



TI 2007-052/3

Tinbergen Institute Discussion Paper

The Impact of Effect Size Heterogeneity on Meta-Analysis: A Monte Carlo Experiment

Mark J. Koetse¹

Raymond J.G.M. Florax²

Henri L.F. de Groot^{1,3}

¹ *Department of Spatial Economics, Vrije Universiteit Amsterdam;*

² *Department of Agricultural Economics, Purdue University, West Lafayette, USA;*

³ *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>.

The Impact of Effect Size Heterogeneity on Meta-Analysis: A Monte Carlo Experiment

Mark J. Koetse^{a,*}, Raymond J.G.M. Florax^{a,b}, Henri L.F. de Groot^{a,c}

^a *Department of Spatial Economics, Vrije Universiteit Amsterdam, The Netherlands*

^b *Department of Agricultural Economics, Purdue University, West Lafayette, United States*

^c *Tinbergen Institute, Amsterdam, The Netherlands*

Abstract

In this paper we use Monte Carlo simulation to investigate the impact of effect size heterogeneity on the results of a meta-analysis. Specifically, we address the small sample behaviour of the OLS, the fixed effects regression and the mixed effects meta-estimators under three alternative scenarios of effect size heterogeneity. We distinguish heterogeneity in effect size variance, heterogeneity due to a varying true underlying effect across primary studies, and heterogeneity due to a non-systematic impact of omitted variable bias in primary studies. Our results show that the mixed effects estimator is to be preferred to the other two estimators in the first two situations. However, in the presence of random effect size variation due to a non-systematic impact of omitted variable bias, using the mixed effects estimator may be suboptimal. We also address the impact of sample size and show that meta-analysis sample size is far more effective in reducing meta-estimator variance and increasing the power of hypothesis testing than primary study sample size.

JEL-codes: C12; C15; C40

Key words: Effect size heterogeneity; meta-analysis; Monte Carlo simulation; fixed effects regression estimator; mixed effects estimator

* Corresponding author: Department of Spatial Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands. Tel: + 31 (0)20 598 6168, Fax: + 31 (0)20 598 6004, E-mail: mkoetse@feweb.vu.nl.

1. Background

Meta-analysis is a form of research synthesis in which previously documented empirical results are combined or re-analysed in order to increase the power of statistical hypothesis testing. Some proponents maintain that meta-analysis can be viewed as *quantitative* literature review (Stanley, 2001), while others assert that meta-analysis can be used to pinpoint aspects critical to the future development of theory (Goldfarb, 1995; Rosenthal and DiMatteo, 2001). The method was originally developed and applied in experimental medicine, but soon extended to other academic areas. Meta-analysis is currently also gaining ground in economics. Important contributions in this field are, among others, Smith and Huang (1995), Card and Krueger (1995), Görg and Strobl (2001), Bateman and Jones (2003) and Weichselbaumer and Winter-Ebmer (2005). Although there has been a wide increase in its application, meta-analysis is still surrounded with various methodical difficulties. For example, in economics, data constraints as well as the desire to be ‘different’ lead to varying sets of control variables across studies, inducing omitted variable bias in at least a subset of the existing empirical studies. Moreover, since the true data generating process is most likely unknown, different effect size measures are reported in primary studies and these are pooled in a meta-analysis sample. In Koetse et al. (2005) Monte Carlo experiments are used in order to investigate the consequences of these two particular problems.

In this study we provide a more general analysis and aim to analyse the impact of effect size heterogeneity on the results of a meta-analysis. Specifically, we investigate heterogeneity in effect size variance, heterogeneity due to random variation of the true underlying effect across primary studies, and heterogeneity due to a non-systematic impact of omitted variable bias across primary studies. The first problem leads to inherent heteroskedasticity in the meta-analysis sample, with potential consequences for meta-estimator efficiency. The differences between the latter two issues is subtle and will be discussed in detail in the next section. We address the small sample behaviour of three estimators under these three situations. We first use a simple OLS estimator, which does not in any way control for effect size heterogeneity. The sec-

ond estimator is the fixed effects regression estimator which incorporates heterogeneity in effect size variance by weighting the meta-analysis data with the standard error of the estimates. The third and final estimator is the mixed effects estimator, which not only incorporates heterogeneity in effect size variance, but also accounts for potential random variation of the true underlying effect by estimating the variance of the underlying population. We use the bias, mean squared error and size and power as indicators of estimator performance.

The remainder of this paper is organised as follows. The next section discusses three sources of effect size heterogeneity in more detail. Section 3 describes the experimental design, while in Section 4 we present and discuss in detail the simulation results. Furthermore, in Section 5 we systematically increase the sample size of both the primary studies and the meta-analyses, which gives us the opportunity to draw inferences on the asymptotic properties of the estimators. Section 6 concludes.

2. Sources and characteristics of effect size heterogeneity

Heterogeneity in effect size precision and random variation in the true underlying effect across primary studies may have substantial consequences for the results of a meta-analysis. To illustrate the potential problems, let T_s be the estimate of the true effect size θ_s from primary study s . This estimate is generally assumed to be normally distributed, such that:

$$T_s \sim N(\theta_s, \sigma_s^2), \tag{1}$$

where σ_s^2 is generally referred to as the within-study variance. Within-study variance generally varies across primary studies, causing heteroskedasticity in a meta-analysis sample. Important sources of heterogeneity in within-study variance are differences in the sample sizes used in primary studies and differences in model specifications and data type. Ultimately, the consequences of heterogeneity in within-study variance on the results of a meta-analysis are potentially serious. Crucial is the fact that, assuming a standard OLS estimation, effect sizes with a higher variance get as much weight as effect sizes with a lower variance. Therefore, OLS is not

efficient, i.e., does not attain the minimum estimated variance, and the variance estimator is biased. The optimal way to correct for this problem is to weight the effect sizes with their respective variances. Since the actual effect size variance is unknown in practice, meta-analyses commonly use the effect size variance estimated by the primary model, which is a good approximation unless sample sizes in primary studies are exceptionally small (see Hedges, 1994, p. 287). A second problem is related to the characteristics of the true underlying effect size θ_s . After the *systematic* variation in effect sizes is controlled for by including dummy variables in the meta-model specification, basically two assumptions on the nature of the remaining *non-systematic* effect size variation exist. An often used assumption is that effect size variation is due solely to sampling error in the underlying primary studies, and that the true effect size θ_s is constant across primary studies, i.e., $\theta_s = \theta$. An alternative assumption is that the remaining variation is partly due to a random variation of the true underlying effect size across primary studies, such that:

$$\theta_s \sim N(\theta, \tau^2), \tag{2}$$

where τ^2 represents the variance of the underlying population, generally referred to as the between-study variance. Third, a difficulty associated with non-systematic variation in a meta-analysis sample is that it is unclear a priori whether it is due to random variation of the true effect across primary studies, or due to a non-systematic impact of misspecifications in primary studies. For instance, it is very likely that the bias in effect sizes due to omitted variables in primary studies is different for every primary study. This means that part of the omitted variable bias is systematic and may be picked up by a dummy variable, and that part of the bias in the meta-analysis sample is random. The difference between random variation due to omitted variable bias and random variation due to a random varying true underlying effect is not the fact that the source of the random variation is different. In fact, after controlling for the systematic part of the effect size variation, the result in both situations is a random effect size distribution around zero. The difference lies in the fact that random variation of the true underlying effect

causes randomness of each effect size in the meta-analysis sample, whereas random variation due to omitted variable bias only causes randomness of effect sizes from misspecified primary studies. Since this may have serious consequences for the optimal weight structure of a meta-estimator, our goal is to investigate whether the two sources of random effect size variation have different consequences for the results of a meta-analysis.

3. Experimental design

The data generation process (DGP) of our simulation exercises follows closely the DGP of Koetse et al. (2005), and consists of four steps: generating the primary data; estimating the primary models; performing the meta-analyses using the estimated effect sizes and characteristics of the primary studies as inputs; analysing the small sample performance of the meta-estimators. These four steps are discussed in detail below.¹

3.1 Generating the primary data

The true underlying primary model is an 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, α , β_0 and β_1 are parameters, and ε is an error term. In our model, β_0 is the parameter of interest, i.e., the true underlying effect. We draw β_0 randomly from a normal distribution with mean μ and between-study variance τ^2 , and set μ equal to 1 and 0 in order to analyse the cases with and without true effect. We set both α and β_1 equal to 1, while the error term ε is normally distributed with mean 0 and variance σ^2 . Furthermore, the sample size of the primary model is fixed at 500 and the number of replications for each primary study combination is 5,000. The variable x is generated, once, according to:

¹ The computer programs used for the analyses in this paper are written in Gauss 8.0, and are available upon request from the authors.

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

where ϑ is drawn from a uniform (0,1) distribution. In order to be able to induce omitted variable bias in a primary study we relate x to z by generating z according to:

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

where λ is a parameter and ψ is an error term drawn from a uniform (0,1) distribution (ψ , ϑ and ε are independent). Note that the potential bias induced in the estimate of β_0 when z is excluded from the primary model does not only increase with the correlation coefficient, but also with the variance of z (see Koetse et al., 2005). Obviously, when $\lambda = 0$, the correlation between x and z is zero, implying that the bias in β_0 when z is excluded from the primary model is zero as well. However, when we increase the value of λ , both the correlation between x and z and the variance of z are increased, thereby invariantly increasing the bias in the estimate of β_0 . In fact, the bias is proportional to λ .

The main issues analysed in this paper revolve around effect size heterogeneity. First, we increase heterogeneity in within-study variance via the error term in primary studies. This error term is normally distributed with mean 0 and variance σ^2 , which we vary systematically between 1 and 10 with increments of 1. Second, we investigate the cases where the true effect size β_0 is both fixed and random across primary studies. In the former case, β_0 is fixed and constant within each meta-analysis, so we set between-study variance $\tau^2 = 0$. In the random effect case we set $\tau^2 > 0$. Note that in this case τ^2 is fixed within a meta-analysis, implying not that β_0 is fixed, but that the distribution from which the true underlying effect is drawn is identical for each effect size within a single meta-analysis. Therefore, in order to investigate the impact of between-study variance on the results of a meta-analysis, we vary τ^2 systematically across (not within) meta-analyses, varying its value between 0 and 2 with increments of 0.2. The third issue deals with effect size heterogeneity due to a non-systematic impact of omitted variables across primary studies. We implement this issue by systematically varying λ , the parameter that determines the amount of bias due to omitted variables in a primary study. Specifically,

we draw λ from a normal distribution with mean 1 and variance υ^2 . The latter takes on a value of 0 when the effect of omitted variables is purely systematic, and a value of 4 when part of the impact of omitted variables is random. Further details on the DGP's that are used to analyse the three issues described above are given in the relevant subsections in Section 4.

3.2 Estimating the primary models

Our approach is different from other Monte-Carlo studies in meta-analysis (see, e.g., Oswald and Johnson, 1998; Sanchez-Meca and Marin-Martinez, 1997, 1998; Bijmolt and Pieters, 2001; Field, 2001; Kuhnert and Böhning, 2007) in that we explicitly incorporate the stage of the *primary data analysis*. Besides the fact that this allows us to explicitly introduce omitted variable bias in primary studies, we may also introduce erroneous effect size operationalisations and assess their impact 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.² The log-linear model is given by:

$$\ln(y) = \hat{\alpha} + \hat{\beta}_0 \ln(x) + \hat{\beta}_1 \ln(z) + \hat{\varepsilon}. \quad (6)$$

We estimate this model by OLS, which produces $\hat{\alpha}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ as estimates of α , β_0 and β_1 , respectively. The parameter of interest is the double-log elasticity of $\ln(y)$ on $\ln(x)$, given by $\eta = \hat{\beta}_0$. This elasticity is correctly estimated given our data generating process; by construction, it is constant across the entire primary data-set. The standard error of the elasticity is simply the standard error of $\hat{\beta}_0$. In order to induce omitted variable bias we use two primary model specifications, i.e., the correctly specified primary model in equation (6) and a misspecified version of this model from which $\ln(z)$ is excluded as an explanatory variable. The latter model induces omitted variable bias in $\hat{\beta}_0$ when $\lambda \neq 0$. An alternative elasticity estimate is obtained by estimating the linear primary model specification, which reads looks as:

² Of course, the choice of the true underlying model is rather arbitrary, i.e., we also could have chosen the linear model as the true underlying model. However, we see no reason why the results presented later on in this paper would change when our choice of true underlying model would have been different.

$$y = \hat{\alpha}^* + \hat{\beta}_0^* x + \hat{\beta}_1^* z + \hat{\varepsilon}^*. \quad (7)$$

Using OLS to estimate this model produces $\hat{\alpha}^*$, $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$ as estimates of α , β_0 and β_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, for say primary study m , as $\eta_m = \hat{\beta}_{0m}^* (\bar{x}_m / \bar{y}_m)$. In reality the estimation of different effect size measures may occur frequently, simply because the true underlying model is unknown and researchers may assume an erroneous model specification. The argument for using the ratio of mean values as the evaluation point is that most primary studies that estimate a point-elasticity do this at the sample mean.³ To calculate the standard error of this elasticity we use the Delta method (see Greene, 2000, p. 359-360), which in this case means that for primary study m we have $se(\eta_m) = se(\hat{\beta}_{0m}^*) (\bar{x}_m / \bar{y}_m)$. As before, in order to induce omitted variables bias we use the model specification in equation (7) and a specification from which z is excluded as an explanatory variable. The latter model induces omitted variable bias in $\hat{\beta}_0^*$ when $\lambda \neq 0$.

3.3 Specification of the meta-estimators

The primary aim in this paper is to compare the small sample performance of three meta-estimators under the three regimes of effect size heterogeneity introduced in subsection 3.1. The elasticities produced by the primary model estimations are used as the dependent variable in our meta-analyses. The amount of primary study misspecification in a meta-analysis sample is set at a moderate level; both the proportion of point-elasticities and the proportion of effect sizes from studies with omitted variables bias in the meta-analysis is fixed at 50%. We furthermore perform separate analyses for $\mu = 0$ and $\mu = 1$. Within these restrictions the elasticities are randomly sampled from the 5,000 primary study replications. Finally, the meta-analysis sample size is 50 and the number of meta-analysis replications is equal to 10,000. Our first and simplest

³ A common alternative is to use the median of the data on x and y , but, given the fact that the choice for the point of evaluation in the data is arbitrary to a certain extent, other points in the data-set are valid as well.

model is a meta-regression model with dummy variables in order to correct for primary study misspecifications. This model is given by:

$$\eta_s = \hat{\delta}_0^1 + \hat{\delta}_1^1 D_s^{pe} + \hat{\delta}_2^1 D_s^{ov} + \hat{\xi}_s^1, \quad (8)$$

where η_s is a vector of elasticities, D_s^{pe} is a dummy variable equal to one if the elasticity is a point-elasticity, D_s^{ov} is a dummy variable equal to one if the primary study is estimated without z among the explanatory variables, and $\hat{\delta}_0^1$, $\hat{\delta}_1^1$ and $\hat{\delta}_2^1$ are the estimated parameters. The model is estimated by OLS, with $\hat{\delta}_0^*$ an estimate of the true underlying effect μ . Furthermore, $\hat{\delta}_1^*$ and $\hat{\delta}_2^*$ are the estimated parameters on the dummy variables that should pick up the systematic impact of point-elasticities and omitted variable bias.

We subsequently test the performance of the fixed effects regression estimator and the mixed effects estimator, which are used to account for inherent heteroskedasticity in meta-analysis. The way in which these estimators account for this is by weighting the meta-analysis data with a measure of effect size precision, the ideal measure being the within-study variance. However, since the true within-study variances are unknown, the estimated variances of the primary study effect sizes are generally used for this purpose. The fixed effects regression model is given by (see Sutton et al., 2000a):

$$\eta_s/w_s = \hat{\delta}_0^2 (1/w_s) + \hat{\delta}_1^2 (D_s^{pe}/w_s) + \hat{\delta}_2^2 (D_s^{ov}/w_s) + \hat{\xi}_s^2/w_s, \quad (9)$$

where w_s is the weight of the effect size from study s , given by the standard error of the elasticity. The transformed model is estimated by OLS, producing $\hat{\delta}_0^2$ as an estimate of the true underlying effect μ , and $\hat{\delta}_1^2$ and $\hat{\delta}_2^2$ as parameter estimates on the dummy variables. Since the estimator is slightly different from the standard fixed effects regression estimator in meta-analysis, a modification of the resulting standard errors is necessary (see Hedges, 1994). The correct standard errors are given by $se^* = se/\sqrt{msr}$, where se is the standard error of the estimated meta-effect given by the computer program, and msr is the mean squared residual of the meta-analysis (see Hedges, 1994, p. 296).

The third meta-model is the mixed effects model. The difference between this model and the fixed effects regression model is that the latter assumes that the true underlying effect size is a fixed effect, whereas the mixed effects model assumes that the true effect size varies between primary studies and is drawn from a population of effect sizes with mean μ and between-study variance τ^2 . The mixed effects model makes an explicit distinction between within-study variance and between-study variance, which has obvious consequences for the model's weight structure. Since the between-study variance τ^2 is unknown it has to be estimated by the model. For this purpose we use a maximum likelihood estimator (see Sutton et al., 2000; Brockwell and Gordon, 2001). The log-likelihood is given by:

$$\text{LogL} = -0.5 \sum_{s=1}^S \left[\left(\eta_s - \hat{\delta}_0^3 - \hat{\delta}_1^3 D_s^{pe} - \hat{\delta}_2^3 D_s^{pv} \right)^2 / (\hat{\tau}^2 + w_s^2) + \ln(\hat{\tau}^2 + w_s^2) \right], \quad (10)$$

where $\hat{\delta}_0^3$, $\hat{\delta}_1^3$, $\hat{\delta}_2^3$ and $\hat{\tau}^2$ are the estimated parameters. The variable of interest $\hat{\delta}_0^3$ is an estimate of the mean μ of the underlying population of true effect sizes, and $\hat{\tau}^2$ is an estimate of the effect size population variance τ^2 . Observe that the model in equation (10) reduces to the model in equation (9) when $\hat{\tau}^2 = 0$.⁴

3.4 Assessing small sample performance

The parameters of interest are the true underlying effect size μ and the meta-estimates $\hat{\delta}_0^1$, $\hat{\delta}_0^2$ and $\hat{\delta}_0^3$. The central issue is now how well the meta-estimators recover the value of the population effect size μ , in terms of both size and statistical significance, in the presence of effect size heterogeneity. Effect size heterogeneity may affect the meta-estimates on several dimensions. 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 μ . Though the impact of misspecifications on the effect sizes may average out, in which

⁴ In this paper we induce systematic variation in the underlying effect size due to omitted variable bias and different elasticity measures. We therefore use the *fixed effect regression model* and the *mixed effects model*. These models' counterparts, i.e., models that assume that there is no systematic variation, are generally referred to as the *fixed effects model* and the *random effects model*, respectively.

case estimator bias is equal to zero, the variance of the estimators may still be substantial. We therefore 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 estimators, and measures the average distance of the estimate to the true parameter, i.e., the smaller the MSE, the closer the estimate will be to the true parameter, on average. The third and final indicator is the proportion of statistically significant results (SIG) of the meta-estimators. Formally, for $\hat{\delta}_0^1$ these indicators are given by:

$$\text{BIAS}(\hat{\delta}_0^1) = E(\hat{\delta}_0^1 - \beta_0) \approx \frac{1}{R} \sum_{r=1}^R (\hat{\delta}_0^1 - \mu)_r, \quad (11)$$

$$\text{MSE}(\hat{\delta}_0^1) = E(\hat{\delta}_0^1 - \mu)^2 = \text{BIAS}(\hat{\delta}_0^1)^2 + \text{var}(\hat{\delta}_0^1) \approx \frac{1}{R} \sum_{r=1}^R (\hat{\delta}_0^1 - \mu)_r^2, \quad (12)$$

$$\text{SIG}(\hat{\delta}_0^1) = \frac{1}{R} \sum_{r=1}^R I(|t_{n-k}| > t_{crit})_r, \quad (13)$$

where $r = 1, 2, \dots, R$ indexes the meta-analyses replications.⁵ In equation (13) 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 $\mu = 0$ and $H_0: \mu = 0$, we are interested in the probability of a Type I error, i.e., the probability that an estimator erroneously rejects H_0 . Therefore, when $\mu = 0$, SIG corresponds to the proportion of Type I errors. From now on we will refer to this as the size of the statistical test on the meta-estimates. Alternatively, when $\mu = 1$, 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 $\mu = 1$, SIG corresponds to (1 – probability of a Type II error), or the power of the statistical test. Since erroneously rejecting the null-hypothesis requires a considerably larger confidence interval than erroneously accepting the null-hypothesis, the two indicators are not reciprocal and provide different types of in-

⁵ The performance indicators for $\hat{\delta}_0^2$ and $\hat{\delta}_0^3$ are obtained by replacing $\hat{\delta}_0^1$ by $\hat{\delta}_0^2$ and $\hat{\delta}_0^3$ in equations (11), (12) and (13).

formation on statistical significance. This is the most important reason why we distinguish between a zero ($\mu = 0$) and a non-zero ($\mu = 1$) true underlying effect size. However, the two tests are clearly related, since decreasing standard errors simultaneously cause a decrease in size and an increase in power, *ceteris paribus*.

4. Simulation results

In this section we analyse the performance of the three meta-estimators under various forms of effect size heterogeneity. In subsection 4.1 we analyse the impact of increasing degrees of heteroskedasticity in the meta-analysis sample. We address the consequences of increasing between-study variance in subsection 4.2, while subsection 4.3 investigates the impact of non-systematic effects of omitted variables in primary studies.

4.1 Increasing heterogeneity in within-study variance

In this subsection we analyse the impact of increasing primary study error variance and of increasing the heterogeneity of primary study error variance (heteroskedasticity) on the results of a meta-analysis. In the experimental design we only vary the primary study error variance and keep constant all other parameters. Specifically, between-study variance $\tau^2 = 0$ and omitted variable bias is constant across primary studies, i.e., $\lambda = 1$ and $\nu^2 = 0$, in which case the mixed effects estimator should reduce to the fixed effects regression estimator. Our design is such that primary study estimator variance is comparable to actual practice. Primary studies are estimated with an error variance ranging from 1 to 10, with increments of 1. For these error variance values, correctly specified primary studies display average R^2 values ranging from 0.38 to 0.06, respectively. In our opinion these R^2 values are reasonable compared to the values found in many areas of economic research.

In the figures below the vertical axis represents the bias and mean squared error of the estimators and the size or power of the statistical tests on the meta-estimates. Along the horizontal axis we measure the degree of heteroskedasticity. We distinguish between ten cases. The first

case is the case with no heteroskedasticity; all effect sizes in the meta-analysis are drawn from primary studies with error variance 1. From the second case up to the tenth case we systematically increase the average error variance and the degree of heteroskedasticity, by systematically increasing the proportion of effect sizes drawn from studies with a higher error variance by 10%. In Table 1 we present the resulting proportions of effect sizes drawn from studies with a prespecified error variance for each of the ten cases. Note that for each case both the average effect size variance and the degree of heteroskedasticity are higher than in the previous cases.

<<< *Insert Table 1* >>>

In Figure 1 we present the performance of the three estimators on the three indicators for the ten cases, representing, simultaneously, an increasing degree of heterogeneity in within-study variance and an increasing average error variance. The fixed effects regression and the mixed effects estimator produce identical result, with a small difference in the size of the statistical tests on the meta-estimates. Apparently, the mixed effects model correctly estimates a zero between-study variance, in which case it reduces to the fixed effects regression estimator. Considering the fact that the horizontal axis also represents an increase in the average error variance of primary studies, the figure shows that increasing effect size variance has no systematic impact on the bias of the meta-estimate, which is in line with theory. It also systematically increases the variance of all three estimators, which is clear both from the increase in the mean squared error and from the decrease in power. The increase in fixed effects regression and mixed effects variance is limited, however.

Most interesting is that under increasing heteroskedasticity the variance of OLS deteriorates rapidly *vis-à-vis* the variance of the fixed effects regression and the mixed effects estimator. However, judging by the size, this is more than compensated by the fact that OLS produces

wider confidence intervals.⁶ The power, on the other hand, is not affected. Not having the standard errors of effect sizes in a meta-analysis precludes the use of fixed effects regression and mixed effects models. In conclusion, since OLS is highly inefficient under effect size heterogeneity, not having the standard errors of effect sizes in a meta-analysis may have serious consequences.

<<< *Insert Figure 1* >>>

4.2 Increasing between-study variance

In this subsection we introduce a random effect size and systematically increase the variance of the random effect size population. Specifically, we increase between-study variance τ^2 from 0 to 2 with increments of 0.2. With respect to heteroskedasticity we replicate the situation in the tenth case in the previous subsection, i.e., maximum heteroskedasticity and average error variance. Values of other variables and parameters remain unchanged. The results of increasing between-study variance are presented in Figure 2. The vertical axis again measures the bias and mean squared error of the estimators and the size or power of the statistical tests on the meta-estimates. The horizontal axis measures the absolute value of the between-study variance.

Increasing between-study variance has no systematic impact on the bias while estimator variance increases substantially, judged by the increase in mean squared error and decrease in power for each of the three estimators. Although the fixed effects model uses erroneous weights when between-study variance is larger than zero, the effects of this weight structure on estimator variance are not clear *a priori* (see Koetse, 2006, p. 65). Our results clearly show that the variance of the fixed effects regression estimator increases *vis-à-vis* the mixed effects estimator variance. Note also that the size associated with the OLS and mixed effects estimators is around its nominal level, and that the increase in size associated with the fixed effects regression model is a result of the increase in the mean squared error and the narrow confidence intervals pro-

⁶ See also Higgins and Thompson (2004) for an analysis of Type I error rates on *non-relevant* study characteristics under various sources of effect size heterogeneity.

duced by this estimator. The fact that the power associated with the mixed effects model deteriorates *vis-à-vis* its fixed effects counterpart, is a direct result of the fact that the latter estimator produces substantially narrower confidence intervals under increasing between-study variance.

A somewhat surprising result at first sight is that the mean squared error of and the size associated with the OLS estimator slowly converge to their mixed effects counterparts. Although this may seem strange, the result follows directly from a comparison of the weight structures used in the estimators. When between-study variance increases, its magnitude relative to within-study variance increases as well. As a consequence, within-study variance becomes less and less important in the weight structure of the mixed effects model. The central point is now that between-study variance is equal for each effect size in the meta-analysis, implying that, under increasing between-study variance, the weight structure of the mixed effects model tends towards a structure in which each effect size gets an equal weight. Since the OLS estimator gives each effect size an equal weight by definition, the estimates produced by the two estimators converge under increasing between-study variance. Also the size associated with OLS is smaller than its mixed effects counterpart. Since the mean squared error of the OLS estimator is higher in all circumstances, this implies that OLS confidence intervals are substantially wider than mixed effects confidence intervals.

<<< *Insert Figure 2* >>>

4.3 Non-systematic impact of omitted variable bias

As discussed in Section 2, effect size variation may be caused by other factors than pure random variation of the true effect across primary studies. Up till now we have assumed that the bias due to omitted variables, if present, is constant across primary studies. The necessary conditions for this assumption to hold in reality are implausible at least. In this subsection we therefore alter this assumption. For each primary-study replication we draw λ , the parameter that determines the amount of bias due to omitted variables in primary studies, from a normal distribution

with mean 1 and variance v^2 , which we fix at 4. This means that part of the omitted variable bias is systematic, which should be picked up by the dummy variable D^{ov} , and that part of the bias in the meta-analysis sample is random. The difference between random effect size heterogeneity due to omitted variable bias and the random effect size heterogeneity introduced in the previous subsection is not due to the fact that the sources of random effect size variation are different. In fact, after controlling for the systematic part of the effect size variation, the result in both situations is a random effect size distribution around zero.⁷ The difference lies in the fact that random variation of the true underlying effect causes randomness of each effect size in the meta-analysis sample, whereas random variation due to omitted variable bias only causes randomness of effect sizes from misspecified primary studies. Potential differences between the two sources of random effect size variation should therefore show when we vary the proportion of effect sizes with omitted variable bias in the meta-analysis sample.⁸ We systematically increase this proportion from 0.05 to 0.95 with increments of 0.1. We induce maximum heteroskedasticity, set between-study variance $\tau^2 = 0$, and meta-analysis sample size is 150 for purposes of presentation (results are identical for smaller and larger sample sizes). Values of other variables and parameters remain unchanged. Results are shown in Figure 3.

The bias of the OLS and mixed estimators is affected slightly when the true underlying effect is equal to one, and OLS variance is substantially higher than the variance of the other two estimators. For relatively small proportions of effect sizes with omitted variable bias in the meta-analysis sample, the mixed effects variance is slightly below the fixed effects regression variance. However, when the proportion of biased effect sizes increases above the 50% level, the variance of the mixed effects estimator starts to increase relative to its fixed effects regres-

⁷ The systematic part of the variation under random effect size heterogeneity due to omitted variable bias is picked up by D^{ov} , while under random variation of the true underlying effect it is picked up by the constant in the meta-model.

⁸ Note that the results and patterns identified in the previous subsection do not change when we vary the proportion of effect sizes with omitted variable bias. Therefore, if the patterns found in this subsection are dependent on this proportion, we can conclude that the two sources of random effect size variation have different consequences for the small sample performance of the three meta-estimators.

sion counterpart. Under this regime of effect size heterogeneity, the mixed effects estimator erroneously assigns the estimated between-study variance to all estimates, and effect sizes from correctly specified primary models get a weight that is too low. Finally, the size associated with the fixed effects model is still substantially above the nominal level, whereas the power is again superior to the power associated with the OLS and mixed effects estimators, which decreases rapidly at high proportions of effect sizes with omitted variable bias in the meta-analysis sample. The latter can only be partly attributed to an increasing estimator variance, implying that the estimated variance of both OLS and mixed effects is substantially upwards biased in these situations. Given the fact that the source of random effect size variation is not known empirical applications, our findings show that, under circumstances that are not uncommon in reality, using the mixed effects estimator may not be optimal.

<<< *Insert Figure 3* >>>

5. Impact of sample size

Since both primary study sample size and meta-analysis sample size may go to infinity, there are two types of asymptotics to meta-estimators (see Hedges and Olkin, 1985). Although the total sample size, i.e., the sum of all primary study sample sizes, may remain unchanged, primary study and meta-analysis sample size may have totally different effects on the results of a meta-analysis. In this section we therefore analyse meta-estimator performance under increasing primary study and meta-analysis sample size.

First, we systematically increase the sample size of the primary studies from 100 to 1000, with increments of 100, and fix the meta-analysis sample size at 25. Second, we increase meta-analysis sample size systematically from 25 to 250, with increments of 25, and keep primary study sample size fixed at 100. We thus can distinguish between ten cases with varying primary study and meta-analysis sample size, but with an equal number of total underlying observations

in each case.⁹ This way we can clearly observe the differential impact of meta-analysis sample size and primary study sample size on the results of a meta-analysis. For simplicity we only present results for the mixed effects estimator, since the patterns for the three meta-estimators are identical. All primary studies have an error variance of 5, i.e., there is no heteroskedasticity, and between-study variance τ^2 is equal to 2. We keep the impact of omitted variable bias fixed across primary studies ($\lambda = 1$ and $\nu^2 = 0$) and the proportion of effect sizes with omitted variable bias and point-elasticities in the meta-analysis sample is 0.5 in both cases. Results are presented in Figure 4.

The figure convincingly shows that increasing the sample size of a meta-analysis is far more effective in reducing the variance of the estimators and narrowing down the confidence intervals. The reason for this result is that deviations of effect sizes from their true underlying value are more and more averaged out when the sample size of the meta-analysis increases. Although these deviations also decrease when the sample size in a primary study increases, they are averaged out to a far lesser extent when the sample size of the meta-analysis remains relatively small. Of course, these results do not imply that the sample size of primary studies does not matter for the outcome of a meta-analysis – it does (especially at very small sample sizes). However, the results do show that relatively large meta-analyses with underlying primary studies with a relatively small number of observations are more efficient and produce narrower confidence intervals than relatively small meta-analyses with underlying studies with a relatively small number of observations.

<<< *Insert Figure 4* >>>

⁹ For instance, in the first case, primary study sample size is 100 and meta-analysis sample size is 25 in both situations, resulting in 2,500 underlying observations. In the second to tenth case primary study sample size increases with 100 under increasing primary study sample size, while meta-analysis sample size increases with 25 under increasing meta-analysis sample size. Therefore, with each case the number of total underlying observations increases with 2,500 under both regimes.

6. Discussion and conclusions

This paper uses Monte-Carlo simulation to investigate the impact of effect size heterogeneity on the results of a meta-analysis. Specifically, we address the performance of the OLS, the fixed effects regression and the mixed effects meta-estimators under three sources of effect size heterogeneity, i.e., heterogeneity in effect size variance, heterogeneity due to a varying true underlying effect across primary studies, and heterogeneity due to a non-systematic impact of omitted variable bias across primary studies.

Our results show that increasing heterogeneity in effect size variance has a detrimental effect on the performance of the OLS estimator compared to the other two estimators. Although the bias is not systematically affected, especially the variance of the OLS estimator deteriorates *vis-à-vis* the variance of the other two estimators. This pattern changes considerably when we allow the true underlying effect to vary randomly across primary studies. Increasing the variance of the population of random effect sizes increases the variance of all three estimator, but especially the variance of the fixed effects estimator, which deteriorates rapidly *vis-à-vis* the variance of the other two estimators. Fixed effects also has a downward biased variance estimator and produces too narrow confidence intervals. This leads to a size that is way off and only a slightly larger power. Alternatively, when random effect size variation is due to a non-systematic impact of omitted variable bias, the mixed effects variance increases *vis-à-vis* its fixed effects regression counterpart for increasing proportions of misspecification. In addition, although the size of test on the fixed effects regression estimate is still above the nominal level, the power of the test on the mixed effects estimate decreases rapidly for very high proportions of misspecification. Since the source of random effect size variation is unknown in reality, our findings show that using the mixed effects estimator in empirical applications of meta-analysis is not uncontested.

Finally, meta-analysis sample size is far more effective in reducing meta-estimator variance than primary study sample size. We show that even for relatively small increases in meta-analysis sample size, the quality of the outcome of a meta-analysis is substantially improved,

even when effect size heterogeneity is high. The crucial factor here is that random effect size deviations from the true underlying effect are averaged out more and more under increasing meta-analysis sample size. Therefore, although the various types of effect size heterogeneity may have substantial detrimental effects on the small sample performance of meta-estimators, effect size deviations from the true underlying effect average out at sample sizes that are common in practice.

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).

References

- Bateman IJ, Jones AP. 2003. Contrasting Conventional with Multi-Level Modeling Approaches to Meta-Analysis: Expectation Consistency in U.K. Woodland Recreation Values. *Land Economics* **79**: 235–258.
- Bijmolt THA, Pieters RGM. 2001. Meta-Analysis in Marketing when Studies Contain Multiple Measurements. *Marketing Letters* **12**: 157–169. DOI: 10.1023/A:1011117103381.
- Brockwell SE, Gordon IR. 2001. A Comparison of Statistical Methods for Meta-Analysis. *Statistics in Medicine* **20**: 825–840. DOI: 10.1002/sim.650.
- Card D, Krueger AB. 1995. Time-Series Minimum-Wage Studies: A Meta-Analysis. *American Economic Review* **85**: 238–243.
- Field AP. 2001. Meta-Analysis of Correlation Coefficients: A Monte Carlo Comparison of Fixed- and Random-Effects Methods. *Psychological Methods* **6**: 161–180.
- Goldfarb RS. 1995. The Economist-as-Audience Needs a Methodology of Plausible Inference. *Journal of Economic Methodology* **2**: 201–222.
- Görg H, Strobl E. 2001. Multinational Companies and Productivity Spillovers: A Meta-Analysis. *Economic Journal* **111**: 723–740. DOI: 10.1111/1468-0297.00669.
- Hedges LV. 1994. Fixed Effects Models. In *The Handbook of Research Synthesis*, Cooper H, Hedges LV (eds). Russell Sage Foundation: New York.
- Hedges LV, Olkin I. 1985. *Statistical Methods for Meta-Analysis*. Academic Press: Orlando, Florida.
- Higgins JPT, Thompson SG. 2004. Controlling the Risk of Spurious Findings from Meta-Regression. *Statistics in Medicine* **23**: 1663–1682. DOI: 10.1002/sim.1752.
- Koetse MJ. 2006. Determinants of Investment Behaviour: Methods and Applications of Meta-Analysis. Dissertation Vrije Universiteit. Tinbergen Institute Research Series no. 374. Vrije Universiteit: Amsterdam.

- Koetse MJ, Florax RJGM, de Groot HLF. 2005. Correcting for Primary Study Misspecifications in Meta-Analysis. Tinbergen Institute Discussion Paper 05-029/3. Tinbergen Institute: Amsterdam.
- Kuhnert R, Böhning D. 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.
- Oswald FL, Johnson JW. 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.
- Rosenthal R, DiMatteo MR. 2001. Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews. *Annual Review of Psychology* **52**: 59–82. DOI: 10.1146/annurev.psych.52.1.59.
- Sanchez-Meca J, Marin-Martinez F. 1997. Homogeneity Tests in Meta-Analysis: A Monte Carlo Comparison of Statistical Power and Type I Error. *Quality & Quantity* **31**: 385–399. DOI: 10.1023/A:1004298118485.
- Sanchez-Meca J, Marin-Martinez F. 1998. Weighting by Inverse Variance or by Sample Size in Meta-analysis: A Simulation Study. *Educational and Psychological Measurement* **58**: 211–220. DOI: 10.1177/0013164498058002005.
- Smith VK, Huang JC. 1995. Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models. *Journal of Political Economy* **103**: 209–227.
- Stanley TD. 2001. Wheat from Chaff: Meta-Analysis as Quantitative Literature Review. *Journal of Economic Perspectives* **15**: 131–150.
- Sutton AJ, Abrams KR, Sheldon TA, Song F. 2000. *Methods for Meta-Analysis in Medical Research*. John Wiley and Sons: New York.
- Weichselbaumer D, Winter-Ebmer R. 2005. A Meta-Analysis of the International Gender Wage Gap. *Journal of Economic Surveys* **19**: 479–511. DOI: 10.1111/j.0950-0804.2005.00256.x.

Table 1: Proportion of effect sizes from primary studies with a pre-specified error variance in ten different cases

| Case | Value of error variance | | | | | | | | | |
|------|-------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 100% | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 2 | 90% | 10% | -- | -- | -- | -- | -- | -- | -- | -- |
| 3 | 80% | 10% | 10% | -- | -- | -- | -- | -- | -- | -- |
| 4 | 70% | 10% | 10% | 10% | -- | -- | -- | -- | -- | -- |
| 5 | 60% | 10% | 10% | 10% | 10% | -- | -- | -- | -- | -- |
| 6 | 50% | 10% | 10% | 10% | 10% | 10% | -- | -- | -- | -- |
| 7 | 40% | 10% | 10% | 10% | 10% | 10% | 10% | -- | -- | -- |
| 8 | 30% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | -- | -- |
| 9 | 20% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | -- |
| 10 | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% |

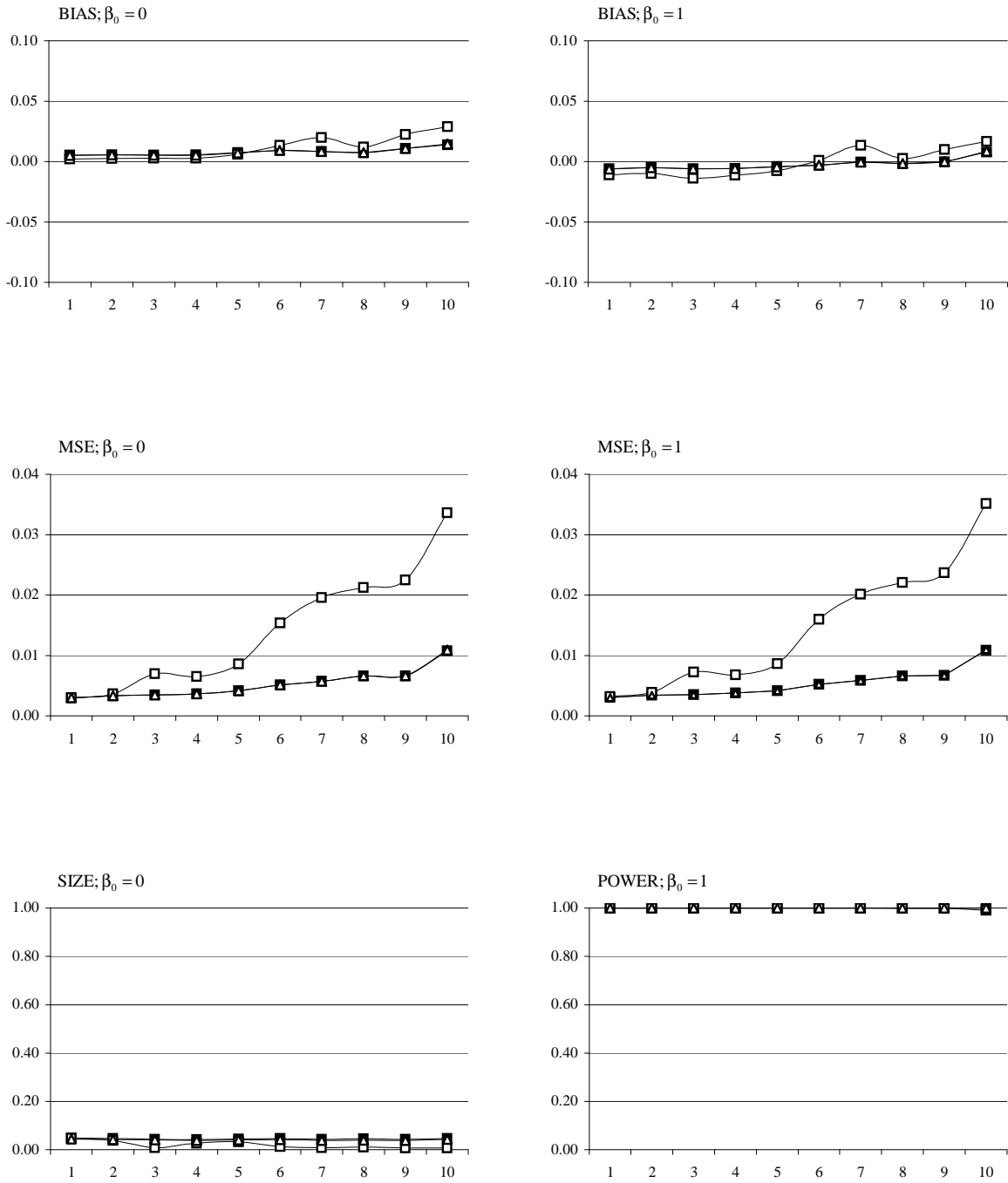


Figure 1: BIAS (top), MSE (middle), and SIG (bottom) for the case where the fixed population effect size $\mu = 0$ (left) and $\mu = 1$ (right), against the degree of heteroskedasticity in the meta-sample along the horizontal axis ($\tau^2 = 0$ and $\nu^2 = 0$). The different lines pertain to the OLS (white square), the fixed effects regression (black square) and the mixed effects estimator (white triangle).

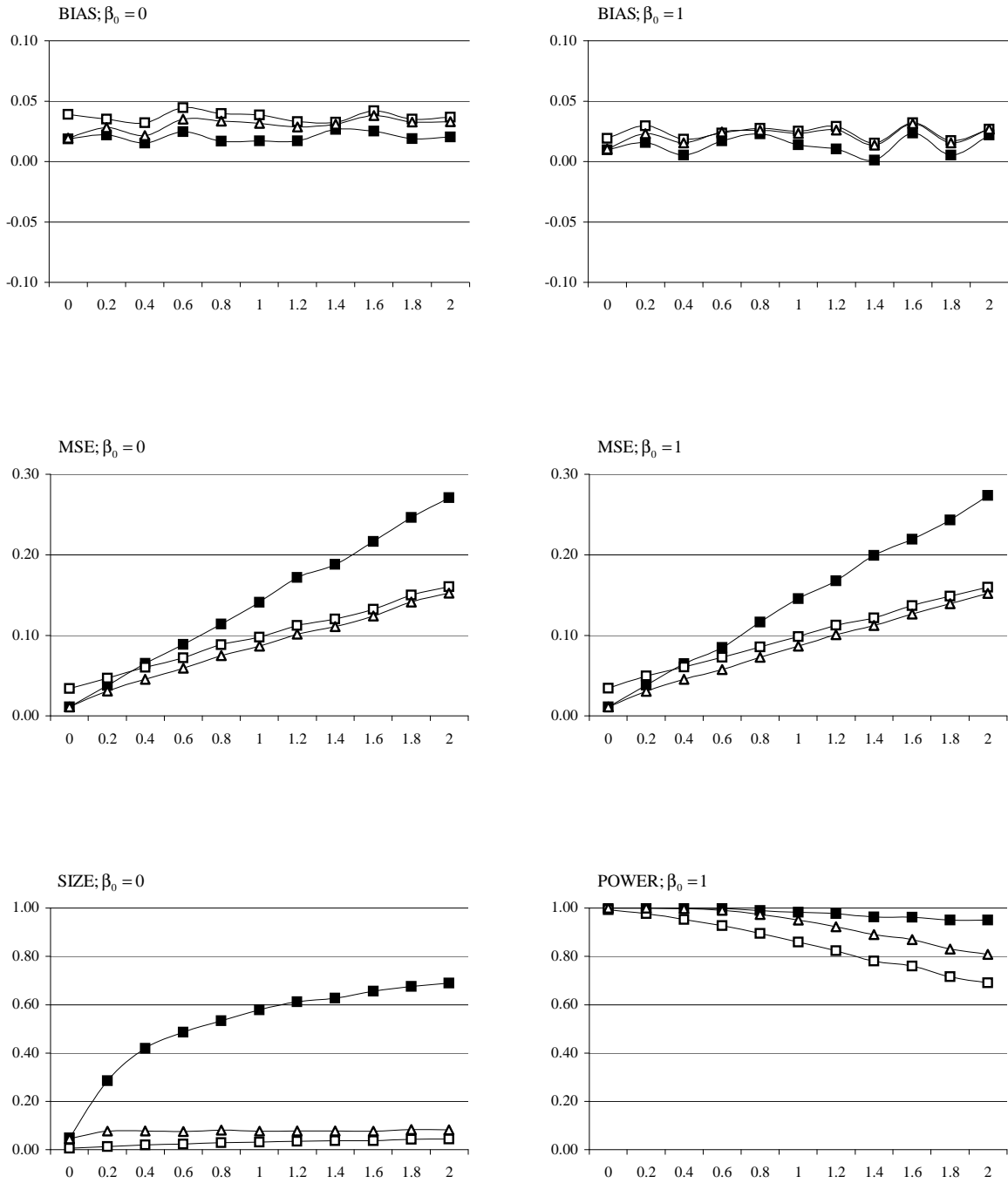


Figure 2: BIAS (top), MSE (middle), and SIG (bottom) for the case where the mean random effect size $\mu = 0$ (left) and $\mu = 1$ (right), against between-study variance τ^2 along the horizontal axis in absolute values (no heteroskedasticity and $v^2 = 0$). The different lines pertain to the OLS (white square), the fixed effects regression (black square) and the mixed effects estimator (white triangle).

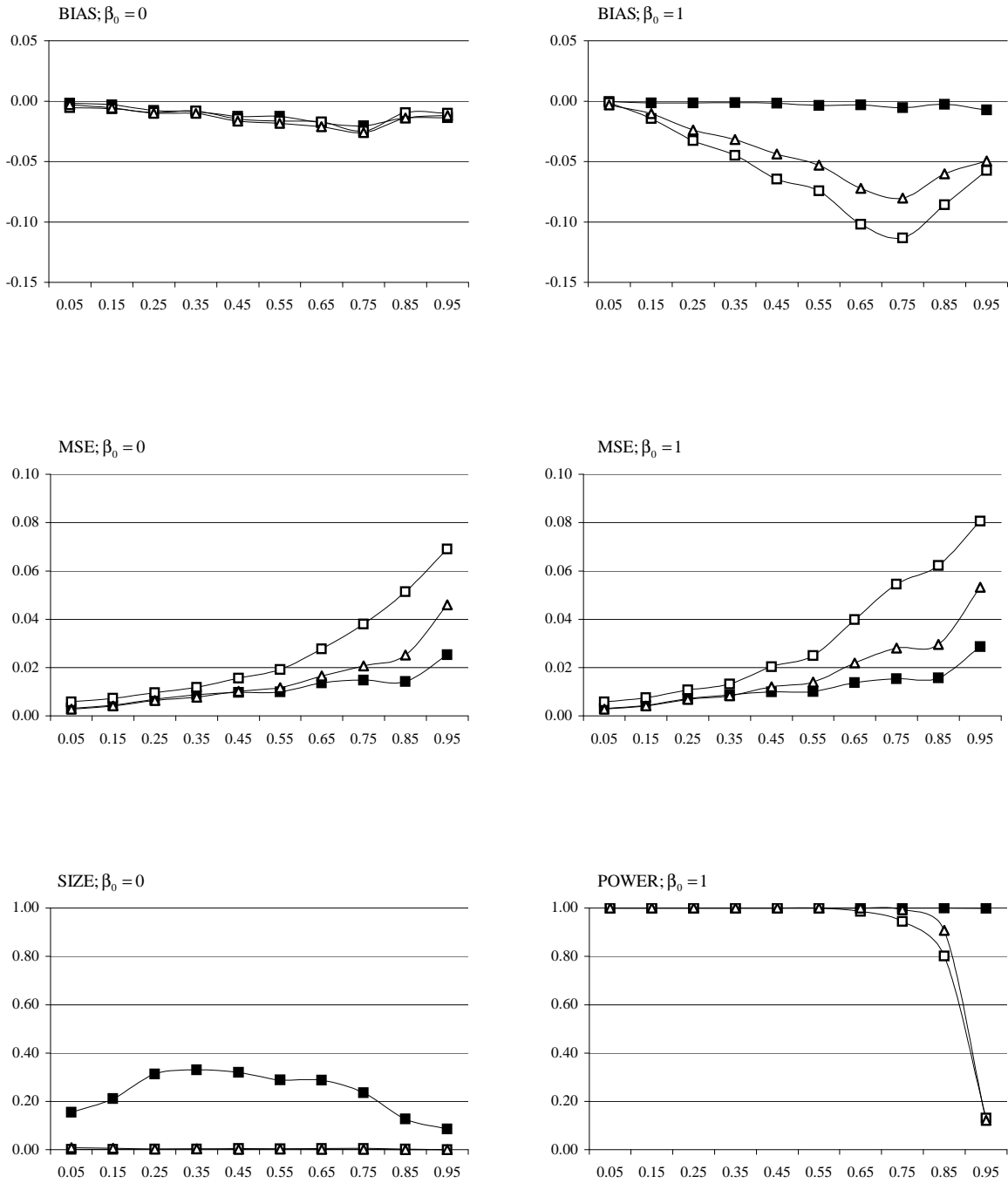


Figure 3: BIAS (top), MSE (middle), and SIG (bottom) for the case where the fixed population effect size $\mu = 0$ (left) and $\mu = 1$ (right), against an increasing proportion of effect sizes with omitted variable bias in the meta-analysis sample (maximum heteroskedasticity and $\tau^2 = 0$). The different lines pertain to the OLS estimator (white square), the fixed effects regression estimator (black square) and the mixed effects estimator (white triangle).

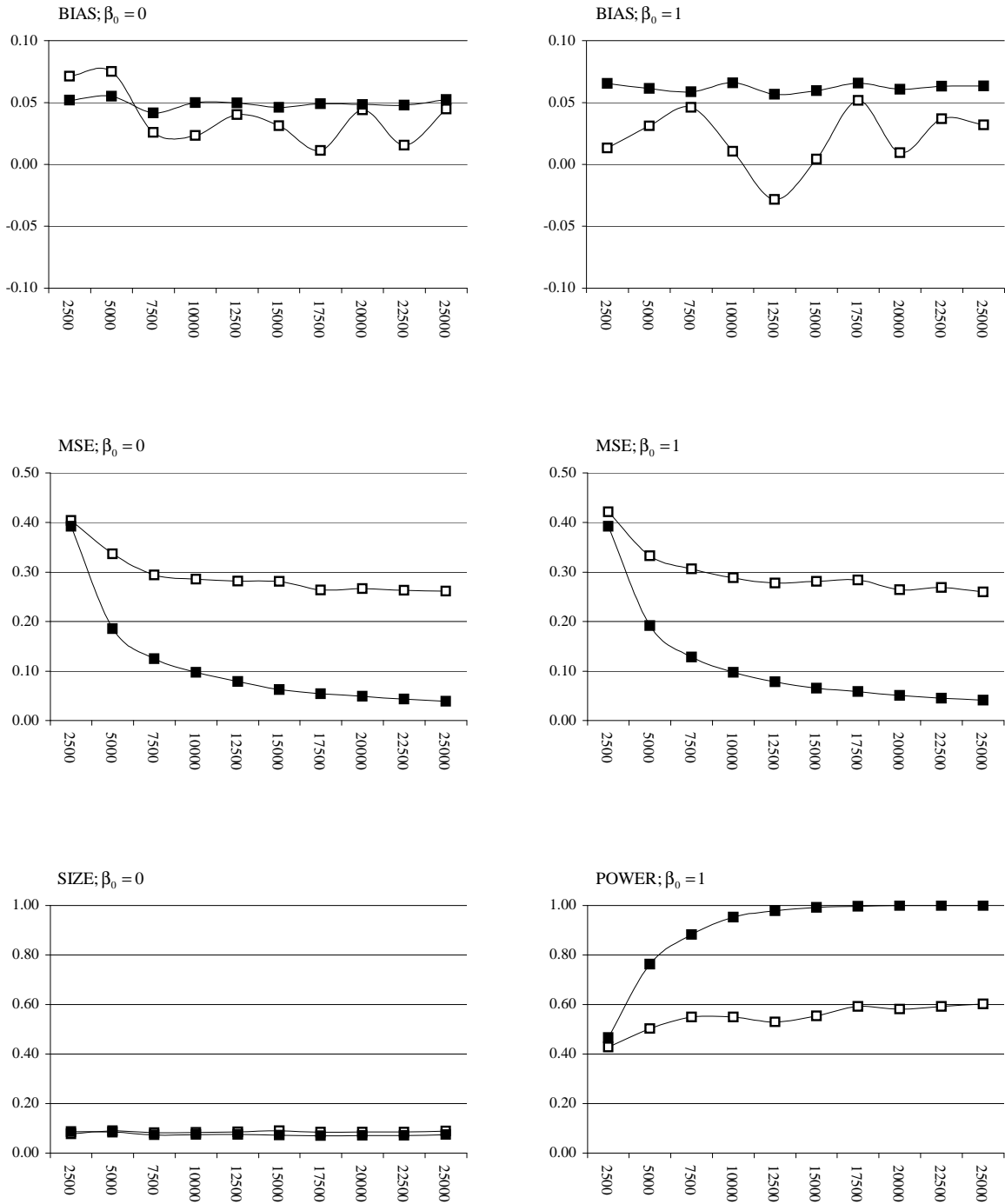


Figure 4: BIAS (top), MSE (middle), and SIG (bottom) for the case where the fixed population effect size $\mu = 0$ (left) and $\mu = 1$ (right), against the total number of underlying observations (no heteroskedasticity and $\tau^2 = 2$). The different lines pertain to the mixed effects estimator under increasing primary study sample size (white square) and the mixed effects estimator under increasing meta-analysis sample size (black square). See main text for further details.