Publication and Data-Mining Bias in Observational Research: Improving Transparency and Reliability of Inferences by Means of Meta-Regressions

PRELIMINARY DRAFT

STEPHAN B. BRUNS[1]

Economic Policy Research Group, University of Kassel, Nora-Platiel-Str. 5, 34109 Kassel, Germany.

JOHN P. A. IOANNIDIS

Departments of Medicine, Health Research and Policy, and Statistics, and Meta-Research Innovation Center at Stanford, Stanford University, 1265 Welch Rd, Medical School Office Building Room X306, Stanford, CA 94305, USA.

Abstract: Meta-regression analysis may help to improve the reliability of inferences in various fields of economics that depend on observational designs and multi-variable adjusted effects. We show that meta-regression analysis can correctly identify which primary regression specifications provide evidence for non-zero underlying coefficients of the variable of interest. These results hold even if the analyzed literature is completely distorted by data-mining and publication bias. Whether non-zero underlying coefficients are caused by genuine effects or omitted-variable biases has to be discussed by the use of economic theory. Meta-regressions may enhance transparency of model selection decisions on the corresponding presence of non-zero underlying coefficients and provide a basis for theory-guided inference on genuine effects.

JEL classification: C12, C18, C40

Keywords: Meta-regression, meta-analysis, observational research, data-mining bias, publication bias, Monte-Carlo simulation.

[1] Corresponding author: Stephan Bruns, Economic Policy Research Group, University of Kassel, Nora-Platiel-Str. 5, 34109 Kassel, Germany, +49 (0) 561 804 2709, s.b.bruns@gmail.com. Stephan Bruns is grateful to the German Science Foundation (DFG) for financial support through the program DFG-GK-1411: “The Economics of Innovative Change”.

1. INTRODUCTION

Meta-regression models that aim to control for various biases are increasingly implemented in observational research in empirical economics (e.g., Doucouliagos et al., 2012; Rusnak et al., 2013; Bruns et al., 2014). Observational research in economics, much like in many other scientific disciplines, utilizes mainly the regression framework to estimate the effect of interest while controlling for a potentially large set of possibly confounding variables. The flexibility of this research design and the corresponding wide range of obtainable estimates has been discussed in the seminal works of Hendry (1980), Leamer (1983), and Sims (1988). This flexibility may lead to a large number of false-positive findings, when control variables are selected in such a way as to obtain nominally significant results for the regression coefficient of the variable of interest. The problem of false-positives due to such analytical flexibility can pervade very diverse scientific fields (Ioannidis, 2005), but may be a major problem for empirical economics in particular (Ioannidis and Doucouliagos, 2013), due to the ubiquitous use of complex regression models with multiple adjusting variables to choose from. The increasing application of experimental designs in economics may help improve the credibility of research findings in the field (Angrist and Pischke, 2010; Banerjee and Duflo, 2009) but there are limitations (Leamer, 2010; Keane, 2010). Many economic research questions may remain difficult or impossible to address with experimental designs, especially in macroeconomics (Stock, 2010). Hence, observational research is likely to remain important at least in macroeconomic research.

Accordingly, we probe here meta-regression analysis as an instrument to account for biases and improve the reliability of inferences from such observational research. Meta-regression analysis in economics was originally proposed to identify the sources of heterogeneity in observational research (Stanley and Jarrell, 1989). The regression coefficients of primary studies are regressed on primary study characteristics, such as variable definitions, estimation techniques, data sets, and functional forms. Subsequently, meta-regression analysis was further developed to identify genuine empirical effects while controlling for potential publication bias (Card and Krueger, 1995; Stanley, 2001; 2008). A genuine effect is characterized by a non-zero underlying coefficient that is consistently estimated by the primary regression. Accordingly, the t-value of an estimated regression coefficient should be associated with the precision of the estimate or the degrees of freedom if a genuine effect is present. This association can be utilized to test for genuine effects. Simulations have shown that meta-regression models can identify genuine effects in literatures which are distorted by publication bias (Stanley, 2008; Moreno et al., 2009). However, other types of bias in primary regressions can also cause non-zero underlying coefficients that are consistently estimated by the primary regressions. Of particular concern is data-mining bias, which is produced when specific control (adjusting) variables are dropped from the multivariable analysis so as to generate omitted-variable biases in order to obtain significant results for the estimated regression coefficient of interest.

In this paper, we suggest addressing data-mining bias by controlling for primary regression specifications in the meta-regression in order to assess which primary regression specifications provide support for non-zero underlying coefficients. We discuss and evaluate the performance of meta-regression models in identifying non-zero underlying coefficients in the presence of publication and data-mining bias for a variety of research scenarios prevalent in empirical economics and the social sciences in general. Our results reveal that meta-regression models can single out whether a given primary regression specification provides evidence for a non-zero underlying coefficient for the variable of interest. Hence, transparency of model selection decisions on the corresponding estimates may be increased and genuine effects may be identified by theoretical considerations that discuss whether a given theory is supported by a non-zero underlying coefficient in an adequately specified primary regression. Our results also imply that the current practice of meta-regression analysis in economics may be misleading as the issue of data-mining bias is not adequately addressed.

Section 2 discusses the flexibility of observational research designs and the corresponding uncertainty about the reliability of inferences, particularly in the presence of incentives to search for specific results, e.g. theory-confirming results. Section 3 presents the meta-regression models and the approach how we suggest addressing data-mining bias. Section 4 and 5 present the designs and results of a controlled and worst-case simulation, respectively. Section 6 discusses the implications of our findings and section 7 concludes.

2. FLEXIBILITY IN OBSERVATIONAL RESEARCH DESIGNS

Empirical research is embedded in an incentive system that is likely to cause severe biases in published empirical findings. The publishing process is characterized by selection for statistically significant results and often spectacular and extreme results are rewarded with publication in the top journals (Young et al., 2008). Moreover, strong theoretical presumptions about the relation between two variables can foster a search for theory-confirming results (Card and Krueger, 1995). Once findings become rooted in the literature, authors may then try to adjust their empirical findings to conform with the anticipated expectations of reviewers (Frey, 2003). These biases may exaggerate the economic significance of an effect or, in the worst case, provide systematic support to false theories. The low acceptance rate of leading economic journals contributes further pressure toward providing significant and theory-conforming results to start or advance a career in academia (Card and DellaVigna, 2013). The search for specific results need not necessarily result from a conscious manipulation of the estimate of interest, but might be rather the outcome of an unconscious 'playing with the data'. The large degree of flexibility of observational research designs lends itself to obtaining basically almost any estimate desired (Leamer, 1983; Leamer and Leonhard, 1983).

Consider a theory that states X causes Y and the corresponding data generating process (DGP) is:

$$Y = \beta X + \boldsymbol{Z}\boldsymbol{\gamma}' + \epsilon \,, \qquad (1)$$

where β is the effect of interest, $\boldsymbol{Z}$ is a $1\text{x}n$ vector of control variables, $\boldsymbol{\gamma}$ is a $1\text{x}n$ vector of coefficients, and $E[\epsilon \mid X, \boldsymbol{Z}] = 0$. Suppose β is actually zero, but authors try to obtain a positive and statistically significant estimate of β.[2] Toward fulfilling this aim, many adjustments can be made in the estimation procedure, including the variable definitions, the functional forms, and the estimation techniques. The set of control variables is of particular interest in this decision process as modifications are simple to implement and explanations for opting for a specific set of control variables are generally easy to find. Suppose an empirical literature with $i = 1, \dots, k$ independent studies that estimate β by using $s = 1, \dots, v$ different primary regression specifications where each primary regression specification is estimated multiple times by different primary studies ($v \ll k$). The estimated primary regressions become:

$$Y_i = \hat{\beta}_{is} X_i + \boldsymbol{Z_{is}}\boldsymbol{\hat{\gamma}'_{is}} + \hat{\epsilon}_{is} \,, \qquad (2)$$

where $\hat{\beta}_{is}$ is the OLS estimate of β, $\boldsymbol{Z_{is}}$ is a $1\text{x}q_s$ vector with a subset of control variables of $\boldsymbol{Z}$ with q_s as the amount of control variables of primary regression specification s and $\boldsymbol{\hat{\gamma}_{is}}$ is a $1\text{x}q_s$ vector of estimated coefficients corresponding to $\boldsymbol{Z_{is}}$. A given study can thus 'play with the data' by including and excluding control variables from the primary regression to obtain omitted-variable biases. We denote the exploitation of omitted-variable biases to achieve positive and significant $\hat{\beta}_{is}$ as data-mining bias, irrespective of whether it is conscious or unconscious. This bias is of great concern in empirical economics. The extreme bounds analysis of Leamer and Leonard (1983) specifically addresses this issue by advocating reporting the fragility of $\hat{\beta}_{is}$ with respect to changes in the set of control variables. Asymptotically, experimental designs do not suffer from this data-mining bias due to the randomization of X with respect to $\boldsymbol{Z}$ (Leamer, 1983). A second source of bias in $\hat{\beta}_{is}$ is the selection of positive and significant results from those results provided by sampling variability. In an extreme case, only those 5% of $\hat{\beta}_{is}$ are published that are significant by chance, whereas the remaining 95% remain in the file-drawer (Rosenthal, 1979). We denote this bias as publication bias

[2] Without loss of generality we assume throughout the paper that the desired estimate is positive rather than negative.

and it is considered to diminish with sample size. The precision of $\hat{\beta}_{is}$ increases with sample size and the probability of a large estimate by chance decreases.

In observational research, data-mining bias and publication bias are the two main instruments that can be utilized to select positive and significant $\hat{\beta}_{is}$ for publication. We summarize the properties of a primary literature under publication selection in the following proposition.

Proposition 1. *Let an empirical literature estimate β in (1) by the use of regressions as expressed in (2), then:*

$$\hat{\beta}_{is} = \beta + \beta_{is}^{OVB} + \hat{\beta}_{is}^{SV},$$

where OVB denotes omitted-variables bias and SV sampling variability. Further, define $\beta + \beta_s^{OVB}$ as the underlying coefficient of primary regression specification s. Each study i consistently estimates the underlying coefficient of the respective primary regression specification:

$$\operatorname*{plim}_{m_i \to \infty} \hat{\beta}_{is} = \beta + \beta_s^{OVB},$$

where m_i is the sample size of study i. Let publication selection for positive and statistically significant $\hat{\beta}_{is}$ (PS) by means of data-mining bias and publication bias as discussed above affect the empirical literature. It follows:

$$E\left[\frac{\hat{\beta}_{is}}{\widehat{se}_{is}}\middle|PS\right] = E\left[\frac{\beta}{\widehat{se}_{is}}\middle|PS\right] + E\left[\frac{\beta_{is}^{OVB}}{\widehat{se}_{is}}\middle|PS\right] + E\left[\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}\middle|PS\right] > t_{0.975,m_i-q_s-1},$$

where $\widehat{\mathrm{se}}_{\mathrm{is}}$ is the estimated standard error of $\hat{\beta}_{is}$, $E\left[\frac{\beta_{is}^{OVB}}{\widehat{se}_{is}}\middle|PS\right]$ *denotes data-mining bias, and* $E\left[\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}\middle|PS\right]$ *denotes publication bias.*

Proof 1. *See Appendix A1.*

As a result, the findings comprising the published literature are distorted if authors utilize data-mining and publication bias to obtain a positive and significant $\hat{\beta}_{is}$. In this case, we can expect $E\left[\frac{\hat{\beta}_{is}}{\widehat{se}_{is}}\middle|PS\right] > t_{0.975,m_i-q_s-1}$, even if $\beta = 0$. Meta-regression analysis aims to integrate the published $\hat{\beta}_{is}$ in order to identify genuine effects while explicitly addressing publication bias and data-mining bias.

3. META-REGRESSION MODELS

3.1 Fixed-Effects Model

The benchmark of our analysis is the weighted average of $\hat{\beta}_{is}$ which is known as the fixed-effects model in meta-analysis:

$$\hat{\beta}_{is} = \alpha + u_i \quad (3)$$

which is estimated by weighted least squares (WLS) with weights equal to the inverse of the variance of $\hat{\beta}_{is}$. This gives more weight to estimates which are based on larger sample sizes and are more precise. In the absence of biases, $\hat{\alpha}$ is a consistent estimate of the genuine effect β (Sutton et al., 2000). However, $\hat{\alpha}$ becomes biased in the presence of publication selection for positive and significant $\hat{\beta}_{is}$. In contrast to randomized experiments, each $\hat{\beta}_{is}$ consistently estimates the underlying coefficient of the respective primary regression specification ($\beta + \beta_s^{OVB}$) rather than β and, consequently, meta-regression analysis in observational research cannot directly infer whether $\beta = 0$ or $\beta \neq 0$. But, meta-regressions can address data-mining bias by including dummy variables for each primary regression specification in the meta-regression allowing the meta-analyst to transparently discuss by the use of theory whether a non-zero underlying coefficient is caused by a genuine effect ($\beta \neq 0$) or by an omitted-variable bias alone ($\beta_s^{OVB} \neq 0$ while actually $\beta = 0$). The weighted average of $\hat{\beta}_{is}$ can be augmented to:

$$\hat{\beta}_{is} = \sum_{s=1}^{v} \alpha_s D_s + u_i \ , \qquad (4)$$

where $D_s = 1$ if study i uses primary regression specification s, namely $\boldsymbol{Z_{is}}$ as control variables, and zero otherwise. Accordingly, H_0: $\alpha_s = 0$ tests for a non-zero underlying coefficient for each primary regression specification *s*. Whether a non-zero underlying coefficient is caused by a genuine effect or omitted-variable bias has to be discussed on a case-by-case basis, and decisions may have to be made based on theoretical or other prior considerations which may or may not be valid. Though adding dummy variables for each primary regression specification helps addressing data-mining bias, publication bias can still bias $\hat{\alpha}_s$ as an estimate of the underlying coefficient.

Proposition 2. *Let an empirical literature be distorted by publication selection for positive and significant $\hat{\beta}_{is}$ as discussed in Proposition 1, then the weighted mean of $\hat{\beta}_{is}$ (3) provides a biased estimate of the genuine effect (β):*

$$E[\hat{\alpha}|PS] = \beta + E\left[\frac{\sum_{i=1}^{n}\frac{\beta_{is}^{OVB}}{\widehat{se}_{is}^2}}{\sum_{i=1}^{n}\frac{1}{\widehat{se}_{is}^2}} \mid PS\right] + E\left[\frac{\sum_{i=1}^{n}\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}^2}}{\sum_{i=1}^{n}\frac{1}{\widehat{se}_{is}^2}} \mid PS\right].$$

The augmented weighted mean of $\hat{\beta}_{is}$ (4) provides the meta-analyst an unbiased estimate of the underlying coefficient of each primary regression specification in the absence of publication bias, but it is still biased in the presence of publication bias:

$$E[\hat{\alpha}_s|PS] = \beta + \beta_s^{OVB} + E\left[\frac{\sum_{i=1}^{n} D_s\,\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}^2}}{\sum_{i=1}^{n} D_s\,\frac{1}{\widehat{se}_{is}^2}} \mid PS\right].$$

Proof 2. *See Appendix A1.*

3.2 Precision-Effect Test

Based on Egger et al. (1997), Stanley (2001; 2008) suggests the Precision-Effect Test (PET) to synthesize empirical findings in economics:

$$t_{is} = \delta + \alpha\frac{1}{\widehat{se}_{is}} + u_i\,, \qquad (5)$$

where t_{is} is the t-value of $\hat{\beta}_{is}$. This model is considered to be largely robust with respect to data-mining and publication bias (Stanley, 2008). Accordingly, H_0: $\alpha = 0$ is interpreted as a direct test for a genuine effect. We, again, suggest augmenting the PET model to estimate the underlying coefficient of each primary regression specification:

$$t_{is} = \sum_{s=1}^{v}(\delta_s D_s + \alpha_s D_s\frac{1}{\widehat{se}_{is}}) + u_i\,. \ (6)$$

If a non-zero underlying coefficient is present, the t-value follows a non-central t-distribution with a centrality parameter that is an increasing function of precision. In the absence of a non-zero underlying coefficient, the t-distribution is randomly distributed around zero without relation to precision. As pure publication bias does not cause the underlying effect to be non-zero, PET is robust with respect to this bias. Specifically, if positive and significant $\hat{\beta}_{is}$ are systematically selected from those results offered by sampling variability, we expect the t-value to be constant rather than increasing with precision. Hence, H_0: $\alpha_s = 0$ tests whether the use of primary regression specification s supports a non-zero underlying coefficient corrected for publication bias.

Proposition 3. *Let an empirical literature be distorted by publication selection for positive and significant $\hat{\beta}_{is}$ as discussed in Proposition 1 and let the primary sample sizes be sufficiently large to ensure that the t_{is} are approximately normally distributed. Then, the PET model (5) provides a biased estimate of the genuine effect (β):*

$$E[\hat{\alpha}|PS] = \beta + E\left[\frac{Cov\left(\frac{\beta_{is}^{OVB} + \hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}, \frac{1}{\widehat{se}_{is}}\right)}{Var\left(\frac{1}{\widehat{se}_{is}}\right)}|PS\right].$$

For the augmented PET model (6), three cases can be distinguished:

i. *If $\beta + \beta_s^{OVB} = 0$, publication selection for positive and significant $\hat{\beta}_{is}$ can be only achieved by publication bias. Then:*

$$E[\hat{\alpha}_s|PS] = \beta + \beta_s^{OVB}.$$

ii. *If $\beta + \beta_s^{OVB} > 0$ and additional publication bias is not required to achieve positive and significant $\hat{\beta}_{is}$, then:*

$$E[\hat{\alpha}_s|PS] = \beta + \beta_s^{OVB}.$$

iii. *If $\beta + \beta_s^{OVB} > 0$ and additional publication bias is required to achieve positive and significant $\hat{\beta}_{is}$, then:*

$$E[\hat{\alpha}_s|PS] = \beta + \beta_s^{OVB} + E\left[\frac{Cov\left(D_s\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}, D_s\frac{1}{\widehat{se}_{is}}\right)}{Var\left(D_s\frac{1}{\widehat{se}_{is}}\right)}|PS\right].$$

Proof 3. *See Appendix A1.*

In the presence of publications selection for positive and significant $\hat{\beta}_{is}$, case i and ii are mainly prevalent in the empirical literature. Case iii is a special case of ii where a primary regression specification is characterized by a small $\beta + \beta_s^{OVB}$ in combination with studies with a low degree of precision. In this case, publication bias has to be used to achieve positive and significant $\hat{\beta}_{is}$ for low degrees of precision, whereas publication bias is not necessary for larger degrees of precision.

The use of $\widehat{se}_i$ as independent variable is known to introduce bias as $\widehat{se}_i$ is only an estimate of se_i (Macaskill et al., 2001). This is in particular a problem if the primary studies are based on small samples and $\widehat{se}_i$ is not precisely estimated. Furthermore, PET can be only applied if the same measurement units for X and Y are used across primary studies (Becker and Wu, 2007).

3.3 p-Value Tests

The prior models use the t-value as the dependent variable, either implicitly in the fixed effects model by weighting $\hat{\beta}_{is}$ with the inverse of its variances or explicitly in PET. However, the t-value provides a non-constant relation to the p-value, as the variance of the t-distribution decreases with degrees of freedom (df). We do not expect necessarily an increasing t-value if df increase and a genuine effect is present. In an extreme case, we may observe a set of $\hat{\beta}_{is}$ that provide decreasing p-values with increasing df, but constant t-values across the observed range of df. Bruns (2013) proposes the use of probit-transformed p-values as dependent variable to overcome this issue:

$$z_{is} = \begin{cases} \Phi^{-1}\left(\frac{p_i}{2}\right), & if\ \hat{\beta}_{is} < 0, \\ \Phi^{-1}\left(1 - \frac{p_i}{2}\right), & if\ \hat{\beta}_{is} \geq 0\,, \end{cases}$$

where p_i is the p-value of a two-sided t-test of $\hat{\beta}_{is}$ and Φ^{-1} is the probit transformation. As a result, z_{is} provides values of the standard normal distribution with exactly the same p-value as the corresponding original t-value. Hence, z_{is} ensures a constant relation to the level of significance. The p-Value Test with Precision (pVT-se) is given by:

$$z_{is} = \sum_{s=1}^{v} (\delta_s D_s + \alpha_s D_s\ \frac{1}{\widehat{se}_{is}}) +\ u_i, \qquad (7)$$

where $H_0\colon \alpha_s = 0$ tests whether the use of primary regression specification s supports a non-zero underlying coefficient corrected for publication bias. This model requires also the same units of measurement for X and Y across studies (Becker and Wu, 2007) and it is also biased (Macaskill et al., 2001). Bruns (2013) further extends this meta-regression model to the p-Value Test with Degrees of Freedom (pVT-df) to overcome these issues:

$$z_{is} = \sum_{s=1}^{v} (\delta_s D_s + \alpha_s D_s\ \sqrt{df_{is}}) +\ u_i\,, \qquad (8)$$

where $H_0\colon \alpha_s = 0$ tests whether the use of primary regression specification s supports a non-zero underlying coefficient corrected for publication bias. The use of df avoids the error-in-measurement bias and is also independent of the used units of measurement. Model (7) and (8) relax the assumption of Proposition 3 that the primary sample sizes have to be sufficiently large to ensure that the t_{is} are normally distributed. The reminder of the proposition holds for both models.

4. CONTROLLED SIMULATION

4.1 Design

We analyze to what extent meta-regression models can single out which primary regression specification supports a non-zero underlying coefficient while adjusting for publication bias. For this purpose, we employ a controlled simulation that allows us to vary a single parameter at a time in order to identify the determinants of the performance of the presented meta-regression models.

Each meta-regression model integrates k primary studies. We consider $k = 20, 40, 60, 80, 160, 320$ to cover meta-regression analyses of small to very large literatures that can be typically observed in economics (Doucouliagos and Stanley, 2011). We draw the sample size of each primary study $i = 1, \dots, k$ from a gamma distribution with scale parameter $\frac{\sigma^2}{(\mu-10)}$ and shape parameter $\frac{(\mu-10)^2}{\sigma^2}$, rounded the obtained value to the next integer, and added 10. Hence, 10 is the smallest primary study sample size and we can vary mean and variance of the gamma distribution independently. We consider $\mu = 30, 60, 120, 240, 480$ as well as $\sigma^2 = 15^2,\ 30^2, 60^2,\ 120^2, 240^2$ to cover a large variety of potential primary sample size distributions. We analyze those combinations of μ and σ for which $\mu \geq 2 * \sigma$ holds to ensure that the truncation at 10 does not prevent μ from adequately presenting the true mean. The sample size distribution of the primary studies is right skewed for small sample sizes and more symmetric for larger sample sizes, just as we can observe in empirical economics. The DGP of each study i is:

$$Y_i = \beta\, X_i + \gamma\, Z_i + \ \varepsilon_i \ (9)$$

with $\varepsilon_i \sim N(0,1)$ and $X_i \sim N(0,1)$. The dependence between the independent variables is modeled as $Z_i = X_i + v_i$ with $v_i \sim N(0,1)$. β is the effect of interest and we consider $\beta = 0, 0.2, 0.4, 0.8$ to evaluate size and power for very small to large genuine effects. Each primary study estimates the following regression:

$$Y_i = \hat{\beta}_{is} X_i + \hat{\epsilon}_{is}\,. \quad (10)$$

We model data-mining bias by introducing a fixed omitted-variable bias of $\gamma = 0.2$ for $\frac{k}{2}$ studies in (9), whereas the data for the remaining $\frac{k}{2}$ studies are generated with $\gamma = 0$. As a result, those $\frac{k}{2}$

studies for which $\gamma = 0$ holds provide unbiased $\hat{\beta}_{is}$, whereas those $\frac{k}{2}$ studies for which $\gamma = 0.2$ holds provide biased $\hat{\beta}_{is}$.

Publication selection for positive and significant $\hat{\beta}_{is}$ is implemented by resampling (9) until the first h% of $\hat{\beta}_{is}$ are positive and significant, whereas the remaining (100-h)% are estimated without resampling. We consider $h = 0, 25, 50, 75, 100$. Publication selection is applied symmetrically to both cases, namely that with fixed omitted-variable bias and without fixed omitted-variable bias. For those $\frac{k}{2}$ studies for which $\gamma = 0$, publication selection is purely driven by publication bias, achieved by resampling (9) until the first h% of studies provide positive and significant $\hat{\beta}_{is}$ solely by chance. For those $\frac{k}{2}$ studies for which $\gamma = 0.2$, publication selection need not rely on publication bias alone, but can also utilize the fixed omitted-variable bias to obtain positive and significant $\hat{\beta}_{is}$. Thus, the outcome is a primary literature with the same amount of correctly estimated and misspecified studies and both groups of studies are affected by the same degree of publication selection. The dummy variables for the correct and misspecified primary regression specifications are modeled as $D_{s=correct} = 1$ if $\gamma = 0$ and zero otherwise and $D_{s=DM} = 1$ if $\gamma = 0.2$ and zero otherwise.

We not only evaluate the performance of meta-regression models that control for primary regression specifications, but also the performance of meta-regression models that pool primary regression specifications irrespective of whether they are either correctly specified or misspecified. This scenario mirrors the current practice of meta-regression analysis in economics where non-zero underlying coefficients are largely interpreted as direct evidence for genuine effects (Stanley, 2008) and, consequently, primary control variables are only accounted for in a non-systematic way. Hence, the rejection of $\mathrm{H_0}: \alpha_{\mathrm{pooled}} = 0$ is understood as a direct test of a genuine effect although this rejection may be also caused by omitted-variable bias.

Meta-regression models can include further variables to take other sources of heterogeneity into account (Stanley and Doucouliagos, 2012). Given that it is unlikely to cover all sources of heterogeneity, we base the inference in our simulations on heteroskedasticity-consistent standard errors that are designed for small samples (Long and Ervin, 2000) as meta-analysis may often face small sample sizes.

4.2 Results

Apart from the fixed-effects model, all meta-regression models that control for primary regression specifications provide a reasonable performance in identifying non-zero underlying coefficients. The *likelihood ratio positive* (LR+) of $\mathrm{H}_0\colon \alpha_{\mathrm{correct}} = 0$ for a small genuine effect of 0.2 is presented in Figure 1 measuring the ratio of the probabilities of a true-positive finding and of a false-positive finding. Given rejection of $\mathrm{H}_0\colon \alpha_{\mathrm{correct}} = 0$, larger values of LR+ indicate a larger probability that a non-zero underlying coefficient indeed exists. Even for a very small genuine effect of 0.2, LR+ is typically >3 and exceeds 10 in most cases for a genuine effect of 0.4 (Fig. A2.1). LR+ values are higher with larger primary sample size variance, meta-analytic sample size, and size of the genuine effect, and with smaller mean of the primary sample size distribution.

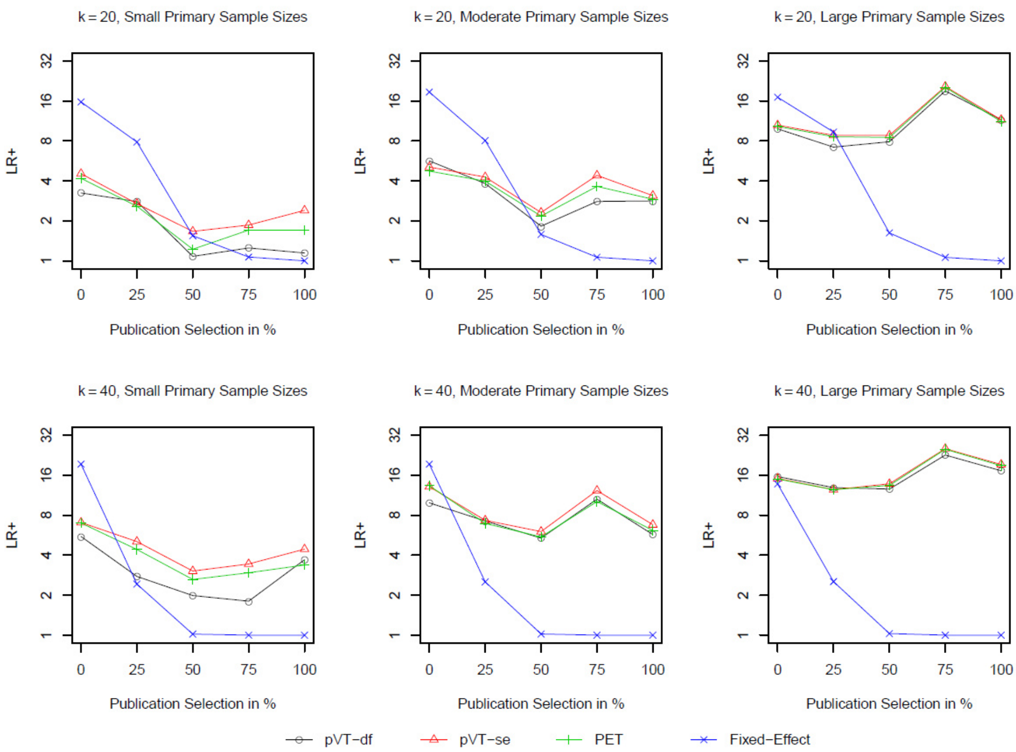

Figure 1: Likelihood ratio positive ($LR^+ = \frac{power}{size}$) for $\mathrm{H}_0\colon \alpha_{\mathrm{correct}} = 0$ with $\beta = 0.2$. LR+ is shown for small ($\mu = 60$ and $\sigma^2 = 30^2$), medium ($\mu = 120$ and $\sigma^2 = 60^2$), and large ($\mu = 480$ and $\sigma^2 = 240^2$) primary sample size distributions and very small ($k/2 = 10$) and small ($k/2 = 20$) meta-analytic sample sizes.

The *likelihood ratio negative* (LR-) of $\mathrm{H}_0\colon \alpha_{\mathrm{correct}} = 0$ for a small genuine effect of 0.2 is presented in Figure 2 measuring the ratio of the probabilities of a false-negative finding and of a true-negative finding. Given acceptance of $\mathrm{H}_0\colon \alpha_{\mathrm{correct}} = 0$, smaller values of LR- indicate a larger probability that the underlying coefficient is indeed zero. For a small genuine effect of 0.2 the LR- can be large indicating that the meta-regression models may have difficulties with detecting small non-zero underlying coefficients reflecting case three in Proposition 3. However, this problem is resolved for large primary sample sizes and for a genuine effect of 0.4 (Fig. A2.2).

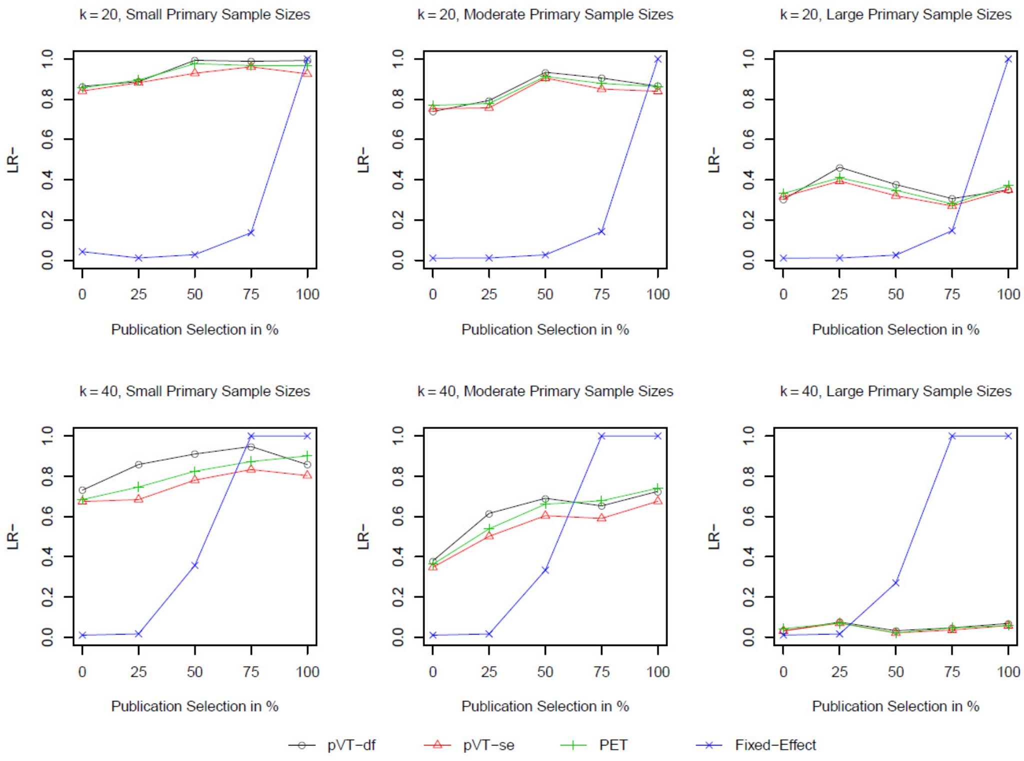

Figure 2: Likelihood ratio negative ($LR^{-} = \frac{1-power}{1-size}$) for $\mathrm{H_0}\colon \alpha_{\mathrm{correct}} = 0$ with $\beta = 0.2$. LR- is shown for small ($\mu = 60$ and $\sigma^2 = 30^2$), medium ($\mu = 120$ and $\sigma^2 = 60^2$), and large ($\mu = 480$ and $\sigma^2 = 240^2$) primary sample size distributions and very small ($k/2 = 10$) and small ($k/2 = 20$) meta-analytic sample sizes.

For $\mathrm{H_0}\colon \alpha_{\mathrm{DM}} = 0$, LR+ and LR- are essentially the same as for $\mathrm{H_0}\colon \alpha_{\mathrm{correct}} = 0$ (Fig. A3.1 and Fig. A3.2). The reason is that in both cases a non-zero underlying coefficient exists either caused by a genuine effect or by an omitted-variable bias.

The rejection rates of $\mathrm{H_0}\colon \alpha_{\mathrm{pooled}} = 0$ for the absence of a genuine effect and the presence of a small omitted-variable bias are shown in Figure 3. If $\mathrm{H_0}\colon \alpha_{\mathrm{pooled}} = 0$ is interpreted as a direct test for a genuine effect, the rejection rates in the absence of a genuine effect are the type I errors. Accordingly, meta-regression models that pool primary regression specifications and interpret a non-zero underlying coefficient as direct support of a genuine effect provide highly inflated type I errors in the presence of data-mining bias.

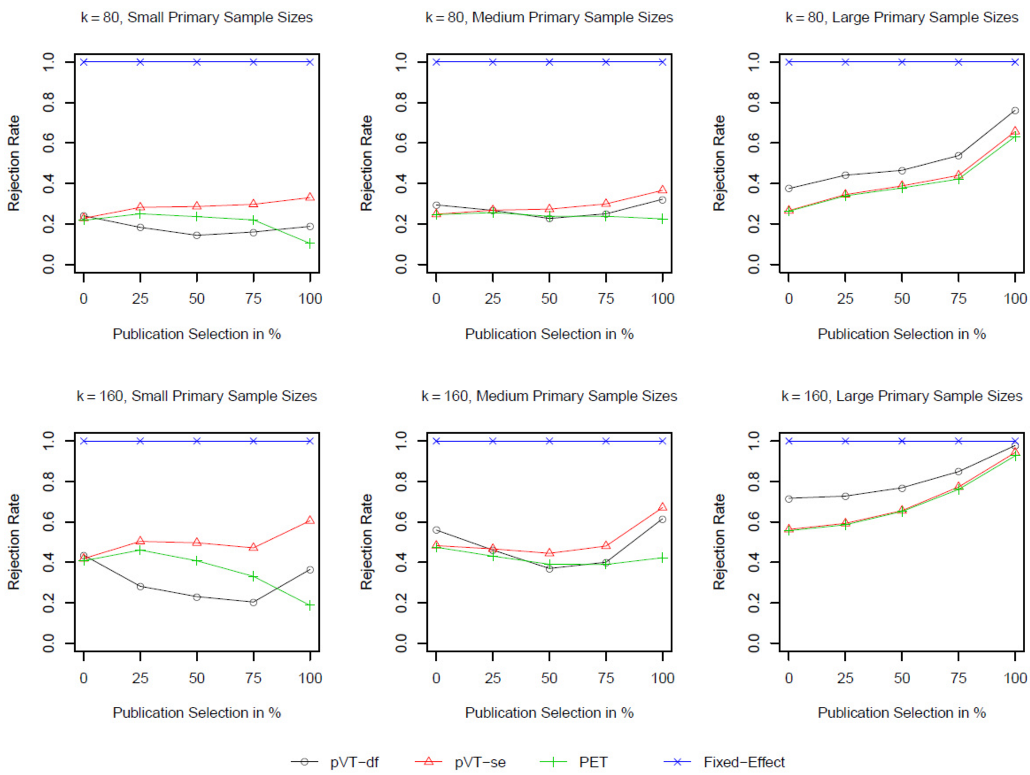

Figure 3: Rejection rates of $\mathrm{H}_0\colon \alpha_{\text{pooled}} = 0$ with $\beta = 0$ and $\gamma = 0.2$. Rejection rates are shown for small ($\mu = 60$ and $\sigma^2 = 30^2$), medium ($\mu = 120$ and $\sigma^2 = 60^2$), and large ($\mu = 480$ and $\sigma^2 = 240^2$) primary sample size distributions and large ($k = 80$) and very large ($k = 160$) meta-analytic sample sizes.

5. WORST-CASE SIMULATION: THEORY-CONFIRMATION BIAS

5.1 Design

The results of the controlled simulation show that those meta-regression models that include dummy variables to control for different primary regression specifications provide good performance in identifying which particular primary regression specification indeed supports a non-zero underlying coefficient. We also evaluate the performance of these models for a worst-case scenario to assess the limitations of these models. For this purpose, we consider an empirical literature that suffers from theory-confirmation bias. Strong theoretical presumptions may force authors to provide theory-confirming results (Card and Krueger, 1995). If these results become established, reviewers and editors may tend to reject results that deviate, causing authors to search for, and give greater weight to, theory-conforming results (Frey, 2003). Suppose a theory states that X causes Y and the corresponding DGP is:

$$Y = \beta X + 0.2\, Z_1 + 0.4\, Z_2 + 0.2\, V_3 + 0.2\, V_4 + \epsilon\,, \qquad (11)$$

where $X = 0.5\,C + 0.5\,v_X$ with $v_X \sim N(0,1)$, $Z_m = 0.5\,C + 0.5\,v_m$ for $m = 1,2$, $V_m = 0.2\,C + 0.8\,v_m$ for $m = 3,4$ with $v_m \sim N(0,1)$, $C \sim N(0,1)$, and $\epsilon \sim N(0,1)$. Each independent variable is a weighted sum of a common component (C) and an *iid* component to model the dependence between these variables. The coefficient of interest is β and the desired result is a positive and significant $\hat{\beta}_{is}$ in order to confirm the theory that states X causes Y. V_3 and V_4 are variables which are understood to be unquestionably relevant for Y, whereas the relevance of Z_1 and Z_2 is a matter of debate. We model small to large omitted variable biases by either omitting Z_1 and Z_2 individually or both simultaneously.

Authors start with estimating the true regression including X, Z_1, Z_2, V_3 and V_4. If $\hat{\beta}_{is}$ is positive and significant by chance, study i is published with 100%. The probability of a non-significant $\hat{\beta}_{is}$ or a negative and significant $\hat{\beta}_{is}$ being published is specified to be 1% as these results contradict the existing theory. If the result is not published, authors drop $Z_1$from the regression to create the smallest omitted-variable bias available. If $\hat{\beta}_{is}$ becomes positive and significant, the result is published with a probability of 80%, as some reviewers may ask whether the estimated regression includes all relevant variables. If $\hat{\beta}_{is}$ is not yet positive and significant, the probability of it being published will again only be 1%. If the result is not published, researchers add Z_1 again and drop Z_2 to exploit the next larger omitted-variable bias of the questionable variables. If $\hat{\beta}_{is}$ becomes positive and significant, the result is published with a probability of 60% and with a probability of 1% otherwise. Finally, Z_1 and Z_2 are dropped simultaneously to obtain the largest omitted-variable bias possible. If $\hat{\beta}_{is}$ becomes positive and significant, the result is published with 40% and otherwise with 1%. If positive and significant results still elude the authors, the results remain completely unpublished. This search for significant results is not necessarily a conscious process, but it can be also thought of as 'playing with the data'. The outcome is either way a biased literature that makes use of data-mining bias and publication bias to achieve a biased estimate that confirms the theory that states X causes Y.

We consider two underlying genuine effects for this theory-confirmation search process. First, we set $\beta = 0$ indicating that the theory is completely false, although empirical evidence suggesting that this theory is true will be overwhelming. For this scenario, we consider the nominal significance level to be 0.05. Second, we consider $\beta = 0.2$ to investigate how the tendency towards theory confirmation impacts the performance of meta-regression models when there is a very small effect of X on Y that may, in fact, be economically irrelevant. In this case, the search for theory-confirming results exaggerates the size and significance of this effect to highlight its economic importance. Therefore, we set the nominal significance level to 0.001 for this scenario.

We evaluate PET, pVT-df and pVT-se due to their reasonable performance in the controlled simulation. We consider very small ($\mu = 30$ and $\sigma^2 = 15^2$), medium ($\mu = 120$ and $\sigma^2 = 60^2$), and moderate ($\mu = 240$ and $\sigma^2 = 120^2$) primary sample size distributions. We meta-analyze those empirical literatures that have at least 20 observations per primary regression specification to achieve a suitable degree of reliability. Accordingly, each meta-regression analysis faces always a literature with four different primary regression specifications, each of them used by at least 20 primary studies. Therefore, we choose the meta-analysis sample size to be $= 160, 320, 640$.

We compare meta-regression models that control for each primary regression specification by dummy variables with meta-regressions that do not account for different primary regression specifications and simply pool all estimates. Those meta-regressions that pool the estimates are again supposed to reflect the current practice of meta-regression analysis in economics where meta-regressions control for primary control variables in a non-systematic way. Our focus is on the comparison of the test for a non-zero underlying coefficient in the correctly specified primary regression ($H_0: \alpha_{correct} = 0$) with $H_0: \alpha_{pooled} = 0$.

5.2 Results

If no genuine effect is present and $\mathrm{H_0: \alpha_{pooled}} = 0$ is understood as a direct test of a genuine effect, then the rejection rate of $\mathrm{H_0: \alpha_{pooled}} = 0$ measures the type I errors. Figure 4 presents the rejection rates of $\mathrm{H_0: \alpha_{pooled}} = 0$ in the absence of a genuine effect as well as the rejection rates of $\mathrm{H_0: \alpha_{correct}} = 0$. The pooled meta-regression models provide highly inflated type I errors, whereas those meta-regression that control for primary regression specifications by the use of dummy variables provide adequate type I errors for medium and moderate primary sample size distributions. For small primary sample size distributions, only pVT-df can provide adequate type I errors due to the error-in-measurement bias in the other meta-regression models as they use the estimated standard error of the regression coefficient as independent variable.

Figure 5 illustrates that as soon as a genuine effect is present ($\beta = 0.2$) those models that account for primary regression specifications provide a reasonable power ($\mathrm{H_0: \alpha_{correct}} = 0$). Those models that pool all primary regression specifications ($\mathrm{H_0: \alpha_{pooled}} = 0$) provide again high rejection rates. Therefore, these models are unable to distinguish between the presence of a genuine effect and the solely presence of omitted-variable biases as both cause a non-zero underlying coefficient.

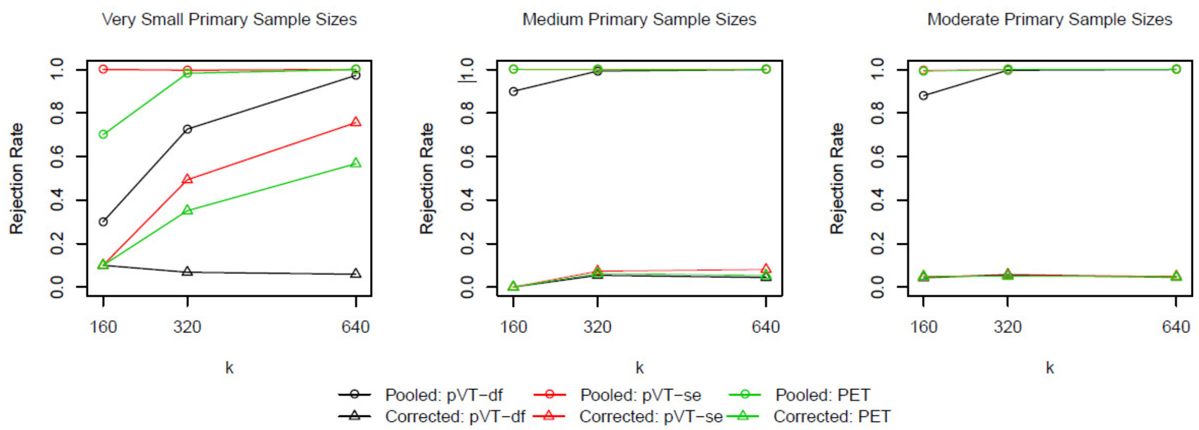

Figure 4: Rejection rates of $\mathrm{H_0}: \alpha_{\text{correct}} = 0$ (triangles) and $\mathrm{H_0}: \alpha_{\text{pooled}} = 0$ (circles) in the absence of a genuine effect $\beta = 0$. Rejection rates are shown for different meta-analytic sample sizes (k) and for very small ($\mu = 30$ and $\sigma^2 = 15^2$), medium ($\mu = 120$ and $\sigma^2 = 60^2$), and moderate ($\mu = 240$ and $\sigma^2 = 120^2$) primary sample size distributions.

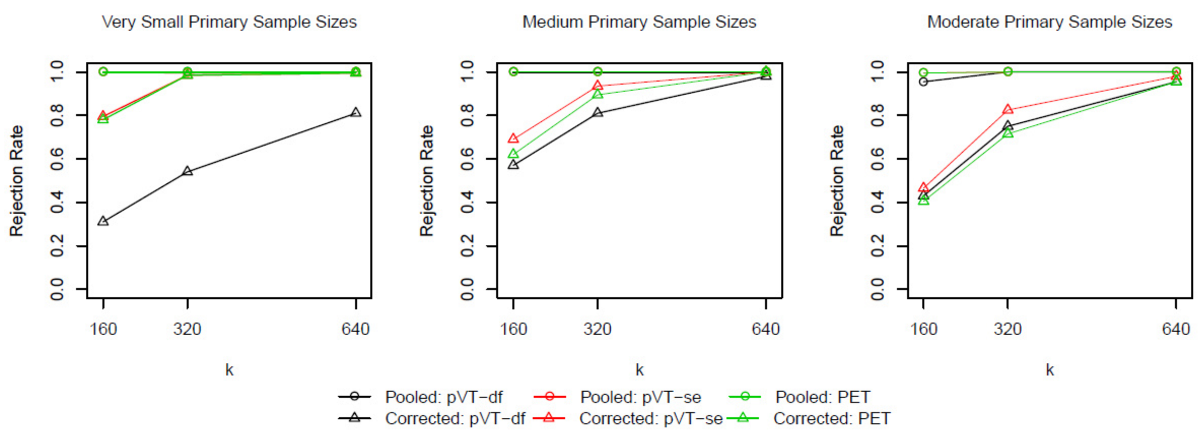

Figure 5: Rejection rates of $\mathrm{H_0}: \alpha_{\text{correct}} = 0$ (triangles) and $\mathrm{H_0}: \alpha_{\text{pooled}} = 0$ (circles) in the presence of a small genuine effect $\beta = 0.2$. Rejection rates are shown for different meta-analytic sample sizes (k) and for very small ($\mu = 30$ and $\sigma^2 = 15^2$), medium ($\mu = 120$ and $\sigma^2 = 60^2$), and moderate ($\mu = 240$ and $\sigma^2 = 120^2$) primary sample size distributions.

The results of the worst-case simulation of theory-confirmation bias support the findings of our controlled simulation. Even if a primary literature provides almost only positive and significant results, but without a genuine underlying effect, meta-regressions that control for primary regression specifications can identify the presence and absence of non-zero underlying coefficients, whereas meta-regressions that do not control for primary regressions specifications almost always suggest the presence of a non-zero underlying coefficient given data-mining bias.

6 DISCUSSION

The simulation results support that meta-regression models can address publication bias defined as the selection of positive and significant results from those results offered by sampling variability. This finding is consistent with prior simulation results (Stanley, 2008; Moreno et al., 2009). However, pure publication bias is likely to be dominated by data-mining bias in empirical economic research as the variation of control variables, or model assumptions in general, can be much easier implemented than the estimation of the same model for new, different samples. The simulation results further support that meta-regression models that control for each primary regression specification by the use of dummy variables can identify which primary regression specifications provide support for non-zero underlying coefficients. This finding helps to address data-mining bias in observational research as the transparency of the relation between model selection decisions and the corresponding presence of non-zero underlying coefficients can be greatly increased. However, data-mining bias cannot be resolved by meta-regression models as a non-zero underlying coefficient can be caused by either a genuine effect or omitted-variable bias. Distinguishing between these two sources of a non-zero underlying coefficient requires theoretical speculations and may be valid or invalid.

Currently, a non-zero underlying coefficient, measured by the association between the t-statistic and precision, is largely understood as direct support for a genuine empirical effect (Stanley, 2008). Therefore, current meta-regression analyses account for primary control variables in a non-systematic way, rather than accounting for the conditional nature of the estimated regression coefficient by controlling for entire primary regression specifications in the meta-regression. Our analysis shows how misleading the outcome of these meta-regression models can be in the presence of data-mining bias. As the estimated regression coefficients of correctly specified and misspecified primary regressions are pooled, these meta-regression models almost always point to the presence of a non-zero underlying coefficient without providing the ability to understand which primary regression specification is actually the source of this non-zero underlying coefficient. As the conditions under which a non-zero underlying coefficient occurs cannot be unambiguously identified, the meta-analyst is unable to discuss the presence of genuine effects even if strong theoretical considerations could be utilized.

We suggest that meta-regression models should provide a comprehensive overview of which primary regression specifications support non-zero underlying coefficients enhancing the transparency of model selection decisions on the corresponding estimates of interest. Wider appreciation and adoption of such methods may also increase pressure on authors of primary studies to increase transparency of their model selection. Meta-regression analysis may discuss whether a given theory is supported in adequately specified primary regressions or whether support can be only found in primary regressions that may suffer from omitted-variable bias. Such an analysis may greatly increase our understanding of a given empirical field by identifying which conditions lead to non-zero underlying coefficients and whether these conditions reasonably mirror a given theory. However, we should acknowledge that the convincing identification of genuine effects in the presence of data-mining and publication bias heavily relies on the strengths of the discussed theories. Theoretical considerations may be valid or invalid and particularly in fields in which we conduct meta-analysis theory may be often weak.

Most importantly, meta-regressions need to be applied carefully and with transparency. Otherwise, they may suffer from similar data-mining biases as the primary studies which they synthesize. Poorly conducted meta-regression analysis can be dangerous as the authority of meta-analytic results claim to be superior to the results of primary studies and increased precision may be spurious (Egger et al., 1998). Our suggestion to control for entire primary regression specifications in the meta-regression allows us to address data-mining bias in a transparent way, but it is limited to empirical literatures that provide plenty of the same primary regression specifications. The common habit of providing a set of (selective) regression models to proof the robustness of the estimate of interest eases this data requirement.

7 CONCLUDING REMARKS

Observational research designs, particularly utilized by empirical economics and the social sciences in general, suffer from a large degree of analytical flexibility and the corresponding wide range of obtainable estimates. Accordingly, scientists may face substantial uncertainty about the reliability of inferences, particularly in the presence of incentives towards selecting specific results for publication, e.g. theory-confirming results. Meta-regressions may be helpful to address the issue of data-mining and publication bias in observational research designs. We suggest controlling for each primary regression specification in the meta-regression allowing the meta-analyst to infer which primary regression specifications provide evidence for non-zero underlying coefficients. The results of both the controlled and worst-case simulation support that meta-regression models are capable to distinguish between primary regressions specifications with a non-zero underlying coefficient and those without a non-zero underlying coefficient in the presence of publication bias. Hence, meta-regression models may help to address the issue of data-mining bias by increasing the transparency of model selection decisions on the corresponding presence of non-zero underlying coefficients. However, whether such a non-zero underlying coefficient is caused by a genuine effect or omitted-variable bias can only be determined by theoretical speculations.

REFERENCES

Angrist, J. D. and Pischke, J. S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24(2)*:3–30.

Banerjee, A. V. and Duflo, E. (2009). The experimental approach to development economics. *Annual Review of Economics 1*:151-178.

Becker, B. J. and Wu, M. J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science 22(3)*:414–429.

Bruns, S. B. (2013). Identification of genuine effects in observational research by means of meta-regressions. *Jena Economic Research Papers 2013-040*, Friedrich Schiller University Jena and Max Planck Institute of Economics.

Bruns, S. B., Gross, C. and Stern, D. I. (2014). Is there really granger causality between energy use and output? *The Energy Journal 35(4)*:101-134.

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

Card, D. and DellaVigna, S. (2013). Nine facts about top journals in economics. *NBER Working Paper Series, Working Paper 18665*.

Doucouliagos, C. and Stanely, T. D. (2011). Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys 27(2)*:316–339.

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

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

Egger, M., Schneider, M. and Smith, G. D. (1998). Spurious precision? Meta-analysis of observational studies. *British Medical Journal 316(7125)*:140–144.

Frey, B. S. (2003). Publishing as prostitution? - Choosing between one’s own ideas and academic success. *Public Choice 116(1-2)*:205–223.

Hendry, D. F. (1980). Econometrics - alchemy or science? *Economica 47(188)*:387–406.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine 2(8)*:e124.

Ioannidis, J. P. A. and Doucouliagos, C. (2013). What's to know about the credibility of empirical economics? *Journal of Economic Surveys 27(5)*:997-1004.

Keane, M. P. (2010). A structural perspective on the experimentalist school. *Journal of Economic Perspectives 24(2)*:47-58.

Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review 73(1)*:31–43.

Leamer, E. E. (2010). Tantalus on the road to asymptopia. *Journal of Economic Perspectives 24(2)*:31-46.

Leamer, E. E. and Leonhard, H. (1983). Reporting the fragility of regression estimates. *Review of Economics and Statistics 65(2)*:306–317.

Long, J. S. and Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *American Statistician 54(2)*:217–224.

Macaskill, P., Walter, S. D. and Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine 20(4)*:641–654.

Moreno, S. G., et al. (2009) Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Medical Research Methodology 9(1)*:2.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin 86(3)*:638–641.

Rusnak, M., Havranek, T. and Horvath, R. (2013). How to solve the price puzzle? A meta-analysis. *Journal of Money, Credit and Banking 45(1)*:37-70.

Sims, C. A. (1988). Uncertainty across models. *American Economic Review 78(2)*:163–167.

Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *Journal of Economic Perspectives 15(3)*:131-150.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection*. *Oxford Bulletin of Economics and Statistics 70(1)*:103–127.

Stanley, T. D. and Doucouliagos, H. (2012). Meta-regression analysis in economics and business. Routledge, London.

Stanley, T. D. and Jarrell, S. B. (1989). Meta-regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys 3(2)*:161–170.

Stock, J. H. (2010). The other transformation in econometric practice: Robust tools for inference. *Journal of Economic Perspectives 24(2)*:83–94.

Sutton, A. J., et al. (2000). Methods for meta-analysis in medical research. J. Wiley, New York.

Young, N. S., Ioannidis, J. P. A. and Al-Ubaydli, O. (2008). Why current publication practices may distort science. *PLoS Medicine 5(10)*:e201.

Appendix

A1: Proofs

Proof of Proposition 1:

Based on the DGP given in (1) and the corresponding primary regressions given in (2), it follows:

$$E[Y|X,\boldsymbol{Z_{is}}] = \beta X + \boldsymbol{Z_{is}}\,\boldsymbol{\gamma'_{is}} + E\left[\boldsymbol{Z_{ip}}\,\boldsymbol{\gamma'_{ip}}\middle|X,\boldsymbol{Z_{is}}\right] + E[\epsilon|X,\boldsymbol{Z_{is}}], \qquad (P1)$$

where $\boldsymbol{Z_{ip}}$ is the complementary subset to $\boldsymbol{Z_{is}}$ such that $\boldsymbol{Z_{ip}}$ is a 1x$p_S$vector of those control variables which are not included in $\boldsymbol{Z_{is}}$ with $p_S = n - q_S$. If study i estimates regression (2) with control variables $\boldsymbol{Z_{is}}$ and omission of the relevant variables $\boldsymbol{Z_{ip}}$, the omitted-variable biases become $E\left[\boldsymbol{Z_{ip}}\,\boldsymbol{\gamma'_{ip}}\middle|X,\boldsymbol{Z_{is}}\right]$**.** For each variable in $\boldsymbol{Z_{ip}}$, we can write $E\left[Z_{ig}\,\gamma_{ig}\middle|X,\boldsymbol{Z_{is}}\right]$ with $g = 1,\dots,p_s$ and it follows:

$$E\left[Z_{ig}\,\left(\gamma_{ig} + \left(\boldsymbol{W}'Z_{ig}\right)^{-1}\boldsymbol{W}'\epsilon\right)\middle|X,\boldsymbol{Z_{is}}\right]$$

with $\boldsymbol{W} = \{X,\boldsymbol{Z}\}$. This leads to:

$$E\left[Z_{ig}\gamma_{ig}\middle|X,\boldsymbol{Z_{is}}\right] = \left(\beta_{ig}^{ovb}X + \boldsymbol{Z_{is}}\boldsymbol{\gamma_{isg}^{ovb'}} + \boldsymbol{E}\left[u_{ig}|X,\boldsymbol{Z_{is}}\right]\right)\gamma_{ig} \quad (P2)$$

with $E\left[u_{ig}|X,\boldsymbol{Z_{is}}\right] = \beta_{ig}^{sv1}X + \boldsymbol{Z_{is}}\boldsymbol{\gamma_{isg}^{sv1'}}$ and to:

$$E\left[Z_{ig}\left(\boldsymbol{W}'Z_{ig}\right)^{-1}\boldsymbol{W}'\epsilon|X,\boldsymbol{Z_{is}}\right] = \beta_{ig}^{sv2}X + \boldsymbol{Z_{is}}\boldsymbol{\gamma_{isg}^{sv2'}}. \quad (P3)$$

The omitted-variable biases are based on β_{ig}^{ovb} measuring the conditional association between X and Z_{ig} and γ_{ig} measuring the relevance of Z_{ig} for Y. Sampling variability of β_{ig}^{ovb} is represented by β_{ig}^{sv1} and sampling variability of γ_{ig} by β_{ig}^{sv2}. We can further write:

$$E[\epsilon|X,\boldsymbol{Z_{is}}] = \beta_{is}^{sv3}X + \boldsymbol{Z_{is}}\boldsymbol{\gamma_{is}^{sv3'}}, \qquad (P4)$$

where β_i^{sv3} represents the sampling variability of estimating β. Inserting P2, P3, and P4 in P1 leads to:

$$E[Y|X,\boldsymbol{Z_{is}}] = \left(\beta + \beta_{is}^{sv3} + \sum_{g=1}^{p_s} \beta_{ig}^{ovb}\gamma_{ig} + \beta_{ig}^{sv1}\gamma_{ig} + \beta_{ig}^{sv2}\right)X$$

$$+ \boldsymbol{Z_{is}}\left(\boldsymbol{\gamma'_{is}} + \boldsymbol{\gamma_{is}^{sv3'}} + \sum_{g=1}^{p_s} \boldsymbol{\gamma_{isg}^{ovb'}}\gamma_{ig} + \boldsymbol{\gamma_{isg}^{sv1'}}\gamma_{ig} + \boldsymbol{\gamma_{isg}^{sv2'}}\right).$$

We can define $\beta_{is}^{OVB} = \sum_{g=1}^{p_s} \beta_{ig}^{ovb}\gamma_{ig}$ and $\beta_{is}^{SV} = \beta_{is}^{sv3} + \sum_{g=1}^{p_s} \beta_{ig}^{sv1}\gamma_{ig} + \beta_{ig}^{sv2}$, then:

$$\hat{\beta}_{is} = \beta + \beta_{is}^{OVB} + \hat{\beta}_{is}^{SV}.$$

If the sample size of a given primary study, m_i, increases, the probability to obtain large $\hat{\beta}_{is}^{SV}$ vanishes:

$$\begin{aligned}\operatorname*{plim}_{m_i\to\infty} \hat{\beta}_{is} &= \operatorname*{plim}_{m_i\to\infty}\left(\beta + \beta_{is}^{OVB} + \hat{\beta}_{is}^{SV}\right)\\ &= \beta + \beta_s^{OVB} + \operatorname*{plim}_{m_i\to\infty}\left(\hat{\beta}_{is}^{SV}\right)\\ &= \beta + \beta_s^{OVB}.\end{aligned}$$

If publication selection for positive and significant $\hat{\beta}_{is}$ distorts the literature, then:

$$\left[\frac{\hat{\beta}_{is}}{\widehat{se}_{is}}\middle|PS\right] > t_{0.975,m_i-q_s-1}$$

and from this follows:

$$E\left[\frac{\beta}{\widehat{se}_{is}}\middle|PS\right] + E\left[\frac{\beta_{is}^{OVB}}{\widehat{se}_{is}}\middle|PS\right] + E\left[\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}\middle|PS\right] > t_{0.975,m_i-q_s-1} \blacksquare$$

Proof of Proposition 2:

The weighted average of $\hat{\beta}_{is}$ (3) under publication selection is:

$$E[\hat{\alpha}|PS] = E\left[\frac{\sum_{i=1}^{n}\frac{\hat{\beta}_{is}}{\widehat{se}_{is}^2}}{\sum_{i=1}^{n}\frac{1}{\widehat{se}_{is}^2}} \mid PS\right]$$

$$= E\left[\frac{\sum_{i=1}^{n}\frac{\beta+\beta_{is}^{OVB}+\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}^{2}}}{\sum_{i=1}^{n}\frac{1}{\widehat{se}_{is}^{2}}} \mid PS\right]$$

$$= \beta + E\left[\frac{\sum_{i=1}^{n}\frac{\beta_{is}^{OVB}+\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}^{2}}}{\sum_{i=1}^{n}\frac{1}{\widehat{se}_{is}^{2}}} \mid PS\right]$$

$$= \beta + E\left[\frac{\sum_{i=1}^{n}\frac{\beta_{is}^{OVB}}{\widehat{se}_{is}^{2}}}{\sum_{i=1}^{n}\frac{1}{\widehat{se}_{is}^{2}}} \mid PS\right] + E\left[\frac{\sum_{i=1}^{n}\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}^{2}}}{\sum_{i=1}^{n}\frac{1}{\widehat{se}_{is}^{2}}} \mid PS\right].$$

The augmented weighted average of $\hat{\beta}_{is}$ (4) under publication selection is:

$$E[\hat{\alpha}_s|PS] = E\left[\frac{\sum_{i=1}^{n} D_s\frac{\hat{\beta}_{is}}{\widehat{se}_{is}^{2}}}{\sum_{i=1}^{n} D_s\frac{1}{\widehat{se}_{is}^{2}}} \mid PS\right]$$

$$= \beta + E\left[\frac{\sum_{i=1}^{n} D_s\frac{\beta_{is}^{OVB}}{\widehat{se}_{is}^{2}}}{\sum_{i=1}^{n} D_s\frac{1}{\widehat{se}_{is}^{2}}} \mid PS\right] + E\left[\frac{\sum_{i=1}^{n} D_s\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}^{2}}}{\sum_{i=1}^{n} D_s\frac{1}{\widehat{se}_{is}^{2}}} \mid PS\right]$$

$$= \beta + \beta_{s}^{OVB} + E\left[\frac{\sum_{i=1}^{n} D_s\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}^{2}}}{\sum_{i=1}^{n} D_s\frac{1}{\widehat{se}_{is}^{2}}} \mid PS\right]. \blacksquare$$

Proof of Proposition 3:

The estimate of $\hat{\alpha}$ in the PET model (5) under publication selection is:

$$E[\hat{\alpha}|PS] = E\left[\frac{Cov\left(\frac{\hat{\beta}_{is}}{\widehat{se}_{is}},\frac{1}{\widehat{se}_{is}}\right)}{Var\left(\frac{1}{\widehat{se}_{is}}\right)}|PS\right]$$

$$= E\left[\frac{Cov\left(\frac{\beta+\beta_{is}^{OVB}+\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}},\frac{1}{\widehat{se}_{is}}\right)}{Var\left(\frac{1}{\widehat{se}_{is}}\right)}|PS\right]$$

$$= \beta + E\left[\frac{Cov\left(\frac{\beta_{is}^{OVB} + \hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}, \frac{1}{\widehat{se}_{is}}\right)}{Var\left(\frac{1}{\widehat{se}_{is}}\right)} | PS\right].$$

The estimate of $\hat{\alpha}_s$ in the augmented PET model (6) under publication selection is:

$$E[\hat{\alpha}_s|PS] = E\left[\frac{Cov\left(D_s\frac{\hat{\beta}_{is}}{\widehat{se}_{is}}, D_s\frac{1}{\widehat{se}_{is}}\right)}{Var\left(D_s\frac{1}{\widehat{se}_{is}}\right)} | PS\right]$$

$$= \beta + E\left[\frac{Cov\left(D_s\frac{\beta_{is}^{OVB}}{\widehat{se}_{is}}, D_s\frac{1}{\widehat{se}_{is}}\right)}{Var\left(D_s\frac{1}{\widehat{se}_{is}}\right)} | PS\right] + E\left[\frac{Cov\left(D_s\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}, D_s\frac{1}{\widehat{se}_{is}}\right)}{Var\left(D_s\frac{1}{\widehat{se}_{is}}\right)} | PS\right]$$

$$= \beta + \beta_s^{OVB} + E\left[\frac{Cov\left(D_s\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}, D_s\frac{1}{\widehat{se}_{is}}\right)}{Var\left(D_s\frac{1}{\widehat{se}_{is}}\right)} | PS\right].$$

In the first case with $\beta + \beta_s^{OVB} = 0$, publication selection for positive and significant $\hat{\beta}_{is}$ can be only achieved by publication bias. For a given primary regression specification, $\widehat{se}_{is}$ decreases with the primary sample size and, thus, the probability to achieve large $\hat{\beta}_{is}^{SV}$ decreases as well. Specifically, in the presence of pure publication bias we expect:

$$E\left[D_s\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}} | PS\right] > t_{0.975, m_i - q_s - 1} \approx 2.$$

It follows:

$$Cov\left(D_s\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}, D_s\frac{1}{\widehat{se}_{is}}\right) \approx 0$$

and

$$E[\hat{\alpha}|PS] = \beta + \beta_s^{OVB}.$$

In the second case with $\beta + \beta_s^{OVB} > 0$, additional publication bias is not necessary to achieve positive and significant $\hat{\beta}_{is}$ apart from the special case that is discussed in the third case. It follows:

$$Cov\left(D_s\frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}, D_s\frac{1}{\widehat{se}_{is}}\right) \approx 0$$

and

$$E[\hat{\alpha}|PS] = \beta + \beta_s^{OVB}.$$

The third case is a special case of the second case where $\beta + \beta_s^{OVB} > 0$, but $\beta + \beta_s^{OVB}$ is very small and primary studies with low precision are present for this primary regression specification. In this special case, publication bias is required to achieve positive and significant $\hat{\beta}_{is}$ for small degrees of precision but not for larger degrees of precision. It follows:

$$Cov\left(D_s \frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}, D_s \frac{1}{\widehat{se}_{is}}\right) < 0$$

and

$$E[\hat{\alpha}_s|PS] = \beta + \beta_s^{OVB} + E\left[\frac{Cov\left(D_s \frac{\hat{\beta}_{is}^{SV}}{\widehat{se}_{is}}, D_s \frac{1}{\widehat{se}_{is}}\right)}{Var\left(D_s \frac{1}{\widehat{se}_{is}}\right)} |PS\right]. \blacksquare$$

A2: LR+ and LR- for a Genuine Effect of $\boldsymbol{\beta = 0.4}$

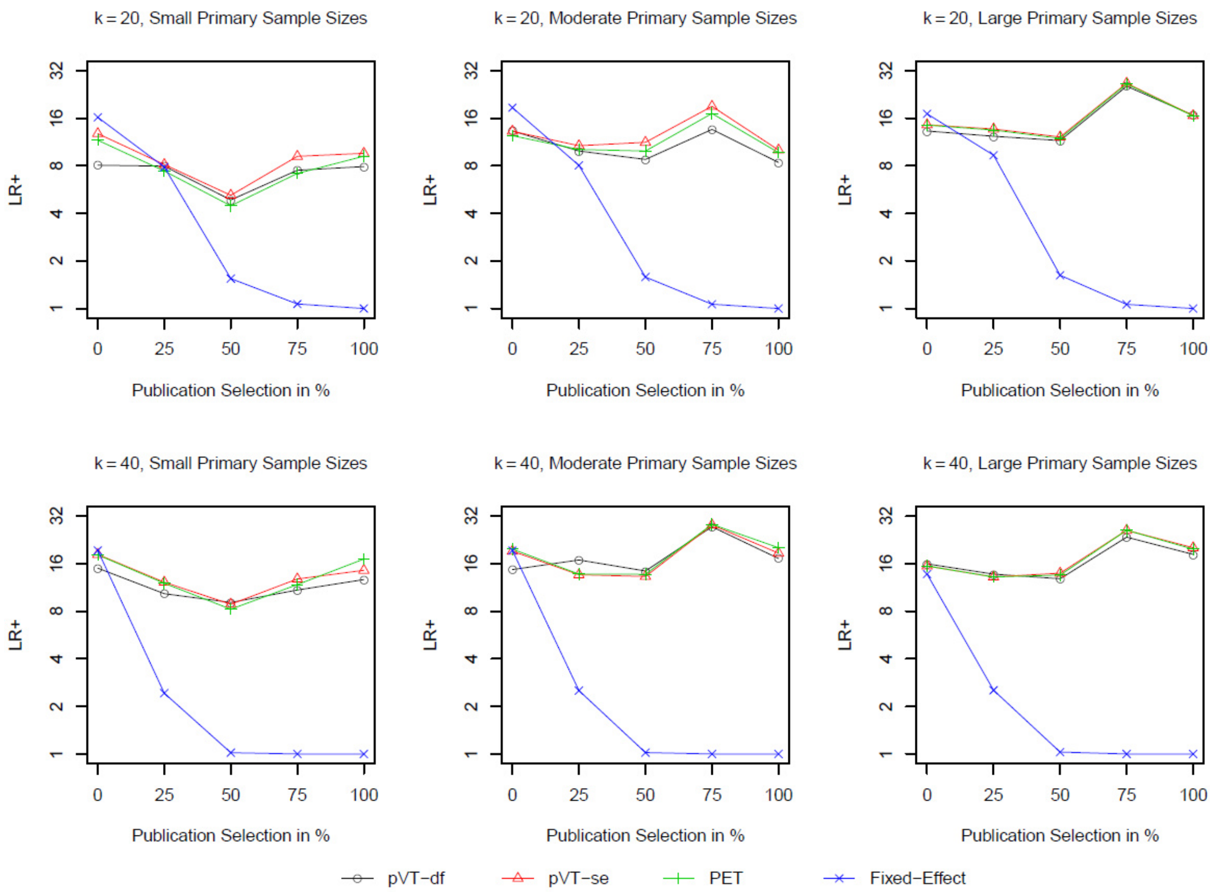

Figure A2.1: Likelihood ratio positive ($LR^{+} = \frac{power}{size}$) for $\mathrm{H}_0\colon \alpha_{\text{correct}} = 0$ with $\beta = 0.4$. LR+ is shown for small ($\mu = 60$ and $\sigma^2 = 30^2$), medium ($\mu = 120$ and $\sigma^2 = 60^2$), and large ($\mu = 480$ and $\sigma^2 = 240^2$) primary sample size distributions and very small ($k/2 = 10$) and small ($k/2 = 20$) meta-analytic sample sizes.

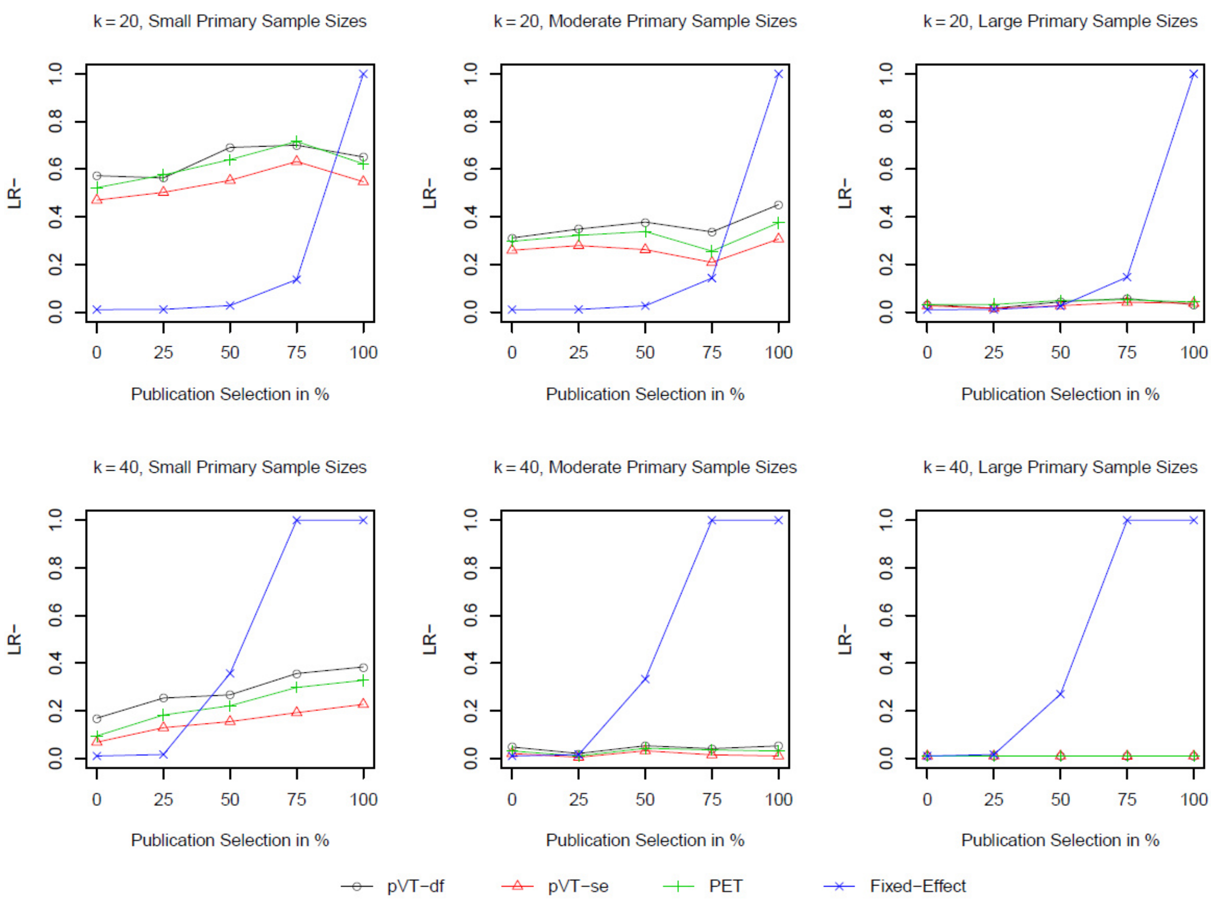

Figure A2.2: Likelihood ratio negative ($LR^- = \frac{1-power}{1-size}$) for $\mathrm{H}_0\!: \alpha_{\text{correct}} = 0$ with $\beta = 0.4$. LR- is shown for small ($\mu = 60$ and $\sigma^2 = 30^2$), medium ($\mu = 120$ and $\sigma^2 = 60^2$), and large ($\mu = 480$ and $\sigma^2 = 240^2$) primary sample size distributions and very small ($k/2 = 10$) and small ($k/2 = 20$) meta-analytic sample sizes.

A3: LR+ and LR- for a Fixed Omitted-Variable Bias of $\gamma = 0.2$

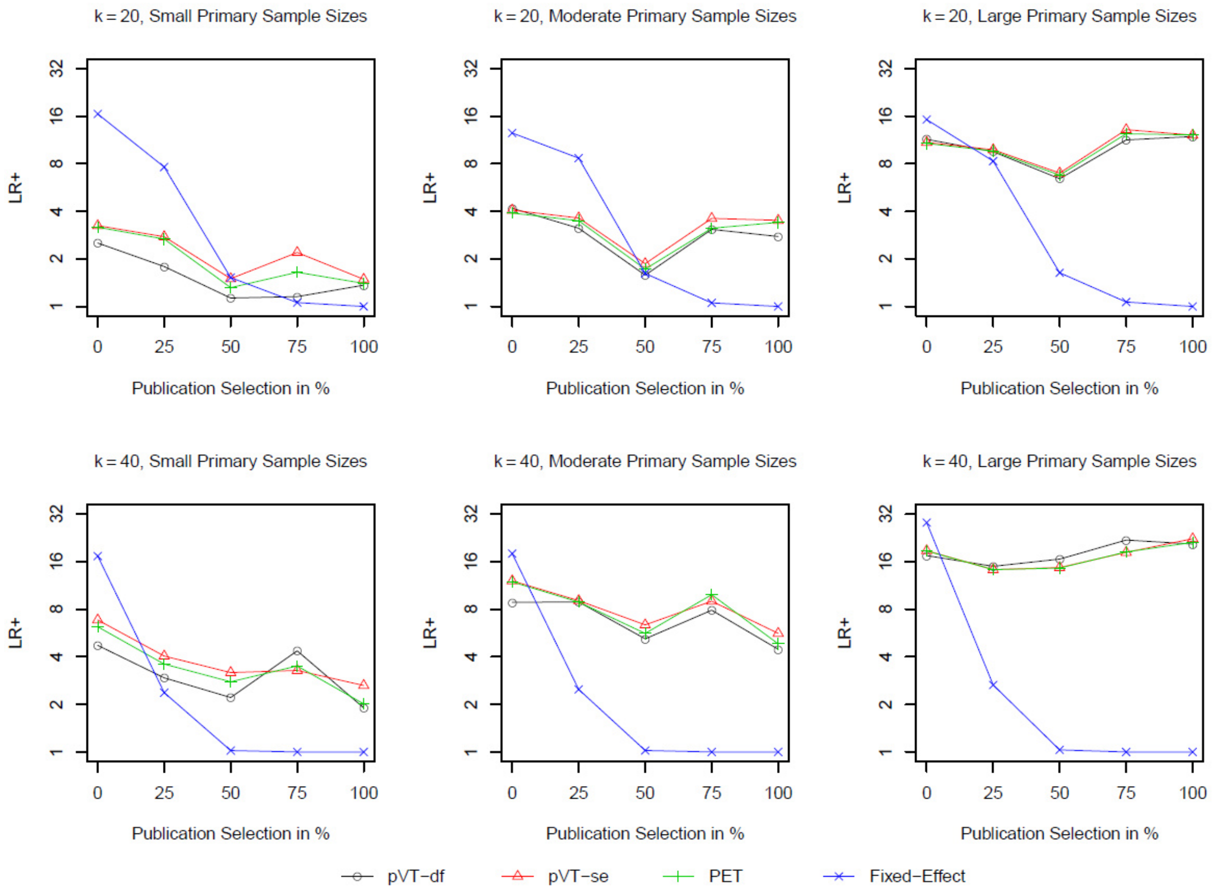

Figure A3.1: Likelihood ratio positive ($LR^+ = \frac{power}{size}$) for $\mathrm{H}_0\colon \alpha_{\mathrm{DM}} = 0$ with $\gamma = 0.2$. LR+ is shown for small ($\mu = 60$ and $\sigma^2 = 30^2$), medium ($\mu = 120$ and $\sigma^2 = 60^2$), and large ($\mu = 480$ and $\sigma^2 = 240^2$) primary sample size distributions and very small ($k/2 = 10$) and small ($k/2 = 20$) meta-analytic sample sizes.

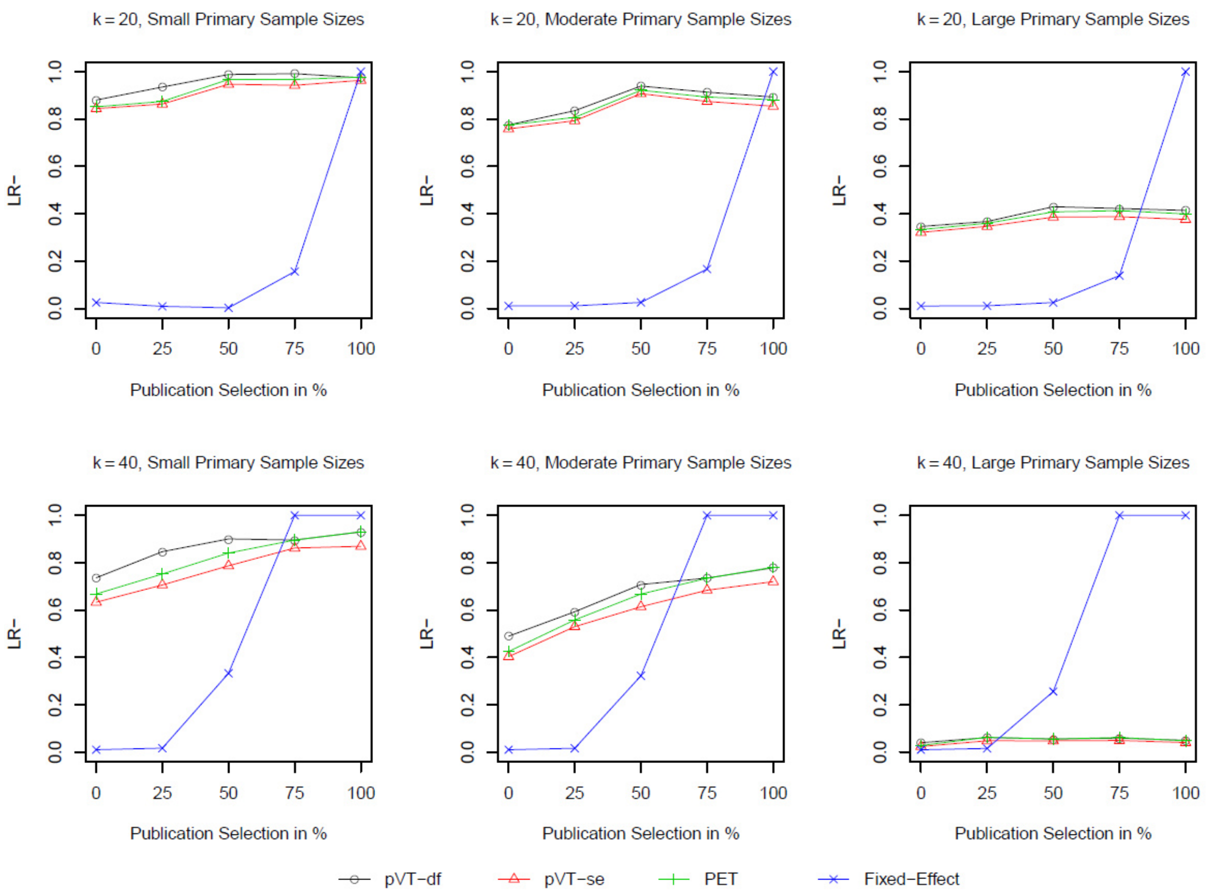

Figure A3.2: Likelihood ratio negative ($LR^- = \frac{1-power}{1-size}$) for $\mathrm{H}_0\colon \alpha_{\mathrm{DM}} = 0$ with $\gamma = 0.2$. LR- is shown for small ($\mu = 60$ and $\sigma^2 = 30^2$), medium ($\mu = 120$ and $\sigma^2 = 60^2$), and large ($\mu = 480$ and $\sigma^2 = 240^2$) primary sample size distributions and very small ($k/2 = 10$) and small ($k/2 = 20$) meta-analytic sample sizes.