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, February 11, 2021

Daniel Bartusek

Abstract

The abstract should concisely summarize the contents of a thesis. Since potential readers should be able to make their decision on the personal relevance based on the abstract, the abstract should clearly tell the reader what information he can expect to find in the thesis. The most essential issue is the problem statement and the actual contribution of described work. The authors should always keep in mind that the abstract is the most frequently read part of a thesis. It should contain at least 70 and at most 120 words (200 when you are writing a thesis). Do not cite anyone in the abstract.

JEL Classification	F12, F21, F23, H25, H71, H87
Keywords	temperature, weather, climate, suicide, suicidality
Title	Effect of Temperature on Suicide - Meta Analysis

Abstrakt

text text

Klasifikace JEL	F12, F21, F23, H25, H71, H87
Klíčová slova	teplota, počasí, klima, sebevraždy
Název práce	Vliv teploty na míru sebevražd: Meta-Analýza

Acknowledgments

The author is grateful especially to doc. PhDr. Zuzana Havránková Ph.D., prof. Lars Hansen, prof. David Zilberman, and participants at several conferences and seminars for their comments. The usual caveat applies.

Slightly modified **??** has been awarded the Czech Economic Society prize for young authors. This thesis is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 681228. Part of this thesis was written while the author was at the Spanish Prime Minister's Economic Bureau.

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 39. Advisor: doc. PhDr. Zuzana Havránková Ph.D.

Contents

List of Tables

List of Figures

Acronyms

FDI Foreign Direct Investment

MNC Multinational Company

PIP Posterior inclusion probability

IRR Incidence Rate Ratio

FAT Funnel-asymmetry test

PET Precisionâ€effect test

FE Fixed-effect

RE Random-effect

BMA Bayesian model averaging

FMA Frequentist model averaging

VIF Variance inflation factor

Master's Thesis Proposal

Author	Daniel Bartušek
Supervisor	doc. PhDr. Zuzana Havránková Ph.D.
Proposed topic	Effect of Temperature on Suicide - Meta Analysis

Research question and motivation Globally, suicide is the leading cause of violent death (WHO, 2019). Although the number of suicide deaths is in decline (see Naghavi et al., 2019, or Weiland Desai, 2019), WHO estimates about 800 thousand people kill themselves each year—which is more than those murdered in homicides and killed in wars combined. Why do suicides happen? Researchers, such as Deisenhammer et al. (2004), argue that no individual suicide can be causally related to a single event, but one can show that the risk of suicide increases with a number of 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 (Nock et al., 2008). 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., 2015). The message about the effect and its significance, however, becomes less clear when one searches through the literature examining the impact of ambient temperature. Some studies suggest that colder temperatures are associated with higher rate of suicide (Kim et al., 2016; Page et al., 2007), while some studies suggest the exact opposite (Preti Miotto, 2000; Salib, 1997). Several others, including Dixon et al. (2007) or Tsai Cho (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) construct a quantitative literature review, so-called meta-analysis. Based on the subsample of 13 observations of the effect in question, they conclude that the relationship is positive and postulate that a 1Â°C increase in temperature is significantly associated with a 1

In my thesis, I want to build on Gao et al. (2019) and examine the relationship between the temperature and suicide rates more thoroughly. I will ask three main questions: first, how large is the effect beyond biases, if any; second, what drives the magnitude of the effect; and third, what are the economic implications stemming from an increased risk associated with this effect. To accomplish this goal, I plan to enlarge the sample of estimates from Gao et al. (2019), use the 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 to tackle the economics behind the lost lives.

Expected Contribution I want to dig deeper than the previous meta-analysis of Gao et al. (2019). My contribution will be fourfold. First, I want to enlarge their sample to make sure the statistical validity of my results holds and to incorporate any new or possibly missing information (including that coming from unpublished studies). Second, I want to use the modern meta-analysis techniques to account for publication bias, including not only linear (Stanley, 2005) but also non-linear ones (for example the selection model by Andrews and Kasy, 2019, who build on famous Hedges, 1992; or the clever extension of Stanley, 2010, by Furukawa, 2019, who exploits the trade-off between bias and variance). Third, I want to analyze heterogeneity in the estimates via regression analysis (not only via testing the presence of variance in various subsamples), or more precisely, using model averaging techniques including Bayesian (Steel, 2020) and frequentist approach (Hansen, 2002). Fourth, I want to construct a synthetic study that would take a subjectively chosen best practice in the literature (the subjectivity comes from identifying what is the most preferable choice for data and methodology to study the effect in the literature) to estimate the €śtrue€ť best-practice value of the effect. Finally, I plan to translate the effect into monetary value.

Methodology I suspect that in pursuit of not mixing €śapples with oranges€ť, Gao et al. (2019) used rather strict inclusion criteria on which estimates to collect and which not to collect. Such criteria left them with very low number of observations to begin with. Instead of concentrating strictly on medical databases, I will construct a search query for Google Scholar, which should be superior to any subgroup of databases, and search for all the available studies that Gao et al. (2019) potentially left out, including those published after July 2018. In case the number of studies (and observations) remains low, I will loosen the original definition of the effect (incidence rate ratio) and recalculate the effects to a standardized measure of partial correlation coefficients (Doucouliagos, 2011). The measure is not directly interpretable but retains the ordinality of the original effect; thus, I will still be able to comment on whether the effect is large or small enough to matter (and analyze the heterogeneity, for that matter).

To determine the presence of publication bias, I will also collect some measure of precision the studies report and recalculate it to a common metric (standard errors, confidence intervals, possibly number of observations used in the primary studies). I will use the usual techniques used in economics to test for publication bias, including the funnel plot (Egger, 1997) and the funnel asymmetry test (Stanley, 2005) which is based on the assumption that in the absence of publication bias, estimates are randomly distributed around the â€śtrueâ€ť effect and independent of their standard error. Apart from the usual techniques, I will apply several weighting schemes to check whether the results are not driven by larger studies or less precise estimates (Havranek, 2015). Moreover, I will use the latest non-linear tests for publication bias, including the weighted average of adequately powered by Ioannidis et al. (2017), selection model by Andrews Kasy (2019), and stem-based method by Furukawa (2019).

To account for the heterogeneity across estimates, I will code different aspects of study design related to data, methodology, and publication characteristics. To test whether some element of the heterogeneity in the literature drives the effects systematically, I will apply Bayesian model averaging (BMA). BMA accounts for model uncertainty inherently present in any meta-analysis; it runs many regressions with all possible combinations of the explanatory variables and averages the estimated coefficients across models, weighting them by their goodness-of-fit (more in Steel, 2020). To check the robustness of my results, I will apply different priors on models and regression coefficients, and include frequentist model averaging (Hansen, 2007) along with the preferred BMA benchmark.

Outline

1. Introduction: brief motivation, contribution and main findings

2. On the estimates of the incidence rate ratio

   - about suicides and why is measuring the effect important (citing the most prominent studies)
   - how is the ratio calculated and what does it imply
   - previous meta-analysis and what new could this thesis bring to the discussion

3. Data

   - search query and inclusion criteria

- basic summary statistics on final sample and what does it tell us

4. Publication bias
   - what is it and is there any reason to expect it in this type of literature
   - visual and statistical testing with linear and non-linear tests for selection bias
5. Why estimates vary?
   - results of BMA and frequentist MA, discussion of results
   - best-practice estimate and sensitivity analysis
6. Economic implications: recalculating the best-practice effect into money using the statistical value of life
7. Concluding remarks: summary of my findings and their implications for policy and future research.

Core bibliography

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

Deisenhammer E.A., Kemmler G. P. Parson (2003): Association of meteorological factors with suicide. Acta Psychiatrica Scandinavica 108(6), 455-459.

Dixon, P. G., McDonald, A. N., Scheitlin, K. N., Stapleton, J. E., Allen, J. S., Carter, W. M., Holley, M. R., Inman, D. D. J. B. Roberts (2007): Effects of temperature variation on suicide in five U.S. counties, 1991-2001. International Journal of Biometeorology, 51(5), 395-403.

Doucouliagos, C. (2011). How Large is Large? Preliminary and Relative Guidelines for Interpreting Partial Correlations in Economics, Economics Series No. 2011/5, Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.

Doucouliagos, C., Stanley, T.D. M. Giles (2012): Are estimates of the value of a statistical life exaggerated? Journal of Health Economics, 31(1), 197-206.

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

Fountoulakis, K. N., Chatzikosta, I., Pastiadis, K., Zanis, P., Kawohl, W., Kerkhof, A. J. F. M., €¦ Bech, P. (2016): Relationship of suicide rates with climate and economic variables in Europe during 2000-2012. Annals of General Psychiatry, 15, art. 19.

Furukawa, C. (2019): €śPublication Bias under Aggregation Frictions: Theory, Evidence, and a New Correction Method.€ť Unpublished paper, MIT.

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

Hansen, B. (2007): Least Squares Model Averaging. Econometrica, 75(4), 1175-1189.

Havranek, T. (2015): Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting. Journal of the European Economic Association, 13(6), 1180-1204.

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

Ioannidis, J. P., T. D. Stanley H. Doucouliagos (2017): €śThe Power of Bias in Economics Research.€ť Economic Journal, 127(605), 236-265.

Kim, Y., Kim, H., Honda, Y., Guo, Y. L., Chen, B. Y., Woo, J. M., K. L. Ebi (2016): Suicide and ambient temperature in East Asian countries: A time-stratified case-crossover analysis. Environmental Health Perspectives, 124(1), 75-80.

Kim, Y., Myung, W., Won, H. H., Shim, S., Jeon, H. J., Choi, J., Carroll B. J. D. K. Kim (2015): Association between air pollution and suicide in South Korea: A nationwide study. PloS ONE, 10(2), art. e0117929.

Naghavi, M. (2019). Global, regional, and national burden of suicide mortality 1990 to 2016: Systematic analysis for the Global Burden of Disease Study 2016. BMJ, 364, art. l94.

Nock, M. K., Borges, G., Bromet, E. J., Cha, C. B., Kessler, R. C., S. Lee (2008): Suicide and suicidal behavior. Epidemiologic Reviews, 30(1), 133-154.

Page, L. A., Hajat, S., R. S. Kovats (2007): Relationship between daily suicide counts and temperature in England and Wales. The British Journal of Psychiatry, 191(2), 106-112.

Preti A. P. Miotto (2000): Influence of method on seasonal distribution of attempted suicides in Italy. Neuropsychobiology 41, 62-72.

Salib, E. (1997): Elderly suicide and weather conditions: Is there a link? International Journal of Geriatric Psychiatry, 12(9), 937-941.

Stanley, T. D. (2005): Beyond Publication Bias. Journal of Economic Surveys, 19(3), 309-345.

Stanley, T. D., Jarrell, S. B. H. Doucouliagos (2010): Could It Be Better to Discard 90% of the Data? A Statistical Paradox. The American Statistician, 64, 70-77.

Steel, M. F. J. (2020): Model Averaging and its Use in Economics. Journal of Economic Literature, forthcoming.

Tsai, J. F. W. Cho (2012): Temperature change dominates the suicidal seasonality in Taiwan: A time-series analysis. Journal of Affective Disorders, 136(3), 412-418.

Weiland, M. N. Desai (2019): Geography of loss-a global look at the uneven toll of suicide. Science: Brain Behavior, doi:10.1126/science.aaz2303.

WHO (2019): 10 facts about violence prevention. World Health Organization, Geneva. Online at www.who.int/features/factfiles/violence/en/ [Accessed on May 4, 2020].

Chapter 1

Data

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.* 2020a). The query used was adjusted to feature majority of studies used in the quantitative analysis of the meta-study by Gao *et al.* (2019). Final form of the query is: ('temperature' OR 'temperatures' OR 'climate' OR 'climatic' OR 'weather')AND('suicide' OR 'suicidal' OR 'suicidality').

Query returned approximately 850 000 results, out of which the first 200 were examined. Additional sources included 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. Since only 138 correlation coefficients were gathered, we decided to set the following inclusion criteria for the quantitative analysis:

- Effect presented in the study can be recalculated to the Incidence Rate Ratio (IRR).

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

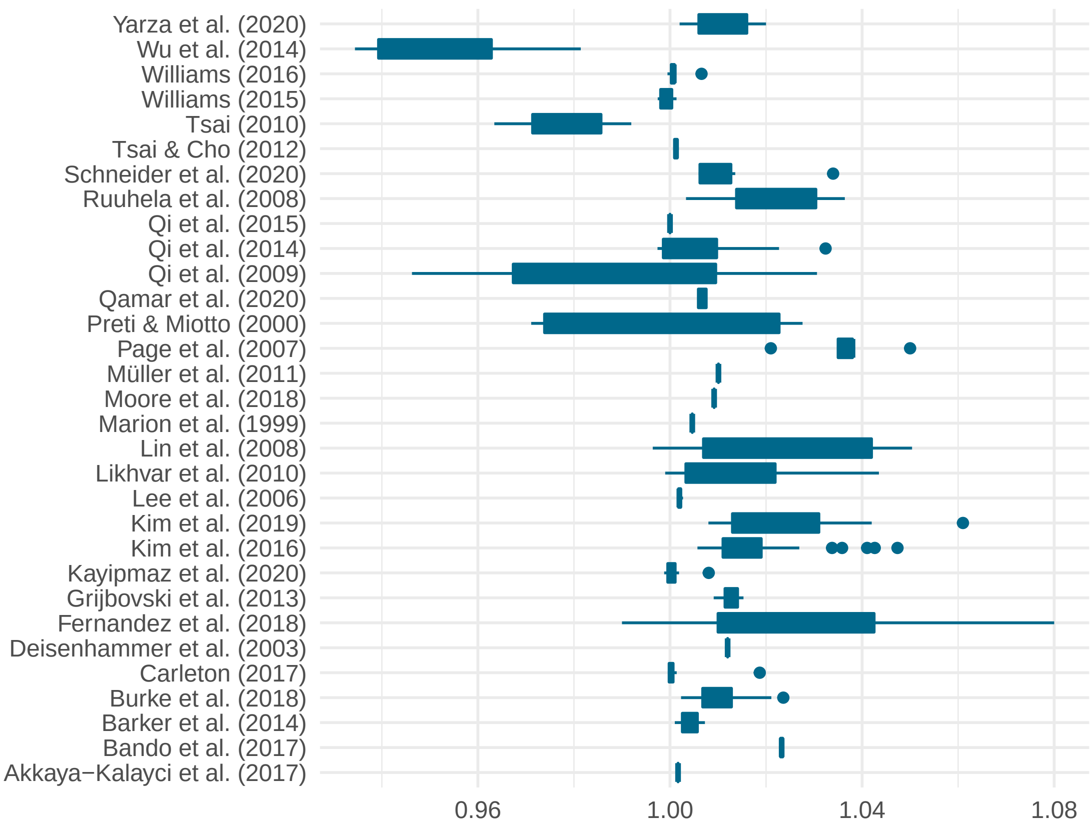
- The final estimates has to be calculated using minimum, mean or maximum temperature. This condition was used to restrict studies which used

> maximum or minimum suicide-temperature (MaxST or MinST). There were several studies which 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 normal temperature measures. Therefore, they were also left out.

Contradictory to conventional criteria, we did not omit effect without reported standard errors. There were only 16 such estimates. The method used for estimating these standard errors is described later in the Method section.

After this process, the dataset was restricted to 31 studies and 186 data points. For comparison, Gao *et al.* (2019) only used 16 studies. The effects are dispersed between the studies, as well as within (Figure 1.1).

Figure 1.1: Market equilibrium



1.1 IRR

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 presented in 'Data' section, the most optimal strategy appears to be standardizing to IRR. Rate ratio utilizes the incidence rate in groups exposed and unexposed to the effect in question (CDC 2012). In our case the effect is a marginal move in temperature, and by incidence we mean the suicide rate. Suicide rate is usually presented as the number of suicides per 100 000. Most of the studies in this thesis report the suicide rates, making the use of incidence rate ratio practical. Moreover, the meta-analysis from Gao *et al.* (2019) also uses this measure, which will be convenient for comparison. In total, 5 types of effect were identified Figure 1.2.

The formula for incidence ratio is:

$$IRR = \frac{IncidenceRate_{exposed}}{IncidenceRate_{unexposed}} = \frac{SR_{temp}}{SR_{norm}}$$

Where 'temp' symbolizes the suicide rate under the estimated effect of 1° C rise in temperature, and 'norm' stands for the original reported suicide rate in the study. Since our final dataset reports the effects in only 5 different measures, we will show the standardizing method for each of them:

Incidence rate ratio IRR is already in our preferred form. The only standardization necessary is in case that the reported effect is per a temperature change not equal to 1° C.

Relative Risk There are inconsistencies in use of risk ratio and rate ratio. Nevertheless, risk ratio has similar interpretation to rate ratio in our case. It is also calculated as the proportion of cumulative incidence rates. Therefore, the standardization will be the same as with rate ratio.

Odds Ratio The IRR can by calculated from odds ratio in this manor. Odds ratio has similar values for IRR, when the prevalence in the population is low, and the odds ratio is close to 1, which is our case (Zhang & Kai 1998).

$$IRR = \frac{OR}{1 - SR_{norm} + SR_{norm} * OR}$$

Regression coefficient Most studies reported their result as the association between suicide and temperature increase. Depending on whether the effect is measured in relation to suicide rate (SR) or daily suicide count(DS), we use:

$$IRR = \frac{SR_{norm} + association}{SR_{norm}} = 1 + \frac{x}{SR_{norm}}$$

$$IRR = \frac{DS_{norm} + association}{DS_{norm}} = 1 + \frac{x}{DS_{norm}}$$

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

$$IRR = \frac{SR_{norm} + x}{SR_{norm}} = 1 + \frac{x}{SR_{norm}} = 1 + percent\_increase$$

There were 36 estimates, which included suicide attempts in their analysis, 31 of which require the reported suicide rate for recalculation to IRR. Naturally attempted suicide has substantially higher rate, than completed suicide. Should we use this value in our calculations, IRR 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.

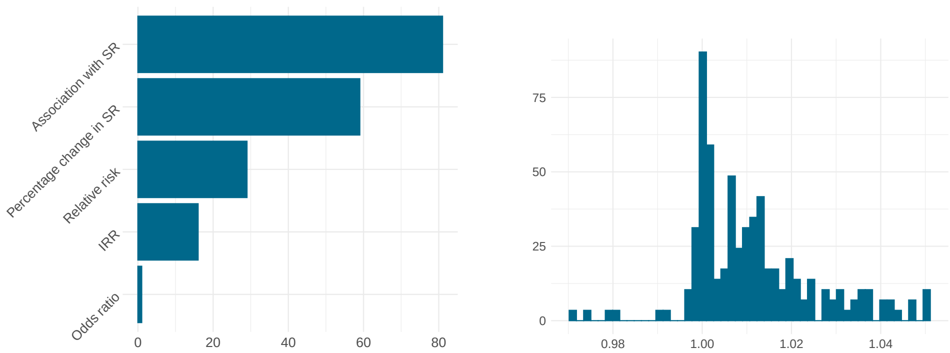
1.2 SE

Given the heterogeneity of the studies, computing the standard error of IRR could not be performed using only one method. In fact, 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 use of Delta Method depends on the method used to calculate the IRR (presented above). Namely for the regression coefficient transformation, we use:

$$se(IRR) = var(f(SR_{norm}, x))^{1/2} = var(1 + \frac{x}{SR_{norm}})^{1/2} = \frac{1}{SR_{norm}} * se(x)$$

- 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:



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

Figure 1.2: Estimates distribution

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

- When neither the original standard error, nor the confidence interval was reported, the standard error of IRR had to be calculated using the relationship between t-statistics and the effect.

$$se(IRR) = \frac{\hat{\beta} - \beta_0}{t_{\hat{\beta}}} = \frac{\hat{\beta} - 1}{t_{\hat{\beta}}}$$

Due to high variance in our data, winsorization was considered. Winsorization is a method for treating outliers in data by making them less extreme. Since there is not enough estimates gathered, this method would be preferred compared to simply trimming the outliers. However, leaving outliers untreated could possibly distort further analysis. Fortunately, even winsorization on the 5% level did not statistically change the distribution, nor did it alter the mean effect. Thus, our data will be left in its original form.

Chapter 2

PUB BIAS

Without any data-based knowledge, the common guess would be that consecutive days of cold temperatures must increase the suicide rates. With this notion in mind, early researches of this effect could be conducting with this notion in mind, and possibly alter their methods. For example, they could select the model which gives results that are most in accordance with their view. Moreover, the journal might be more keen to publish statistically significant results, that support the current notion (Card & Krueger 1995).

Better methodologies With better methodologies being developed and systematical reviews being published, the conventional view started to be questioned. Review by Deisenhammer (2003) was perhaps most influential in challenging the classical notion.

> 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)

Deisenhammer comments further on heterogeneity of methodology, and study-specific characteristics, such as seasonality and data granularity, which could have impacted the results of reviewed studies. Deisenhammer's review gave rise to more studies analysing this relationship, majority of which reported positive link. Nevertheless, the motive to publish effect of selected sign likely did not disappear, it merely changed direction. This issue can be observed by plotting the effect values. We see a sudden jump in values over the 1.00 threshold, suggesting that negative effects are underreported Figure 1.2.

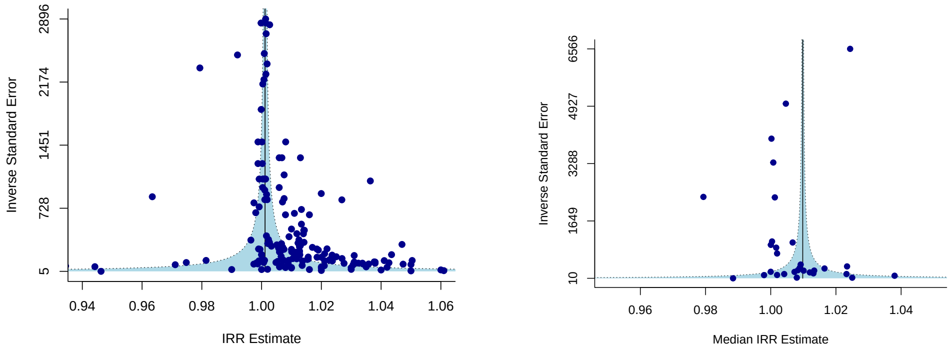
This issue is called publication bias. Researches are motivated to produce statistically significant results of certain sign. When study fails to reject the null hypothesis, due to large standard errors, the conductors will attempt to collect more data or restrict their sample size in order to make their results significant. On the contrary, effects with lower magnitude that are significant do not meet this constraint(Card & Krueger 1995; Brodeur *et al.* 2018), and will be published. Another case could be that effects failing to reject the null hypothesis will stay unpublished, which is 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 statements for health economic journals has 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 significant results decreased by 18%. Moreover, study by Brodeur *et al.* (2018), which compared results published in top economic journals, shows that certain methods produce systematically bigger estimates than other. For illustration in our data, a simple mean estimate yields result of 1.012, while study-weighted mean drops to 1.007. Thus, it is necessary to correct this meta-analysis for possible publication bias.

Perhaps the most convenient method for detecting publication bias is simply plotting the data as a funnel plot. When estimates are plotted against the inverse of their respective standard errors, we can expect the most precise values to be densely distributed around the true effect, while 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". Left plot in Figure 2.1 shows the right skewness of the estimates, as well as the cut-off at 1. The bias is not that apparent in the right plot, where only median points from each study is plotted. This indicates that the right skewness relies heavily on a subset of studies reporting many estimates. Nevertheless, it is insufficient to make a decision solely by subjective judgement of the funnel plot. Thus, it is needed to compute the bias numerically using state of the art baseline methods.

2.0.1 Baseline methods

As previously mentioned, publication bias happens, when researches attempt to lower their standard errors in order to make their estimates significant. Consequently, we can expect the standard error to be correlated with the effect



(a) Precision (b) Median Precision per Study

Figure 2.1: Funnel plot

itself. When we regress the estimate on standard errors, the intercept represents the 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:

$$IRR_{ij} = \beta_0 + \beta_1 * SE(IR\hat{R}ij) + u_{ij}$$

Where IRR_{ij} is the i-th estimate from j-th study. We can then perform a Funnel-asymmetry test (FAT) to identify the publication bias by rejecting null hypothesis that the beta coefficient of SE is 0. Moreover, Precisionâ€effect test (PET) provides further confirmation of a genuine effect beyond this bias. Since our final standardized effect is in the form of IRR with mean equal to 1, we are going to test null hypothesis that the intercept is not 1 (Stanley 2005; Stanley *et al.* 2010).

However, since large sample size also decrease 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 weighting scheme is the inverse of the number estimates produced by a single study, since heterogeneity of studies could also affect the effects. The range of estimates from a single study in this meta-analysis ranges from 1 to 33. By applying this weight to the regression, we assure that every study impact 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, having one common true effect. Thus, sampling error can only arise from 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 weighting schemes. The results are presented in Table 2.1, along with their standard errors. The standard errors are clustered on study level in order to account for the within-study correlation. Non-clustered errors introduce false precision levels of models.

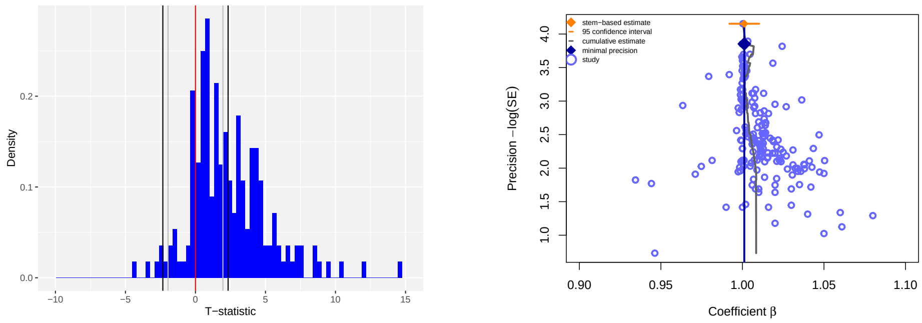
Every method reports the mean estimate to be more than 1, suggesting that increase in temperature does increase the risk of suicide. Nevertheless, each estimate is below 1.01, which is an estimate reported by Gao *et al.* (2019), and has not been corrected for publication bias. PET on methods which use some form of precision weighting does not reject the null hypothesis of no effect of temperature. Although only two out of our five methods identified statistically significant presence of publication bias, we can still consider this a credible result. The reason is that RE is more balance by putting weight on smaller studies. Results of FE would have been more plausible, if studies came from identical environment, were performed by the same researchers, and used common methods, which is not the case.

Table 2.1: Linear tests of publication bias

	OLS	**FE**	**BE**	**Precision**	**Study**
SE	0.271	1.482	0.845*	0.907*	-0.138
Publication bias	(0.452)	(1.176)	(0.352)	(0.273)	(0.532)
Constant	1.009*	1.003	1.006**	1.003	1.009
Mean beyond bias	(0.004)	(0.002)	(0.002)	(0.002)	(0.004)
Studies	31	31	31	31	31
Observations	186	186	186	186	186

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Aforementioned methods assume 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 then 1 (Bom & Rachinger 2019). There could also be bias in the standard error due to random sampling error introduced by the researchers (Stanley



(a) T-statistic distribution (b) Stem-based method

Figure 2.2: Funnel plot

2005). For this reason, we also apply several new methods, which account for publication bias, while assuming non-linearity.

Perhaps the simplest method of non-linear publication bias estimation is the Top10 developped by Stanley *et al.* (2010), whose work shows that if levels of statistical significance do affect the chance of paper being published, the sample of effects is indeed not representative, and any statsitical computation method will be biased. Therefore, meta-analyst might be better of to simply discard 90% of data based on lower levels of precision, and leave the rest with higher precision. Stanley *et al.* (2010) however recognizes, that this method contradicts the traditional Central Limit Theorem, and rather presents the work with intention to highlight this issue. When applied to our dataset, the final estimate is significantly higher than in other methods, both linear and non-linear. This is due to the fact, that one study by Fernández-Niño *et al.* (2018) report both high estimates and high precision. As a result, all 11 out of the 19 estimates, which passed the threshold, are in the sample. Nevertheless, we report the result along with other more sophisticated methods.

It is also possible to treat non-linearities in publication bias using weighted distribution theory. Instead of choosing one weighting scheme for the whole range of values, step function will be used to apply different weights will to every interval of reported p-values (Hedges 1992). The cut-offs in these interval are set as the conventionally reported values of 0.001, 0.01, 0.05. The correct functional form of these weights is determined with maximum likelihood function. Jumps at the conventional values are not apparent in our data (Figure 2.2).

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, assuming the right-skewness of estimates. Eligible studies 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 36 adequately powered estimates.

Last method we chose is The Stem Based Method (Furukawa 2019), 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 3 out of 186 meets the criteria of this method Figure 2.2.

Apart from the Top10 method, results of non-linear methods confirm the previous findings stating that the mean estimate corrected for publication bias will be less than 1.01. 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 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 know as the PEESE method (Stanley 2005).

Table 2.2: Non-linear tests of publication bias

	Hedges (1992)	**Stanley et al (2010)**	**Ioannidis et al (2017)**	**Furukawa (2019)**
Mean beyond bias	1.006***	1.022**	1.003***	1.008
	(0.003)	(0.007)	(0.00004)	(0.005)

2.0.2 Extensions

Although the robustness of aforementioned methods for publication bias detection is proven by a number of meta-analysis, validity of the results can be further strengthened by employing methods, which address the potential heterogeneity of standard errors.

Endogeneity can be corrected for by choosing an instrumental variable. Most convenient IV in our case is the inverse of the square root of the sample size, since standard error decreases as sample size increases, while the model of choice is likely not chosen with respect to number of observations.

Recent technique by the name of p-uniform* also addresses the endogeneity (van Aert & van Assen 2020). P-uniform* assumes uniform distribution of p-values around the true underlying effect. The technique compares segments of the p-curve by recomputing the p-values for different true mean effects. The mean effect with p-values distribution closest to the normal distribution is then reported. Results of the IV regression and P-uniform* technique are presented in Table 2.3. Both methods report the mean estimate to be less than 1.01, although the values are bigger, than in the baseline methods.

Table 2.3: Treatment of endogeneity in standard errors

	IV	**p-uniform***
Mean beyond bias	1.005	1.009***
	(0.003)	(0.001)
Studies	31	31
Observations	186	186

Last method we use is called the Caliper test, developed by Gerber *et al.* (2008). Contrary to the baseline methods, Caliper test does not assume a relationship between the main effect and the standard error. To use this method, it is necessary to identify key values of t-statistic. In our case, these values will be set as 0 and 1.96, as we have right-skewed data. It is not feasible to make similar tests for the -1.96 t-statistic due to low number of data points around that value. Caliper test than compares frequencies around these key values to identify sudden jumps, which would indicate the presence of publication bias. Since we have low number of observations, it is necessary to set large enough calipers. Our caliper levels were set to comprise approximately 30 observations. Nevertheless, all six of our tests failed to identify publication bias, since the frequencies on both sides of the thresholds have similar distribution, as visible in Figure 2.2.

To conclude, most of the methods applied in this meta-analysis identified the presence publication bias, and report the corrected mean estimate to be lower than by Gao *et al.* (2019), although the relationship between temperature and suicide remains positive.

Table 2.4: Caliper tests corresponding to 5% significance thresholds

Caliper size	**0.5**	**0.6**	**0.7**
Threshold: 0	0.052 (0.094)	0.063 (0.089)	0.083 (0.083)
Observations	29	32	36
Threshold: 1.96	-0.017 (0.094)	-0.016 (0.091)	0.014 (0.086)
Observations	29	31	35

Chapter 3

Heterogeneity

Although publication bias proved affect the underlying mean, there are other factors in means of study characteristic, that could still conceal the true estimate. We distinguish potential sources of heterogeneity to categories such as the data origin, methods used, publication characteristics, or factors, which the original study controls for. Every factor from these categories could be of significant effect for the final mean estimate. Therefore, it is recommended to deploy methods which identify heterogeneity using context of the estimates.Â¨

The first obstacle in terms of heterogeneity specification is that not every variable affects the estimate. Our data consists of 37 variables which could be of importance for the mean estimate. However, should every study characteristic be included, we risk overfitting our model and introducing collinearity, which reduces 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.

First step is to treat colinearity. 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 variables until the maximum value of Variance inflation factor (VIF) is under 10. By removing variables with high VIF and ambiguous relationship to IRR, we cut down the number of observations to 20. Using this procedure, we obtain the maximum VIF value of 4,68, and maximum correlation equal to 0.635.

It is not feasible to manually select the correct variables from the rest. Given our 20 independent variables, we would have to run 2Θ20 $\approx$ 1000000 different combinations, which would take an immense amount of time. This issue, called model uncertainty, can be addressed using Bayesian model averaging (BMA)

(Eicher *et al.* 2011) and Frequentist model averaging (FMA) (Hansen 2007).

BMA does not require the concrete set of independent variables to be chosen in advance. Instead BMA runs a set number of models, and assigns to each model its posterior model probability, which increases with 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.* 2020b). Markov chain Monte Carlo algorithm is then used to traverse only the part of model with high PIP(Madigan *et al.* 1995). Variables with PIP higher then artificially set threshold is then included in the final set of significant variables. The coefficient for these variables is calculated as the weighted average of coefficient in previously run models, using the posterior model probabilities as weights (Havranek *et al.* 2018).

To be able 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 strong conviction for some of the variables affecting the main estimate. G-prior are set 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 advanced. Regarding model priors, the dillution prior will be used. Compared to the uniform alternative, dillution prior treats multicollinearity between the variables.

Nonetheless, it is appropriate to compute BMA using other modifications. Apart from increasing the number of iterations the choice of priors in also BMA 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. Regarding our choice of uniform model prior, models with mean number of variables are most common in the uniform distribution, while models with very little or close to all variables include will not be represented well.

1)Another form of model prior is to set the probability equal, regardless of the count of variables included.

3) BMA precision weighted -increased collineartity, cant comment on results other than magnitude 4) add new vars (+-ss,se,male,violent..)

Results of the former modifications should be accompanied by other specifications, which do not stem from BMA. For that reason, we deploy FMA with robust standard errors is a convenient alternative. FMA includes all explanatory variables, and weights them.

3.0.1 Variables

Data characteristics This category comprises the environment and sample of the respective studies, We include the number of observations in original studies to treat potential small sample bias. The time range and the midpoint year of every study will be included. We deploy a dummy variable equal to one, if the the data is in form of panel data or pooled cross-section, and equal to 0, if in form of time series. Regarding data granularity, more than half of the studies use daily suicide data. Therefore, we use a dummy with baseline category for studies with daily data, while studies using weekly, monthly or annual data will will have this dummy variable equal to 1. Some estimates only used data for elderly population, which is also accounted for using a dummy variable. Another dummy variable will denote, whether the study used only completed suicides, or includes attempts as well. Regarding socioeconomic and environmental characteristics, we gathered state-level median expenditures, suicide rates, minimum, mean and maximum temperature, Since temperature changes are likely experienced more severely in some climatic zones, we include the absolute value of latitude. Lastly,a variable for the gender proportion of the sample, as well as a dummy variable indicating, whether the suicides in sample were violent or non-violent only, were also gathered for some studies. Though we did not manage to collect these characteristics for enough studies, so these will not be included in the main procedure.

Statistical approach We include a dummy equal to 1, if study has not reported confidence interval or standard error for its estimate.

Specification Studies use minimum, mean, as well as maximum temperature values to model the association with suicide rates, which will also be represented by dummies. Dummy variables will also be deployed to denote, whether study controls for seasonality, day in the week holidays, daylight, rainfall, atmospheric pressure, humidity, ICD reporting, an the age and gender in its sample, Potentially important dummy variable will control for the usage of lagged weather characteristics in association with present suicide rates. Some studies used a temperature threshold, below which the temperatures would not be included in the model. This is also encoded with a dummy variable.

Publication characteristics We also include the number of citations of each paper, and the impact factor of the journal, in which the study was published. Lastly, we use a dummy variable equal to 1, if the association of temperature with suicide was the main object in the study.

George, 2010: collinearity problem -> compute the determinant (denoted by |R|) of the correlation matrix and then use the determinant as a weight for computing model probability -> lot of collinearity will be penalized and will contribute with less weight into the final posterior coefficients

Havranek 2019: weighted average of all models: The weight increases with data fit, but decreases with model complexity (given the same fit, a regression with 4 variables will get more weight than a regression with 5 variables). So, think of adjusted R-squared as an intuitive weight for model averaging. Itâ€™s not an optimal weight, but you get the idea.

Best practice, estimation: For each variable deemed useful by the BMA exercise, that is, with PIP larger than 0.5, we plug in a preferred value, a sample minimum or a sample maximum, or, in the case of no preference, a sample mean. Then we compute a linear combination of regression parameters and obtain the value of the partial correlation coefficient conditional on our definition of best practice.

Robustness check - First, we report the results of BMA when employing alternative priors (gâ€prior and model size). Second, we present the results for unweighted regressions with the same priors for BMA as in the baseline estimation in Section 5.. Third, we only use frequentist methods (OLS and fixed effects). Fourth, we use inverseâ€variance weights, which are more common in metaâ€analysis.

Include model formula from sigma page 25

Statistical approach: Dummy: method Dummy: variable definition Dummy: lin/non-lin

Publication chars: journal, cite the recursive discounted RePEc impact factor of the outlet? Dummy: main/side finding (seasonality, different main objective= side) About the plot: The columns in the figure denote the individual regression models, whereas their width indicates the modelsâ€™ posterior probabilities. The variables are sorted by their PIP in descending order. If the sign of a variable's regression coefficient is positive, it is denoted by blue colour (darker in grayscale). Conversely, if the sign of a variable's coefficient is negative, it is coloured in red. Where a variable is excluded from a model, the corresponding cell is left blank. The horizontal axis measures the cumulative model probabilities: the models that are the most successful in explaining the heterogeneity in the estimates of the competition effect are on the left, and we can see that they include less than a half of all the variables.

Table 3.1: Definitions and summary statistics of explanatory variables

Variable	Description	Mean	Std. dev.	WM
Variable definition				
IRR Estimate	Incidence rate ratio associated with 1 C increase in temperature	1.011	0.0011	1.008
Standard error	Standard error of the IRR estimate	0.008	0.0009	0.007
Data characteristics				
No. of obs.	The logarithm of the number of observations used in the regression	9.12	0.13	8.69
Panel data	= 1 if panel data are used in study (reference category: time-series)	0.28	0.03	0.21
Daily data	= 1 if study used daily data of suicide and temperature (reference category: weekly, monthly, annualy)	0.69	0.03	0.55
Suicide rate	Rate of suicide per 100 000	11.35	0.54	11.19
Specification				
Complete suicides	=1 if study uses only completed suicides in analysis (reference category: both suicide attempts)	0.8	0.03	0.73
Elderly sample	=1 if if study used elderly sample in analysis (reference category: no age restrictions)	0.06	0.02	0.09
Season control	=1 if study controls for seasonality	0.72	0.03	0.68
Daylight control	= 1 if study controls for	0.44	0.04	0.51
Rainfall control	= 1 if study controls for	0.43	0.04	0.61
Humidity pressure	= 1 if study controls for humidity or atmospheric pressure	0.52	0.04	0.4
Lagged temp control	= 1 if study allows lagged forms of temperature in analysis (reference category: only direct association)	0.28	0.03	0.37
Temp threshold	= 1 if study analyzes only temperatures above certain threshold (reference category: no restrictions on temperature data)	0.06	0.02	0.08
Mean temperature	Mean temperature reported in study	16.24	0.45	15.19
Min temperature used	= 1 if study uses minimum temperature in analysis of the relationship	0.13	0.02	0.17
Max temperature used	= 1 if study uses minimum temperature in analysis of the relationship	0.07	0.03	0.12

Continued on next page

Table 3.1: Definitions and summary statistics of explanatory variables (continued)

Variable	Description	Mean	Std. dev.	WM
Median expenditure	The logarithm of median expenditures in country, where the study was conducted	6.51	0.06	6.61
Lattitude	The absolute value of latitude in country, where the study was conducted	33.11	0.96	36.29
Statistical approach				
Nonlinear method	= 1 if study used model, which allows non-linear relationship between temperature and suicide rate	0.77	0.03	0.74
IRR reported	= 1 if study reports the effect in rate ratio or relative risk (reference category: association with suicide rate)	0.25	0.03	0.26
Publication characteristics				
Citation	Number of time the study has been cited (Google Scholar citation)	51.06	3.49	6.61
Midpoint	Mean year of the data used minus the earliest mean year in the data	17.42	0.69	16.87

Note: Collected from published studies estimating the elasticity of substitution between capital and labor. When dummy variables form groups, we mention the reference category.

Chapter 4

Conclusion

The conclusion should briefly summarize the problem statement and the general content of the work and the emphasize on the main contribution of the work.

When writing the conclusion keep in mind that some readers may not have gone through the whole thesis, but have jumped directly to the conclusion after having read the abstract in order the decide on the personal relevance of the thesis. Therefore, the conclusion should be self contained, which means that a reader should be able to understand the essence of the conclusion without having to read the whole thesis.

The conclusion typically ends with an outlook that describes possible extensions of the presented approaches and of planned future work.

Bibliography

(2012): "Principles of epidemiology."

van Aert, R. C. & M. van Assen (2020): "Correcting for publication bias in a meta-analysis with the p-uniform* method." *Manuscript submitted for publication Retrieved from: https://osfio/preprints/bitss/zqjr92018.[Google Scholar]* .

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 (2018): "Methods matter: P-hacking and causal inference in economics." .

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.

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

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.

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.

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

Gao, 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 (2020a): "Measuring capital-labor substitution: The importance of method choices and publication bias." *An online appendix* pp. 1–49.

Gechert, S., T. Havranek, Z. Irsova, & D. Kolcunova (2020b): "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.

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."

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.

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

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

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

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.

ZHANG, J. & F. Y. KAI (1998): "What's the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes." *Jama* **280(19)**: pp. 1690–1691.

Appendix A

Title of Appendix A

Text text text text text text text text text text text text text text text. Text text text text text text text text text text. Text text text text text text text text text text text text text text text. Text text text text text text text text text text text text text text. Text text text text text text text text text text. Text text text text text text text text text text text text text text text.

Text text text text text text text text text text text text text text text. Text text text text text text text text text text. Text text text text text text text text text text text text text text text. Text text text text text text text text text text text text text text text. Text text text text text text text text text text. Text text text text text text text text text text text text text text text.

Text text text text text text text text text text text text text text text. Text text text text text text text text text text. Text text text text text text text text text text text text text text text. Text text text text text text text text text text text text text text text. Text text text text text text text text text text. Text text text text text text text text text text text text text text text.

Text text text text text text text text text text text text text text text. Text text text text text text text text text text. Text text text text text text text text text text text text text text text. Text text text text text text text text text text text text text text text. Text text text text text text text text text text. Text text text text text text text text text text text text text text text.

Appendix B

Project's website

You can create a special website for your project which contains empirical data and MatLab/R/Stata source codes, see meta-analysis.cz/sigma, for example. Stating in your thesis that the data and source codes are available upon request is enough but please, have them prepared for such requests. The faculty does not allow enclosed DVD.

- File 1: Master's thesis
- File 2: Empirical data
- File 3: Source codes