

Koetse, Mark J.; Florax, Raymond J.G.M.; de Groot, Henri L.F.

**Working Paper**

## Correcting for Primary Study Misspecifications in Meta-Analysis

Tinbergen Institute Discussion Paper, No. 05-029/3

**Provided in Cooperation with:**

Tinbergen Institute, Amsterdam and Rotterdam

*Suggested Citation:* Koetse, Mark J.; Florax, Raymond J.G.M.; de Groot, Henri L.F. (2005) : Correcting for Primary Study Misspecifications in Meta-Analysis, Tinbergen Institute Discussion Paper, No. 05-029/3, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/86400>

**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.*



TI 2005-029/3

Tinbergen Institute Discussion Paper

# Correcting for Primary Study Misspecifications in Meta-Analysis

*Mark J. Koetse<sup>a,c</sup>*

*Raymond J.G.M. Florax<sup>a,b</sup>*

*Henri L.F. de Groot<sup>a,c</sup>*

<sup>a</sup> *Department of Spatial Economics, Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam,*

<sup>b</sup> *Regional Economics Applications Laboratory (REAL), University of Illinois, Urbana-Champaign, USA,*

<sup>c</sup> *Tinbergen Institute.*

**Tinbergen Institute**

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

**Tinbergen Institute Amsterdam**

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

**Tinbergen Institute Rotterdam**

Burg. Oudlaan 50

3062 PA Rotterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at  
<http://www.tinbergen.nl>.

# Correcting for Primary Study Misspecifications in Meta-Analysis

Mark J. Koetse<sup>a</sup>, Raymond J.G.M. Florax<sup>b,c,d</sup>, Henri L.F. de Groot<sup>b,d</sup>

<sup>a</sup> Corresponding author: Institute for Environmental Studies (IVM), VU University  
Amsterdam, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands, t: + 31 (0)20 598  
3194, e: mark.koetse@vu.nl

<sup>b</sup> VU University Amsterdam, Department of Spatial Economics, Amsterdam, The Netherlands

<sup>c</sup> Purdue University, Department of Agricultural Economics, West Lafayette, United States

<sup>d</sup> Tinbergen Institute, Amsterdam, The Netherlands

## **Abstract**

In non-experimental sciences the errors associated with model misspecifications in primary studies carry over to meta-analysis. We use Monte Carlo simulations to analyse the effects of these misspecifications on results of a meta-analysis using a meta-estimator that calculates a simple average effect and a meta-estimator that includes dummy variables to control for primary study misspecification. The results show that using the dummy variable model goes a long way in mitigating the negative effects of error propagation on the bias and mean squared error of the meta-estimator, and the size and the power of statistical tests. Although primary study misspecifications can usually be observed and controlled for in a meta-analysis, the more complex interactions between these observed characteristics are far more difficult to control for in practice. Our results show that these interactions may, however, substantially affect the outcomes of a meta-analysis. When meta-analysis is used to look for a ‘true’ effect rather than for analysing variation in outcomes, our results provide an argument for study selection on model quality to avoid the impact of error propagation in meta-analysis.

*JEL-codes:* C12, C15, C40

*Key words:* meta-analysis, misspecification, omitted variable bias, incorrect functional form, Monte Carlo simulation

## 1. Introduction

Although meta-analysis was originally developed in the experimental sciences, it is currently also gaining ground in non-experimental sciences, especially in economics.<sup>1</sup> The switch from the experimental to the non-experimental context that prevails in most areas in economics induces specific methodical problems. First, in an experimental research design, sampling of sizeable experimental and control groups mitigates the need for control variables. The design can as a result be largely standardised. This is very different in economics, where slight changes in research design are often viewed as an innovation over earlier work.<sup>2</sup> Typically, data constraints as well as the desire to be ‘different’ lead to varying sets of control variables across studies. This obviously induces omitted variable bias and/or multicollinearity in at least a subset of the available primary studies. Second, the true data generating process (DGP) is generally unknown in economics, implying that meta-analyses in economics typically contain a mix of correctly and incorrectly specified effect size measures.

Although the impact of these misspecifications on the outcomes of a meta-analysis can to some extent be derived analytically, many of the analytical effects are difficult to obtain. Moreover, since most meta-analyses in economics have to deal with relatively limited sample sizes, assessing the small sample performance of meta-analysis models is in most cases more interesting than analysing their asymptotic characteristics through analytical inference. In this paper we use Monte Carlo simulation to analyse the impact of primary model misspecifications on the results of a meta-analysis, and assess the mitigating effects of standard procedures to control for these issues.

The remainder of this paper is organised as follows. Section 2 discusses the issues of omitted variable bias and erroneous effect size measures in more detail. In Section 3 we present the data generating process of our simulation experiment. The results are presented in Sections 4 and 5. In Section 4 we present some of the main results graphically, while Section 5 discusses the results of a response surface analysis. Section 6 concludes.

---

<sup>1</sup> For popular contemporary contributions in economics see, among others, Feld and Heckemeyer (2011), Kluge (2010), Doucouliagos and Paldam (2009), Disdier and Head (2008) and Viscusi and Aldy (2003). For an overview of the meta-analysis literature in environmental economics see Nelson and Kennedy (2009).

<sup>2</sup> In medicine, or more specifically epidemiology, the same tendency is apparent in so-called observational studies, which are generally considered less homogeneous than randomised clinical trials (see Sutton et al., 2000).

## 2. Primary model misspecification

In this paper we test the impact of two of the most common primary model misspecifications, i.e., incorrect functional form and omitted variables. An often used effect size in economics is the elasticity.<sup>3</sup> In general, elasticities can be measured as a point-elasticity or, alternatively, the true elasticity can be assumed constant across the demand or supply curve and can be directly derived from a double-log specification. Since the true data generating process is unknown, economic meta-analyses typically contain a mix of both types of elasticities (see, e.g., Dalhuisen et al., 2003; Koetse et al., 2011).<sup>4</sup> In economics slight changes in research design are often viewed as an innovation over earlier work, leading to varying sets of control variables across studies. This obviously induces omitted variable bias in a subset of the primary studies. In a strict sense, effect sizes estimated with different sets of control variables cannot be assumed to represent identical population effect sizes (see Keef and Roberts, 2004). Especially when omitted variables in primary model lead to a bias in effect sizes, the results of a meta-analysis may be affected in several ways. To show this, assume the *true* underlying model is given by:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \quad (1)$$

where  $y$  is the dependent variable,  $\beta_1$  and  $\beta_2$  are coefficients on exogenous explanatory variables  $x_1$  and  $x_2$ , respectively, and  $\varepsilon$  is an error term. When this model is estimated by OLS, producing  $\hat{\beta}_1$  as an estimate of  $\beta_1$ , then under well known conditions (see Kennedy, 2003, pp. 15-17) we get  $E(\hat{\beta}_1) = \beta_1$ . Our interest is in the situation where the model in equation (1) is estimated without  $x_2$  as an explanatory variable. In this case we get  $\hat{\beta}_1^{ov}$  as an estimate of  $\beta_1$ . The expected value of  $\hat{\beta}_1^{ov}$ , in terms of the correlation coefficient between  $x_1$  and  $x_2$ , is given by (see Greene, 2000, p. 334):

$$E(\hat{\beta}_1^{ov}) = \beta_1 + \beta_2 r_{12} \frac{\sigma(x_2)}{\sigma(x_1)}, \quad (2)$$

where  $\sigma(x_1)$  and  $\sigma(x_2)$  are the sample standard deviations of  $x_1$  and  $x_2$ , respectively, and  $r_{12}$  is the sample correlation coefficient between  $x_1$  and  $x_2$ . Assuming  $\beta_2 \neq 0$ , equation (2)

---

<sup>3</sup> An elasticity represents the percentage change in a certain variable due a one-percent change in another variable.

<sup>4</sup> An interesting study on comparability of effect sizes is Peterson and Brown (2005), who analyse the effects of including control variables for study characteristics in a meta-analysis of correlation coefficients.

clearly shows that  $\hat{\beta}_1^{ov}$  is unbiased only when  $r_{12} = 0$  or  $x_2$  is a constant. The sign of the bias in this model is equal to the sign of the correlation coefficient times the sign of  $\beta_2$ ; generally, both are unknown in practice. We are therefore left with a bias in  $\hat{\beta}_1^{ov}$  of unknown sign and size. As the model in (1) is expanded to include more regressors and is again estimated without  $x_2$ , the situation deteriorates invariantly. In this case the extended version of (2) involves multiple coefficients, multiple correlation coefficients of unknown sign and size, and multiple variance terms of unknown size. The only possible up-turn of omitted variables is that the variance of  $\hat{\beta}_1^{ov}$  may be smaller than the variance of  $\hat{\beta}_1$ . The net effect on the mean squared error of the estimator, which is a composite of the bias and variance of the estimator, is ambiguous. It is therefore possible that we end up with a biased but more precise estimator (Greene, 2000, pp. 335-337).

The consequences of omitted variables and using incorrect effect size measures in primary models for meta-analysis are potentially serious. Obviously, an ideal situation in meta-analysis would be to have a set of studies that have identical and correct model specifications and use similar effect size measures. Not surprisingly, such circumstances are often found in experimental designs rather than the quasi-experimental or non-experimental designs typically used in economics. Fortunately, the differences in model specifications and effect size measures are observable and can hence be controlled for in meta-analysis. The usual approach here is to include dummy variables in the meta-model specification. In the next section we describe the set-up of the Monte Carlo simulations that are used to test the impact of primary model misspecifications on the results of a meta-analysis and to analyse the extent to which dummy variables in the meta-model specification can mitigate the propagation of errors in primary studies.

### 3. Experimental design

The experimental simulations are comparable to the ones used in Koetse et al. (2010), and consist of five steps: generating primary study data, estimating primary study models, sampling primary studies effect sizes, specifying and estimating meta-analysis models, and assessing meta estimator performance. In the five subsections below we discuss these steps in detail.<sup>5</sup>

---

<sup>5</sup>The computer programs used for the analysis in this paper are written in Gauss 8.0, and are available from the authors upon request.

### 3.1 Generating primary study data

In order to induce omitted variable bias in primary study effect sizes we generate the primary data according to a pre-specified ‘true’ model, and re-estimate the relationship using a correctly specified and a misspecified primary model, inducing bias in the latter case. It is clear that a relatively large true underlying model with, for instance, ten explanatory variables, would allow us to induce bias from various sources by systematically excluding relevant explanatory variables from the primary model specification. The amount of bias induced in our estimate of interest depends to a large extent on the degree of multicollinearity between the explanatory variables in the true underlying model. A small true underlying model (i.e., a model with a limited number of explanatory variables) can, however, be used without loss of generality. As shown in the previous section, the bias induced in a primary model estimate due to omitting a single variable depends on the correlation between the included and the excluded variable and their respective variances. In analysing the impact of various levels of omitted variable bias, the number of explanatory variables in the true underlying model does not matter, and we therefore use a ‘small’ model. The ‘true’ underlying model used for *primary data generation* is a general unrestricted Cobb-Douglas function of the form:

$$y = e^{\alpha} x^{\beta_0} z^{\beta_1} e^{\varepsilon}, \quad (3)$$

where  $y$  is a stochastic variate,  $x$  and  $z$  are exogenous variables,  $\alpha$ ,  $\beta_0$  and  $\beta_1$  are parameters, and  $\varepsilon$  is an error term. In the Cobb-Douglas model the elasticity of  $y$  on  $x$  equals  $\beta_0$  over the entire data range. We set  $\beta_0$  equal to 0 and 1, thereby including the cases of a zero and a non-zero population effect size. We set both  $\alpha$  and  $\beta_1$  equal to one, while the error term  $\varepsilon$  is normally distributed with mean zero and a variance that is equal to 1, 5 or 10. This error term is generated for each replication. Furthermore, since in actual practice primary studies vary in size, we use sample sizes of 50, 100, 500 and 1,000. The way in which we generate  $x$  and  $z$  is crucial for our analyses. The variable  $x$  is generated, once, according to:

$$x = e^{\mathfrak{g}}, \quad (4)$$

where  $\mathfrak{g}$  is a vector drawn from a random uniform (0,1) distribution. We systematically vary the relation between  $x$  and  $z$  by generating  $z$  according to:



$$z = x^\lambda e^\psi, \quad (5)$$

where  $\lambda$  is a parameter,  $\psi$  is an error term drawn, once, from a uniform (0,1) distribution, and  $\psi$ ,  $\vartheta$  and  $\varepsilon$  are independent. The potential bias induced in the estimate of  $\beta_0$  when  $z$  is excluded from the primary model does not only increase with the bivariate correlation between  $x$  and  $z$ , but also with the variance of  $z$ , as is shown by equation (2). Obviously, when  $\lambda = 0$ , the correlation between  $x$  and  $z$  is zero, implying that the bias in  $\beta_0$  when  $z$  is excluded from the primary model is zero as well. However, when we increase the value of  $\lambda$ , both the correlation between  $x$  and  $z$  and the standard deviation of  $z$  are increased, thereby invariantly increasing the bias in the estimate of  $\beta_0$ . We therefore distinguish between three cases, i.e., no correlation ( $\lambda = 0$ ), moderate correlation ( $\lambda = 0.5$ , average correlation coefficient between  $x$  and  $z$  is around 0.45), and high correlation ( $\lambda = 1.5$ , average correlation coefficient between  $x$  and  $z$  is around 0.80).<sup>6</sup>

### 3.2 Estimating primary study models

Our approach is different from other Monte Carlo studies in meta-analysis (see, e.g., Bijmolt and Pieters, 2001; Field, 2001; Kuhnert and Böhning, 2007; Oswald and Johnson, 1998; Sanchez-Meca and Marin-Martinez, 1997, 1998;) in that we explicitly incorporate the stage of the *primary data analysis* (see also Koetse et al., 2010). Besides the fact that this gives us the opportunity to explicitly introduce omitted variable bias in primary study effect sizes, it also allows us to investigate the impact of including erroneous effect size measures on the results of a meta-analysis. Specifically, we use the data generated by the model in equation (3) to estimate a log-linear model, which is mathematically equivalent to the model in (3), and an alternative linear model.<sup>7</sup> The log-linear model is given by:

$$\ln(y) = \alpha + \beta_0 \ln(x) + \beta_1 \ln(z) + \varepsilon. \quad (6)$$

We estimate this model by OLS, which gives us  $\hat{\alpha}$ ,  $\hat{\beta}_0$  and  $\hat{\beta}_1$  as estimates of  $\alpha$ ,  $\beta_0$  and  $\beta_1$ , respectively. The parameter of interest is the elasticity of  $y$  with respect to  $x$ , given by  $\eta_s = \hat{\beta}_{0s}$  for primary study  $s$ . Given our DGP this elasticity can be viewed as the ‘true’

<sup>6</sup> Average correlation coefficients are obtained by running 5,000 replications of  $x$  and  $z$  from equations (4) and (5).

<sup>7</sup> Of course, the choice of the true underlying model is arbitrary to some extent, i.e., we also could have chosen the linear model as the true underlying model. Although this would affect our results to some extent, there is no reason to suspect that this would change the results presented in this paper in a qualitative sense.

elasticity. In order to induce omitted variables in our simulations, we use two primary model specifications, i.e., the correctly specified primary model in (6) and a misspecified version of this model from which  $\ln(z)$  is excluded as an explanatory variable. As discussed in the previous section, the latter model induces omitted variable bias in  $\hat{\beta}_0$  when  $\lambda \neq 0$ .

An alternative elasticity estimate is obtained by estimating an incorrect linear primary model specification, which looks as:

$$y = \alpha^* + \beta_0^*x + \beta_1^*z + \varepsilon^*. \quad (7)$$

Using OLS to estimate this model produces  $\hat{\alpha}^*$ ,  $\hat{\beta}_0^*$  and  $\hat{\beta}_1^*$  as estimates of  $\alpha^*$ ,  $\beta_0^*$  and  $\beta_1^*$ , respectively. In this linear model we estimate the intrinsically non-linear relationship between  $y$ ,  $x$  and  $z$ , and compute a point-elasticity at the sample mean as  $\eta_s = \hat{\beta}_{0s}^* (\bar{x}_s / \bar{y}_s)$  for primary study  $s$  (a common alternative is to use the median of the data on  $x$  and  $y$ ). In reality the estimation of different effect size measures may occur frequently, simply because the true underlying model is unknown and researchers may build their analysis on a misspecified model. As before, in order to induce omitted variables we use the model specification in (7) and a specification from which  $z$  is excluded as an explanatory variable. The latter model again induces omitted variable bias in  $\hat{\beta}_0^*$  when  $\lambda \neq 0$ . Strictly speaking, when using an incorrect functional form, omitted variables does not by definition lead to bias, i.e., the effects of incorrect functional form and model misspecification may cancel out. This is highly unlikely, however. We analyse these interactions in more detail later on in Section 5.

The experimental design for primary study characteristics includes two values for the true underlying effect  $\beta_0$ , three values for  $\lambda$ , four sample sizes, three values for the error variance  $\varepsilon$ , two functional forms for the primary model producing a point- and a double-log elasticity, and two separate primary model specifications, each with a different set of explanatory variables. Ultimately, we end up with  $2 \times 3 \times 4 \times 3 \times 2 \times 2 = 288$  different primary study types, for each of which we run 10,000 replications.

### 3.3 Sampling primary study effect sizes

In this section we describe how we mimic the study retrieval process, determining which study types end up in a meta-analysis sample. This procedure is relevant for our response surface analysis, which is presented in Section 5 and consists of a number of steps. In the first step we split the sample of studies into separate sub-samples, because some of the primary

study characteristics are held constant within a single meta-analysis. First, within each meta-analysis the true value of the population effect size  $\beta_0$  should be identical. If not, the expected outcome of the meta-analysis is unknown and we cannot sensibly assess meta-estimator performance (see Section 3.5). The sample of effect sizes is split into two separate sub-samples; the value of the true population effect size in the first sub-sample is zero, while for the second sub-sample it is equal to one. Second, in order to clearly demonstrate the impact of several levels of omitted variable bias, we fix the value of  $\lambda$  in a single meta-analysis. Therefore, each of the two sub-samples given above is subdivided into three sub-samples based on the value of  $\lambda$ . Ultimately, we end up with six separate sub-samples. Considering the fact that we have a total of 288 different primary studies types, we have  $288/6 = 48$  different primary study types within each sub-sample, each with 10,000 replications.

In the second step, we randomly draw from the available pool of primary studies within each sub-sample. However, if the sampling procedure is based on randomly drawing *studies* (without replacement), we would asymptotically obtain meta-analyses with *study characteristics* that are on average ‘fixed’ in the same proportions with which they were generated in the experimental design. This implies that, across meta-analysis replications, we end up with very little variation in primary study characteristics. We therefore do not use a sampling procedure based on randomly drawing *studies*, but on randomly drawing *study characteristics*. Although this procedure is slightly more complicated, it safeguards that across meta-analysis replications sufficient variation in primary study characteristics is available. The sampling procedure is discussed in detail below.

For each meta-analysis replication within each sub-sample we compute a probability of sampling one of the available 48 primary study types. This is done by randomly drawing a sampling probability for each *study characteristic* from a uniform (0,1) distribution. By multiplying the sampling probabilities for each study characteristic, we obtain a sampling probability for each *study type*. When we multiply the resulting sampling probabilities with the sample size of the meta-analysis, we obtain the absolute number of studies of each *study type* to be included in the meta-analysis. The absolute number of studies per study type are drawn randomly from the 10,000 replications that are available for each study type. For a sufficient number of meta-analysis replications, this procedure ensures maximum variation in study characteristics across replications.

To give an example of the procedure discussed above, suppose we have 6 different study types from 2 study characteristics – omitted variable bias (OVB, 2 alternatives) and

primary model error variance (VAR, 3 alternatives). We draw the probability for sampling a study with omitted variables bias from a uniform (0,1) distribution, for example 0.35. The probability of sampling a study without omitted variable bias is thus equal to 0.65. Furthermore, the probability of sampling a study with a certain disturbance variance is determined as follows. For each disturbance variance we draw probabilities from a uniform (0,1) distribution, for example 0.25, 0.60 and 0.40. Since the sum of probabilities is not equal to 1, we divide each probability by the sum of the three probabilities, resulting in three sampling probabilities that do sum to 1, in this case 0.20, 0.48 and 0.32. The sampling probability and the absolute number of replications for each of the 6 different study types are given in Table 1. In this example, the meta-analysis consists of 100 primary studies. Note that the sampling probabilities per study type and the number of studies per study type sum to 1 and 100, respectively.

<<< *Insert Table 1* >>>

### 3.4 Meta-estimator specification

The primary aim in this paper is to compare the performance of two meta-estimators for different meta-sample characteristics. The first model is a rather naïve meta-regression specification in which we take the mean of the estimated effect sizes as an estimator of the population effect size  $\beta_0$ . This model looks as:

$$\eta_s = \delta_0 + \xi_s, \quad (8)$$

where  $\eta_s$  is a vector of primary study elasticities,  $\xi_s$  is an error term, and  $s = 1, 2, \dots, S$  indexes the primary studies. The model is estimated by OLS, producing  $\hat{\delta}_0$  as an estimate of the true underlying effect size  $\beta_0$ . The estimate and its variance are given by:

$$\hat{\delta}_0 = \frac{1}{S} \sum_{s=1}^S \eta_s, \quad (9)$$

and

$$\text{var}(\hat{\delta}_0) = \text{var}\left(\frac{1}{S} \sum_{s=1}^S \eta_s\right) = \frac{1}{S^2} \sum_{s=1}^S \text{var}(\eta_s) = \frac{1}{S^2} \sum_{s=1}^S \sigma_{\eta_s}^2, \quad (10)$$

respectively. In the latter equation  $\sigma_{\eta_s}^2$  represents the variance of elasticity  $\eta$  from primary study  $s$ . The variance in (10) shows one of the main general advantages of meta-analysis. As  $S$  increases, the variance of the meta-estimator decreases rapidly.<sup>8</sup>

The impact of omitted variables on this meta-estimator can to a certain extent be derived analytically. Assuming the magnitude of the omitted variable bias is constant across misspecified primary studies, which apart from sampling error it is in our case, the bias for the naïve meta-estimator in equation (8) is given by:

$$\text{bias}(\hat{\delta}_0) = (1-c) \times \text{bias}(\eta_i) + c \times \text{bias}(\eta_j), \quad (11)$$

where  $\eta_i$  is an effect size from a correctly specified primary model,  $\eta_j$  is an effect size from a misspecified primary model, and  $c$  is the proportion of effect sizes in the meta-analysis sample from misspecified models. Equation (11) simply states that the bias of the naïve estimator is a weighted average of the bias in the subsample of non-biased estimates and the bias in the subsample of biased estimates, with the subsample proportions as weights. Since  $\text{bias}(\eta_i)$  is approximately equal to zero, the bias of the meta-estimator is equal to  $c \times \text{bias}(\eta_j)$ . Because  $\text{bias}(\eta_j)$  is not known, simulation is necessary to show the actual magnitude of the bias under plausible parameter values. When using erroneous effect-size measures, in our case a point-elasticity instead of a double-log elasticity, analytical inference is not straightforward, so simulation is needed to obtain the relevant insights.

In the second model we apply the approach that is commonly used in meta-analysis in economics to account for primary study misspecifications, which is including dummy variables in order to control for observable differences between primary study characteristics. We include two dummy variables; the first dummy is equal to one when the effect size is obtained from a primary model with omitted variables, the second dummy is equal to one when the effect size is a point-elasticity.<sup>9</sup> The second meta-model looks as:

$$\eta_s = \delta_0^* + \delta_1^* D_s^{pe} + \delta_2^* D_s^{ov} + \xi_s^*, \quad (12)$$

---

<sup>8</sup> Note that this variance is different from the variance obtained by OLS estimation of the naïve model. See Hedges (1994) for details on the differences between the two models.

<sup>9</sup> In practice, values for dummy variables are known because characteristics such as data type, functional form and model specification can be clearly observed in studies. However, it is generally difficult to assess whether excluding a specific variable in a model specification actually represents omitted variable bias, or which functional form is the right one. In most cases a meta-analysis can only show whether these issues have a systematic effect or not.

where  $\eta_s$  is a vector of primary study elasticities,  $D_s^{pe}$  is a dummy variable equal to one if the effect size is a point-elasticity,  $D_s^{pv}$  is a dummy variable equal to one if the primary study is estimated without  $z$  among the explanatory variables, and  $\xi_s^*$  is an error term. This model is also estimated by OLS, producing  $\hat{\delta}_0^*$  as an estimate of the true underlying effect size  $\beta_0$ . Furthermore,  $\hat{\delta}_1^*$  and  $\hat{\delta}_2^*$  are estimates of the two dummy variables that should pick up the impact of omitted variable bias and using point-elasticities, thereby reducing to some extent the impact of these two misspecifications on the estimate of interest, in this case  $\hat{\delta}_0^*$ .

One of the most important assumptions underlying the second model is that the relationship between differences in primary study characteristics on the one hand, and differences in primary study effect sizes on the other, is linear and constant. If this holds, then the changes in effect sizes due to primary model misspecifications should be largely accounted for by the dummy variables. However, if the impact of misspecification in primary studies is different for each study, the effects may to a certain extent average out in the meta-analysis, especially in large samples. In this case the impact is not or only partly picked up by the dummy variables, leaving the bias unaffected but increasing the variance of the meta-estimator in equation (12). In fact, when the differences average out completely, the model in (12) reduces to a slightly less efficient version of the model in (8).

### 3.5 Assessing meta-estimator performance

The central issue in this paper is how well the meta-estimators recover the value of the population effect size  $\beta_0$  in terms of size and statistical significance in the presence of primary study misspecifications. Specifically, we compare the naïve meta-regression estimate  $\hat{\delta}_0$  from the model in equation (8) to  $\hat{\delta}_0^*$ , the estimate from the meta-regression model with dummy variables to control for primary study misspecifications as given in equation (12).

As discussed previously, the misspecifications may affect the meta-estimates on several dimensions, i.e., the estimate itself, the variance of the estimate, and as a result the statistical significance of the estimate. We therefore use three different performance indicators to investigate the impact. First, the bias (BIAS) of the estimates measures the difference between the average value of the estimates and  $\beta_0$ . As we have argued in the previous subsection, a problem with this indicator is that the impact of misspecifications on the effect sizes may average out, in which case bias is equal to zero. However, the variance of the estimate may still be substantial. In order to account for this we also use the mean squared error (MSE) of the estimate as a performance indicator. This second indicator combines the bias and the

variance of the estimate, and measures the average distance of the estimate to the true parameter, i.e., the smaller the MSE, the closer, on average, the estimate will be to the true parameter. The third and final indicator is the proportion of statistically significant results (SIG) of the meta-estimators. Formally, for  $\hat{\delta}_0$  these indicators are given by:

$$\text{BIAS}(\hat{\delta}_0) = E(\hat{\delta}_0 - \beta_0) \approx \frac{1}{R} \sum_{r=1}^R (\hat{\delta}_0 - \beta_0)_r, \quad (13)$$

$$\text{MSE}(\hat{\delta}_0) = E(\hat{\delta}_0 - \beta_0)^2 = \text{BIAS}(\hat{\delta}_0)^2 + \text{var}(\hat{\delta}_0) \approx \frac{1}{R} \sum_{r=1}^R (\hat{\delta}_0 - \beta_0)_r^2, \quad (14)$$

$$\text{SIG}(\hat{\delta}_0) = \frac{1}{R} \sum_{r=1}^R I(|t_{n-k}| > t_{crit})_r, \quad (15)$$

where  $r = 1, 2, \dots, R$  indexes the meta-analyses replications. The performance indicators for  $\hat{\delta}_0^*$  are obtained by replacing  $\hat{\delta}_0$  by  $\hat{\delta}_0^*$  in equations (13), (14) and (15). In equation (15),  $I$  is an indicator function equal to one if the absolute  $t$ -value of the meta-estimate is greater than a pre-specified critical  $t$ -value, denoted by  $t_{crit}$ , and 0 otherwise. We apply two-sided significance tests using a 5% significance level. When  $\beta_0 = 0$  and  $H_0: \beta_0 = 0$ , we are interested in the probability of a Type I error, i.e., the probability that an estimator erroneously rejects  $H_0$ . Therefore, when  $\beta_0 = 0$  SIG corresponds to the proportion of Type I errors, or the *size* of the statistical test on the meta-estimates (see Kennedy, 2003, pp. 77-78). Alternatively, when  $\beta_0 = 1$ , and under the same null-hypothesis, we are interested in the probability of a Type II error, i.e., the probability that the statistical test on the meta-estimate erroneously accepts  $H_0$ . When  $\beta_0 = 1$ , SIG corresponds to  $(1 - \text{probability of a Type II error})$ , or the *power* of the statistical test (see Kennedy, 2003, pp. 77-78). Type I and Type II errors provide different types of information on statistical significance, which is the most important reason why we distinguish between a zero ( $\beta_0 = 0$ ) and a non-zero ( $\beta_0 = 1$ ) true underlying effect size.

#### 4. Simulation results: Graphical illustrations

In this section we illustrate the isolated impact of the two primary study misspecifications introduced in the previous section. For these purposes we systematically vary the proportion of misspecified primary studies in the meta-analysis, and keep constant all the other variables

and parameters introduced in the previous section. Specifically, the results in this section are generated for primary studies with sample size 500, a normally distributed error term  $\varepsilon$  with mean zero and variance 5, and for  $\lambda$  equal to 0.5 ( i.e., correlation between  $x$  and  $z$  is approximately 0.45, see Section 3.1). Finally, the number of replications for the primary studies is 5,000, the sample size of the meta-analysis is 100 and the number of meta-analysis replications is 10,000. We distinguish between the situation where the true underlying effect size is zero and one, respectively. Because in each experiment in this section we only vary a single parameter, we do not have to use the rather cumbersome sampling procedure discussed in subsection 3.4. We simply run our meta-analyses for the desired values of the parameter to be varied. In the following two subsections we analyse the impact of omitted variable bias and the impact of erroneous effect size measures, respectively.

#### 4.1 Impact of omitted variable bias

In this subsection we analyse the impact of omitted variable bias on the performance of the meta-models specified in equations (8) and (12). In Figure 1 we show the three performance indicators for both models. On the horizontal axis, we measure the extent of misspecification by the proportion of effect sizes in the meta-sample that are affected by omitted variable bias, which we systematically increase from 1% to 99% with increments of 1%. The vertical axis represents the bias and mean squared error of the meta-estimators and the size or power of the statistical tests on the meta-estimates. In order to isolate the impact of omitted variable bias, we include only a small number of point-elasticities in the meta-sample; the proportion is fixed at 5%.<sup>10</sup>

The figure shows that, as was also discussed in Section 3.4, increasing the proportion of estimates from misspecified studies in the meta-analysis sample leads to a linear increase in meta-estimator bias. More interesting is the fact that meta-estimator bias induced by omitted variables in primary models can be substantial even for a moderate correlation between  $x$  and  $z$ . Indeed, when assessing the sum of squares of the bias the mean squared error is replicated almost exactly, implying that the variance of the estimate from the naïve model does not increase with this proportion. There is no apparent trade-off between the bias and variance of the meta-estimator in this case. The increasing bias in the meta-estimate of the naïve meta-model also implies that the number of Type I errors (or the size of the statistical test) increases

---

<sup>10</sup> We use 5% instead of 0% because the latter would lead to models in which one of the dummy variables is excluded because it would have no variation.



rapidly. This result clearly shows that the naïve meta-estimator is not useful since it virtually always rejects the null hypothesis that there is no real effect, regardless whether this is true or not.

The figure also shows that the impact of omitted variable bias in primary studies on the bias of the meta-estimator is nearly fully controlled for by the inclusion of dummy variables in the meta-model specification. The bias of the meta-estimator is close to zero in all cases. When drawing a line for the sum of squares of the bias it is below the mean squared error depicted in the graphs. This implies that the variance of the meta-estimate from the model with dummy variables increases with  $c$ , in contrast to the naïve model. This result can be explained by realising that  $\hat{\delta}_0^*$  is simply the mean of effect sizes from correctly specified primary models. Since the number of such effect sizes decreases when going along the horizontal axis, there is less and less opportunity for *random deviations* of these effect sizes to average out. The increase in the mean squared error or variance obviously also explains the increase in the size of the statistical test on the meta-estimate from the model with dummy variables. However, although the size of the test is higher than the critical 5% level, it is still substantially better than the size of the statistical test on the naïve meta-estimate. Finally, the increase in estimator variance does not cause a decrease in power. The power of the test for both estimators is always equal to one in our simulations (the black square in the figures is hidden behind the white square).

Summarising, the naïve estimator is biased when effect sizes from primary studies with omitted variable bias are present in the meta-analysis. Moreover, this bias increases substantially for increasing proportions of effect sizes that suffer from omitted variable bias. The bias is not mitigated by a decrease in estimator variance, and the size of the statistical test is considerably larger than the critical level. The estimator with dummy variables is unbiased, but displays a slightly increasing variance at high proportions of misspecification. Although this increase in estimator variance negatively affects the size of the statistical test on the meta-estimate, the size is reasonably close to the critical level, while the power associated with the dummy model is not affected. This implies that any source of systematic variation or bias can be effectively controlled for in a meta-analysis by means of dummy variables. It also means that when certain sources of systematic bias are not controlled for, for example because they are not observed or because there are too few observations, propagation of errors from primary study misspecifications in meta-analysis may be quite substantial.

<<< *Insert Figure 1* >>>

#### 4.2 Impact of incorrect effect size measures

In this subsection we analyse the impact of incorrect effect size measures on the results of a meta-analysis. Similar to the approach in the previous subsection, the vertical represents the bias and mean squared error of the estimators, and size or power of the statistical tests. On the horizontal axis we measure the extent of primary study misspecification, representing the proportion of point-elasticities included in the meta-sample. As before, we systematically increase this proportion from 1% to 99%, with increments of 1%. In order to isolate the impact of point-elasticities, the number of studies containing omitted variable bias in the meta-sample is kept small and is fixed at 5%. Results are presented in Figure 2.

The inclusion of incorrect effect size measures has a substantial impact on the bias and mean squared error of the naïve meta-estimator. Also the size is larger than the nominal level if the percentage of misspecified results is smaller than 50%, but decreases as the proportion of erroneous effect size measures in the meta-analysis increases. The decrease in size implies that the increase in estimator variance, as indicated by the mean squared error, is more than offset by an increase in the estimated variance. The latter apparently has no consequences for the power in this case; the power of the test for both estimators is always equal to one (the black square in the figures is hidden behind the white square).

The inclusion of a dummy variable in the meta-model specification mitigates the impact of incorrect effect size measures on the bias of the meta-estimator. This holds for the mean squared error as well, except for high proportions of misspecification in the meta-analysis because of the limited potential for averaging out of random sampling error. The increase in actual variance, indicated by the increase in mean squared error, is more than offset by the increase in estimated variance, judging by the decrease in the size, but the power decreases rapidly from a certain point onwards.

In conclusion, the naïve meta-estimator is biased when erroneous effect sizes are included in the meta-analysis; the dummy estimator is not. The dummy estimator also outperforms the naïve estimator in terms of mean squared error, except for high proportions of misspecified studies in the meta-analysis. The size associated with the naïve estimator is not adequate, but converges to the critical 5% level due to widening confidence intervals for increasing proportions of misspecification.

<<< *Insert Figure 2* >>>

## 5. Response surface analysis

In order to arrive at more generally valid conclusions about the behaviour of the two meta-estimators, we proceed by analysing the bias, mean squared error and the size and power as dependent variables in a response-surface analysis. We estimate separate response surface regressions for each value of the true population effect size  $\beta_0$  and for each of the two meta-estimators. We use the procedure given in subsection 3.3 to sample primary study types into the meta-analysis. Using this procedure, the study characteristics that are randomly sampled are primary study sample size, effect size measure, model specification (omitted variable bias), and magnitude of primary data error variance. For details on the values of the parameters we refer to Section 3. The sample size of the meta-analysis takes on values of 50, 100 and 500, and the value of  $\lambda$ , determining the extent of bias when the primary model has omitted variables, is kept constant *within* each meta-analysis. Finally, we run 10,000 meta-analysis replications in order to obtain sufficient accuracy. The number of observations for each of the response surface regressions, resulting from the use of three different meta-analysis sample sizes, three values of  $\lambda$ , and 10,000 meta-analysis replications, is equal to 90,000.

The set of exogenous variables in each response surface contains a constant, two dummy variables for meta-analysis sample size equal to 50 and 100, implying that the effects of a meta-analysis sample size equal to 500 is subsumed in the constant term, and the average sample size of primary studies in the meta-sample divided by 1,000. We furthermore include the percentage of primary studies included in the meta-analysis with a specific error variance, and the percentage of point-elasticities in the meta-analysis. Similarly, for each value of  $\lambda$  we include the percentage of effect sizes affected by omitted variables. We also include the percentage of effect sizes affected by both omitted variables (regardless of the value of  $\lambda$ ) and incorrect functional form in order to analyse potential interaction effects between primary study misspecifications.<sup>11</sup>

Given the variation in data generation discussed above, we suspect that the residual term of the response surface regression is heteroskedastic, and that some clusters of covariances are

---

<sup>11</sup> Note that we actually coded the percentage of studies in which an effect size is affected by both incorrect functional form and omitted variable bias, implying that this variable measuring interaction effects is strictly speaking not an interaction term. The problems associated with interaction terms in non-linear models identified by Ai and Norton (2003) are therefore circumvented.

non-zero, simply because the data used for the meta-analyses within these clusters are similar. Both issues render OLS estimation inefficient. We define clusters according to two dimensions of variation. Specifically, we use meta-analysis sample size and the number of point-elasticities in a meta-analysis, measured in deciles, to determine the clusters. We estimate the response surface analyses on the bias and the mean squared error utilising the clustered Huber-White sandwich estimator, thereby simultaneously correcting for heteroskedasticity between clusters and for within-cluster dependence (see Williams, 2000; Wooldridge, 2002, Section 13.8.2). For size and power we estimate binary probit models and use a similar robustness correction.

The results, in terms of the estimated coefficients (also for the probit models), are presented in Table 2. The amount of explained variation for the naïve estimator is large, but is very limited for the dummy estimator, with an exception for the probit model on the power. Although at first glance this appears worrisome, it actually shows that the dummy estimator recovers the true underlying effect size so well that there is very little left to explain; the bulk of the variation in the results from the meta-analyses is due to random error. A related observation is that the dummy estimator outperforms the naïve estimator on most accounts. Specifically, the bias and mean squared error of the dummy estimator are substantially less affected by primary study misspecifications than the bias and mean squared error of the naïve estimator. As a result, the size associated with the dummy estimator is less affected by misspecifications as well, although the effects of omitted variables clearly increase with  $\lambda$ . This implies that including dummies goes a long way in controlling for error propagation of primary study misspecifications, but that an increase in effect sizes that suffer from omitted variables will still substantially increase the number of Type I errors made.

<<< *Insert Table 2* >>>

Although the dummy estimator largely picks up the systematic and isolated effects of primary study misspecifications, increasing the number of effect sizes that suffer from both misspecifications clearly has a large effect on the bias and especially on the mean squared error of the dummy estimator. Also the size increases substantially due to these interaction effects. These interactions can be controlled for in principle, i.e., by including a dummy in the meta-specification for those effect sizes that suffer from both misspecifications, but in practice there are two substantial problems. First, it is highly questionable whether interaction

effects are picked up by the additional dummy since mainly the mean squared error of our dummy meta-estimator used in this study is affected. Second, recognising a primary study misspecification in actual practice is difficult. Generally we only observe differences between primary studies without knowing, for example, the true underlying model or the correct model specification. This makes that the number of interactions to be tested increases exponentially with the number of control variables in the meta-analysis model.

Increasing the error variance in primary studies by and large has the expected effects, although results for the naïve estimator are counterintuitive. Results from a model that includes an interaction term between error variance and incorrect functional form show that error variance and primary model misspecifications may have strong combined effects, and coefficients on error variance for the naïve estimator are as expected in this model. The sample size of the meta-analysis has no impact on the bias of the meta-estimators, but substantially decreases estimator variance and estimated standard errors. This obviously leads to an increase in the power of the test for both estimators.<sup>12</sup> However, judging by the substantial increase in size of the statistical test, the decrease in variance of the naïve estimator is more than offset by the decrease in standard errors. This does not hold for the dummy estimator because of the more substantial decrease in estimator variance; the size of the test is not or only marginally affected. As expected, increasing the sample size of primary studies decreases the bias and mean squared error of the dummy estimator, leading to a decrease in size and an increase in the power. An unusual result is that the bias of the naïve estimator increases with increasing sample size in primary studies. We are likely picking up some unexplained variation in this case. Indeed, when we include an interaction term to control for the combined impact of erroneous effect size measures and primary study sample size, the coefficients on sample size show the expected signs. In conclusion, the results show that increasing primary study and meta-analysis sample size slightly enhance the performance of the naïve estimator, while they substantially decrease the variance of the dummy estimator.

---

<sup>12</sup> The probit model on POWER for the naïve estimator did not converge when the 'Meta-analysis sample = 100' dummy was included, because the dependent variable contained too little variation. We therefore excluded this dummy from the model, implying 'Meta-analysis sample size = 500' and 'Meta-analysis sample size = 100' are both subsumed in the constant term.

## 6. Conclusions

In this study we analyse the impact of common primary model misspecifications on the results of a meta-analysis, and the effectiveness of standard procedures that are used to correct for these issues. We induce two types of model misspecifications, i.e., omitted variables and incorrect functional form, and distinguish between two meta-estimators. The first estimator is a 'naïve' estimator, for which the meta-estimate is simply the average value of the effect sizes. The second estimator uses dummy variables to correct for observable misspecifications in primary studies. Although some of the effects can be obtained analytically, they are generally qualitative in nature. We use Monte Carlo simulations to obtain quantitative information on the direction and the magnitude of the effects under plausible parameter values. We assess the small sample performance of the estimators and use the bias and mean squared error of the estimators, and the size and power of the statistical tests on the meta-estimates, as performance indicators.

Our results show that, although the isolated effects of misspecifications on the naïve meta-estimator can be substantial, including dummy variables in the meta-model specification to correct for these aberrations goes a long way in reducing the effects of misspecifications on the bias and mean squared error of the meta-estimator. Results on the statistical tests show that the naïve meta-estimator is not useful since it virtually always rejects the null hypothesis that there is no real effect, regardless of whether this is true or not. The deteriorating performance of the dummy estimator for large proportions of effect sizes suffering from misspecification is purely due to random error, but clearly demonstrates that sufficient correct effect size measures are necessary for accurately measuring the 'true' effect. Meta-analyses for which this is not the case may produce estimates that are seriously off the mark. Response surface analyses also show that, although the dummy estimator picks up most of the negative effects of primary study misspecifications in isolation, the interaction between misspecifications may substantially and negatively affect outcomes of a meta-analysis as well. Error propagation of primary study misspecifications in meta-analysis may therefore be quite substantial, because in actual practice these complex interactions between observed primary study characteristics are difficult to control for. When meta-analysis is used to look for a 'true' effect rather than for analysing variation in outcomes, our results provide an argument for study selection on model quality to avoid the impact of error propagation in meta-analysis.

## Acknowledgements

This research is supported through the program ‘Stimulating the Adoption of Energy-Efficient Technologies’, funded by the Netherlands Organisation for Scientific Research (NWO) and the Dutch Ministry of Economic Affairs (SenterNovem). The usual disclaimer applies.

## References

- Ai C, EC Norton, 2003, Interaction Terms in Logit and Probit Models, *Economics Letters* **80**, 123–129.
- Bijmolt THA, RGM Pieters, 2001, Meta-Analysis in Marketing when Studies Contain Multiple Measurements, *Marketing Letters* **12**, 157–169.
- Dalhuisen JM, RJGM Florax, HLF de Groot, P Nijkamp, 2003, Price and Income Elasticities of Residential Water Demand: A Meta-Analysis, *Land Economics* **79**, 292–308.
- Disdier A, K Head, 2008, The Puzzling Persistence of the Distance Effect on Bilateral Trade, *Review of Economics and Statistics* **90**, 37–48.
- Doucouliafos H, M Paldam, 2009, The Aid Effectiveness Literature: The Sad Result of 40 Years of Research, *Journal of Economic Surveys* **23**, 433–461.
- Feld LP, JH Heckemeyer, 2011, FDI and Taxation: A Meta-Study, *Journal of Economic Surveys* **25**, 233–272.
- Field AP, 2001, Meta-Analysis of Correlation Coefficients: A Monte Carlo Comparison of Fixed- and Random-Effects Methods, *Psychological Methods* **6**, 161–180.
- Greene WH, 2000, *Econometric Analysis*, Prentice Hall International, Inc., Upper Saddle River, New Jersey.
- Hedges LV, 1994, Fixed Effects Models, in: H Cooper, LV Hedges (eds.), *The Handbook of Research Synthesis*, Russell Sage Foundation, New York, pp. 285–299.
- Keef SP, LA Roberts, 2004, The Meta-Analysis of Partial Effect Sizes, *British Journal of Mathematical and Statistical Psychology* **57**, 97–129.
- Kennedy P, 2003, *A Guide to Econometrics*, The MIT Press, Cambridge, Massachusetts.
- Kluve J, 2010, The Effectiveness of European Labor Market Programs, *Labour Economics* **17**, 904–918.
- Koetse MJ, HLF de Groot, RJGM Florax, 2011, A Meta-Regression Analysis of the Investment-Uncertainty Relationship, in: RJGM Florax, HLF de Groot, P Mulder (eds.),

- Improving Energy Efficiency Through Technology: Trends, Investment Behaviour and Policy Design*, Edward Elgar, Cheltenham, pp. 176–204.
- Koetse MJ, RJGM Florax, HLF de Groot, 2010, Consequences of Effect Size Heterogeneity for Meta-Analysis: A Monte Carlo Study, *Statistical Methods and Applications* **19**, 217–236.
- Kuhnert R, D Böhning, 2007, A Comparison of Three Different Models for Estimating Relative Risk in Meta-Analysis of 3 Clinical Trials Under Unobserved Heterogeneity, *Statistics in Medicine* **26**, 2277–2296.
- Nelson JP, PE Kennedy, 2009, The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment, *Environmental and Resource Economics* **42**, 345–377.
- Oswald FL, JW Johnson, 1998, On the Robustness, Bias, and Stability of Statistics from Meta-Analysis of Correlation Coefficients: Some Initial Monte Carlo Findings, *Journal of Applied Psychology* **83**, 164–178.
- Peterson RA, SP Brown, 2005, On the Use of Beta Coefficients in Meta-Analysis, *Journal of Applied Psychology* **90**, 175–181.
- Sanchez-Meca J, F Marin-Martinez, 1997, Homogeneity Tests in Meta-Analysis: A Monte Carlo Comparison of Statistical Power and Type I Error, *Quality & Quantity* **31**, 385–399.
- Sanchez-Meca J, F Marin-Martinez, 1998, Weighting by Inverse Variance or by Sample Size in Meta-analysis: A Simulation Study, *Educational and Psychological Measurement* **58**, 211–220.
- Sutton AJ, KR Abrams, TA Sheldon, F Song, 2000, *Methods for Meta-Analysis in Medical Research*, John Wiley and Sons, New York.
- Viscusi WK, JA Aldy, 2003, The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World, *Journal of Risk and Uncertainty* **27**, 5–76.
- Williams RL, 2000, A Note on Robust Variance Estimation for Cluster-Correlated Data, *Biometrics* **56**, 645–646.
- Wooldridge JM, 2002, *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge, Massachusetts.



Table 1. Sampling procedure example; generating sampling probabilities for 6 different study types  
(meta-analysis sample size = 100)

<i>Study type</i>	Sampling probability per <i>study characteristic</i>		Sampling probability per <i>study type</i>	Number of studies per <i>study type</i>
	P (OVB)	P (VAR)		
OVB – VAR 1	0.35	0.20	0.07	7
OVB – VAR 2	0.35	0.48	0.17	17
OVB – VAR 3	0.35	0.32	0.11	11
No OVB – VAR 1	0.65	0.20	0.13	13
No OVB – VAR 2	0.65	0.48	0.31	31
No OVB – VAR 3	0.65	0.32	0.21	21

Table 2. Response surface analysis on the bias, mean squared error and the size and power, for the naïve meta-estimator and the dummy meta-estimator, and for  $\beta_0 = 0$  and  $\beta_0 = 1$ <sup>a,b</sup>

Estimator Population effect size Exogenous ↓ \ Dependent variable →	Naïve meta-estimator						Dummy meta-estimator					
	$\beta_0 = 0$			$\beta_0 = 1$			$\beta_0 = 0$			$\beta_0 = 1$		
	BIAS	MSE	SIZE	BIAS	MSE	POWER	BIAS	MSE	SIZE	BIAS	MSE	POWER
Constant <sup>c</sup>	0.012 (3.66)	-0.133 (-31.2)	-0.520 (-7.78)	0.043 (8.62)	-0.133 (-22.6)	4.661 (19.8)	-0.005† (-0.88)	-0.190 (-3.48)	-1.517 (-22.0)	-0.006† (-1.15)	-0.200 (-3.43)	9.128 (13.2)
OVB $\lambda = 0$ (in %)	0.035 (7.27)	0.270 (35.4)	0.065† (0.39)	0.070 (9.16)	0.405 (23.3)	3.268 (4.14)	-0.074 (-6.98)	0.012† (0.19)	0.030† (0.34)	-0.034 (-3.69)	-0.006† (-0.09)	-6.534 (-9.06)
OVB $\lambda = 0.5$ (in %)	0.249 (154.7)	0.216 (66.3)	2.075 (20.3)	0.266 (121.6)	0.254 (51.9)	3.370 (7.99)	-0.017 (-4.43)	0.031† (1.00)	0.103 (2.50)	-0.003† (-0.71)	0.024† (0.72)	-3.278 (-9.24)
OVB $\lambda = 1.5$ (in %)	0.483 (243.8)	0.588 (134.9)	2.563 (20.8)	0.465 (121.7)	0.533 (50.1)	1.027 (3.57)	0.018 (4.92)	0.099 (4.58)	0.285 (10.9)	0.003† (0.81)	0.105 (4.56)	-2.360 (-9.97)
Point-elasticity (in %)	0.013 (3.14)	0.064 (10.2)	-0.337 (-4.10)	-0.189 (-60.2)	0.091 (12.7)	-1.768 (-6.00)	-0.029 (-3.74)	0.040† (0.45)	-1.859 (-12.2)	-0.008† (-1.31)	0.021† (0.20)	-8.222 (-10.4)
Point-elasticity AND OVB (in %)	-0.095 (-15.2)	-0.208 (-28.5)	-0.805 (-3.43)	-0.051 (-11.6)	-0.458 (-41.2)	-0.810† (-0.65)	0.155 (9.72)	0.407 (2.53)	1.630 (6.76)	0.068 (5.16)	0.487 (2.65)	7.470 (8.82)
Meta-analysis sample size = 100 (dummy)	0.000† (0.52)	0.010 (5.64)	-0.755 (-15.7)	-0.001† (-0.53)	0.009 (3.08)	-- (--)	0.004 (2.19)	0.087 (3.86)	-0.052† (-1.71)	0.000† (0.14)	0.095 (4.25)	-1.230 (-5.80)
Meta-analysis sample size = 50 (dummy)	0.002† (1.44)	0.026 (7.78)	-1.305 (-17.7)	0.000† (0.19)	0.025 (8.37)	-1.253 (-9.60)	0.001† (0.29)	0.173 (6.29)	-0.066 (-2.37)	0.001† (0.63)	0.185 (6.05)	-1.971 (-10.4)
Mean sample size primary studies (000's)	0.000 (4.26)	0.000† (1.12)	0.001 (14.7)	0.000 (8.19)	0.000† (0.66)	0.002 (12.1)	0.000 (-3.07)	0.000 (-4.94)	0.000 (-5.24)	0.000 (-4.74)	0.000 (-4.80)	0.001 (14.6)
Disturbance variance = 5 (in %)	-0.016 (-5.68)	-0.009 (-2.17)	-0.555 (-10.4)	-0.061 (-9.32)	-0.019 (-4.16)	-1.340 (-5.92)	0.027 (3.82)	0.143 (5.63)	0.150 (2.81)	0.029 (5.10)	0.158 (5.54)	-0.831 (-18.6)
Disturbance variance = 10 (in %)	-0.035 (-8.25)	-0.024 (-4.23)	-1.165 (-16.9)	-0.140 (-9.82)	-0.049 (-5.89)	-2.618 (-13.1)	0.048 (5.88)	0.304 (5.43)	0.177 (3.64)	0.055 (6.37)	0.318 (5.23)	-1.713 (-28.4)
NOBS	90,000	90,000	90,000	90,000	90,000	90,000	90,000	90,000	90,000	90,000	90,000	90,000
R <sup>2</sup> -adj. (McFadden for SIZE and POWER)	0.894	0.859	0.594	0.877	0.802	0.400	0.014	0.062	0.111	0.003	0.047	0.428

<sup>a</sup> The results for the bias and mean squared error are presented with t-statistics (in parentheses) based on clustered Huber-White corrected standard errors. Clusters are based on the sample size of the meta-analysis and deciles of the proportion of point-elasticities in the meta-analysis. The results for size and power are binary probit estimates, utilising the same corrective device for robustness. A † indicates that the coefficient is not statistically significant at a 5 percent critical significance level.

<sup>b</sup> % indicates the variable is operationalised as the percentage of primary studies included in the meta-sample for which the description holds.

<sup>c</sup> The effects of 'Meta-analysis sample size = 500' and 'Disturbance variance = 1 (%)' are subsumed in the constant term.