

Bruns, Stephan B.

**Working Paper**

## Identifying genuine effects in observational research by means of meta-regressions

Jena Economic Research Papers, No. 2013-040

**Provided in Cooperation with:**

Max Planck Institute of Economics

*Suggested Citation:* Bruns, Stephan B. (2013) : Identifying genuine effects in observational research by means of meta-regressions, Jena Economic Research Papers, No. 2013-040, Friedrich Schiller University Jena and Max Planck Institute of Economics, Jena

This Version is available at:

<https://hdl.handle.net/10419/98446>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



# JENA ECONOMIC RESEARCH PAPERS



# 2013 – 040

## **Identifying Genuine Effects in Observational Research by Means of Meta-Regressions**

**by**

**Stephan B. Bruns**

[www.jenecon.de](http://www.jenecon.de)

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact [markus.pasche@uni-jena.de](mailto:markus.pasche@uni-jena.de).

Impressum:

Friedrich Schiller University Jena  
Carl-Zeiss-Str. 3  
D-07743 Jena  
[www.uni-jena.de](http://www.uni-jena.de)

Max Planck Institute of Economics  
Kahlaische Str. 10  
D-07745 Jena  
[www.econ.mpg.de](http://www.econ.mpg.de)

© by the author.

# IDENTIFYING GENUINE EFFECTS IN OBSERVATIONAL RESEARCH BY MEANS OF META-REGRESSIONS

Stephan B. Bruns<sup>1,2</sup>

## Abstract

Meta-regression models are increasingly utilized to integrate empirical results across studies while controlling for the potential threats of data-mining and publication bias. We propose extended meta-regression models and evaluate their performance in identifying genuine empirical effects by means of a comprehensive simulation study for various scenarios that are prevalent in empirical economics. We can show that the meta-regression models here proposed systematically outperform the prior gold standard of meta-regression analysis of regression coefficients. Most meta-regression models are robust to the presence of publication bias, but data-mining bias leads to seriously inflated type I errors and has to be addressed explicitly.

*Keywords:* Meta-regression, meta-analysis, publication bias, data mining, Monte Carlo simulation

*JEL classification:* C12, C15, C40

---

<sup>1</sup>Max Planck Institute of Economics, Kahlaische Str. 10, 07745 Jena, Germany. Phone: +49 3641 686 871, fax: +49 3641 686 868, e-mail: [s.b.bruns@gmail.com](mailto:s.b.bruns@gmail.com)

<sup>2</sup>I would like to thank Chad Baum, Alessio Moneta, David I. Stern, Tom D. Stanley and Benjamin Volland for helpful comments. I am indebted to Wolfgang Viechtbauer for suggestions and discussions regarding the substance of the paper. I am grateful to the German Science Foundation (DFG) for financial support through the program DFG-GK-1411: "The Economics of Innovative Change".

## 1 Introduction

This paper proposes extended meta-regression models and assesses their performance in identifying genuine empirical effects by means of a comprehensive simulation study for various scenarios that are prevalent in empirical economics. Meta-regression analysis aims to enhance the credibility of observational research by evaluating the authenticity and size of empirical effects while accounting for the threats of data-mining and publication bias. Glass (1976) describes meta-analysis as the "analysis of analyses" that takes stock of the increasing amount of empirical findings beyond the potentially misleading summaries of narrative surveys. In medical research, meta-analysis is widely applied to integrate distinct clinical trials to increase the certainty of treatment effects. The most basic application is the synthesis of odds ratios, each stemming from a randomized control trial, by calculating the weighted average with weights equal to the inverse of the variances of the odds ratios (Sutton et al., 2000, pp. 58).

Integrating empirical findings in economics is more challenging, however, as the majority of studies make use of observational research designs. The relation between two variables is estimated by utilizing the regression framework to control for a potentially large set of possibly confounding variables. Leamer (1983) points out that a properly randomized treatment ensures in expectations that an unbiased estimate of the effect of interest will be attained. On the contrary, observational designs suffer from the uncertainty caused by whether all possibly confounding variables are indeed accounted for by control variables included in the regression. The set of control variables can be expected to differ across empirical economic studies, resulting in methodological heterogeneity in the estimation of the effect of interest. This heterogeneity leads to a wide range of empirical results that casts doubts about the credibility of empirical economics reliant on the use of observational data (Hendry, 1980; Leamer, 1983; Sims, 1988). Improvements in the credibility of empirical economics stem primarily from the increasing use of experimental research designs (Angrist and Pischke, 2010). However, the types of economic questions that can be addressed by the use of experimental designs may be limited. This limitation is an issue for macroeconomics in particular (Stock, 2010). The purpose of this study is to complement measures that directly enhance the credibility of empirical estimates, such as experimental research designs, with meta-regression models that facilitate the integration of empirical estimates in order to identify genuine effects *ex post* and distinguish them from estimates that are only statistically significant due to biases.

Meta-regression analysis has been suggested in the seminal work by Stanley and Jarrell (1989) as an approach to explain the large heterogeneity observed in empirical economic findings. The *t*-value of a regression coefficient is regressed on study characteristics such as methodology, data sources, and variable definitions to identify the main sources of heterogeneity in the estimates of the effect of interest. Subsequently, Card and Krueger (1995) and Stanley (2001, 2008) augment meta-regression models to test for the presence of underlying genuine effects while controlling for heterogeneity and publication bias. The underlying idea is to relate the *t*-value of a regression coefficient to its precision, defined as the inverse of the standard error of the regression coefficient or approximated by degrees of freedom. If sample size increases and the estimate of the coefficient becomes more precise, the *t*-value increases in the presence of a genuine effect.

Meta-regression analysis is increasingly applied to empirical economics in order to identify the authenticity and size of empirical effects. Card and Krueger (1995) and Doucouliagos and Stanley (2009) find no evidence of an adverse employment effect of minimum wages, although both economic theory and published empirical findings support this conclusion. They find evidence of publication selection for significant and theory-confirming results which may be induced by strong theoretical presumptions about an adverse employment effect of minimum wages. A further example of meta-regression analysis is Doucouliagos et al. (2012) who provide a publication bias and heterogeneity-corrected estimate of the value of a statistical life which is 70-80% smaller than suggested by a simple average of the published literature.

We propose the use of improved meta-regression models that relate a measure of statistical significance rather than the  $t$ -value to precision or degrees of freedom. The  $t$ -value of a regression coefficient does not necessarily increase even if a genuine effect is present and sample size increases. In specific, an increasing sample size causes the variance of the  $t$ -distribution to decrease, thereby requiring smaller  $t$ -values for the same level of significance. Hence, we propose to relate the probit-transformed  $p$ -values either to precision or degrees of freedom and define these models as  $p$ -value tests (pVT). The probit-transformed  $p$ -values guarantee a constant relation to the level of significance. We evaluate the comparative performance of the pVT models to existing meta-regression models in identifying genuine effects in the presence of data-mining and publication bias by means of a comprehensive simulation study. We can show that pVT models systematically outperform all other meta-regression models. Our simulations also reveal that publication bias can be perfectly handled by most meta-regression models. However, all meta-regression models are not robust with respect to data-mining bias. This finding extends the current view on meta-regression models as being robust with respect to data-mining bias (Stanley, 2008).

Section 2 provides an overview of the existence of data-mining and publication bias in observational research and presents the meta-regression models. Section 3 presents the simulation design and the results. Section 4 discusses the implications of our findings for meta-regression analyses in economics. Section 5 concludes.

## 2 Meta-Regression Analysis of Regression Coefficients

### 2.1 Genuine Empirical Effects, Data-Mining, and Publication Bias

The majority of empirical economic research uses the regression framework to estimate conditional associations between variables that stem from observational data. This research design is characterized by a high degree of flexibility and, as a result, the corresponding range of obtainable estimates is wide (Hendry, 1980; Leamer, 1983; Sims, 1988). Sources of this flexibility include the choice of estimation techniques, functional forms, variable definitions, and, in particular, the sets of control variables included. Variations in the sets of control variables may introduce omitted-variable biases in the estimates of the effect of interest. To illustrate this, suppose a theory that states  $X$  causes  $Y$  and the corresponding data generating process (DGP) is:

$$Y = \beta_0 + \beta_1 X + \mathbf{Z}\boldsymbol{\delta}' + \epsilon, \quad (1)$$

where  $\beta_1$  is the coefficient of interest,  $\mathbf{Z} = [Z_1, \dots, Z_h]$  is a vector of  $h$  control variables with coefficients  $\boldsymbol{\delta} = [\delta_1, \dots, \delta_h]$ , and  $\epsilon \sim N(0, 1)$ . Let us define  $\mathbf{Z}_q$  as a subsample of  $q$  variables from  $\mathbf{Z}$  and  $\mathbf{Z}_p$  as the complement to  $\mathbf{Z}_q$  so that each variable of  $\mathbf{Z}$  is either in  $\mathbf{Z}_q$  or  $\mathbf{Z}_p$ . Let  $\boldsymbol{\delta}_q$  and  $\boldsymbol{\delta}_p$  be the corresponding coefficients of  $\mathbf{Z}_q$  and  $\mathbf{Z}_p$ , respectively. Consider  $i = 1, \dots, k$  independent studies estimating the following regressions:

$$Y_i = \hat{\beta}_{0i} + \hat{\beta}_{1i}X_i + \mathbf{Z}_{qi}\boldsymbol{\delta}'_{qi} + \hat{\epsilon}_i, \quad (2)$$

where  $\hat{\beta}_{0i}$ ,  $\hat{\beta}_{1i}$ , and  $\boldsymbol{\delta}'_{qi}$  are the ordinary least squares (OLS) estimates of  $\beta_0$ ,  $\beta_1$ , and  $\boldsymbol{\delta}'_{qi}$  for study  $i$ . The set of control variables  $\mathbf{Z}_{qi}$  is specific to study  $i$ . Some studies may use the same  $\mathbf{Z}_{qi}$ , but in general we observe a variety of  $\mathbf{Z}_{qi}$  across studies. Let us define  $\mathbf{Q}_i = [X_i, \mathbf{Z}_{qi}]$  so that  $\mathbf{Q}_i$  contains all variables which are present in study  $i$ . Accordingly, the OLS estimate of  $\hat{\beta}_{1i}$  becomes:

$$\hat{\beta}_{1i} = \beta_1 + (\mathbf{Q}'_i X_i)^{-1} \mathbf{Q}'_i \mathbf{Z}_{pi} \boldsymbol{\delta}'_{pi} + (\mathbf{Q}'_i X_i)^{-1} \mathbf{Q}'_i \epsilon, \quad (3)$$

where the second term of the right-hand side is well-known from illustrating omitted-variable bias (e.g. Greene, 2008, pp. 133). The omission of the relevant variables  $\mathbf{Z}_{pi}$  biases  $\hat{\beta}_{1i}$ , dependent on the conditional relation between  $X_i$  and  $\mathbf{Z}_{pi}$  as well as the relevance of  $\mathbf{Z}_{pi}$  for  $Y$ , measured by  $\boldsymbol{\delta}_{pi}$ . The expectation of the third term on the right-hand side is usually assumed to be zero as  $E[\epsilon] = 0$  is a key assumption of the regression framework. If omitted-variable biases occur randomly, e.g.  $[(\mathbf{Q}'_i X_i)^{-1} \mathbf{Q}'_i \mathbf{Z}_{pi} \boldsymbol{\delta}'_{pi}] \sim N(0, 1)$ , and  $E[\epsilon] = 0$  indeed holds, the synthesis of  $\hat{\beta}_{1i}$  across studies in order to identify genuine underlying effects is straightforward and we can learn a lot about  $\beta_1$  by simply averaging the  $\hat{\beta}_{1i}$ .

However, empirical research is characterized by a selection for statistically significant results. It has been shown that the vast majority of published  $p$ -values cluster just below the widely used significance thresholds for both the leading general-interest journals (Ridley et al., 2007) and top economics journals (Brodeur et al., 2013). Results have to be significant, but they also have to confirm the theory or hypothesis presented in the paper. Fanelli (2010) shows that the probability that a paper finds support for its hypothesis is extremely high across all research disciplines and much higher for social sciences compared to natural sciences. A reason for this might be that social sciences can select from a wider range of empirical estimates due to the flexibility of observational research designs compared to experimental research designs. The pressure to provide significant and hypothesis-confirming results is accelerated by the combination of decreasing acceptance rates in economic top journals and the need to publish in these journals to start or advance an academic career (Card and DellaVigna, 2013). As a result, Young et al. (2008) compare the publication process to the winner's curse in auction theory. The most spectacular or exaggerated results are rewarded with publication in the top journals and the scientific community rather than the author is cursed. In extreme cases, strong theoretical presumptions may force authors to search for theory-confirming results (Card and Krueger, 1995). As soon as potentially false theories become established, empirical research may be characterized by selection for results which cohere with the anticipated expectations of reviewers (Frey, 2003) rather than falsifying the false theory. The combination of flexible observational research designs and incentives to select for specific results may introduce severe distortions in empirical economic studies.

Suppose a publication selection for positive and significant  $\hat{\beta}_{1i}$ . In such a case, only a biased subset of  $\hat{\beta}_{1i}$  may be observable in the published literature, whereas a potentially large amount of  $\hat{\beta}_{1i}$  estimated in the process of conducting research remains unpublished. Accordingly, we may observe:

$$E[\hat{\beta}_{1i}|PS] = \beta_1 + E[(Q_i'X_i)^{-1}Q_i'Z_{pi}\delta_{pi}'|PS] + E[(Q_i'X_i)^{-1}Q_i'\epsilon|PS], \quad (4)$$

where PS denotes publication selection for positive and significant  $\hat{\beta}_{1i}$ . We can expect that  $E[(Q_i'X_i)^{-1}Q_i'Z_{pi}\delta_{pi}'|PS] > 0$  if authors vary the set of control variables until positive and significant  $\hat{\beta}_{1i}$  are obtained. We denote the search for positive and significant  $\hat{\beta}_{1i}$  by varying the set of control variables as data-mining bias. Such a search process need not even necessarily be a conscious manipulation of the estimate of the effect of interest, but might result from a naive and unconscious experimentation with the data or as a consequence of facing limitations on data availability. Leamer (1983) and Leamer and Leonard (1983) highlight the importance of this data-mining bias as the key source of the low credibility of observational research.<sup>1</sup> They also illustrate that  $E[(Q_i'X_i)^{-1}Q_i'Z_{pi}\delta_{pi}'] = 0$  holds for randomized experiments, given that randomization ensures that  $X$  is uncorrelated with  $Z$ . As a result, randomized experiments offer, at least asymptotically, a solution to data-mining bias by providing an unbiased  $\hat{\beta}_{1i}$ .

A further source of bias in the published  $\hat{\beta}_{1i}$  is the selection of positive and significant  $\hat{\beta}_{1i}$  from those results provided by sampling variability. In an extreme case, we may observe only those 2.5% of  $\hat{\beta}_{1i}$  that are positive and significant by chance, whereas the 95% insignificant  $\hat{\beta}_{1i}$  as well as the 2.5% negative and significant  $\hat{\beta}_{1i}$  remain in the file-drawer (Rosenthal, 1979). This bias is mirrored by  $E[(Q_i'X_i)^{-1}Q_i'\epsilon|PS] > 0$  in (4) and is usually referred to as publication bias. Given that data-mining bias is addressed by randomization, experimental designs may primarily be forced to focus on publication bias in order to select for significant results. In empirical economics, however, publication bias is likely to be dominated by data-mining bias, as the observational research design allows authors to search for positive and significant  $\hat{\beta}_{1i}$  by varying assumptions of the estimated model and not necessarily by estimating the same model in various samples.

We evaluate the performance of meta-regression models in identifying genuine effects with regards to three main cases in which empirical findings are distorted by selection for positive and significant  $\hat{\beta}_{1i}$ . In the following, we denote those regressions utilized by studies that conduct the original estimation as primary regressions, whereas meta-regressions synthesize the findings of primary regressions ex post. First, we consider a case in which the literature estimates the correct primary regression, but is distorted by the existence of different degrees of publication bias. Publication bias selects the positive and significant results among those results offered by sampling variability. This bias vanishes with sample size, due to the increasing precision of  $\hat{\beta}_{1i}$  and, thereby, decreasing probability of obtaining large values of  $\hat{\beta}_{1i}$  by chance. This pattern helps the meta-regression models to account for this bias. Second, we consider a case with data-mining

<sup>1</sup>It should be noted that data-mining, understood as general-to-specific modeling without a particular variable of interest, can reveal the true DGP if correctly applied (Hendry and Krolzig, 2005; Hoover and Perez, 1999). Our definition of data-mining bias considers the misuse of data-mining.

bias in which authors search for positive and significant  $\hat{\beta}_{1i}$  by using various omitted-variable biases. The size of the omitted-variable bias that is required to obtain a positive and significant  $\hat{\beta}_{1i}$  vanishes as sample size increases. If authors desire a positive  $\hat{\beta}_{1i}$  that is simply significant at the 5% level and they want to use a bias that is as small as possible, larger biases will appear in small samples and smaller biases in larger samples. In this second case, the actual size of the effect is constantly decreasing with sample size and the emphasis is on statistical significance of  $\hat{\beta}_{1i}$ . Third, we consider a case in which data-mining bias is present, though the probability of observing large biases does not vanish with sample size. This case establishes how particular regression specifications might become prevalent in the literature and are applied irrespective of sample sizes. In this third case, the size of  $\hat{\beta}_{1i}$  is not decreasing with sample size and can remain constant.

In summary, we evaluate the performance of various meta-regression models in identifying genuine empirical effects for these three cases, each reflecting a different example of the conscious and unconscious strategies available to obtain desired estimates of the effect of interest.

## 2.2 Fixed-Effects Model

The unadjusted fixed-effects model serves as the benchmark for our analysis:

$$\hat{\beta}_{1i} = \alpha + u_i, \quad (5)$$

estimated by weighted least squares (WLS) where the weights are equal to the inverse variance of  $\hat{\beta}_{1i}$ . This weighting procedure gives smaller weights to imprecisely estimated  $\hat{\beta}_{1i}$  and larger weights to more precisely estimated  $\hat{\beta}_{1i}$  (e.g. Sutton et al., 2000, pp. 58).  $H_0 : \alpha = 0$  tests for a non-zero weighted mean of  $\hat{\beta}_{1i}$ . The fixed-effects model does not control for publication bias. In other words, if authors select for positive and significant  $\hat{\beta}_{1i}$  among those results provided by sampling variability, the weighted mean becomes biased.

The fixed-effects model as well as the following meta-regression models can control for heterogeneity in  $\hat{\beta}_{1i}$  by introducing control variables in the meta-regression. We may observe genuine heterogeneity in terms of different  $\beta_1$  for different countries, firms, households, or across time and methodological heterogeneity may be present in the estimates due to different choices in the process of conducting research, specifically regarding estimation techniques, variable definitions, data sources, and the sets of primary control variables included. Our simulations focus on methodological heterogeneity in terms of primary control variables that vary across studies as this source of heterogeneity is of particular concern for the credibility of observational research in economics (Leamer, 1983; Leamer and Leonard, 1983). We aim to evaluate the robustness of meta-regression models with respect to this data-mining bias, so that our simulations do not consider meta-regressions that explicitly control for varying sets of primary control variables, e.g. by including dummy variables in the meta-regression.

An alternative to the fixed-effects model is the random-effects model which models the under-



lying effect to be drawn from a random distribution:

$$\hat{\beta}_{1i} = \alpha + \mu_i + u_i, \quad (6)$$

again estimated by WLS where the weights are equal to the inverse variance of  $\hat{\beta}_{1i}$  and  $\mu_i \sim N(0, v)$  is the random effect. A fundamental limitation of this model is the assumption of strict exogeneity, i.e.  $E[u_i|C] = 0$ , where  $C$  is an arbitrary control variable in the meta-regression (Wooldridge, 2006, pp. 493). This assumption may hold in the meta-analysis of randomized experiments, but the meta-analysis of observational research designs has to include a large amount of control variables to adequately control for methodological and genuine heterogeneity. However, it is very likely that some sources of heterogeneity are not perfectly accounted for by the control variables included in the meta-regression. This disregarded heterogeneity will enter into the random effect and is likely to be correlated with the control variables present in the meta-regression. As a result, the assumption of strict exogeneity is violated. In actual fact, this assumption is basically never met in observational data and, for this reason, it is common to pair random-effects models with fixed-effects models to justify their use (Wooldridge, 2006, pp. 497). Consequently, we focus our analysis on the fixed-effects model.

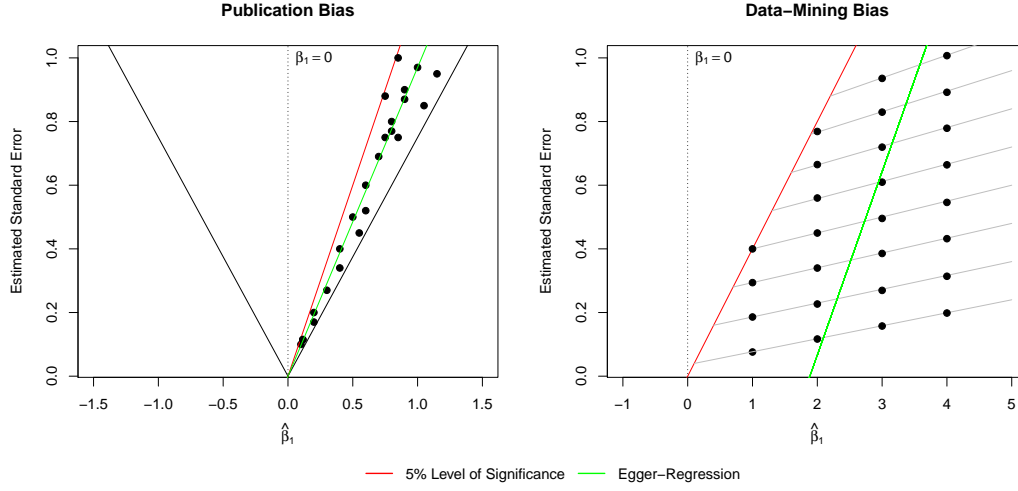
Disregarded heterogeneity is additionally problematic given the potential introduction of heteroskedasticity. We address this issue by the use of heteroskedasticity-consistent standard errors. As meta-analysis is often confronted by limitations of sample size, we utilize the standard errors of Long and Ervin (2000) which are designed for use with small samples rather than the more well-known Huber-White errors (Huber, 1967; White, 1980) which provide inflated type I errors for small samples.

### 2.3 Precision-Effect Test (PET)

Publication selection is a sample-selection rule and if information about the non-published results were available, the two-step Heckman (1979) approach could be applied (Stanley and Doucouliagos, 2007). Unfortunately, we do not have any information about the non-published results. In fact, thousands of estimations might be conducted before those results are obtained that are presented in the literature. However, there are patterns that can help control for publication selection. If the primary regression is correctly specified,  $\hat{\beta}_{1i}$  is an unbiased and consistent estimate of  $\beta_1$ . We can expect that  $\hat{\beta}_{1i}$  is a precise estimate of  $\beta_1$  if the sample size is large and  $\hat{se}_{1i}$  small. On the contrary, if the sample size is small and  $\hat{se}_{1i}$  large, sampling variability yields a wide range of  $\hat{\beta}_{1i}$  with substantial uncertainty over the size of the genuine effect. Publication bias towards positive and significant  $\hat{\beta}_{1i}$  requires larger  $\hat{\beta}_{1i}$  as  $\hat{se}_{1i}$  increases. Consequently, we observe an association between  $\hat{\beta}_{1i}$  and  $\hat{se}_{1i}$  in the presence of publication bias and  $\hat{\beta}_{1i}$  becomes an increasingly precise estimate of  $\beta_1$  as  $\hat{se}_{1i}$  decreases. These patterns can help us to identify the genuine effect, even in the extreme case where all observable  $\hat{\beta}_{1i}$  are positive and significant, but  $\beta_1$  is actually zero.

We may also expect an association between  $\hat{\beta}_{1i}$  and  $\hat{se}_{1i}$  in the presence of data-mining bias. First, large  $\hat{se}_{1i}$  require large omitted-variable biases to obtain positive and significant  $\hat{\beta}_{1i}$ . Second, keeping sample size constant,  $\hat{se}_{1i}$  increases with the size of the omitted-variable bias. Therefore,

Figure 1: Illustration of Data-Mining and Publication Bias



*Notes:* The left graph shows the range of results that can be obtained by sampling variability (black lines). All  $\hat{\beta}_{1i}$  between the red line and the black line on the right represent estimates that are positive and statistically significant and can be obtained by publication bias. The use of the Egger-regression allows us to take the systematic relation between the estimated standard error ( $\hat{se}_{1i}$ ) and  $\hat{\beta}_{1i}$  into account. The intersection of the x-axis and the Egger-regression thus represents the publication bias-corrected estimate of  $\hat{\beta}_{1i}$  that correctly suggests  $\beta_1 = 0$ . A simple average of the  $\hat{\beta}_{1i}$  would suggest, however, that  $\beta_1 > 0$ . The right graph illustrates the presence of data-mining bias. Holding sample size constant, the grey lines represent different values of the biased  $\hat{\beta}_{1i}$  if the omitted-variable bias is varied from 1 to 4. These lines have a positive slope given that  $\hat{se}_{1i}$  increases with the size of the omitted-variable bias when the sample size is constant. This positive slope and the fact that larger  $\hat{se}_{1i}$  require larger  $\hat{\beta}_{1i}$  to obtain positive and significant  $\hat{\beta}_{1i}$  makes it likely that  $\hat{se}_{1i}$  and  $\hat{\beta}_{1i}$  are also associated in the presence of data-mining bias. Correcting for this association can help to reduce the impact of data-mining bias, but still does not completely resolve this issue *per se*, as  $\hat{\beta}_{1i}$  does not converge to  $\beta_1$  in the presence of omitted-variable biases. In the presented case, the Egger-regression would falsely point to  $\beta_1 = 1.9$ .

we can expect larger omitted-variable biases to be associated with larger  $\hat{se}_{1i}$ . However,  $\hat{se}_{1i}$  decreases with sample size irrespective of the size of the omitted-variable bias, causing  $\hat{\beta}_{1i}$  to estimate the biased  $\beta_1$  with increasing precision. Hence, we can only expect that an association between  $\hat{\beta}_{1i}$  and  $\hat{se}_{1i}$  fully accounts for the presence of data-mining bias in the unlikely event that the omitted-variable bias is constantly decreasing with sample size, such that  $\hat{\beta}_{1i}$  becomes an increasingly precise estimate of  $\beta_1$  as  $\hat{se}_{1i}$  decreases.

The fixed-effects model can be augmented by  $\hat{se}_{1i}$  in order to account for publication bias and partially for data-mining bias by controlling for the association between  $\hat{\beta}_{1i}$  and  $\hat{se}_{1i}$ :

$$\hat{\beta}_{1i} = \alpha + \gamma \hat{se}_{1i} + u_i, \quad (7)$$

again estimated by WLS with weights being equal to the inverse variance of  $\hat{\beta}_{1i}$ . This regression is known as the Egger-regression in medical research (Egger et al., 1997).  $H_0 : \alpha = 0$  tests for the presence of a genuine effect corrected for publication bias and partially corrected for data-mining bias. See Figure 1 for an illustration of this test. However, this form of the Egger-regression is heteroskedastic. Therefore, dividing by  $\hat{se}_{1i}$  leads to:

$$t_{1i} = \gamma + \alpha \frac{1}{\hat{se}_{1i}} + u_i, \quad (8)$$

where  $t_{1i}$  is the  $t$ -value of  $\hat{\beta}_{1i}$ . Stanley (2001, 2008) suggest  $H_0 : \alpha = 0$  to test for a genuine effect in the presence of publication and data-mining bias, referring to this test as the Precision-Effect Test (PET). If a genuine effect is present,  $t_{1i}$  follows a non-central  $t$ -distribution with a non-centrality parameter that is an increasing function of the degrees of freedom or precision. In contrast to this,  $t_{1i}$  should be unrelated to precision in the absence of a genuine effect. PET can handle publication bias and partially data-mining bias, as  $t_{1i}$  is unrelated to  $\frac{1}{\hat{se}_{1i}}$  if  $\hat{\beta}_{1i}$  is constantly decreasing with  $\hat{se}_{1i}$ , so that  $\frac{\hat{\beta}_{1i}}{\hat{se}_{1i}} = t_{1i}$  remains constant.

The Egger-regression and PET are algebraically equivalent and may only differ if we base our inference on the assumption of homoskedastic standard errors for PET and heteroskedasticity-consistent standard errors for the Egger-regression. However, we use heteroskedasticity-consistent standard errors for all meta-regression models to address the likely presence of disregarded heterogeneity in the data. As a result, the remainder of the paper exclusively refers to PET given that this version is more common in the economic literature.

PET is known to be biased due to the use of the estimate of  $se_{1i}$  as an independent variable in the meta-regression. This bias is particularly notable for estimates obtained from small sample sizes (Macaskill et al., 2001). Furthermore, the use of  $\hat{se}_{1i}$  in the meta-regression requires  $X$  and  $Y$  to be measured in the same units across all studies (Becker and Wu, 2007).

## 2.4 Meta-Significance Testing (MST)

Meta-Significance Testing (MST) is proposed by Card and Krueger (1995) and popularized by Stanley (2001, 2005a, 2008). MST is basically a modification of PET as precision is substituted by  $\sqrt{df}$  where  $df$  are the degrees of freedom. If a genuine effect exists,  $t_{1i}$  should increase with  $\sqrt{df_i}$  as the precision of  $\hat{\beta}_{1i}$  increases. However, MST uses a log-log transformation which leads to some undesirable properties:

$$\ln(|t_{1i}|) = \gamma + \alpha \ln(df_i) + u_i, \quad (9)$$

where  $H_0 : \alpha = 0$  tests for the presence of a genuine effect. MST can be upwardly biased as taking the magnitude of  $t_{1i}$  does not allow to distinguish between positive and negative  $t_{1i}$  (Stanley, 2008). In addition to that, the use of log transformations introduce heteroskedasticity and asymmetrically distributed  $u_i$ .

## 2.5 $p$ -Value Tests (pVT-df and pVT-se)

Prior models use the relationship between  $t$ -value and precision or  $df$  to identify genuine effects. However, the variance of the  $t$ -distribution does not remain constant, but instead converges to the variance of the standard normal distribution as  $df$  increase. Hence the size of the  $t$ -value is not constantly linked to the level of significance that we are actually interested to measure. In specific, attaining the same  $p$ -value requires larger  $t$ -values for smaller  $df$  than compared to larger  $df$ .<sup>2</sup> At

<sup>2</sup>For the MST model, focusing on the magnitude of the  $t$ -value, this implies that we should expect  $\alpha < 0$  rather than  $\alpha = 0$  if no genuine effect is present. This is indeed a frequently reported result for the MST model (e.g. Card and

the extreme, we may observe the same  $t$ -values across a range of  $df$ , a fact which prior meta-regression models have tended to interpret as the absence of a genuine effect. However, constant  $t$ -values across a range of  $df$  imply an increasing level of significance, actually suggesting the existence of a genuine effect. This issue can be resolved by using probit-transformed  $p$ -values as the dependent variable in the meta-regression:

$$z_{1i} = \begin{cases} \Phi^{-1}(\frac{p_{1i}}{2}), & \text{if } \hat{\beta}_{1i} < 0, \\ \Phi^{-1}(1 - \frac{p_{1i}}{2}), & \text{if } \hat{\beta}_{1i} \geq 0, \end{cases} \quad (10)$$

where  $p_{1i}$  is the  $p$ -value of a two sided  $t$ -test of  $\hat{\beta}_{1i}$  and  $\Phi^{-1}$  represents the probit-transformation. The new dependent variable of the meta-regression model,  $z_{1i}$ , is now the value of a standard normal distribution with the same  $p$ -value as the original  $t$ -value. Hence, the level of significance is constantly related to the value of  $z_{1i}$ , independent from  $df$  or precision.<sup>3</sup> We can test the relation between  $z_{1i}$  and  $\sqrt{df_i}$  directly rather than using a log-log specification:

$$z_{1i} = \gamma + \alpha \sqrt{df_i} + u_i \quad (11)$$

with  $H_0 : \alpha = 0$  as a test for the presence of a genuine empirical effect. We denote this test as "p-Value Test with  $df$ " (pVT-df).<sup>4</sup> We also consider a modification of PET in which we replace the  $t$ -value with the corresponding  $z_{1i}$ -value in equation (8):

$$z_{1i} = \gamma + \alpha \frac{1}{\hat{se}_{1i}} + u_i. \quad (12)$$

We denote  $H_0 : \alpha = 0$  as "p-Value Test with Precision" (pVT-se). Table 1 provides an overview of the considered meta-regression models and the corresponding tests for the presence of a genuine effect.

Table 1: Overview of Meta-Regression Models

| Method              | Test Regression  | Test of Genuine Effect |
|---------------------|--|------------------------|
| Fixed-Effects Model | $\hat{\beta}_{1i} = \alpha + \epsilon_i$ weighted by $\frac{1}{\hat{se}_{1i}^2}$ | $H_0 : \alpha = 0$     |
| PET                 | $t_{1i} = \gamma + \alpha \frac{1}{\hat{se}_{1i}} + u_i$                         | $H_0 : \alpha = 0$     |
| MST                 | $\ln( t_{1i} ) = \gamma + \alpha \ln(df_i) + u_i$                                | $H_0 : \alpha = 0$     |
| pVT-df              | $z_{1i} = \gamma + \alpha \sqrt{df_i} + u_i$                                     | $H_0 : \alpha = 0$     |
| pVT-se              | $z_{1i} = \gamma + \alpha \frac{1}{\hat{se}_{1i}} + u_i$                         | $H_0 : \alpha = 0$     |

Krueger (1995); Doucouliagos and Paldam (2008)).

<sup>3</sup>Transformations to yield a normal distribution are also suggested to attain a common metric whenever the test statistics are differently distributed (Stanley, 2001, 2005b).

<sup>4</sup>Bruns et al. (2013) and Bruns and Stern (2013) extend the pVT-df model to the meta-regression analysis of Granger-causality tests that are  $F$ - and  $\chi^2$ -distributed through the application of an asymmetric probit-transformation.

### 3 Simulation

#### 3.1 Design

We simulate empirical economic literatures that are affected by publication and data-mining bias in order to evaluate the performance of meta-regression models in identifying genuine empirical effects. We consider the sample sizes of the meta-regression analyses as  $k = 20, 40, 80, 160, 320$ . These values mirror the usual sample sizes of meta-analyses in economics (e.g. Doucouliagos and Stanley, 2011). The primary study sample size of the  $i$ th study is drawn from a gamma distribution with scale parameter equal to  $\sigma^2/(\mu - 10)$  and shape parameter equal to  $(\mu - 10)^2/\sigma^2$ . Thus,  $\mu$  denotes the mean of the primary study sample size distribution and  $\sigma^2$  its variance. We round the obtained value for the primary study sample size to the next integer and add 10 so that 10 is the smallest primary study sample size. The choice of the scale and shape parameters allows us to vary  $\mu$  and  $\sigma^2$  independently. The use of a gamma distribution provides right-skewed primary study sample size distributions for small  $\mu$  and increasingly symmetric ones for larger  $\mu$ , which is what we can observe in empirical economics. We consider  $\mu = 30, 60, 120, 240, 480$  and  $\sigma^2 = 15^2, 30^2, 60^2, 120^2, 240^2$ . Smaller  $\mu$  may be encountered in macroeconomic time-series analysis, whereas larger  $\mu$  may be encountered in microeconomic panel-data analysis.

We evaluate the performance of the various meta-regression models for the three cases discussed in Section 2.1. First, we consider empirical literatures that are distorted by pure publication bias. We generate the observations of each study  $i$  by:

$$Y_i = \beta_1 X_i + \epsilon_i, \quad (13)$$

where  $X_i \sim N(0, 1)$  and  $\epsilon_i \sim N(0, 1)$ . The coefficient of interest is  $\beta_1$  and we consider  $\beta_1 = 0, 0.2, 0.4, 0.8$  to evaluate size and power for various strengths of a genuine effect. Each study  $i$  estimates:

$$Y_i = \hat{\beta}_{1i} X_i + \hat{\epsilon}_i. \quad (14)$$

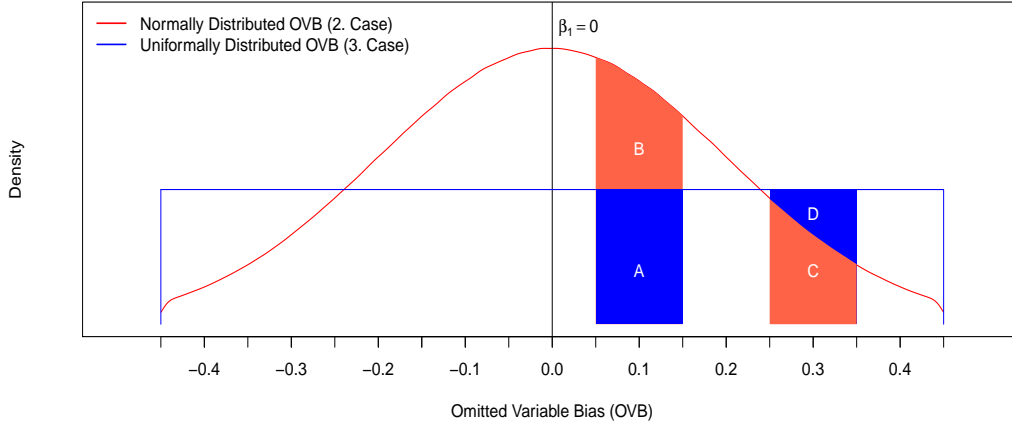
We model publication selection for positive and significant  $\hat{\beta}_1$  by resampling (13) until the first  $c\%$  of primary studies achieve positive and significant  $\hat{\beta}_{1i}$ , where  $c = 25, 50, 75, 100$ . The remaining  $(100 - c)\%$  of the primary studies are not affected by publication bias.

Regarding the second case, we evaluate the performance of the meta-regression models in the presence of data-mining bias, defined as the utilization of omitted-variable biases to obtain a positive and significant  $\hat{\beta}_{1i}$ . The observations of each primary study  $i$  are generated by:

$$Y_i = \beta_1 X_i + \beta_{2i} Z_i + \epsilon_i, \quad (15)$$

where  $X_i \sim N(0, 1)$ ,  $Z_i \sim N(0, 1)$ ,  $\epsilon_i \sim N(0, 1)$ , and  $Z_i = X_i + v_i$  with  $v_i \sim N(0, 1)$ . The coefficient of interest is again  $\beta_1$  and we again consider  $\beta_1 = 0, 0.2, 0.4, 0.8$ .  $\beta_{2i}$  represents the data-mining bias which we model to vanish as sample size increases. For this purpose, we draw  $\beta_{2i}$  from  $N(0, 0.2^2)$ . We truncate the normal distribution to  $|\beta_{2i}| < 0.45$  and round the obtained value to the next decimal place after the decimal point, so that we obtain

Figure 2: Illustration of the Weighting Schemes



*Notes:* The probabilities of using an omitted-variable bias of either 0.1 or 0.3 are exemplary displayed. The probability of obtaining a small bias of 0.1 is considerably larger ( $A + B$ ) compared to the probability of obtaining a large bias of 0.3 ( $C$ ) for the normally distributed omitted-variable bias. On the contrary, the probability of obtaining a bias of 0.1 equals to the probability of obtaining a bias of 0.3 for the uniformly distributed omitted-variable bias, which is mirrored by the area  $A$  and  $C + D$ , respectively.

$\beta_{2i} = -0.4, -0.2, -0.1, 0, 0.1, 0.2, 0.4$ . These values mirror seven different primary regression specifications that are symmetric around zero. Specifications with smaller biases, however, have a higher probability to be given emphasis by the authors. See Figure 2 for an illustration of the weighting scheme. Each study  $i$  again estimates (14) and  $\hat{\beta}_{1i}$  can be biased in these regressions due to  $\beta_{2i} \neq 0$ . We model publication selection for positive and significant  $\hat{\beta}_{1i}$  by generating new  $Y_i$  in (15) if  $\hat{\beta}_{1i}$  is not positive and significant. For this purpose, we draw a new  $\beta_{2i}$  while holding  $X_i$  and  $\epsilon_i$  constant. This procedure models the systematic utilization of various strengths of omitted-variables biases to obtain positive and significant  $\hat{\beta}_{1i}$ . We draw 100 times new values of  $\beta_{2i}$  to ensure that each bias is likely to be tested by the author. If a positive and significant  $\hat{\beta}_{1i}$  is not obtained, the results remain unpublished. This is the publication bias component of data-mining bias. The outcome of this second case is the presence of larger biases in small samples, whereas smaller biases are more likely to appear in larger samples.

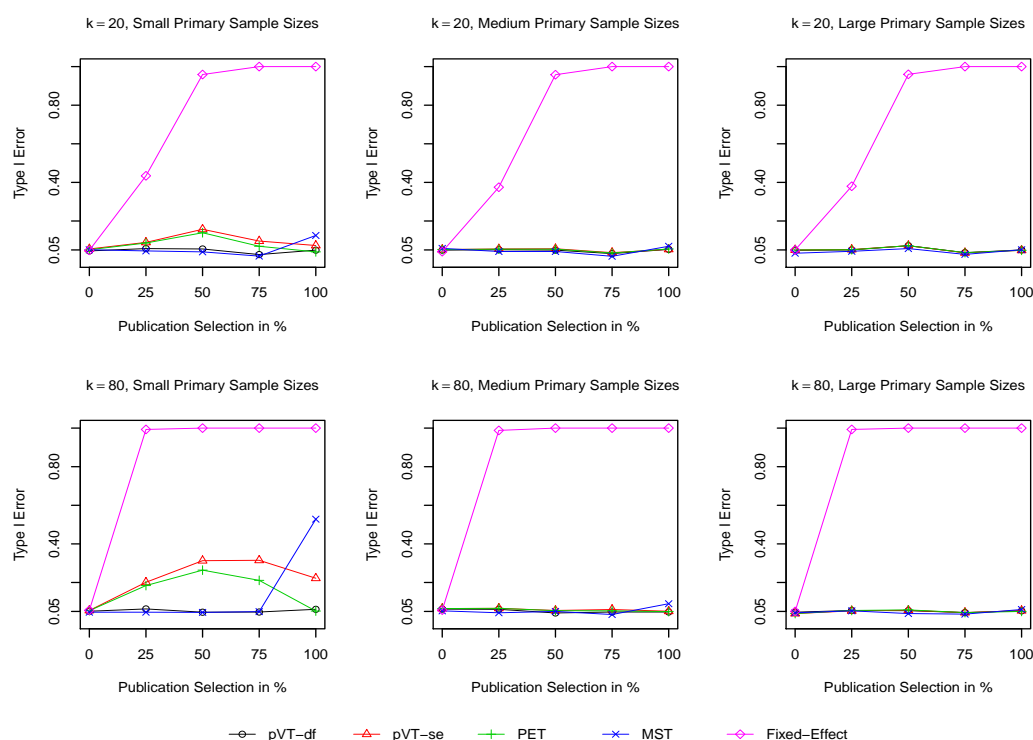
Finally, regarding the third case, we consider data-mining bias that does not vanish as sample size increases. For this purpose, we draw  $\beta_{2i}$  from a uniform distribution in the interval  $[-0.45 < \beta_{2i} < 0.45]$  and round the obtained value again to the next decimal place after the decimal point, such that the range of omitted-variable biases is identical to the previous case, but large biases are now as likely to occur as small biases. See again Figure 2 for an illustration of the weighting scheme.

### 3.2 Results

For the first case of pure publication bias, the relation between the type I errors of  $H_0 : \alpha = 0$  and the degree of publication selection is illustrated in Figure 3. The fixed-effects model does not

control for publication bias and, as a result, provides highly inflated type I errors that are increasing with the degree of publication selection. Those meta-regression models that use precision as an independent variable (pVT-se and PET) show inflated type I errors in small primary sample size distributions. This outcome is due to an imprecise estimation of  $se_{1i}$  in small samples. Regarding medium and large primary sample size distributions, pVT-se and PET provide accurate type I errors close to the nominal level of 0.05. pVT-df provides type I errors close to the nominal level across all scenarios and MST suffers from inflated type I errors only in the presence of small sample size distributions and a publication selection of 100%.

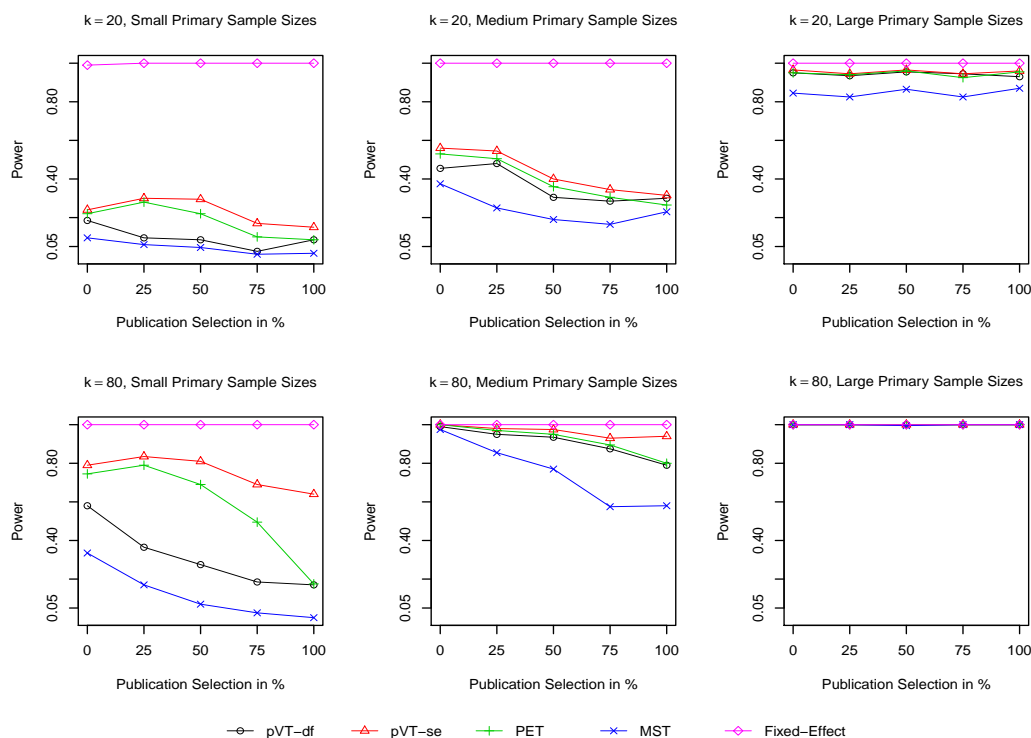
Figure 3: Type I Errors for First Case of Pure Publication Bias



Notes: Type I errors of  $H_0 : \alpha = 0$  (Genuine effect of  $\beta_1 = 0$ ) for first case of pure publication bias. The graphs illustrate the relation between type I errors and the degree of publication selection for both very small meta-analyses ( $k = 20$ ) and moderate meta-analyses ( $k = 80$ ) in combination with a small primary sample size distribution ( $\mu = 30$  and  $\sigma = 15$ ), a medium primary sample size distribution ( $\mu = 120$  and  $\sigma = 60$ ), and a large primary sample size distribution ( $\mu = 480$  and  $\sigma = 240$ ).

The relation between the power of  $H_0 : \alpha = 0$  and the degree of publication selection for the first case of pure publication bias is illustrated in Figure 4. The fixed-effects model has a power close to one across all scenarios. However, this meta-regression model also suffers from highly inflated type I errors, nearly always rejecting  $H_0 : \alpha = 0$  in the presence of publication selection irrespective of whether a genuine effect exists or not. Accordingly, this model is not helpful in identifying genuine effects. The power of pVT-se and PET can be reasonably interpreted for medium and large primary sample size distributions given that these models provide adequate type I errors in these scenarios. pVT-se can systematically outperform PET, while pVT-df systematically outperforms MST across all scenarios. Comparing pVT-se and pVT-df, the power of pVT-se is greater for every scenario, but pVT-df provides adequate type I errors even in small primary

Figure 4: Power for First Case of Pure Publication Bias



Notes: Power of  $H_0 : \alpha = 0$  (Genuine effect of  $\beta_1 = 0.2$ ) for first case of pure publication bias. The graphs illustrate the relation between type I errors and the degree of publication selection for both very small meta-analyses ( $k = 20$ ) and moderate meta-analyses ( $k = 80$ ) in combination with a small primary sample size distribution ( $\mu = 30$  and  $\sigma = 15$ ), a medium primary sample size distribution ( $\mu = 120$  and  $\sigma = 60$ ), and a large primary sample size distribution ( $\mu = 480$  and  $\sigma = 240$ ).

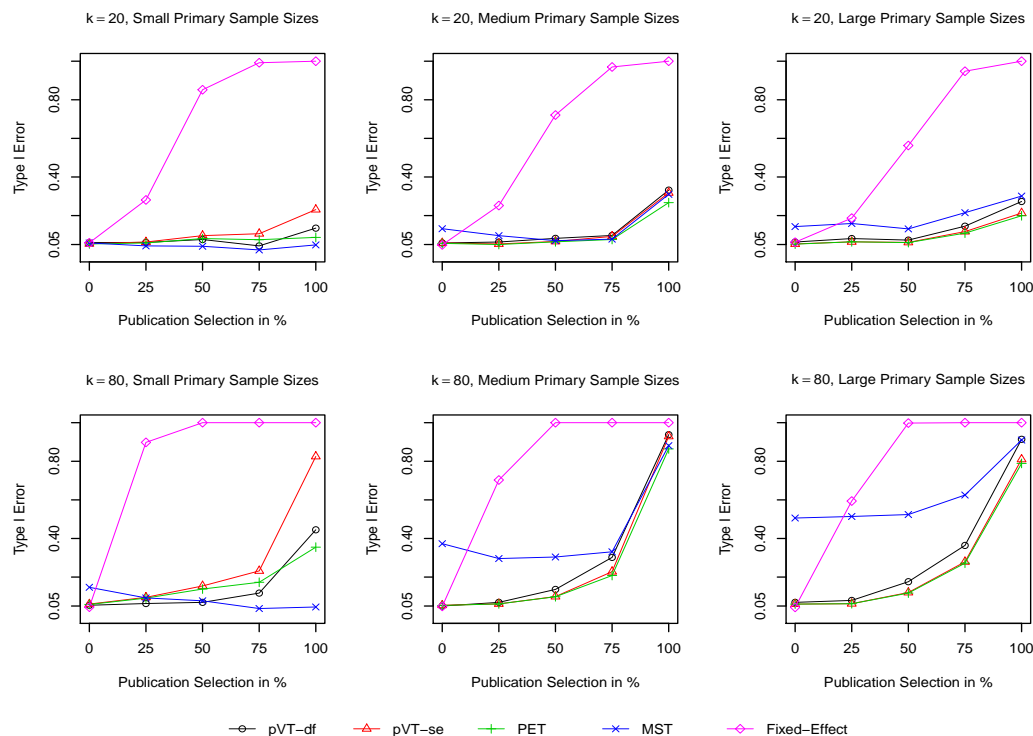
sample size distributions.

In general, the power increases along with the size of the genuine effect ( $\beta_1$ ) and reaches often one even if the genuine effect is small (See appendix for graph). The power also increases with the variance of the primary sample size distribution ( $\sigma$ ) as the variance of the independent variable in the meta-regression model is increased. Furthermore, power decreases with the mean of the primary sample size distribution ( $\mu$ ) as  $df_i$  and  $\hat{se}_{1i}$  are related to  $z_{1i}$  and  $t_{1i}$  in functional forms which imply that changes in  $df_i$  and  $\hat{se}_{1i}$  have a higher impact on  $z_{1i}$  and  $t_{1i}$  for smaller values of  $df_i$  and  $\hat{se}_{1i}$ . Hence, an increasing  $\mu$  is coupled with increasing  $df_i$  and  $\hat{se}_{1i}$ , consequently reducing the variance in  $z_{1i}$  and  $t_{1i}$ .

For the second case of data-mining bias, the relation between the type I errors of  $H_0 : \alpha = 0$  and the degree of publication selection is illustrated in Figure 5. This second case considers a data-mining bias of  $\beta_{2i} \sim N(0, 0.2^2)$  and basically all meta-regression models suffer from inflated type I errors as publication selection becomes more prevalent. The only exception is MST in small primary sample size distributions, for which the type I errors decrease as publication selection increases. The source of this pattern is the use of the  $t$ -value in combination with small  $df$ , as the differences in the variances of the  $t$ -distribution are large for small  $df$ . However, this pattern exists only for small primary sample size distributions. In all other scenarios, MST provides highly inflated type I errors, even irrespective of the degree of publication selection that results from the



Figure 5: Type I Errors for Second Case of Normally Distributed Data-Mining Bias



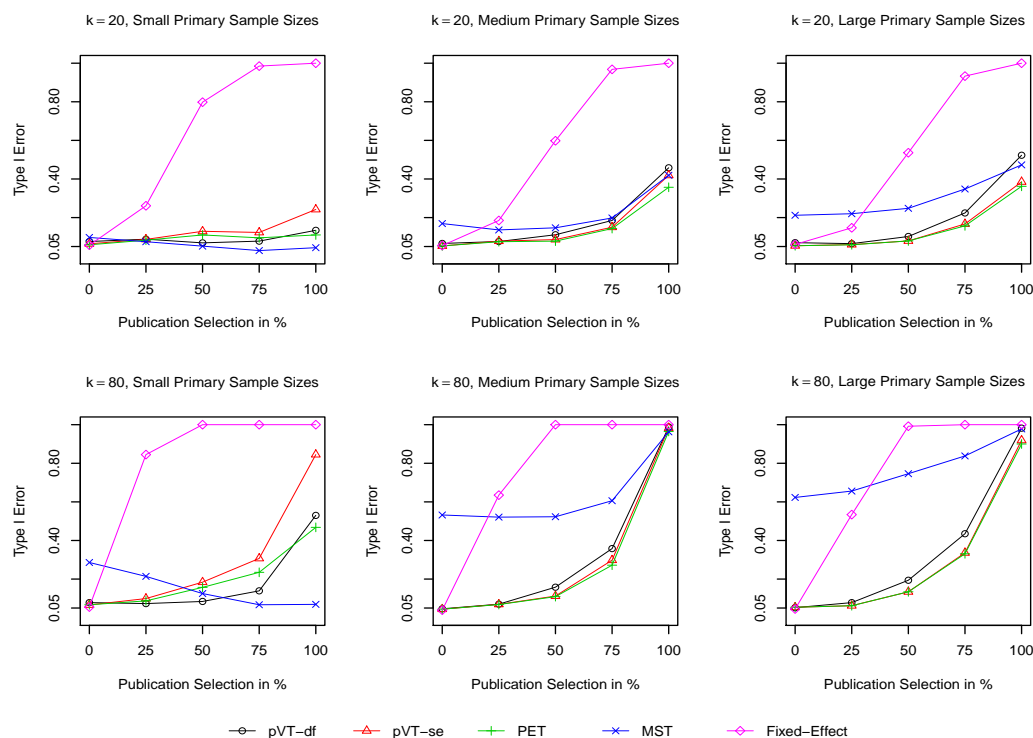
Notes: Type I errors of  $H_0 : \alpha = 0$  (Genuine effect of  $\beta_1 = 0$ ) for second case of normally distributed data-mining bias. The graphs illustrate the relation between type I errors and the degree of publication selection for both very small meta-analyses ( $k = 20$ ) and moderate meta-analyses ( $k = 80$ ) in combination with a small primary sample size distribution ( $\mu = 30$  and  $\sigma = 15$ ), a medium primary sample size distribution ( $\mu = 120$  and  $\sigma = 60$ ), and a large primary sample size distribution ( $\mu = 480$  and  $\sigma = 240$ ).

use of the magnitude of the  $t$ -values. This second case shows that meta-regression models are not robust to the presence of data-mining bias, even if the probability of observing large omitted-variable biases decreases with sample size.

For the third case of data-mining bias, the relation between the type I errors of  $H_0 : \alpha = 0$  and the degree of publication selection is illustrated in Figure 6. For this third case, the data-mining bias is uniformly distributed in the interval  $[-0.45 < \beta_{2i} < 0.45]$ . The type I errors are slightly higher compared to the second case in which the probability of observing large omitted-variable biases decreases with sample size. Both the type I errors in the second and third case increase with the same determinants as power in the first case of pure publication bias. The reason is that the omitted-variable biases are mistaken for genuine effects.

We do not discuss the power of  $H_0 : \alpha = 0$  for those cases with data-mining bias, since all meta-regression models discussed here suffer from severe inflation of type I errors. Given uncertainty regarding the degree of publication selection, meta-regression models must explicitly address data-mining bias and cannot rely on robustness with respect to this type of bias.

Figure 6: Type I Errors for Third Case of Uniformly Distributed Data-Mining Bias



Notes: Type I errors of  $H_0 : \alpha = 0$  (Genuine effect of  $\beta_1 = 0$ ) for third case of uniformly distributed data-mining bias. The graphs illustrate the relation between type I errors and the degree of publication selection for both very small meta-analyses ( $k = 20$ ) and moderate meta-analyses ( $k = 80$ ) in combination with a small primary sample size distribution ( $\mu = 30$  and  $\sigma = 15$ ), a medium primary sample size distribution ( $\mu = 120$  and  $\sigma = 60$ ), and a large primary sample size distribution ( $\mu = 480$  and  $\sigma = 240$ ).

## 4 Implications for Meta-Regression Analyses

We can show that most meta-regression models can identify genuine effects in the presence of pure publication bias. Most models are robust even if the results published in the literature completely distort the underlying genuine effect by only providing positive and significant  $\hat{\beta}_{1i}$ . The models outlined here, pVT-df and pVT-se, systematically outperform the widely used PET and MST. Nonetheless, meta-regression models are generally not robust to the presence of data-mining biases as all discussed meta-regression models provide inflated type I errors that are increasing with the degree of publication selection.

Our finding that most meta-regression models are robust to pure publication bias confirms the results of Stanley (2008). He also shows that PET and MST provide adequate type I errors for the integration of estimated regression coefficients across studies, even if the underlying literature is distorted by publication bias. The fixed-effects model is the exception and provides highly inflated type I errors in the presence of publication bias. This model does not control for publication bias and is therefore only used in the meta-analysis of randomized experiments within conditional estimation strategies. First, the absence of publication bias is evaluated with the Egger-regression by testing  $H_0 : \gamma = 0$  in (7) and, second, either the non-parametric trim and ill method is used to adjust for publication bias if  $\gamma \neq 0$  or a fixed-effects model is used when the Egger-regression

suggests the absence of publication bias ( $\gamma = 0$ ). However, these conditional testing procedures are shown to be inferior to meta-regressions that directly control for publication bias in the context of medical research (Moreno et al., 2009). In empirical economic research, there is not even a valid conditional testing approach, as the use of the Egger-regression to test for publication bias is only valid if heterogeneity in  $\hat{\beta}_{1i}$  is small or absent (Ioannidis, 2008). However, the degree of heterogeneity in empirical economics is large. A valid test for publication bias would be the Egger-regression with control variables that account for heterogeneity, but such a model is equivalent to PET and can be directly utilized to test for the presence of genuine effects.

Some studies have first attempted to ascertain the presence of publication bias in order to then use the fixed-effects model by testing how robust the attained results are to the inclusion of unpublished studies (e.g. Rosenbusch et al., 2013). However, this approach is unlikely to be helpful for empirical economic research, given how data-mining bias can be utilized to obtain nearly any result desired. Accordingly, we do not expect to observe many unbiased and insignificant results at the level of working papers. In other words, the potentially large amounts of estimates that have been conducted to obtain the results that are presented either in working papers or journal articles are likely to remain unobserved. Hence, the use of the fixed-effects model cannot be recommended for the identification of genuine effects in empirical economics. Instead, we should use meta-regression models that directly control for publication bias. This implication is further supported by the finding of Doucouliagos and Stanley (2011) that publication bias is widespread in economics.

Meta-regression models are robust with respect to biases that are closely related to  $se_{1i}$  or  $df_i$ , respectively. This is perfectly true for the selection of significant results from those results that sampling variability offers, namely pure publication bias. Our second case models data-mining bias that has a higher probability to use small omitted-variable biases compared to large omitted-variable biases. As small sample sizes need larger biases to obtain positive and significant  $\hat{\beta}_{1i}$ , we can observe larger omitted-variable biases in smaller samples and smaller omitted-variables biases in larger samples. Although this case mirrors the situation where an omitted-variable bias vanishes with sample size, meta-regression models are not robust to this type of bias. Those models that control for  $se_{1i}$  or  $df_i$  can reduce the inflation of type I errors considerably compared to the fixed-effects model, but are still by no means robust. The reason is that the discrete nature of omitted-variable biases does not allow producing biases that perfectly decrease with sample size. More important, there is no statistical law, such as statistical power for publication bias, that restricts the probability to observe large biases in large sample sizes. As a result, the type I errors are only slightly smaller in the second case that considers vanishing data-mining biases compared to the third case that considers uniformly distributed omitted-variable biases. These findings extend the results of Stanley (2008). He finds PET to be robust with respect to publication selection for positive and significant effects. He models the search for positive and significant effects as the simultaneous selection of a new omitted-variable bias and the resampling of all variables. This case of publication selection implies that authors, facing an insignificant  $\hat{\beta}_{1i}$ , do not only select a new omitted-variable bias but they also estimate the model with the new omitted-variable bias for completely new data. In this case, data-mining bias is dominated by pure publication bias and the robustness of PET with respect to this type of publication selection is consistent with our

finding that meta-regression models are robust with respect to pure publication bias. We simulate cases of publication selection that model authors playing intensively with omitted-variable biases to obtain positive and significant  $\hat{\beta}_{1i}$  for the same data. Only if the desired  $\hat{\beta}_{1i}$  cannot be obtained, new data is considered. For these cases, that we think are more prevalent in empirical economics, meta-regression models suffer from severely inflated type I errors.

The reason why data-mining bias leads to inflated type I errors is that the omitted variable can substantially bias  $\hat{\beta}_{1i}$ , irrespective of degrees of freedom or precision. As a result, the biased coefficient is "consistently" estimated. Bruns and Ioannidis (2013) suggest a solution to this issue. They point out that meta-regressions can only provide evidence of the existence of a non-zero underlying effect, which could equally be the result of a genuine effect or omitted variable bias. They argue that meta-regression models should control for primary regression specifications in the meta-regression. This adjustment can be used to identify which primary regression specification suggests the presence of a non-zero underlying effect. Subsequently, theoretical considerations can be used to determine whether a primary regression is reasonably specified or, conversely, might suffer from omitted-variable bias in order to discuss the presence of genuine effects.

Overall, we can show that pVT-se performs best in identifying genuine effects in the presence of publication bias for medium and large primary sample size distributions. Moreover, pVT-se provides power close to one even for small genuine effects and small meta-analysis sample size. pVT-df provides the best power for small primary sample size distributions among those methods with adequate type I errors in this scenario. pVT-df is also a good alternative to pVT-se if the measurement units of  $X$  and  $Y$  differ across studies. These models can handle publication bias perfectly, but still must explicitly address the issue of data-mining bias.

## 5 Conclusions

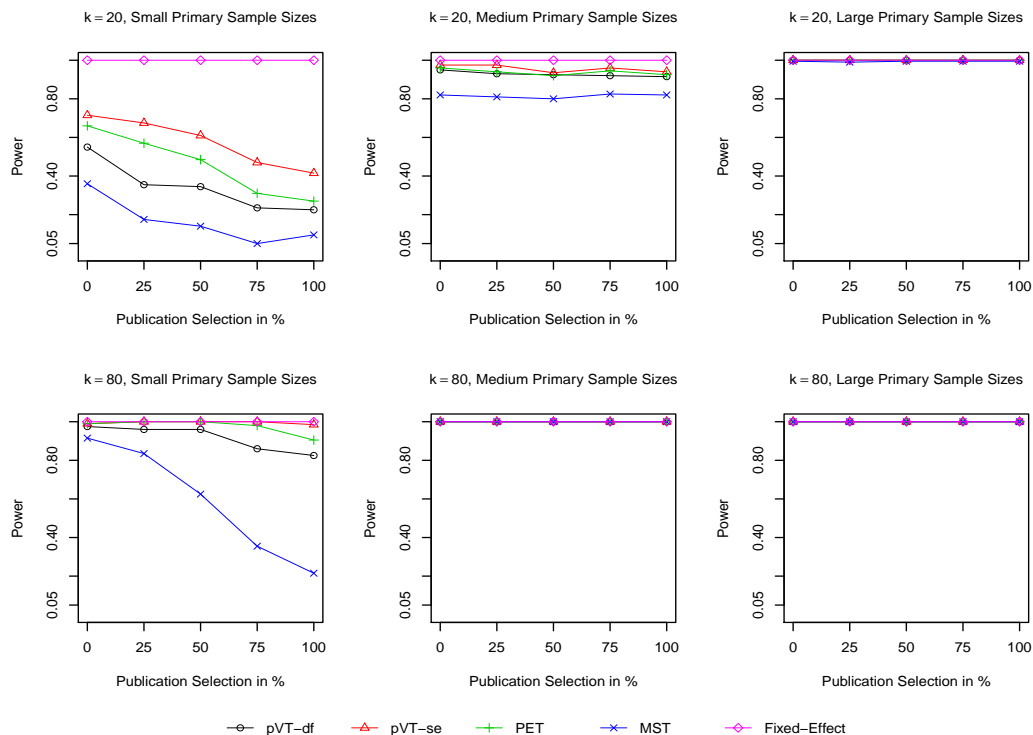
We suggest using the probit-transformed  $p$ -values ( $z_{1i}$ ) as the dependent variable in meta-regression models. The application of this transformation resolves the issue that the  $t$ -value is not constantly related to the  $p$ -value. The pVT-se model that relates  $z_{1i}$  to precision, and is an extension of the PET model, is therefore proposed as solution. We can show that pVT-se systematically outperforms PET across a wide range of research scenarios in the field of economics. We also suggest the pVT-df model that relates  $z_{1i}$  to  $\sqrt{df_i}$ . This model has lower power than pVT-se, but provides type I errors close to the nominal significance level even for small primary sample size distributions. Hence, pVT-df can be recommended for small primary sample size distributions and pVT-se for larger primary sample size distributions. Moreover, pVT-df is a good alternative to pVT-se whenever different units of measurement are used for  $X$  and  $Y$ . Both models are completely robust to the presence of pure publication bias occurring when authors select the positive and significant results among the range of results offered by sampling variability. However, data-mining bias, understood as the systematic utilization of omitted-variable biases to obtain positive and significant results, is a threat to the validity of meta-regression results and has to be addressed with great care. Otherwise, meta-regression analysis may suggest the presence of a genuine effect where no genuine effect is actually present. Accordingly, meta-regression analysis can suffer from the

same shortcomings as narrative surveys and may be even worse as the authority of meta-results claim correctness (Egger et al., 1998). Bruns and Ioannidis (2013) discuss an approach that makes meta-regression models robust with respect to data-mining bias.

Meta-regression analysis is thus a promising instrument to increase the credibility of observational research in empirical economics. It is a complementary tool to the use of techniques, such as experimental designs (Angrist and Pischke, 2010), extreme bounds analysis (Leamer and Leonard, 1983; Sala-I-Martin, 1997), and general-to-specific modeling (Hendry and Krolzig, 2005; Hoover and Perez, 1999) that directly enhance the credibility of primary studies. More methodological work on meta-regression models is required to understand both their potential and limitations in identifying genuine empirical effects in observational research. This includes methods that measure the size of empirical effects to evaluate not only the existence of empirical effects, but also their economic relevance.

## 6 Appendix

Figure 7: Power for First Case of Pure Publication Bias (Small Genuine Effect)



*Notes:* Power of  $H_0 : \alpha = 0$  (Genuine effect of  $\beta_1 = 0.4$ ) for first case of pure publication bias. The graphs illustrate the relation between type I errors and the degree of publication selection for both very small meta-analyses ( $k = 20$ ) and moderate meta-analyses ( $k = 80$ ) in combination with a small primary sample size distribution ( $\mu = 30$  and  $\sigma = 15$ ), a medium primary sample size distribution ( $\mu = 120$  and  $\sigma = 60$ ), and a large primary sample size distribution ( $\mu = 480$  and  $\sigma = 240$ ).

## 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.
- Becker, B. J. and Wu, M.-J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science*, 22(3):414–429.
- Brodeur, A., Lé, M., Sangnier, M., and Zylberberg, Y. (2013). Star wars: The empirics strike back. *IZA Discussion Paper No. 7268*.
- Bruns, S. B., Gross, C., and Stern, D. I. (2013). Is there really granger causality between energy use and output? *Crawford School Research Papers 1307*.
- Bruns, S. B. and Ioannidis, J. P. A. (2013). Meta-regression simulation of genuine effects, data-mining and publication bias. *mimeo*.
- Bruns, S. B. and Stern, D. I. (2013). Meta-granger causality. *mimeo*.
- Card, D. and DellaVigna, S. (2013). Nine facts about top journals in economics. *NBER Working Paper Series, Working Paper 18665*.
- Card, D. and Krueger, A. B. (1995). Time-series minimum-wage studies: A meta-analysis. *American Economic Review*, 85(2):238–243.
- Doucouliagos, C. and Stanley, T. D. (2011). Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys*, 27(2):316–339.
- Doucouliagos, C., Stanley, T. D., and Giles, M. (2012). Are estimates of the value of a statistical life exaggerated? *Journal of Health Economics*, 31(1):197–206.
- Doucouliagos, H. and Paldam, M. (2008). Aid effectiveness on growth: A meta study. *European Journal of Political Economy*, 24(1):1–24.
- Doucouliagos, H. and Stanley, T. D. (2009). Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations*, 47(2):406–428.
- Egger, M., Schneider, M., and Davey Smith, G. (1998). Spurious precision? Meta-analysis of observational studies. *British Medical Journal*, 316(7125):140–144.
- 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.
- Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE*, 5(4):e10068.
- Frey, B. S. (2003). Publishing as prostitution? - Choosing between one's own ideas and academic success. *Public Choice*, 116(1-2):205–223.

- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10):3–8.
- Greene, W. H. (2008). *Econometric Analysis*. Pearson Education LTD, 6 edition.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–61.
- Hendry, D. F. (1980). Econometrics - alchemy or science? *Economica*, 47(188):387–406.
- Hendry, D. F. and Krolzig, H.-M. (2005). The properties of automatic gets modelling\*. *The Economic Journal*, 115(502):32–61.
- Hoover, K. D. and Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *The Econometrics Journal*, 2(2):167–191.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, pages 221–233.
- Ioannidis, J. P. A. (2008). Interpretation of tests of heterogeneity and bias in meta-analysis. *Journal of Evaluation in Clinical Practice*, 14(5):951–957.
- Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, 73(1):31–43.
- Leamer, E. E. and Leonard, H. (1983). Reporting the fragility of regression estimates. *The 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. *The American Statistician*, 54(3):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., Sutton, A. J., Ades, A., Stanley, T. D., Abrams, K. R., Peters, J. L., and Cooper, N. J. (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Medical Research Methodology*, 9(1):2.
- Ridley, J., Kolm, N., Freckelton, R. P., and Gage, M. J. G. (2007). An unexpected influence of widely used significance thresholds on the distribution of reported p-values. *Journal of Evolutionary Biology*, 20(3):1082–1089.
- Rosenbusch, N., Brinckmann, J., and Müller, V. (2013). Does acquiring venture capital pay off for the funded firms? A meta-analysis on the relationship between venture capital investment and funded firm financial performance. *Journal of Business Venturing*, 28(3):335–353.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3):638–641.
- Sala-I-Martin, X. X. (1997). I just ran two million regressions. *American Economic Review*, 87(2):178–183.



- 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. (2005a). Beyond publication bias. *Journal of Economic Surveys*, 19(3):309–345.
- Stanley, T. D. (2005b). Integrating the empirical tests of the natural rate hypothesis: A meta-regression analysis. *Kyklos*, 58(4):611–634.
- 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, C. (2007). Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression. *School Working Paper - Economics Series 11*.
- 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., Abrams, K. R., Jones, D. R., Jones, D. R., Sheldon, T. A., and Song, F. (2000). *Methods for meta-analysis in medical research*. J. Wiley.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838.
- Wooldridge, J. M. (2006). *Introductory econometrics: A modern approach*, volume 3. Cengage Learning.
- Young, N. S., Ioannidis, J. P. A., and Al-Ubaydli, O. (2008). Why current publication practices may distort science. *PLoS Medicine*, 5(10):e201.