including other socioeconomic variables.

Our work mitigates the impact of prolonged temperature deviances from the mean on suicide rates. Nevertheless, the seasonality of suicide has been confirmed by numerous studies, including the ones used in our analysis. Out of the 17 studies used for our analysis, which reported a season peak, 15 identified the peak to be in spring or summer. On the other hand, some studies report a strong positive correlation for above-average temperatures in winter, suggesting that the lack of cold in the winter months is also responsible for suicidality.

There are several limitations associated with our results. Firstly, not enough studies reported the proportions of violent to non-violent, and completed to attempted suicides. These are promising characteristics for explaining the variance in the suicide-temperature association (Lin *et al.* 2008; Tsai & Cho 2012; Wu *et al.* 2014). To make the completed and attempted suicides estimates aggregable, we divided the reported rate of attempted suicide per 100 000 by 25, since for every one completed suicide, there are 25 attempts (McIntosh & Drapeau 2012). We acknowledge that this measure differs in time and place, but it is not an easily obtainable statistic. Another limitation is the inability to account for studies measuring the relationship using the maximum or minimum suicide-temperature. In some cases, such studies produce estimates 20 to 30 times larger than the usual method (Kim *et al.* 2019; Sim *et al.* 2020). Using them would distort our mean estimate. Lastly, the ecological design of the studies made it impossible for us to collect data regarding the psychological and socioeconomic state on the individual level. The socioeconomic status was supplemented using the macro-level median expenditure data for each state. However, this characteristic was not identified as statistically significant in our model averaging procedure.

Bibliography

Ajdacic-Gross, V., C. Lauber, R. Sansossio, M. Bopp, D. Eich, M. Gostynski, F. Gutzwiller, & W. Rössler (2007): "Seasonal associations between weather conditions and suicide - evidence against a classic hypothesis." *American journal of epidemiology* **165(5)**: pp. 561–569.

Akkaya-Kalayci, T., B. Vyssoki, D. Winkler, M. Willeit, N. D. Kapusta, G. Dorffner, & Z. Özlü-Erkilic (2017): "The effect of seasonal changes and climatic factors on suicide attempts of young people." *BMC psychiatry* **17(1)**: pp. 1–7.

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

Barker, A., K. Hawton, J. Fagg, & C. Jennison (1994): "Seasonal and weather factors in parasuicide." *British Journal of Psychiatry* **165(3)**: pp. 375–380.

Blanco-Perez, C. & A. Brodeur (2020): "Publication bias and editorial statement on negative findings." *The Economic Journal* **130(629)**: pp. 1226–1247.

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

Brodeur, A., N. Cook, & A. G. Heyes (2020): "Methods matter: P-hacking and causal inference in economics." *American Economic Review* **110(10)**: pp. 3634–3660.

Brodeur, A., M. Lé, M. Sangnier, & Y. Zylberberg (2016): "Star wars: The empirics strike back." *American Economic Journal: Applied Economics* **8(1)**: pp. 1–32.

Burke, M., F. González, P. Baylis, S. Heft-Neal, C. Baysan, S. Basu, & S. Hsiang (2018): "Higher temperatures increase suicide rates in the united states and mexico." *Nature climate change* **8(8)**: pp. 723–729.

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

Carleton, T. A. (2017): "Crop-damaging temperatures increase suicide rates in india." *Proceedings of the National Academy of Sciences* **114(33)**: pp. 8746–8751.

Deisenhammer, E. (2003): "Weather and suicide: the present state of knowledge on the association of meteorological factors with suicidal behaviour." *Acta Psychiatrica Scandinavica* **108(6)**: pp. 402–409.

Dixon, P. G. & A. J. Kalkstein (2009): "Climate–suicide relationships: A research problem in need of geographic methods and cross-disciplinary perspectives." *Geography Compass* **3(6)**: pp. 1961–1974.

Doucouliagos, C., T. D. Stanley, & M. Giles (2012): "Are estimates of the value of a statistical life exaggerated?" *Journal of health economics* **31(1)**: pp. 197–206.

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

Fernández-Niño, J. A., V. A. Flórez-García, C. I. Astudillo-García, & L. A. Rodríguez-Villamizar (2018): "Weather and suicide: a decade analysis in the five largest capital cities of colombia." *International journal of environmental research and public health* **15(7)**: p. 1313.

Fountoulakis, K. N., I. Chatzikosta, K. Pastiadis, P. Zanis, W. Kawohl, A. J. Kerkhof, A. Navickas, C. Höschl, D. Lecic-Tosevski, E. Sorel *et al.* (2016): "Relationship of suicide rates with climate and economic variables in europe during 2000–2012." *Annals of general psychiatry* **15(1)**: pp. 1–6.

Furukawa, C. (2020): "Publication bias under aggregation frictions: Theory, evidence, and a new correction method: Evidence, and a new correction method." *Working paper, MIT* .

Gao, J., Q. Cheng, J. Duan, Z. Xu, L. Bai, Y. Zhang, H. Zhang, S. Wang, Z. Zhang, & H. Su (2019): "Ambient temperature, sunlight duration, and suicide: A systematic review and meta-analysis." *Science of the total environment* **646**: pp. 1021–1029.

Gechert, S., T. Havranek, Z. Irsova, & D. Kolcunova (2020): "Measuring capital-labor substitution: The importance of method choices and publication bias." *An online appendix* pp. 1–49.

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

Grjibovski, A. M., G. Kozhakhmetova, A. Kosbayeva, & B. Menne (2013): "Associations between air temperature and daily suicide counts in astana, kazakhstan." *Medicina* **49(8)**: p. 59.

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

Havranek, T. (2019): "Why model averaging is useful in meta-analysis." MAER.Net: Blog post, https://www.maer-net.org/post/model-averaging.

Havranek, T., Z. Irsova, K. Janda, & D. Zilberman (2015): "Selective reporting and the social cost of carbon." *Energy Economics* **51**: pp. 394–406.

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

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

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

Helama, S., J. Holopainen, & T. Partonen (2013): "Temperature-associated suicide mortality: contrasting roles of climatic warming and the suicide prevention program in finland." *Environmental health and preventive medicine* **18(5)**: pp. 349–355.

von Hippel, P. T. (2015): "The heterogeneity statistic i 2 can be biased in small meta-analyses." *BMC medical research methodology* **15(1)**: pp. 1–8.

Hong, S. & W. R. Reed (2021): "Using monte carlo experiments to select meta-analytic estimators." *Research Synthesis Methods* **12(2)**: pp. 192–215.

Ioannidis, J. P. A., T. D. Stanley, & H. Doucouliagos (2017): "The Power of Bias in Economics Research." *The Economic Journal* **127(605)**: pp. F236–F265.

Karamouzian, M. & M. Rostami (2019): "Suicide statistics in iran: let's get specific." *American journal of men's health* **13(1)**: p. 1557988318807079.

Kim, Y., H. Kim, A. Gasparrini, B. Armstrong, Y. Honda, Y. Chung, C. F. S. Ng, A. Tobias, C. Íñiguez, E. Lavigne *et al.* (2019): "Suicide and ambient temperature: a multi-country multi-city study." *Environmental health perspectives* **127(11)**: p. 117007.

Kim, Y., H. Kim, Y. Honda, Y. L. Guo, B.-Y. Chen, J.-M. Woo, & K. L. Ebi (2016): "Suicide and ambient temperature in east asian countries: a time-stratified case-crossover analysis." *Environmental health perspectives* **124(1)**: pp. 75–80.

Lee, H.-C., H.-C. Lin, S.-Y. Tsai, C.-Y. Li, C.-C. Chen, & C.-C. Huang (2006): "Suicide rates and the association with climate: a population-based study." *Journal of affective disorders* **92(2-3)**: pp. 221–226.

Likhvar, V., Y. Honda, & M. Ono (2011): "Relation between temperature and suicide mortality in japan in the presence of other confounding factors using time-series analysis with a semiparametric approach." *Environmental health and preventive medicine* **16(1)**: pp. 36–43.

Lin, H.-C., C.-S. Chen, S. Xirasagar, & H.-C. Lee (2008): "Seasonality and climatic associations with violent and nonviolent suicide: a population-based study." *Neuropsychobiology* **57(1-2)**: pp. 32–37.

Madigan, D., J. York, & D. Allard (1995): "Bayesian graphical models for discrete data." *International Statistical Review/Revue Internationale de Statistique* pp. 215–232.

Maes, M., S. Scharpé, R. Verkerk, P. D'Hondt, D. Peeters, P. Cosyns, P. Thompson, F. De Meyer, A. Wauters, & H. Neels (1995): "Seasonal variation in plasma l-tryptophan availability in healthy volunteers: relationships to violent suicide occurrence." *Archives of general psychiatry* **52(11)**: pp. 937–946.

Mann, J. J. (2002): "A current perspective of suicide and attempted suicide." *Annals of internal medicine* **136(4)**: pp. 302–311.

McIntosh, J. L. & C. Drapeau (2012): "Usa suicide 2010: Official final data." *Washington, DC: American Association of Suicidology* .

Millner, A. J., M. D. Lee, & M. K. Nock (2017): "Describing and measuring the pathway to suicide attempts: A preliminary study." *Suicide and Life-Threatening Behavior* **47(3)**: pp. 353–369.

Möller-Leimkühler, A. M. (2003): "The gender gap in suicide and premature death or: why are men so vulnerable?" *European archives of psychiatry and clinical neuroscience* **253(1)**: pp. 1–8.

Müller, H., T. Biermann, S. Renk, U. Reulbach, A. Ströbel, J. Kornhuber, & W. Sperling (2011): "Higher environmental temperature and global radiation are correlated with increasing suicidality - a localized data analysis." *Chronobiology international* **28(10)**: pp. 949–957.

Nordhaus, W. D. (2017): "Revisiting the social cost of carbon." *Proceedings of the National Academy of Sciences* **114(7)**: pp. 1518–1523.

Page, L. A., S. Hajat, & R. S. Kovats (2007): "Relationship between daily suicide counts and temperature in england and wales." *The British Journal of Psychiatry* **191(2)**: pp. 106–112.

Pearce, D. (2003): "The social cost of carbon and its policy implications." *Oxford review of economic policy* **19(3)**: pp. 362–384.

Preti, A. & P. Miotto (2000): "Influence of method on seasonal distribution of attempted suicides in italy." *Neuropsychobiology* **41(2)**: pp. 62–72.

Ruuhela, R., L. Hiltunen, A. Venäläinen, P. Pirinen, & T. Partonen (2009): "Climate impact on suicide rates in finland from 1971 to 2003." *International journal of biometeorology* **53(2)**: pp. 167–175.

Schneider, A., R. Hampel, K.-H. Ladwig, J. Baumert, K. Lukaschek, A. Peters, & S. Breitner (2020): "Impact of meteorological parameters on suicide mortality rates: A case-crossover analysis in southern germany (1990–2006)." *Science of the total environment* **707**: p. 136053.

Sher, L. (2020): "The impact of the covid-19 pandemic on suicide rates." *QJM: An International Journal of Medicine* **113(10)**: pp. 707–712.

Sim, K., Y. Kim, M. Hashizume, A. Gasparrini, B. Armstrong, F. Sera, C. F. S. Ng, Y. Honda, & Y. Chung (2020): "Nonlinear temperature-suicide association in japan from 1972 to 2015: its heterogeneity and the role of climate, demographic, and socioeconomic factors." *Environment international* **142**: p. 105829.

Snowdon, J. & N. G. Choi (2020): "Undercounting of suicides: Where suicide data lie hidden." *Global public health* **15(12)**: pp. 1894–1901.

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

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

Tsai, J.-F. (2010): "Socioeconomic factors outweigh climate in the regional difference of suicide death rate in taiwan." *Psychiatry research* **179(2)**: pp. 212–216.

Tsai, J.-F. & W. Cho (2012): "Temperature change dominates the suicidal seasonality in taiwan: a time-series analysis." *Journal of affective disorders* **136(3)**: pp. 412–418.

Williams, M. N., S. R. Hill, & J. Spicer (2015): "Will climate change increase or decrease suicide rates? the differing effects of geographical, seasonal, and irregular variation in temperature on suicide incidence." *Climatic Change* **130(4)**: pp. 519–528.

World Health Organisation (2019): "Suicide fact sheet." Https://www.who.int/news-room/fact-sheets/detail/suicide/.

Wu, Y.-W., C.-K. Chen, & L.-J. Wang (2014): "Is suicide mortality associated with meteorological and socio-economic factors? an ecological study in a city in taiwan with a high suicide rate." *Psychiatria Danubina* **26(2)**: pp. 0–158.

Appendix A

Studies used in analysis

Akkaya-Kalayci et al. (2017)	Bando et al. (2017)	Barker et al. (2014)
Burke et al. (2018)	Carleton (2017)	Deisenhammer et al. (2003)
Fernandez et al. (2018)	Grijbovski et al. (2013)	Kayipmaz et al. (2020)
Kim et al. (2016)	Kim et al. (2019)	Lee et al. (2006)
Likhvar et al. (2010)	Lin et al. (2008)	Marion et al. (1999)
Moore et al. (2018)	Müller et al. (2011)	Page et al. (2007)
Preti & Miotto (2000)	Qamar et al. (2020)	Qi et al. (2009)
Qi et al. (2014)	Qi et al. (2015)	Ruuhela et al. (2008)
Schneider et al. (2020)	Tsai & Cho (2012)	Tsai (2010)
Williams (2015)	Williams (2016)	Wu et al. (2014)
Yarza et al. (2020)		

Appendix B

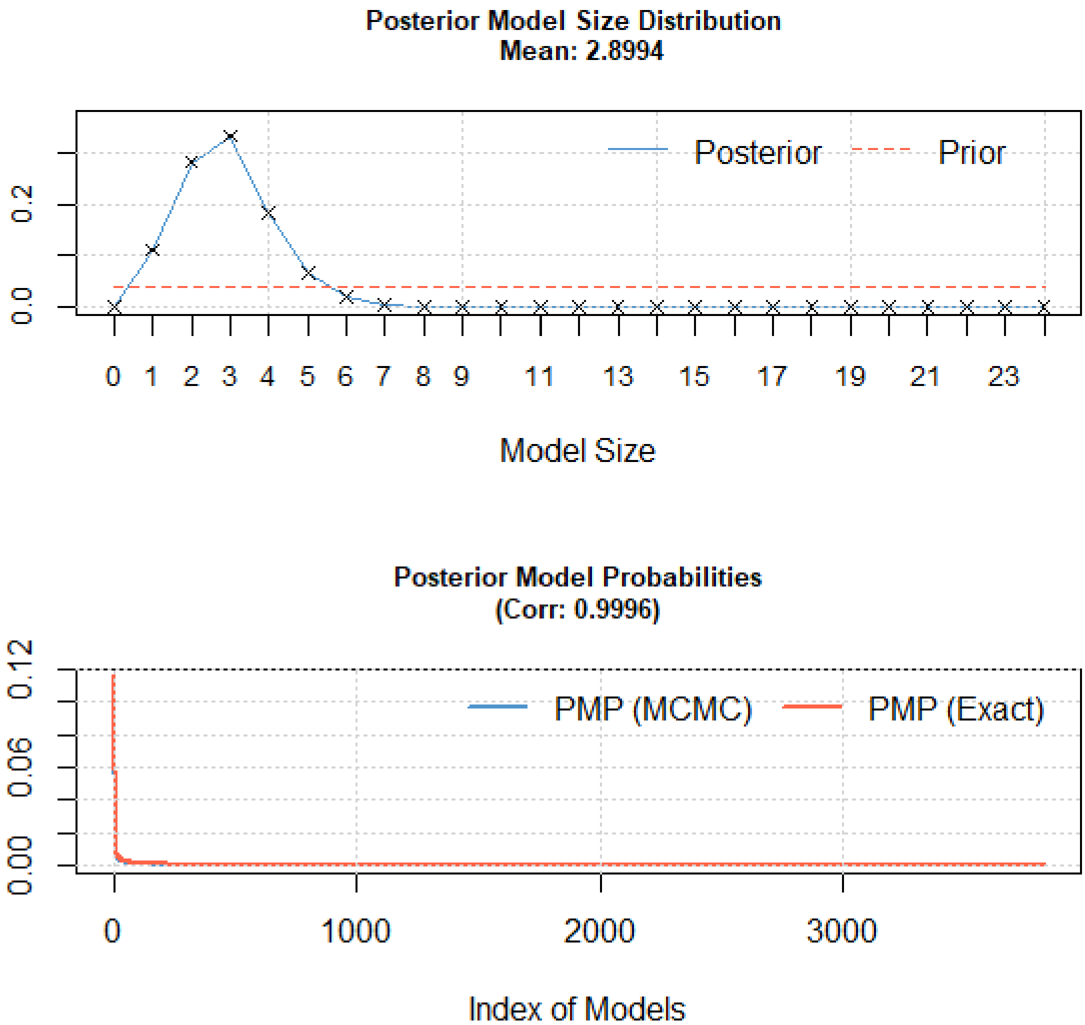
BMA Diagnostics and Robustness Check

Table B.1: Diagnostics of the baseline BMA estimation

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
2.9092	300000	100000	27.53728 secs	57680
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs*
16777216	0.34%	100%	0.9997	186
Model prior	*g-prior*	*Shrinkage-stats*		
Random-dilution / 12	UIP	Av=0.9947		

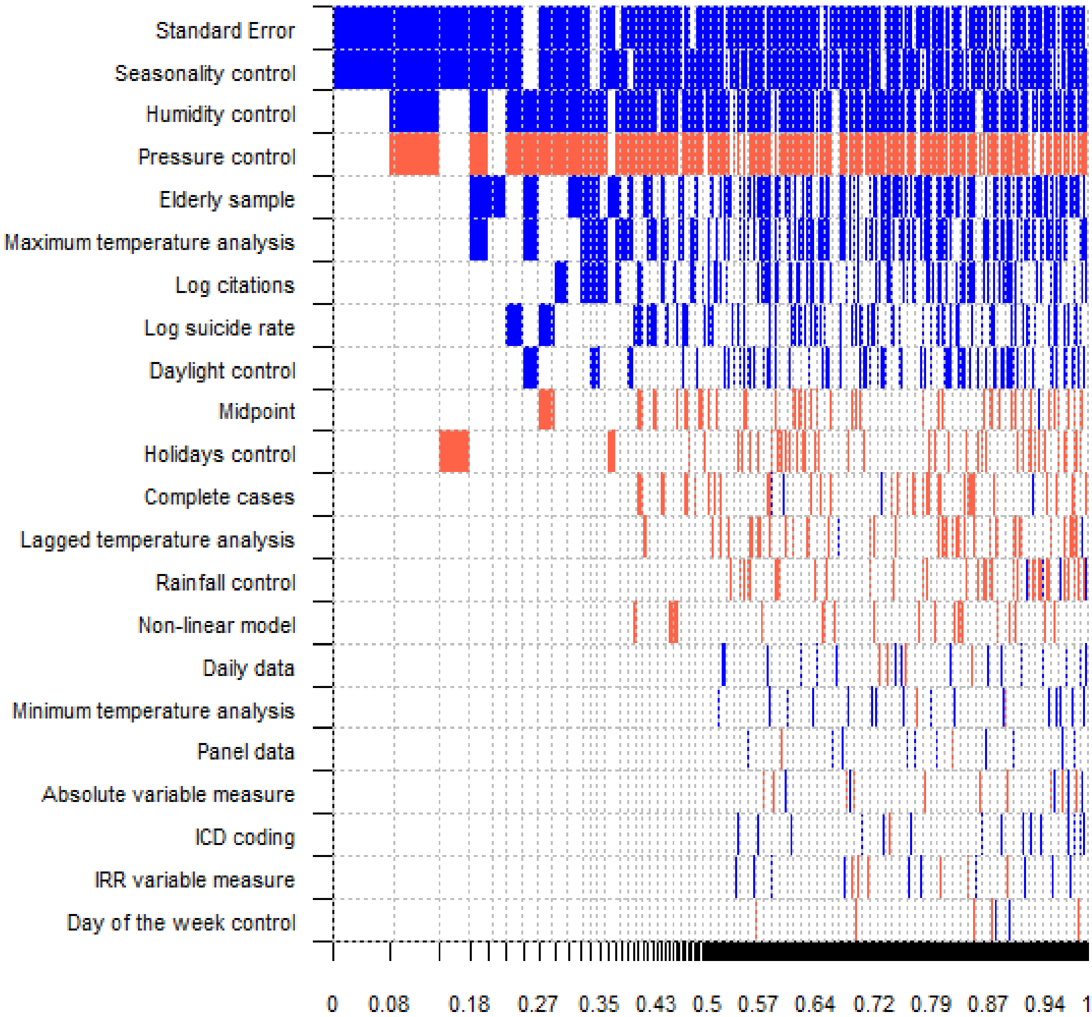
Notes: This table provides the summary of our baseline Bayesian model averaging settings. The model prior has been set as the dilution prior to account for the collinearity in our dataset. This settings corresponds to the results presented in Chapter 5.

Figure B.1: Model size and convergence of the baseline BMA estimation



Notes: This figure shows the posterior model probabilities for different model sizes. Judging by the upper plot, models with less variables are more efficient.

Figure B.2: Bayesian model averaging - weighted data



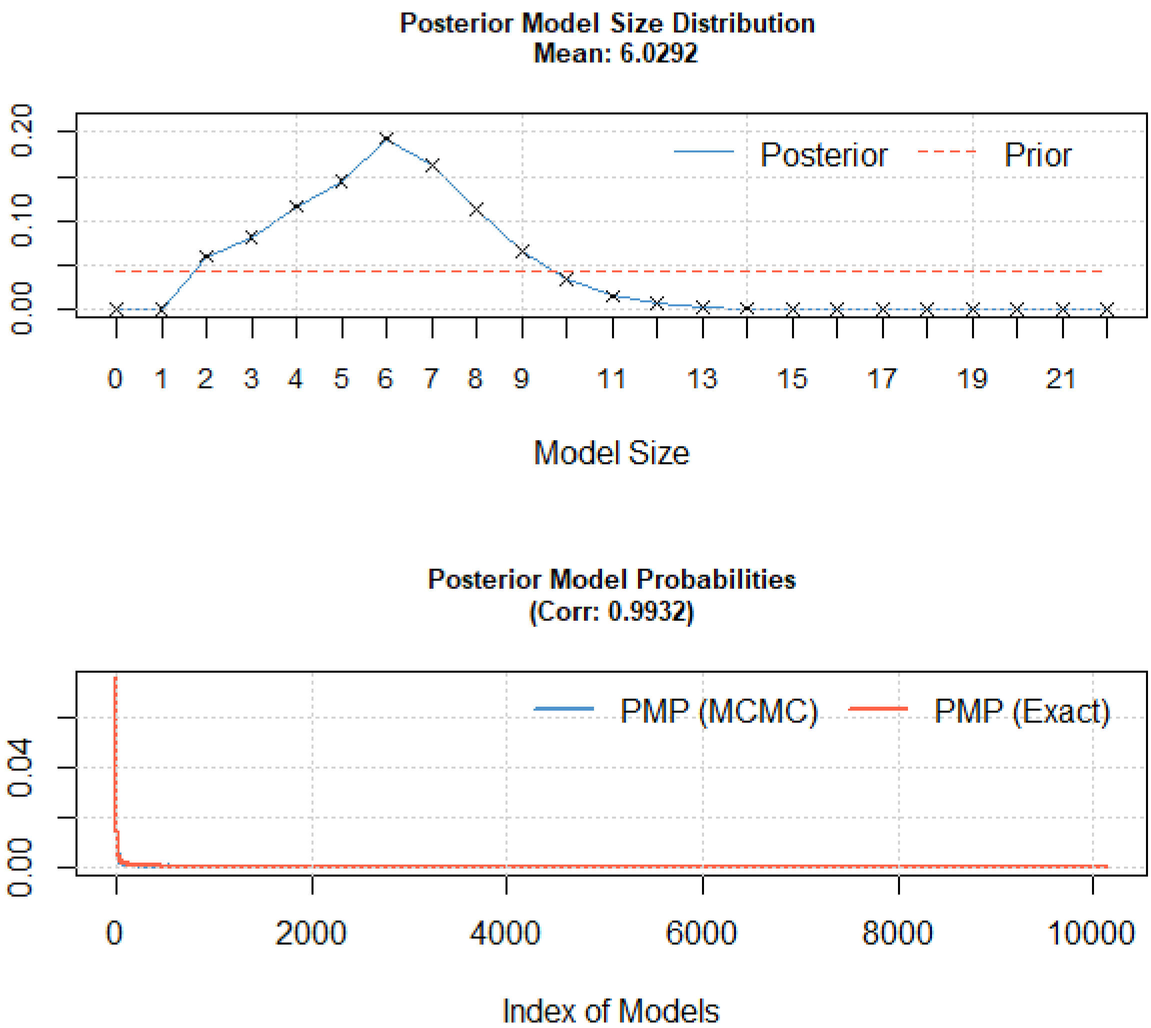
Notes: This figure illustrates the results of our weighted Bayesian model averaging specification. Every row symbolizes a single explanatory variable used in the models, sorted by their posterior inclusion probability. The columns represent individual models. Blue color means that the variables increases the effect, while red color decreases the effect. If the cell is white, the variable was not used in the model. Models are sorted by their posterior model probabilities. The best models are on the left. In the best model, only two variables are included.

Table B.2: Diagnostics of the weighted BMA estimation

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
5.9662	300000	100000	35.15151 secs	70763
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs*
4194304	1.7%	100%	0.9988	186
Model prior	*g-prior*	*Shrinkage-stats*		
Random-dilution / 11	UIP	Av=0.9947		

Notes: This table provides the summary of our weighted Bayesian model averaging settings. The model prior has been set as the dilution prior to account for the collinearity in our dataset.

Figure B.3: Model size and convergence of the weighted BMA estimation



Notes: This figure shows the posterior model probabilities for different model sizes for the weighted specification. In contrast with the baseline approach, weighted specification produces better posterior model probabilities for models with more variables.