Meta-Analysis and Publication Bias:
How Well Does the FAT-PET-PEESE Procedure Work?

Nazila Alinaghi†

W. Robert Reed

Department of Economics and Finance, University of Canterbury

Abstract: This study uses a Monte Carlo analysis to evaluate the performance of the FAT-PET-PEESE (FPP) procedure, a commonly employed approach for addressing publication bias in the economics and business meta-analysis literature. The main three objectives of FPP procedure are: (i) Funnel Asymmetry Testing (FAT) to test whether the sample of estimates is influenced by publication selection bias, (ii) Precision Effect Testing (PET) to test whether there is a genuine non-zero true effect of estimates once the publication bias is accommodated and corrected, and (iii) Precision Effect Estimate with Standard Errors (PEESE) to obtain an improved estimate of the overall mean effect. In this simulation two common types of publication bias including (i) publication bias against insignificant results and (ii) publication bias against wrong-signed (according to associated theory) estimates in three different data environments: Fixed Effects, Random Effects, and Panel Random Effects are considered. Our findings indicate that the FPP procedure performs well in the basic but unrealistic environment of "Fixed Effects", when there is one true effect and sampling error is the only reason why studies produce different estimates. However, once we study its performance in more realistic data environments, where there is heterogeneity in the population effects between and within studies, the FPP procedure becomes unreliable for the first two objectives, and is less efficient than other estimators when estimating overall mean effects. Further, hypothesis tests about the overall, mean effect cannot be trusted.

Keywords: Meta-analysis, publication bias, funnel asymmetry test (FAT), Precision Effect Estimate with Standard Error (PEESE), Monte Carlo, Simulations

JEL Classifications: B41, C15, C18

† Email:nazila.alinaghi@pg.canterbury.ac.nz, nazila_alinaghi2000@yahoo.com

1.1. Introduction

Meta-analysis offers a statistical analysis through which conflicting theoretical and/or empirical findings on a given topic can be summarized and compared. Two main objectives of meta-analysis are (i) to reach a single conclusion about the magnitude and significance of the results, and (ii) to compare findings yielded from various studies and explain potential reasons for the heterogeneity observed among estimates. Meta-analysis has become an increasingly popular method in economics and business. Figure 1.1 depicts a time series bar chart that lists all Web of Science journal articles in economics and business that have the word “meta-analysis” in the title. The trend is clearly upward reflecting the fact that the number of studies applying this tool is increasing over time.

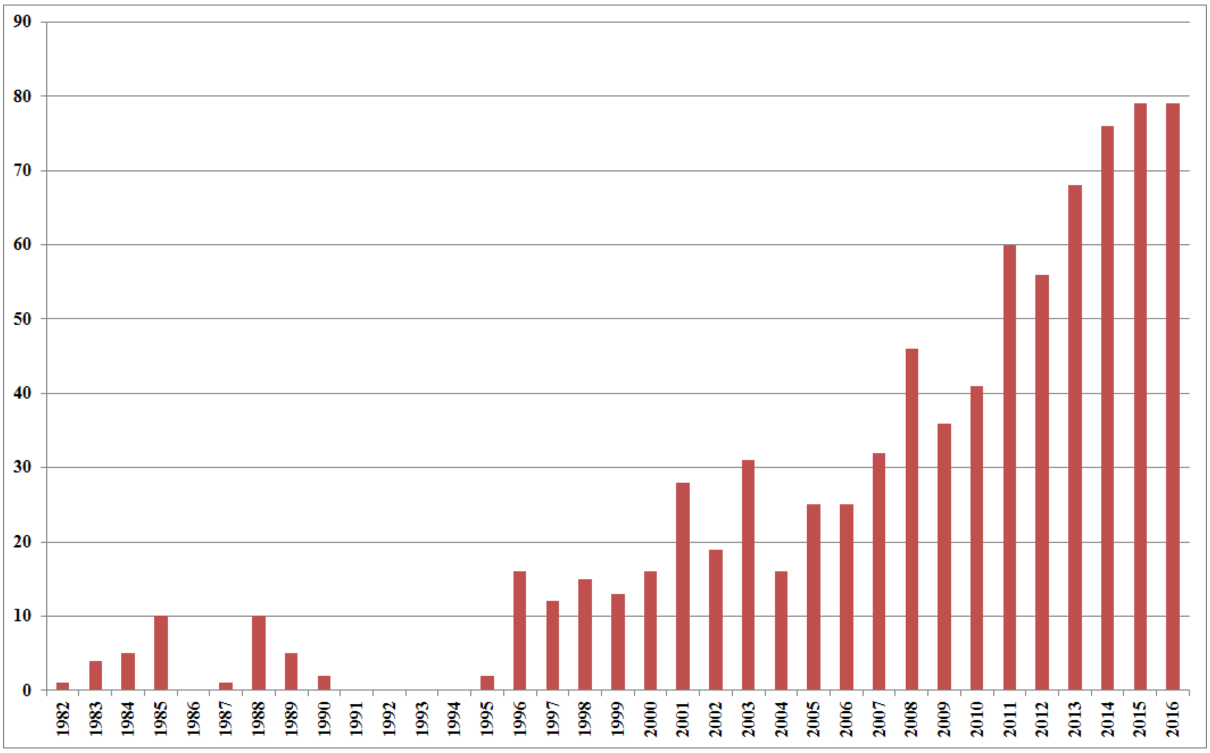

Figure 1.1: Number of Articles in Economics and Business Listed in Web of Science with “Meta-Analysis” in the Title

Note: Web of Science categories are: Economics, Business Finance, Business, Management, Criminology Penology, Urban Studies, and Social Sciences Interdisciplinary (813 articles).

It is well known that publication selection bias or “selectivity bias” can distort the distribution of estimated effects in the literature. Publication bias might happen because there is a tendency amongst researchers, reviewers, and editors to avoid reporting and publishing statistically insignificant estimates or estimates which are inconsistent with well-established theories. As a result, the true effect of the focal predictor on the response variable might be over- or under-estimated. An example of the second type of bias called directional publication bias was provided by Stanley (2005). He documented how the price elasticity of water demand is exaggerated fourfold in the literature as a direct result of publication selection bias. It is generally accepted that positive estimates of price elasticities are inconsistent with theory.

The data used for meta-analysis consists of estimated effects from studies on a particular phenomenon. If the distribution of those effects are distorted, so will be any conclusion derived from them. It is therefore crucial to identify whether the literature on a given topic suffers from publication selection bias and if there is, how it should be corrected.

A common procedure for doing this in the economics and business literature is through the FAT-PET-PEESE procedure (Stanley and Doucouliagos, 2012; 2014a). Figure 1.2 shows the associated four steps procedure. The first is the Funnel Asymmetry Test (FAT) to test whether the sample of estimates is influenced by publication selection bias. It uses Weighted Least Squares (WLS) to regress the estimated effects $\left(\widehat{\alpha}_j\right)$ on a constant term (β_0) and the standard errors of the estimated effects $\left(SE_j\right)$; where weights $\omega_i = \left(\frac{1}{SE_i}\right)$ are applied to correct for heteroskedasticity in the estimates (which is inevitable in meta-analysis). If the estimated coefficient on the standard error variable, $\widehat{\beta}_1$, is statistically significant, then the estimates suffer from publication bias.

Regardless of the results in the previous step, the next step is to conduct a Precision Effect Test (PET) to test whether there is a genuine, non-zero, true effect of estimates once publication bias is accommodated and corrected. It uses the same equation as the FAT, but tests whether $\beta_0 = 0$. If the SE_i variable were not included in the equation and if OLS was used rather than WLS, then the estimate of β_0 would simply be the arithmetic mean of the estimated effects in the literature. Thus, $\hat{\beta}_0$ is an estimate of the overall effect, and the PET tests $\hat{\beta}_0$ for statistical significance, correcting for publication bias.

If the PET fails to reject the null hypothesis of no effect, then $\hat{\beta}_0$ is taken as the estimate of overall effect with the understanding that it is statistically insignificant from zero. In other words, there is not enough evidence to support the existence of any empirical effect. If the PET, however, rejects the null, then one concludes that there is a genuine non-zero true effect. In that case, one estimates a new specification known as the PEESE, or Precision Effect Estimate with Standard Error. The corresponding estimate of β_0 then becomes the "best" estimate of overall effect.

Given the wide application of the FAT-PET-PEESE procedure in the economics and business literature (e.g. Costa-Font; Gemmill, and Rubert (2011); Doucouliagos, Stanley, and Viscusi (2014); Doucouliagos and Paldam (2013); Efendic, Pugh, and Adnett (2011); Haelermans and Borghans (2012); Havránek (2010); Iwasaki and Tokunaga (2014); Laroche (2016); Lazzaroni and Van Bergeijik (2014); Linde Leonard, Stanley, and Doucouliagos (2014); and Nelson (2013)), it is surprising that there have not been any comparative evaulations of its performance. That is the purpose of this chapter.

The three objectives of this study are to evaluate how well the FAT-PET-PEESE procedure (i) correctly detects the existence of publication selection bias, (ii) correctly tests the existence of a genuine non-zero true effect, and (iii) compares with three common meta-analysis estimators that do not correct for publication bias.

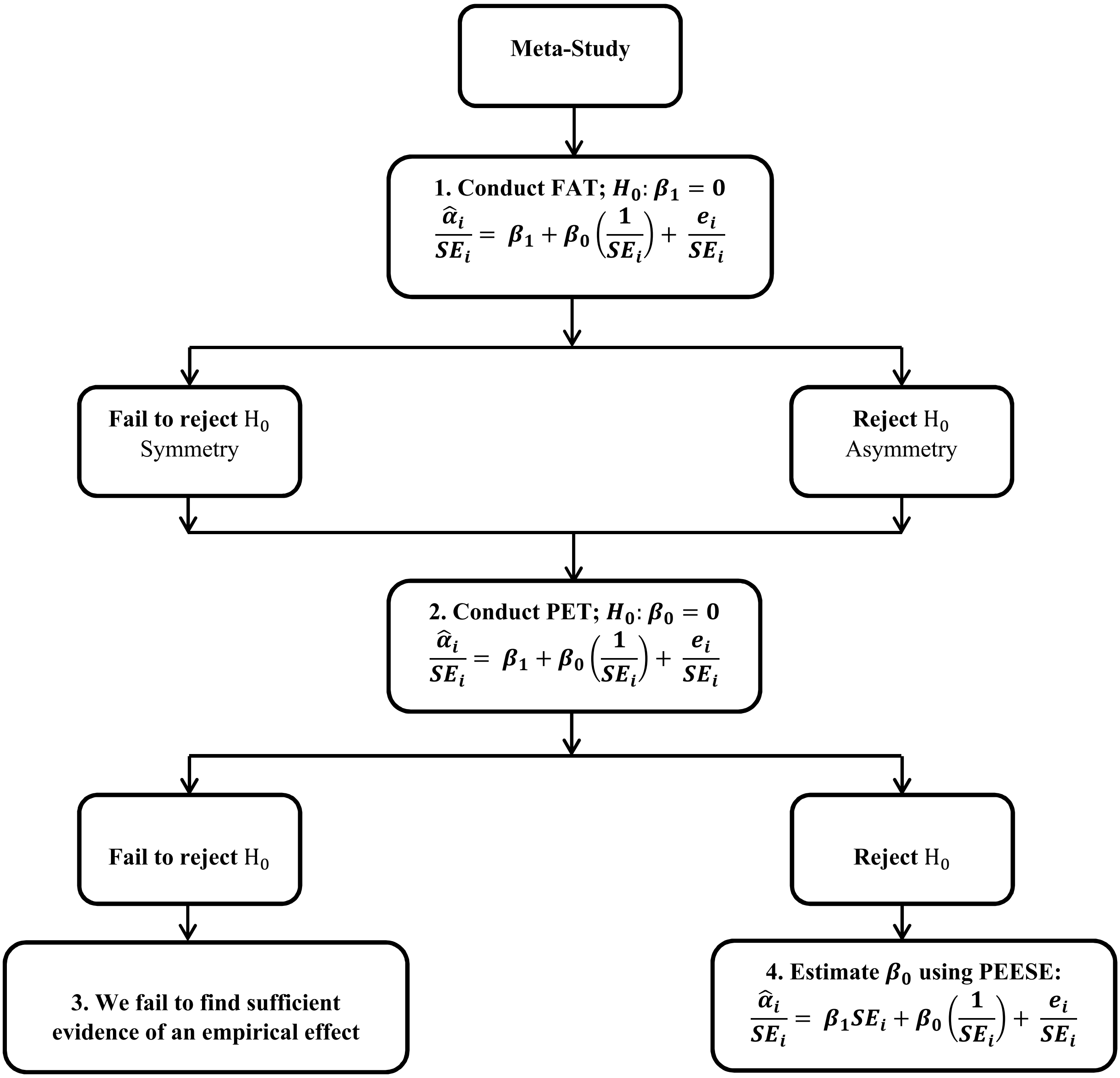

Figure 1.2: FAT-PET-PEESE Procedure

Source: Stanley and Doucouliagos (2012, page 79)

I use Monte Carlo experiments to demonstrate that the FPP procedure does not perform well in the kind of statistical environments one would expect to encounter in economics and business. Section 4.2 describes my Monte Carlo experiments including associated terminology. Section 4.3 describes my experimental design, the simulated datasets used in my analysis, and presents the results. Section 4.4 presents the main conclusions from this research.

1.2. Description of the Monte Carlo Experiments

1.2.1. Publication Bias

It is widely recognized that publication selection bias or "selectivity bias" distorts the distribution of estimated effects that appear in the literature. This arises because there is a tendency amongst researchers, reviewers, and editors to submit or accept manuscripts for publication based upon the direction or the strength of their results. Thus, it is unlikely that papers with statistically insignificant results or results which are not in line with an established theory could get published in a peer-reviewed journal. These studies usually end up sitting, unpublished, in file drawers of researchers. That is why this problem is called the "file-drawer problem".

Two popular types of publication bias modelled in these experiments are: (i) publication bias against insignificant results and (ii) publication bias against wrong-sign results. An example of the latter is a price elasticity where there is a strong presumption that the estimate should be negative, so that positive estimates will find it difficult to get published.

In my analysis, I model the first type of publication bias assuming that there is a tendency in favour of the strength of the results. Therefore if the absolute values of the reported t-statistics associated with the estimates are greater than or equal to 2, then they will get published. Studies with insignificant estimates can still get published, but with a relatively small probability. For the second type of publication bias, I assume that theory predicts the "correct" sign should be positive. The publication process then works against negative estimates. While negative estimates can still get published, they can do so only with a relatively small probability.

1.2.2. Estimators

I use Monte Carlo experiments to compare the performance of seven different estimators. For the sake of comparison, I use estimators studied by Reed (2015).[1] However, the main focus of this study is on a new estimator, the estimator described above as part of the FAT-PET-PEESE procedure, henceforth "FPP". I compare these estimators using three performance measures: Bias, Mean-Squared Error (MSE) as an efficiency test, and Type I error rates associated with testing whether the estimate of α equals its true value. The remainder of this section describes the respective estimators.

The "Unadjusted" estimator. The unadjusted estimator of the mean true effect of x on y is given by OLS estimates of β_0 in the following equation:

$$\hat{\alpha}_{i1} = \beta_0 + e_i, i = 1,2, \dots, M \tag{4.1}$$

where $\hat{\alpha}_{i1}$ is the ith estimated effect of y on x, and M is the number of estimates in the "Post-Publication" bias sample. The unadjusted estimator simply calculates the arithmetic mean of estimated effects across studies. It is used as a benchmark to compare the various meta-analysis (MA) estimators against.

The "Fixed Effects" (FE) estimator. Under this model I assume that there is one true underlying effect size and the only reason for the studies to obtain different estimated effect

[1] The conceptual design for my Monte Carlo experiments is based on Reed (2015), the replicated results are identical.

sizes is due solely to sampling error. This is why Borenstein et al. (2010) call this model the "common-effect model," which conveys the message more precisely. The fixed effect estimator weights all the observations by the inverse of the estimated standard error of $\hat{\alpha}_i$, SE_i. The FE estimator of the mean true effect is the weighted least squares estimate of β_0, except that the residuals are standardized to produce a sample variance of 1.

$$\frac{\hat{\alpha}_{i1}}{SE_i} = \beta_0\left(\frac{1}{SE_i}\right) + \frac{e_i}{SE_i}, i = 1,2,\ldots,M \tag{1.2}$$

The "Weighted Least Squares" (WLS) estimator. The WLS estimator is identical to the FE estimator except that the residual remains unstandardized. It is worthwhile to note that both FE and WLS estimators produce identical estimates of β_0, but the associated standard errors are different.

The "Random Effects" (RE) estimator. While the fixed effects models assume that there is one underlying true effect for all studies, this assumption seems quite implausible for most meta-analyses conducted in economics and business. Thus, under this model I assume that there is a distribution of true underlying effects and the goal is to estimate the mean of this distribution of true effects. The RE estimator is motivated by the assumption that differences in estimated effects across studies are due to (i) sampling error and also (ii) genuine differences in the underlying effects. The second component is represented by τ, which is the "between studies" variance. If the two variances are independent of each other, then,

$$SE(\hat{\alpha}_i) = \sqrt{(SE_i)^2 + \tau^2} = \omega_i \tag{1.3}$$

The RE estimator of the mean true effect is given by weighted least squares estimation of β_0, with weights equal to ω_i:

$$\frac{\hat{\alpha}_{i1}}{\omega_i} = \beta_0\left(\frac{1}{\omega_i}\right) + \frac{e_i}{\omega_i}, i = 1,2,\ldots,M \tag{1.4}$$

The "Precision Effect Testing" (PET) estimator. The PET estimator is designed to report the genuine underlying empirical effect after accommodating and correcting for publication bias. The PET adds the ith study's estimated standard error of the estimated effect, (SE_i), as an explanatory variable to control for publication bias. It then estimates the value of the mean effect as follows:

$$\hat{\alpha}_{i1} = \beta_0 + \beta_1 SE_i + e_i, i = 1,2, \dots, M \quad (1.5)$$

WLS estimation of β_0 provides an estimate of the mean true effect of x on y, weighting by $\left(\frac{1}{SE_i}\right)$, where SE_i is the same term used to correct for publication bias:

$$\frac{\hat{\alpha}_{i1}}{SE_i} = \beta_0 \left(\frac{1}{SE_i}\right) + \beta_1 + \frac{e_i}{SE_i}, i = 1,2, \dots, M \quad (1.6)$$

The "Precision-Effect Estimate with Standard Error" (PEESE) estimator. The PEESE estimator is designed to provide a better estimate of the actual empirical effect corrected for publication bias. What makes this estimator different from the PET is that it replaces SE_i with $(SE_i)^2$ in equation (1.5):

$$\hat{\alpha}_{i1} = \beta_0 + \beta_1 (SE_i)^2 + e_i, i = 1,2, \dots, M \quad (1.7)$$

This yields the following weighted least squares specification:

$$\frac{\hat{\alpha}_{i1}}{SE_i} = \beta_0 \left(\frac{1}{SE_i}\right) + \beta_1 SE_i + \frac{e_i}{SE_i}, i = 1,2, \dots, M \quad (1.8)$$

The last estimator, the FPP estimator, which is the main focus of this chapter, combines the "FAT" with elements of both the "PET" and "PEESE."

The "FAT-PET-PEESE" (FPP) estimator. Three main elements available in the "FAT-PET-PEESE" approach are summarized as follows: (i) identify the existence of publication bias (FAT); (ii) identify the presence of a genuine non-zero "true" effect (PET), corrected for publication bias; and (iii) estimate the magnitude of this "true" effect after accommodating and correcting for publication bias (both PET and PEESE).

The "FAT-PET-PEESE" (FPP) estimator – Step One. The first step involved in implementing the FPP estimator carries out the Funnel Asymmetry Test (FAT). This test is designed to test for publication bias. As can be seen in Equation (1.5), it uses Weighted Least Squares (WLS) to regress the estimated effects ($\hat{\alpha}_i$) on a constant term and the standard errors of the estimated effects (SE_i); where weights $\omega_i = \left(\frac{1}{SE_i}\right)$ are applied to correct for heteroskedasticity in estimates. Whereas the PET focuses on the coefficient β_0 in Equation (1.5), the FAT tests whether the coefficient on the *SE* variable, β_1, is significantly different from zero. If $\hat{\beta}_1$ is significant, then the null hypothesis, H_0: $\beta_1 = 0$, which indicates there is no publication bias, is rejected. Note that the bias can be positive or negative. If the conclusion of the FAT is a failure to reject the null, then there is not enough evidence to support the existence of publication bias. As part of my analysis of the performance of the FAT-PET-PEESE estimator, I will also record the performance of this FAT.

The "FAT-PET-PEESE" (FPP) estimator – Steps Two and Three. Regardless of the results on the previous step, one proceeds to the Precision Effect Test (PET), which is designed to identify whether there is genuine non-zero empirical effect after accounting for publication bias. According to Stanley and Doucouliagos (2012), the reason why the test is called precision effect testing is because β_0 is the coefficient on precision (the inverse of the standard error). It uses the same equation as the FAT, but tests whether H_0: $\beta_0 = 0$.

If the SE_jvariable was not included in the equation and if OLS was used rather than WLS, then the estimate of β_0 would simply be the arithmetic mean of the estimated effects in the literature. Thus, $\hat{\beta}_0$ is an estimate of the overall effect, and the PET tests $\hat{\beta}_0$ for statistical significance, corrected for publication bias.

If one fails to reject the null hypothesis of no effect, $H_0: \beta_0 = 0$. then $\hat{\beta}_0$ is taken as the estimate of overall effect with the understanding that it is statistically indistinguishable from zero. However, if the null is rejected, then one proceeds to Step Four and a new specification is estimated.

The "FAT-PET-PEESE" (FPP) estimator – Step Four. If the previous step determines that the true effect/mean value of the distribution of true effects is statistically different from zero, Stanley and Doucouliagos (2007, 2011) recommend that one estimate Equation (1.8) rather than Equation (1.6). In this case, the associated estimate of β_0 represents the best estimate of overall true effect.

In this study the four-step procedure explained above produces two test results (the FAT and PET), and a single estimate of overall effect (which I identify as the "FPP" estimate). To summarize, each simulation starts with conducting the FAT. Following that, regardless of the results for the FAT, the PET is conducted. If the PET produces a failure to reject conclusion, the coefficient on the precision term (the inverse of the standard error), is taken as the estimate of overall effect. If the PET produces a reject conclusion, the PEESE specification is estimated (cf. Equation 4.8), and the coefficient on the precision term from this specification is taken as the estimate of overall effect. This procedure is represented in the Figure 1.2 flowchart.

1.3. The Experiments

I create three different simulation environments to conduct my Monte Carlo experiments. In the first two simulation environments ("Fixed Effects" and "Random Effects") only one estimate per study is produced. In the last one ("Panel Random Effects"), multiple estimates per study are produced. The latter case is far more realistic, as most studies in the economics and business literature produce more than one estimate of the effect that is being studied.

In the Fixed Effects environment, there is only one underlying true effect and the only reason for the studies to obtain different estimated effect sizes is because of sampling error. In contrast, the true effect of x on y differs across studies in the Random Effects environment. In the last data environment, the true effects are heterogeneous both within and between studies. Given that studies differ in sample sizes, estimation methods, sets of control variables, geographical units, and time periods, the environments that model heterogeneity in true effects across studies a distribution of true effects are most realistic.

In the Fixed Effects environment, the experiments begin by simulating a common true effect. The common true effect, α, is used to generate individual (y, x) observations, from which a single estimate is derived. In the Random Effects environment, the experiments begin by simulating a distribution of true effects that is normally distributed with mean value α. Random draws from this distribution generate study-specific "true effects",α_i. The α_i's are used to generate individual (y, x) observations, from which a single estimate is derived. The Panel Random Effects environment also builds in heterogeneity across regressions within a study. Each of these environments is described in greater detail below.

The estimates derived within each of these environments are then put through a publication bias "filter", with the number of estimates in the post-publication bias sample, M, being determined endogenously. The resulting sample constitutes the sample of estimates available to the hypothetical meta-analyst.

The respective estimators are applied to this post-publication bias sample to produce estimates of α, the true effect in the Fixed Effects environment, and the mean of the distribution of true effects in the Random Effects and Panel Random Effects environments. This process simulates a single meta-analysis study.

The process is repeated to produce 10,000 simulated meta-analysis studies. The estimates for each of the estimators are then aggregated over these simulated studies and compared on the dimensions of Bias, MSE, and Type I error rates.

For each of the tree environments, I run experiments for nine different values of α including: 0 (i.e., no overall effect), 0.5, 1, 1.5, 2, 2.5, 3, 3.5, and 4. When the distribution of true effects is centred on zero, there will be more statistically insignificant estimates, and more wrong-signed estimates, than when the distribution shifts to the right. As a result, the percent of studies excluded by publication bias will be greatest at $\alpha = 0$. As α increase and the distribution shifts to the right, fewer studies are impacted by publication bias. Eventually, for sufficiently large value of α , all studies are "published", and the post-publication bias sample is identical to the pre-publication bias sample. As will be demonstrated below, the consequence of increasing α will differ depending on the nature of the publication bias (statistical insignificance versus wrong-signed (or wrong-direction) estimates).

Performance Tests. Table 1.2 through 4.9 compare seven different estimators across three different performance dimensions: (i) Average Estimate of Mean True Effect, (ii) Mean Squared Error (MSE), and (iii) Type I error rates.

Mean Squared Error (MSE). MSE is one of the three performance dimensions investigated in this study. The MSE measures the average squared difference between the estimator $\hat{\alpha}$ and the parameter α, which represents either the true effect (Fixed Effects environment), or the mean of the distribution of true effects (Random Effects and Panel Random Effects environments).

$$MSE = E(\hat{\alpha} - \alpha)^2 = Var(\hat{\alpha}) + [E(\hat{\alpha}) - \alpha]^2 = Var(\hat{\alpha}) + Bias(\hat{\alpha})^2, \quad (1.9)$$

where

$$Bias\ (\hat{\alpha}) = E(\hat{\alpha}) - \alpha \tag{1.10}$$

This is also called the risk function of an estimator, with $(\hat{\alpha} - \alpha)^2$ comprising a quadratic loss function. Thus, MSE contains two components. The first component measures the variability of the estimator (precision), while the second measures its bias (accuracy). One of the properties of a good estimator is that it should have a relatively small MSE.

Type I Error Rates. Another measure of an estimator's performance is the Type I error rate associated with testing whether the estimate of α equals its true value. In the context of my experiments, the associated null hypothesis is:

$$H_0: \beta_0 = \alpha$$

I test this null at the 95% confidence/5% significance level. As a result, the associated rejection rates should likewise be equal to 5 percent. Type I error rates substantially larger or smaller than 5% indicate that the results from hypothesis testing with the given estimator are not reliable. For example, a Type I error rate equal to 0.89 means that, in my experiments, the true null hypothesis is incorrectly rejected 89% of the time. This compares with an expected rejection rate of 5% given the 5% significance level employed in the tests.

1.3.1. The Fixed Effects Data Environment

Experimental Design. For the Fixed Effects (FE) data environment, the true effect is the same for all studies. The data generating process (DGP) for the experiments in this data environment is given by

$$y_{it} = 1 + \alpha x_{it} + \varepsilon_{it}, t = 1, 2, \dots, T \tag{1.11}$$

In my experiments, I set $T = 100$ observations. In order to generate different coefficient standard errors, I allow the DGP error term to have different variances across studies:

$$\varepsilon_{it} = \lambda_i \cdot NID(0,1), \text{where} \tag{1.12}$$

$$\lambda_i = 0.2 + UID(0,30) \tag{1.13}$$

λ_i controls the variance of the error term. The last specification, Equation (1.13) provides a realistic range of λ_i and ensures that the variance is always nonzero.

Table 4.1 is designed to give a picture of what a typical meta-analysis sample looks like, both before and after publication bias. The specific case that is represented is when the true effect equals 1 ($\alpha = 1$). The top panel represents average sample characteristics of an "empirical literature" consisting of 1000 estimated effects, before publication bias keeps some of them from seeing the light of day. This is the "Pre-Publication Bias" sample. The next two panels respectively report average sample characteristics after imposing the two types of publication bias: publication bias against insignificance and publication bias against wrong-signs. As noted earlier, we assume that theory predicts that the respective effect should be positive (as in a value-of-life study). These are each "Post-Publication Bias" samples, and comprise the samples that the hypothetical meta-analyst analyzes.

When $\alpha = 1$ and there is no publication bias, the (average) median value of estimated effect in the full sample is 1.00, as would be expected. Estimated effects range from an average minimum of -6.85 to an average maximum of 8.92. t-statistics range from an average minimum of -2.69 to an average maximum of 45.62. The median t-value in the full sample is, on average, statistically insignificant.

These estimated effects and corresponding statistics are unobserved to the meta-analyst, as the meta-analyst only sees the estimates that survive publication bias (the "Post-Publication Bias" samples. When $\alpha = 1$ and publication bias is in favour of statistical significance, the average meta-analysis sample reduces to 318 estimated effects. The associated median estimated effect is 1.18 (representing a bias of 18%), and the average median t-statistic has increased from 0.94 in the unbiased, full sample to 2.60, and is now statistically significant. Similar results can be seen when publication bias favours estimates that are positively signed, though the median t-statistic is, on average, not so large as to be significant.

Note that when $\alpha = 1$, both types of publication bias disproportionately omit negative estimated effects, inducing a positive bias in both estimated effects and t-statistics in the post-publication bias samples. Further, both post-publication bias samples look “reasonable”. The estimated ranges of t-statistics/precision are comparable to those typically reported in economics and business.

Table 1.1: Sample Characteristics for a Simulated Meta-Analysis Data Set (Fixed Effects ($\alpha = 1$))

Variable	***Median***	***Minimum***	***P5%***	***P95%***	***Maximum***
Pre-Publication Bias (100 percent of estimates):					
Estimated effect	1.00	-6.85	-1.99	3.99	8.92
t-statistic	0.94	-2.69	-0.96	6.08	45.62
Publication Bias Against Insignificance (31.8 percent of estimates):					
Estimated effect	1.18	-6.74	-0.84	5.19	8.86
t-statistic	2.60	-2.69	-0.45	14.95	45.44
Publication Bias Against Negative Effects (80.5 percent of estimates):					
Estimated effect	1.20	-4.61	0.11	4.26	8.92
t-statistic	1.27	-1.89	0.07	7.33	45.45

Fixed Effects: Performance Tests. Table 1.2 and Table 1.3 compare seven different estimators across the three different performance dimensions mentioned earlier. While Table 1.2 reports the results when publication bias is directed against statistical insignificance, Table 1.3 examines publication bias against wrong-signed estimates. Each of the estimators is studied for a set of mean true effect values (α) ranging from 0.0 to 4.0.

The top panel of Table 1.2 reports the average estimated effects for each of the respective estimators. The first two columns report the value of the true effect (α) and the average percent studies included in the simulated meta-analysis (MA) studies, where the average is taken over 10,000 simulated studies. The first thing to note is that there is a strong relationship between the size of the true effect and the number of studies that survive publication bias against statistical insignificance. When $\alpha = 0$, less than 15% of all studies appear in the meta-analyst's sample. As α increases and the mean of the distribution of estimated effects moves away from zero, more and more studies produce significant estimates. When $\alpha = 4$, approximately three-fourths of all studies survive publication bias and are included in the meta-analyst's sample.

The next column reports results for the Unadjusted estimator. When $\alpha = 0$ and publication bias discriminates against insignificant estimates, the average estimated value of α for the Unadjusted estimator (averaged across the 10,000 simulated MA studies) is 0.01, which is very close to its expected value of 0. The Unadjusted estimator is an unbiased estimator of the true effect when $\alpha = 0$ because sampling error is equally likely to produce significant estimates that are above and below the true effect. However, as α increases, publication bias disproportionately omits studies with estimates below the true effect since, ceteris paribus, studies with small estimates are more likely to have small t-values.

Table 1.2: Comparative Performance of Meta-Analysis Estimators (Fixed Effects/Publication Bias against Insignificance)

α	*Percent*	*Unadjusted*	*PET*	*PEESE*	*FPP*	*FE*	*WLS*	*RE*
Average Estimate of Mean True Effect								
0.0	14.3	0.01	-0.01	0.00	0.00	0.00	0.00	0.00
0.5	23.0	0.92	0.48	0.51	0.51	0.51	0.51	0.56
1.0	31.8	1.57	0.97	1.00	1.00	1.01	1.01	1.05
1.5	40.0	2.14	1.46	1.50	1.50	1.51	1.51	1.54
2.0	47.6	2.67	1.96	2.00	2.00	2.01	2.01	2.03
2.5	54.6	3.17	2.47	2.50	2.50	2.51	2.51	2.53
3.0	61.1	3.65	2.97	2.99	2.99	3.01	3.01	3.02
3.5	67.0	4.12	3.47	3.49	3.49	3.51	3.51	3.51
4.0	72.2	4.58	3.97	3.99	3.99	4.01	4.01	4.01
Mean Squared Error								
0.0	14.3	0.0520	0.0021	0.0016	0.0018	0.0015	0.0015	0.0071
0.5	23.0	0.1935	0.0005	0.0002	0.0002	0.0002	0.0002	0.0035
1.0	31.8	0.3378	0.0012	0.0001	0.0001	0.0002	0.0002	0.0024
1.5	40.0	0.4199	0.0014	0.0001	0.0001	0.0002	0.0002	0.0017
2.0	47.6	0.4504	0.0014	0.0001	0.0001	0.0002	0.0002	0.0012
2.5	54.6	0.4494	0.0012	0.0001	0.0001	0.0001	0.0001	0.0008
3.0	61.1	0.4245	0.0011	0.0001	0.0001	0.0001	0.0001	0.0005
3.5	67.0	0.3845	0.0009	0.0001	0.0001	0.0001	0.0001	0.0003
4.0	72.2	0.3353	0.0007	0.0001	0.0001	0.0001	0.0001	0.0002
Type I Error Rates ($H_0: \beta_0 = \alpha$)								
0.0	14.3	0.05	0.16	0.14	0.12	0.22	0.05	0.01
0.5	23.0	0.82	0.50	0.22	0.22	0.33	0.13	0.73
1.0	31.8	1.00	0.86	0.10	0.10	0.29	0.15	0.71
1.5	40.0	1.00	0.91	0.07	0.07	0.27	0.17	0.64
2.0	47.6	1.00	0.91	0.08	0.08	0.23	0.17	0.52
2.5	54.6	1.00	0.89	0.11	0.11	0.20	0.16	0.40
3.0	61.1	1.00	0.86	0.12	0.12	0.17	0.15	0.29
3.5	67.0	1.00	0.82	0.14	0.14	0.15	0.14	0.21
4.0	72.2	1.00	0.75	0.13	0.13	0.13	0.13	0.15

When $\alpha = 1.0$, the Unadjusted estimator overestimates the mean true effect by approximately 57%. As α increases, fewer and fewer studies are affected by publication bias. While the table does not show that, further increases in α would eventually cause the publication bias associated with the Unadjusted estimator to disappear.

Continuing with the top panel of Table 1.2, I turn my attention to the performances of the six MA estimators. I am particularly interested in the first three estimators, which are specifically designed to address publication bias. The last of these reports the overall effect derived from the four-step, FAT-PET-PEESE procedure ("FPP"). With respect to estimation bias, all three estimators do very well compare to the Unadjusted estimator. When $\alpha = 1$, the mean estimates of the true effect for the PET, PEESE, and FPP estimators are 0.97, 1.00 and 1.00, respectively. When $\alpha = 2$, the estimates are 1.96, 2.00 and 2.00. In fact, for every value of α, the PET reports a bias of 3% to 4%. However, the PEESE and FPP estimators eliminate estimation bias. This success seemingly validates the ability of the PET, PEESE and FPP estimators to correct for publication bias.

However, the next three columns demonstrate that the other MA estimators also perform well, even though they do not explicitly address publication bias. The explanation lies in how the study estimates are weighted. In one way or another, all six of these estimators weight by the inverse of the estimated coefficient's standard error.

Turning to the middle panel of Table 1.2, which focuses on MSE performance, I see that the Unadjusted estimator, unsurprisingly, performs by far the worst. Of the three MA estimators designed to address and accommodate for publication bias, the PEESE and FPP estimators are most efficient. However, the FE and WLS estimators perform are extremely close to these in terms of efficiency.

The bottom panel is the first indication that the respective MA estimators suffer from performance inadequacies, and this includes the FPP estimator. Type I error rates associated with the hypothesis $H_0: \beta_0 = \alpha$ are often far in excess of their expected value of 5%.

In summary, all six MA estimators perform better than the Unadjusted estimator. Among those six, the FPP estimator performs among the best, but struggles when it comes to reliability in hypothesis testing.

Table 1.3 continues investigating the Fixed Effects case, where all studies share a common true effect, but it analyses publication bias that is targeted towards wrong-signed estimates. As evidenced by the top panel, the Unadjusted estimator again produces effect estimates that can be substantially biased. In contrast to publication bias against statistical insignificance, the bias is greatest for small values of α. As α increases, studies estimate fewer negative effects, so more studies get "published". When α is very large (= 4), almost all studies get published (98%), and the Unadjusted estimator correspondingly has a relatively small estimation bias.

Turning now to the PET, PEESE and FPP estimators, FPP estimator is superior. However, depending on whether $\alpha = 0$ or $\alpha > 0$, the relative performance of PET and PEESE are quite different. When $\alpha = 0$, the PET estimator performs very well; when $\alpha > 0$, the PEESE estimator dominates. As before, the other three MA estimators generally also do a good job of eliminating estimation bias; and have MSE performance similar to the PEESE and FPP estimators. None of the estimators except FPP is reliable for hypothesis testing.

Table 1.3: Comparative Performance of Meta-Analysis Estimators (Fixed Effects /Publication Bias against Wrong Sign)

α	*Percent*	*Unadjusted*	*PET*	*PEESE*	*FPP*	*FE*	*WLS*	*RE*
Average Estimate of Mean True Effect								
0.0	55.0	1.01	0.00	0.51	0.00	0.07	0.07	0.08
0.5	71.7	1.22	0.47	0.50	0.50	0.51	0.51	0.51
1.0	80.6	1.56	0.98	1.00	1.00	1.01	1.01	1.01
1.5	86.5	1.94	1.48	1.50	1.50	1.51	1.51	1.51
2.0	90.6	2.34	1.99	2.00	2.00	2.00	2.00	2.00
2.5	93.5	2.76	2.49	2.50	2.50	2.50	2.50	2.50
3.0	95.5	3.20	2.99	3.00	3.00	3.00	3.00	3.00
3.5	97.0	3.65	3.49	3.50	3.50	3.50	3.50	3.50
4.0	98.0	4.11	4.00	4.00	4.00	4.00	4.00	4.00
Mean Squared Error								
0.0	55.0	1.0164	0.0001	0.0028	0.0003	0.0054	0.0054	0.0068
0.5	71.7	0.5245	0.0010	0.0001	0.0001	0.0003	0.0003	0.0003
1.0	80.6	0.3174	0.0006	0.0001	0.0001	0.0002	0.0002	0.0002
1.5	86.5	0.1957	0.0004	0.0001	0.0001	0.0001	0.0001	0.0001
2.0	90.6	0.1179	0.0003	0.0001	0.0001	0.0001	0.0001	0.0001
2.5	93.5	0.0701	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
3.0	95.5	0.0412	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.5	97.0	0.0240	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
4.0	98.0	0.0140	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Type I Error Rates ($H_0: \beta_0 = \alpha$)								
0.0	55.0	1.00	0.08	0.99	0.08	1.00	1.00	1.00
0.5	71.7	1.00	0.86	0.08	0.08	0.45	0.50	0.44
1.0	80.6	1.00	0.71	0.07	0.07	0.22	0.27	0.22
1.5	86.5	1.00	0.53	0.08	0.08	0.13	0.17	0.13
2.0	90.6	1.00	0.37	0.08	0.08	0.08	0.11	0.08
2.5	93.5	1.00	0.25	0.08	0.08	0.08	0.10	0.08
3.0	95.5	0.98	0.17	0.07	0.07	0.06	0.07	0.06
3.5	97.0	0.82	0.12	0.07	0.07	0.06	0.07	0.06
4.0	98.0	0.53	0.10	0.07	0.07	0.05	0.06	0.05

1.3.2. Random Effects Data Environment

Experimental Design. For the Random Effects data environments, I generate heterogeneity in true effects across studies by letting the true effect be normally and independently distributed with mean α and variance 1. In particular, the DGP producing individual observations for study i is given by:

$$y_{it} = 1 + \alpha_i x_{it} + \varepsilon_{it}, t = 1, 2, \dots, T, \text{ where} \tag{1.14}$$

$$\alpha_i = NID(\alpha, 1) \tag{1.15}$$

All the studies have $T = 100$ observations. In order to generate different coefficient standard errors, I allow the DGP error term to have different variances across studies as follows:

$$\varepsilon_{it} = \lambda_i NID(0,1), \text{ where} \tag{1.16}$$

$$\lambda_i = 0.5 + UID(0,30. \tag{1.17}$$

As before, λ_i controls the variance of the error term. The last specification sets the minimum and maximum value for λ_i.

The specific parameter values used in the experiments were selected to simultaneously satisfy four criteria:

1. Produce a realistic range of t-values for the estimated effects.
2. Produce realistic-looking funnel plots.
3. Cause the percent of studies eliminated by publication bias to range between 10 and 90 percent (so all the MA studies are impacted by publication bias to some degree)
4. Produce realistic values of "effect heterogeneity"

"Effect heterogeneity" refers to the differences in true effects across studies.

As discussed earlier, the experiments model two kinds of publication bias: selection against statistical insignificance, and selection against wrong-signed estimates. In both cases, statistically insignificant/wrong-signed estimates are allowed to be included in the post-publication bias sample, but with a relatively low probability. The experiments set this probability at 10 percent.

Table 1.4 gives average sample characteristics for a typical meta-analysis sample in the the Random Effects data environment when $\alpha = 1$. The associated parameter values in Equations (1.16) and (1.17) have been chosen to produce a range of estimated effects and t-statistics similar to those produced in the Fixed Effects data environment (cf. Table 4.1). The additional sample characteristics added to this table is a measure of effect heterogeneity, I^2.

I^2 takes values between 0 and 1 and measures the share of variation in the estimated effects that is not attributed to sampling error (Higgins and Thompson, 2002). As a point of comparison, I^2 values between 70-95% are commonly encountered in economics and business meta-analysis research.

Table 1.4: Sample Characteristics for a Simulated Meta-Analysis Data Set (Random Effects ($\alpha = 1$))

Variable	***Median***	***Minimum***	***P5%***	***P95%***	***Maximum***
Pre-Publication Bias (100 percent of estimates):					
Estimated effect	1.00	-6.24	-2.38	4.37	8.25
t-statistic	0.79	-8.12	-1.48	5.98	31.79
I^2	0.84	0.65	0.74	0.92	0.95
Publication Bias Against Insignificance (32.9 percent of estimates):					
Estimated effect	1.81	-5.90	-2.09	5.64	8.26
t-statistic	2.55	-8.15	-2.29	12.77	31.50
I^2	0.93	0.72	0.87	0.97	0.99
Publication Bias Against Negative Effects (74.7 percent of estimates):					
Estimated effect	1.55	-3.50	0.01	4.76	8.30
t-statistic	1.28	-2.93	0.01	7.40	31.57
I^2	0.79	0.41	0.63	0.90	0.95

Random Effects: Performance Tests. Table 1.5 is the Random Effects analogue to Table 1.2. It shows estimator performance in the presence of publication bias against statistical insignificance. As in the Fixed Effects case, when $\alpha = 0.0$, the Unadjusted estimator is an unbiased estimator of the true effect because sampling error is equally likely to produce significant estimates that are above and below the true effect. However, as α increases, publication bias disproportionately omits studies with estimates below the true effect since, ceteris paribus, studies with small estimates are more likely to have small t-values. When $\alpha = 1.0$, the Unadjusted estimator overestimates the mean true effect by approximately 82 percent. As α increases, fewer and fewer studies are affected by publication bias. While the table does not show that, further increases in α would eventually cause the publication bias associated with the Unadjusted estimator to disappear. Continuing with the top panel of Table 1.5, the PET, PEESE and FPP do very well compared to the Unadjusted estimator.

The FE estimator and its near twin, the WLS estimator perform almost as well as the PET, PEESE, and FPP estimators, even though they do not explicitly correct for publication bias. As before, the explanation lies in how the study estimates are weighted. In one way or another, all of these estimators weight by the inverse of the estimated coefficient's standard error.

The RE estimator consistently overestimates the mean true effect for nonzero value of α. Interestingly, it is tailored to match the data environment in which the simulations are conducted. Despite that fact, it is the most biased estimator of the six MA estimators. This seemingly paradoxical result has been noted by other researchers (Doucouliagos and Paldam, 2013, p.586; Stanley and Doucouliagos, 2012, p.83). On the dimension of estimation bias, when $0 < \alpha < 2$, the PET performs best of all MA estimators. For $\alpha > 2$, the PEESE and FPP estimators perform best.

The middle panel of Table 1.5 focuses on MSE performance, with smaller MSE values indicating greater efficiency. The Unadjusted estimator performs poorly compared to the MA estimators for all values of $\alpha > 0$. Among MA estimators when $\alpha > 0$, the PEESE, FPP and FE/WLS estimators generally perform best.

Finally, when it comes to hypothesis testing, the bottom panel of Table 1.5 suggests that caution is in order. The FE, WLS, and RE estimators all produce type I error rates that are unacceptably large. For example, when $\alpha = 0.0$, the FE and WLS estimators reject the hypothesis that $\alpha = 0.0$ in 89 percent and 47 percent of the tests, despite the fact that the hypothesis is true. This compares with an expected rejection rate of 5 percent given the 5 percent significance level employed in the tests. The PEESE and FPP are substantially better, though they also produce Type I error rates larger than 5 percent when $0.5 \leq \alpha \leq 1.5$. Given these unattractive choices, one might easily be tempted to conclude that the PET estimator is serviceable for hypothesis testing. However, the subsequent results will render this option less tempting.

Table 1.5: Comparative Performance of Meta-Analysis Estimators (Random Effects/Publication Bias against Insignificance)

α	*Percent*	*Unadjusted*	*PET*	*PEESE*	*FPP*	*FE*	*WLS*	*RE*
Average Estimate of Mean True Effect								
0.0	27.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.5	28.7	1.01	0.52	0.59	0.55	0.61	0.61	0.89
1.0	33.0	1.82	1.02	1.12	1.12	1.15	1.15	1.58
1.5	39.1	2.44	1.48	1.60	1.60	1.63	1.63	2.09
2.0	45.9	2.94	1.95	2.06	2.06	2.09	2.09	2.53
2.5	52.8	3.40	2.43	2.52	2.52	2.56	2.56	2.96
3.0	59.2	3.84	2.93	3.00	3.00	3.04	3.04	3.40
3.5	65.1	4.28	3.42	3.49	3.49	3.53	3.53	3.84
4.0	70.4	4.71	3.93	3.99	3.99	4.02	4.02	4.29
Mean Squared Error								
0.0	27.1	0.026	0.059	0.037	0.051	0.036	0.036	0.012
0.5	28.7	0.286	0.056	0.043	0.057	0.044	0.044	0.164
1.0	33.0	0.693	0.049	0.046	0.047	0.050	0.050	0.340
1.5	39.1	0.888	0.043	0.036	0.037	0.041	0.041	0.352
2.0	45.9	0.893	0.042	0.028	0.028	0.032	0.032	0.285
2.5	52.8	0.815	0.046	0.026	0.024	0.027	0.027	0.216
3.0	59.2	0.711	0.044	0.024	0.024	0.024	0.024	0.160
3.5	65.1	0.609	0.044	0.024	0.024	0.023	0.023	0.120
4.0	70.4	0.511	0.042	0.024	0.024	0.022	0.022	0.089
Type I Error Rates ($H_0: \beta_0 = \alpha$)								
0.0	27.1	0.05	0.08	0.07	0.07	0.89	0.47	0.03
0.5	28.7	0.92	0.09	0.12	0.13	0.90	0.55	0.95
1.0	33.0	1.00	0.08	0.16	0.16	0.92	0.64	1.00
1.5	39.1	1.00	0.08	0.13	0.13	0.91	0.65	1.00
2.0	45.9	1.00	0.09	0.09	0.09	0.89	0.61	1.00
2.5	52.8	1.00	0.10	0.07	0.08	0.88	0.59	1.00
3.0	59.2	1.00	0.10	0.07	0.07	0.87	0.60	1.00
3.5	65.1	1.00	0.10	0.07	0.07	0.87	0.59	1.00
4.0	70.4	1.00	0.10	0.07	0.07	0.87	0.61	1.00

Table 1.6 repeats the preceding analysis for the case when publication bias discriminates against negative effect estimates. The Unadjusted estimator again produce substantially biased estimates of the mean true effect, now even when $\alpha = 0$. Unlike the previous case, the MA estimators also produce biased estimates when α is relatively small. For example, when $\alpha = 1$, the associated bias ranges from 21 percent to 49 percent. These biases get smaller as α increases and the proportion of included studies becomes larger.

Table 1.6 tells a story for MSE performance that is similar to Table 1.5. The FE/WLS estimator often performs as well, and sometimes slightly better, than the PET, PEESE and FPP estimators. Interestingly, when $\alpha \geq 3$, the RE estimator is most efficient, despite being the most biased. The explanation has to do with the fact that RE estimates have generally smaller variances than other MA estimators.

Finally as in Table 1.5, Type I error rates for the FE, WLS, and RE estimators are unacceptably large. Unlike Table 1.5, the PET, PEESE and FPP estimators now also have unacceptably large Type I error rates for small values of α.

Summarizing the results for the Random Effects data environment, I find that the MA estimators that do not explicitly correct for publication bias often perform as well, if not better, than those that do. While the MA estimators always reduce estimation bias in our experiments, they do not always eliminate it. In other words, up to this point, there is little that distinguishes the FPP estimator from other MA estimators that do not correct for publication bias.

Table 1.6: Comparative Performance of Meta-Analysis Estimators (Random Effects /Publication Bias against Wrong Sign)

α	*Percent*	*Unadjusted*	*PET*	*PEESE*	*FPP*	*FE*	*WLS*	*RE*
Average Estimate of Mean True Effect								
0.0	55.0	1.26	0.61	0.66	0.65	0.69	0.69	0.91
0.5	65.4	1.52	0.90	0.95	0.95	0.97	0.97	1.18
1.0	74.7	1.81	1.21	1.26	1.26	1.29	1.29	1.49
1.5	82.0	2.12	1.59	1.63	1.62	1.65	1.65	1.85
2.0	87.4	2.48	2.01	2.05	2.05	2.07	2.07	2.24
2.5	91.3	2.86	2.49	2.51	2.51	2.53	2.53	2.66
3.0	94.0	3.27	2.98	3.00	3.00	3.02	3.02	3.11
3.5	95.9	3.70	3.48	3.50	3.50	3.51	3.51	3.58
4.0	97.2	4.15	3.99	4.00	4.00	4.01	4.01	4.06
Mean Squared Error								
0.0	55.0	1.602	0.405	0.461	0.456	0.498	0.498	0.828
0.5	65.4	1.053	0.184	0.218	0.218	0.241	0.241	0.467
1.0	74.7	0.654	0.073	0.087	0.088	0.099	0.099	0.245
1.5	82.0	0.392	0.038	0.037	0.036	0.041	0.041	0.122
2.0	87.4	0.229	0.032	0.024	0.025	0.025	0.025	0.060
2.5	91.3	0.133	0.033	0.023	0.022	0.022	0.022	0.030
3.0	94.0	0.078	0.035	0.023	0.023	0.022	0.022	0.016
3.5	95.9	0.045	0.035	0.024	0.024	0.022	0.022	0.009
4.0	97.2	0.026	0.035	0.024	0.023	0.022	0.022	0.006
Type I Error Rates ($H_0\colon \beta_0 = \alpha$)								
0.0	55.0	1.00	0.89	0.96	0.90	1.00	1.00	1.00
0.5	65.4	1.00	0.74	0.91	0.92	1.00	1.00	1.00
1.0	74.7	1.00	0.29	0.53	0.53	0.98	0.96	1.00
1.5	82.0	1.00	0.10	0.18	0.17	0.91	0.79	1.00
2.0	87.4	1.00	0.08	0.09	0.09	0.87	0.68	1.00
2.5	91.3	1.00	0.08	0.07	0.07	0.87	0.65	0.92
3.0	94.0	1.00	0.08	0.07	0.07	0.87	0.65	0.63
3.5	95.9	0.94	0.08	0.07	0.07	0.87	0.66	0.36
4.0	97.2	0.70	0.08	0.07	0.07	0.87	0.66	0.20

1.3.3. Panel Random Effects Data Environment

Experimental Design. The last set of experiments examines the performance of the respective MA estimators when each study contains multiple regressions/effect estimates. The true effects are modelled as differing both across and within studies.

There is a debate in the literature as to whether MA studies should include all estimates from a study, or just one, or a selected few. To the extent a consensus exists, it is that MA estimators should include all the estimates, but correct for error correlation across estimates within studies (Stanley and Doucouliagos, 2012; Ringquist, 2013).

My Monte Carlo experiments fix the number of pre-publication bias studies at 100, each with 10 estimates per study, where each estimate is based upon 100 observations. True effects are modelled as differing both within and across studies, with the differences within studies, σ_1^2, being smaller than the differences across studies, σ_2^2, such that $var\left(\alpha_{ij} \middle| \alpha_i\right) < var(\alpha_i)$.[2]

$$y_{ijt} = 1 + \alpha_{ij} x_{ijt} + e_{ijt}, t = 1, 2, \ldots, 100, \text{ where} \quad (1.18)$$

$$\alpha_{ij} = \alpha_i + 0.5N(0,1), j = 1, 2, \ldots, 10, \text{ and} \quad (1.19)$$

$$\alpha_i = \alpha + 2N(0,1), i = 1, 2, \ldots, 100 \quad (1.20)$$

The different weights on the standard normal variates in (1.19) and (1.20) are designed to capture the idea that effects are more likely to be similar within a study than across studies.

The error terms are modelled similarly, with error variances again differing both within and across studies, but with most of the variation occurring across studies.

$$e_{ijt} = \lambda_{ij} \cdot NID(0,1), \text{ where} \quad (1.21)$$

[2] In my experiments, σ_1^2 and σ_2^2 are set equal to 0.25 and 4, respectively.

$\lambda_{ij} = \lambda_i + UID(0,1)$, and (1.22)

$\lambda_i = 0.5 + 30 \cdot UID(0,1)$ (1.23)

As in the Random Effects data environment, these DGP parameters are designed to simultaneously satisfy the four criteria listed above.

Publication bias is also treated differently in the panel random effects environment. The experiments assume that the bias works at the level of the study and not the individual estimate. In the case of bias against statistical insignificance, I assume that in order to be published, a study must have most of its estimates (at least 7 out of 10) be statistically significant. If the study meets that selection criterion, all the estimates from that study are "published". If the study does not meet that criterion, none of the estimates from that study are published. An identical "7 out of 10, or more" rule applies to publication bias against wrong-signed estimates.

Another difference has to do with the specification of the MA regressions. I modify equation (1.1) to include multiple estimates per study:

$\hat{\alpha}_{ij1} = \beta + e_{ij}$ (1.24)

Dividing through by the appropriate standard error (either SE_{ij} or $\omega_{ij} = \sqrt{\left(SE_{ij}\right)^2 + \tau^2}$) produces the FE, WLS, and RE estimators as described above.

The PET estimator follows the recommendation of Stanley and Doucouliagos (2012, see (i) Equation 5.5, p.85, and (ii) Equation 5.9, .101):

$\hat{\alpha}_{ij1} = \beta + \sum_i \gamma_i SE_{ij} D_i + e_{ij}$ (1.25)

where D_i is a dummy variable that takes the value 1 for study i and 0 for other studies. Dividing through by SE_{ij} produces the following specification:

$$\frac{\hat{\alpha}_{ij1}}{SE_{ij}} = \beta\left(\frac{1}{SE_{ij}}\right) + \sum_i \gamma_i D_i + \frac{(e_{ij})}{SE_{ij}} \quad (1.26)$$

The panel version of PEESE estimator is given by:

$$\frac{\hat{\alpha}_{ij1}}{SE_{ij}} = \beta\left(\frac{1}{SE_{ij}}\right) + \sum_i \gamma_i SE_i D_i + \frac{(e_{ij})}{SE_{ij}} \quad (1.27)$$

For all estimators except the FE estimator, coefficient standard errors are calculated using a clustered robust procedure to allow for within-study correlation of error term.

The above experimental design is intended to capture the fact that studies typically contain more than one estimate of a given "effect", perhaps because separate regressions are estimated for different subsamples of the data, or because the regression equations differ in their specifications or econometric procedures used. Thus, a realistic study of meta-analysis performance should incorporate this feature.

Table 1.7 give average sample characteristics for a typical meta-analysis sample in the Panel Random Effects data environment when $\alpha = 1$, both pre- and post-publication bias. The associated parameter values in the DGP above have been chosen to produce a range of estimated effects and t-statistics similar to those produced in the Fixed Effects and Random Effects data environments (cf. Tables 4.1 and 4.4). As in Table 4.4, I again report a measure of effect heterogeneity, I^2. As mentioned earlier, I^2 values between 70-95% are common in meta-analysis studies conducted in economics and business. Table 4.7 makes clear that the simulated meta-analysis samples that I use for analysing the performance of the FAT-PET-PEESE estimator "look like" the kinds of meta-analysis samples that researchers apply in practice.

Table 1.7: Sample Characteristics for a Simulated Meta-Analysis Data Set (Panel Data/Random Effects ($\alpha = 1$))

Variable	*Median*	*Minimum*	*P5%*	*P95%*	*Maximum*
Pre-Publication Bias (100 percent of estimates):					
Estimated effect	0.96	-7.50	-3.47	5.47	9.43
t-statistic	0.67	-10.16	-3.17	7.44	20.69
I^2	0.86	0.46	0.68	0.97	0.99
Publication Bias Against Insignificance (21.8 percent of estimates):					
Estimated effect	2.24	-3.03	-2.18	5.73	7.03
t-statistic	3.62	-8.81	-6.08	15.56	20.66
I^2	0.92	0.00	0.71	0.99	1.00
Publication Bias Against Negative Effects (56.2 percent of estimates):					
Estimated effect	2.22	-3.99	-0.84	6.17	9.33
t-statistic	1.75	-2.11	-0.50	11.28	20.66
I^2	0.74	0.11	0.43	0.94	0.98

Panel Random Effects: Performance Tests. Table 4.8 reports performance measures for the respective estimators when publication bias favours estimates that are statistically significant. As before, the Unadjusted estimator provides an unbiased estimate of the mean true effect when $\alpha = 0$. As α increases, publication bias at first worsens, and then eventually starts to improve as more studies are "published". The numerical bias can be quite substantial. For example, when $\alpha = 2.0$, the Unadjusted estimator estimates an average value of 3.36 for α.

With respect to bias, the PET, PEESE and FPP estimators perform best of all MA estimators, with the FPP performing marginally better. For example, when $\alpha = 2.0$, the PET and PEESE estimators produce a mean estimate of α equal to 2.24, compared to 2.37 and 3.31 for the MA estimators that do not correct for publication bias. The FPP estimator produces a least biased estimate of 2.21.

However, superiority on the dimension of bias does not necessarily translate into superiority in MSE performance. While the PET, PEESE, and also FPP estimators are least biased, they are also least efficient among the MA estimators, and sometimes even less efficient than the Unadjusted estimator (cf. $0 \leq \alpha \leq 1$). Among MA estimators, the FE/WLS estimators are generally most efficient, though the RE estimator is best for low values of α.

Finally, when it comes to hypothesis testing, the lesson from the bottom panel of Table 4.8 could perhaps be summarized as "don't". In almost every case, the Type I error rates are so much larger than 5 percent that any results derived from hypothesis testing about the mean true effect should be regarded as highly dubious.

Table 1.8: Comparative Performance of Meta-Analysis Estimators (Panel Random Effects/Publication Bias against Insignificance)

α	*Percent*	*Unadjusted*	*PET*	*PEESE*	*FPP*	*FE*	*WLS*	*RE*
				Average Estimate of Mean True Effect				
0.0	19.2	0.01	0.01	0.01	0.01	0.01	0.01	0.01
0.5	19.9	1.09	0.61	0.61	0.59	0.66	0.66	1.01
1.0	22.0	2.05	1.20	1.21	1.18	1.29	1.29	1.90
1.5	25.2	2.78	1.73	1.74	1.71	1.85	1.85	2.59
2.0	29.5	3.36	2.24	2.24	2.21	2.37	2.37	3.13
2.5	34.7	3.84	2.74	2.75	2.73	2.86	2.86	3.60
3.0	40.4	4.26	3.21	3.21	3.20	3.31	3.31	4.00
3.5	46.4	4.65	3.66	3.67	3.66	3.76	3.76	4.39
4.0	52.8	5.03	4.11	4.12	4.11	4.20	4.20	4.77
				Mean Squared Error				
0.0	19.2	0.506	1.765	1.553	1.629	0.874	0.874	0.443
0.5	19.9	0.796	1.767	1.554	1.628	0.879	0.879	0.655
1.0	22.0	1.435	1.700	1.506	1.577	0.880	0.880	1.111
1.5	25.2	1.866	1.673	1.465	1.548	0.851	0.851	1.387
2.0	29.5	2.000	1.531	1.341	1.415	0.782	0.782	1.428
2.5	34.7	1.916	1.461	1.277	1.338	0.722	0.722	1.312
3.0	40.4	1.671	1.415	1.231	1.281	0.652	0.652	1.094
3.5	46.4	1.397	1.335	1.159	1.198	0.577	0.577	0.874
4.0	52.8	1.126	1.287	1.107	1.138	0.527	0.527	0.670
				Type I Error Rates ($H_0: \beta_0 = \alpha$)				
0.0	19.2	0.05	0.29	0.28	0.25	0.97	0.17	0.05
0.5	19.9	0.15	0.29	0.29	0.29	0.97	0.17	0.14
1.0	22.0	0.43	0.29	0.29	0.30	0.97	0.19	0.37
1.5	25.2	0.71	0.30	0.30	0.31	0.97	0.22	0.62
2.0	29.5	0.89	0.29	0.29	0.29	0.97	0.23	0.80
2.5	34.7	0.95	0.29	0.29	0.29	0.97	0.23	0.88
3.0	40.4	0.98	0.29	0.29	0.29	0.96	0.21	0.90
3.5	46.4	0.98	0.27	0.26	0.27	0.96	0.18	0.89
4.0	52.8	0.98	0.28	0.27	0.27	0.96	0.17	0.84

Table 1.9 provides further support that superiority on the dimension of biasedness does not imply superiority on efficiency. The RE estimator is now either best or close to best on the dimension of MSE for all values of α. Meanwhile, the Unadjusted estimator is more efficient than every MA estimator except the RE estimator. Reliability in hypothesis testing for the estimators continues to be abysmal across the full range of α values.

Table 1.9: Comparative Performance of Meta-Analysis Estimators (Panel Random Effects /Publication Bias against Wrong Sign)

α	*Percent*	*Unadjusted*	*PET*	*PEESE*	*FPP*	*FE*	*WLS*	*RE*
				Average Estimate of Mean True Effect				
0.0	38.4	2.01	1.74	1.74	1.69	1.77	1.77	1.88
0.5	47.7	2.19	1.92	1.92	1.89	1.94	1.94	2.07
1.0	56.8	2.40	2.14	2.15	2.12	2.17	2.17	2.29
1.5	65.6	2.63	2.40	2.40	2.38	2.41	2.41	2.53
2.0	73.6	2.89	2.66	2.66	2.64	2.68	2.68	2.80
2.5	80.6	3.19	3.00	3.00	2.98	3.01	3.01	3.11
3.0	86.2	3.51	3.35	3.35	3.34	3.36	3.36	3.45
3.5	90.6	3.87	3.73	3.73	3.72	3.73	3.73	3.82
4.0	93.9	4.26	4.14	4.14	4.13	4.14	4.14	4.22
				Mean Squared Error				
0.0	38.4	4.090	3.897	3.664	3.591	3.414	3.414	3.592
0.5	47.7	2.900	2.884	2.672	2.648	2.388	2.388	2.513
1.0	56.8	2.002	2.176	1.999	1.997	1.672	1.672	1.709
1.5	65.6	1.312	1.703	1.522	1.540	1.143	1.143	1.106
2.0	73.6	0.830	1.362	1.194	1.221	0.796	0.796	0.689
2.5	80.6	0.507	1.231	1.061	1.084	0.609	0.609	0.418
3.0	86.2	0.299	1.173	1.004	1.027	0.502	0.502	0.245
3.5	90.6	0.172	1.143	0.980	1.000	0.445	0.445	0.144
4.0	93.9	0.105	1.142	0.973	0.989	0.423	0.423	0.092
				Type I Error Rates ($H_0: \beta_0 = \alpha$)				
0.0	38.4	1.00	0.78	0.90	0.78	1.00	0.99	1.00
0.5	47.7	1.00	0.67	0.77	0.75	1.00	0.92	1.00
1.0	56.8	1.00	0.54	0.61	0.61	1.00	0.77	1.00
1.5	65.6	1.00	0.44	0.47	0.48	0.99	0.56	1.00
2.0	73.6	1.00	0.36	0.38	0.38	0.97	0.38	0.98
2.5	80.6	0.97	0.32	0.33	0.33	0.96	0.28	0.87
3.0	86.2	0.79	0.30	0.30	0.30	0.96	0.21	0.60
3.5	90.6	0.49	0.28	0.27	0.27	0.95	0.17	0.34
4.0	93.9	0.26	0.28	0.27	0.27	0.96	0.16	0.18

1.3.4. Funnel Asymmetry and Precision Effect Tests

Almost all MRA studies start by testing whether or not there is a publication bias. This helps meta-analysts to determine whether accommodating and correcting for publication bias is required. Testing $H_0: \beta_1 = 0$ in the following equation is designed to answer this question (FAT).

$$\frac{\hat{\alpha}_{i1}}{SE_i} = \beta_0 \left(\frac{1}{SE_i}\right) + \beta_1 + \frac{e_i}{SE_i} \tag{1.28}$$

Further, testing $H_0: \beta_0 = 0$ is a test for the presence of a genuine non-zero effect (PET). The results from FAT and PET hypothesis testing are reported in Table 1.10.

As noted above, there are six classes of experiments based on the pairing of: (i) type of date environment (Fixed Effects, Random Effects, Panel Randome Effects), and (ii) type of publication bias. The table is divided vertically into three panels according to type of data environment, from least realistic (Fixed Effects) to most realistic (Panel Random Effects). It is divided horizontally into left and right halves based on type of publication bias. The far left column reports the true overall effect, α.

I start with the Fixed Effects data environment, where each study produces only one estimate and there is one population effect underlying all studies. Each cell in the table reports the results of testing 10,000, simualted, post-publication bias, meta-analysis samples. Each meta-analysis sample starts with 1,000 estimates, but not all of these are observed by the meta-analysts due to publication bias.[3] For example, when $\alpha = 0$ and publication bias is directed against insignificance (cf. left side of the table), the average meta-analysis sample contains 143 studies/estimates (14.3 percent).

[3] Note that, for the FE and RE DGPs, there is one estimate per study. However, for the PRE DGP, there are 100 studies, each containing 10 estimates.

Each of these 10,000 meta-analysis samples is tested for publication bias (FAT). As discussed above, under publication bias against insignificance, when $\alpha = 0$, there is no bias in the estimate of the overall effect, so that the null hypothesis is true. The FAT performs very well in this case, producing a rejection rate of 6 percent--close to its 5 percent significance level. In contrast, the PET is oversized with a 16 percent rejection rate. Both the FAT and the PET show excellent power. Rejection rates for the null hypotheses of no publication bias and no effect are 100 percent whenever $\alpha > 0$.

Continuing with publication bias against insignificance (left side of the table), I move down a panel to the more realistic case of Random Effects. While the rejection rates of 0.08 for both the FAT and PET are close to their significance levels when $\alpha = 0$, the tests do not perform as well when $\alpha > 0$. For example, when $\alpha = 0.5$, the FAT rejects the (false) null of no publication bias only about 33 percent of the time. The PET fails to reject the (false) null of no effect approximately 35 percent (=1-0.65) of the time. While the performances of the FAT and PET generally improve as α increases, the tests are not as reliable as they were in the Fixed Effects data environment.

The bottom panel reports results for the most realistic data environment, Panel Random Effects, where studies contain more than one estimate and there is heterogeneity in true effects both within and across studies. Both the FAT and the PET perform substantially worse.[4] When $\alpha = 0$ and publication bias is directed towards statistical insignificance, the FAT rejects the (true) null of no publication bias over half of the time (55 percent). The PET rejects the true null 29 percent of the time, and that rejection rate increases slowly as α gets larger.

[4] Heteroskedasticity-robust standard errors were used when testing hypotheses in the FE and RE cases. Clustered robust standard errors were used in the PRE case.

Moving to the right side of the table and beginning again with the top panel, I see that the FAT again does well within the Fixed Effects data environment. When publication bias is directed against negatively signed estimates and $\alpha = 0$, so that approximately half of all estimates are wrong signed, the FAT rejects the null of no publication bias 100 percent of the time. As α increases, fewer and fewer estimates are eliminated via publication bias, so that publication bias diminishes. Correspondingly the rejection rate from the FAT also falls.

The PET also performs well. When $\alpha = 0$, 45 percent (=100-55) of the estimates are eliminated because of negative signs. This causes the remaining estimates to have a strong positive bias. Even so, the PET is not fooled, and generally leads to the correct conclusion of no effect: The rejection rate of 8 percent is close to its 5 percent significance level. Further, the PET accurately identifies the existence of a nonzero effect 100 percent of the time for all $\alpha > 0$.

As before, the performances of the FAT and PET decline as the data environments become more realistic. Compared to the 100 percent rejection rate for the FAT when $\alpha = 0$ in the Fixed Effects environment, the FAT falls to 62 percent for the same scenario in the Random Effects data environment. Likewise, the PET finds evidence of an effect 90 percent of the time under Random Effects when there is, in fact, no effect ($\alpha = 0$). Things decline further still in the most realistic environment of Panel Random Effects. The FAT is largely insensitive to changes in the degree of publication bias, and the ability of the PET to identify an effect when there really is one is worse. In summary, while the FPP procedure does very well in the basic, unrealistic case of a Fixed Effects data environment, its performance declines substantially when data environments become more realistic.

Table 1.10: Funnel Asymmetry Testing (FAT) and Precision Effect Testing (PET)

	Publication Bias against Insignificant			Publication Bias against Wrong Sign		
Fixed Effects Data Environments						
α	***Percent***	**FAT**	**PET**	***Percent***	**FAT**	**PET**
0.0	14.3	0.06	0.16	55.0	1.00	0.08
0.5	23.0	1.00	1.00	71.7	1.00	1.00
1.0	31.8	1.00	1.00	80.6	1.00	1.00
1.5	40.0	1.00	1.00	86.5	1.00	1.00
2.0	47.6	1.00	1.00	90.6	1.00	1.00
2.5	54.6	1.00	1.00	93.5	0.98	1.00
3.0	61.1	1.00	1.00	95.5	0.81	1.00
3.5	67.0	1.00	1.00	97.0	0.53	1.00
4.0	72.2	1.00	1.00	98.0	0.30	1.00
Random Effects Data Environments						
α	***Percent***	**FAT**	**PET**	***Percent***	**FAT**	**PET**
0.0	27.1	0.08	0.08	55.0	0.62	0.90
0.5	28.7	0.33	0.65	65.4	0.62	1.00
1.0	33.0	0.67	0.99	74.7	0.56	1.00
1.5	39.1	0.79	1.00	82.0	0.48	1.00
2.0	45.9	0.79	1.00	87.4	0.35	1.00
2.5	52.8	0.75	1.00	91.3	0.24	1.00
3.0	59.2	0.69	1.00	94.0	0.17	1.00
3.5	65.1	0.61	1.00	95.9	0.13	1.00
4.0	70.4	0.55	1.00	97.2	0.10	1.00
Panel Random Effects Data Environments						
α	***Percent***	**FAT**	**PET**	***Percent***	**FAT**	**PET**
0.0	19.2	0.55	0.29	38.4	0.45	0.78
0.5	19.9	0.58	0.34	47.7	0.47	0.84
1.0	22.0	0.66	0.46	56.8	0.44	0.87
1.5	25.2	0.72	0.60	65.6	0.43	0.91
2.0	29.5	0.66	0.73	73.6	0.46	0.93
2.5	34.7	0.59	0.83	80.6	0.50	0.95
3.0	40.4	0.66	0.90	86.2	0.46	0.97
3.5	46.4	0.69	0.94	90.6	0.45	0.98
4.0	52.8	0.65	0.97	93.9	0.41	0.99

1.4. Conclusion

This uses Monte Carlo simulation to evaluate the performance of the FAT-PET-PEESE (FPP) procedure, a commonly employed approach for detecting and correcting publication bias in economics and business meta-analyses. The FPP procedure addresses three main objectives: (i) testing whether the sample of estimates is influenced by publication selection bias; (ii) testing whether there is a genuine non-zero true effect of estimates once the publication bias is accommodated and corrected; and (iii) obtaining an estimate of the overall mean effect.

My analysis investigated two types of publication bias: (i) publication bias against insignificant results and (ii) publication bias against wrong-signs. I also considered three data environments: (i) the Fixed Effects data environment where each study only contains one estimated effect, and where there is one true effect underlying all studies, so that all differences in estimated effects are due to sampling error; (ii) the Random Effects data environment where each study still only has one estimated effect, but where there is a distribution of true effects across studies; and (iii) the Panel Random Effects data environment where studies contain multiple estimates and there is heterogeneity in true effects both across estimates and within studies. The Panel Random Effects data environment is the data environment that most realistically models what meta-analysts in business and economics are likely to encounter in their research.

My findings indicate that the FPP procedure performs very well in the basic environment of Fixed Effects. However, in more realistic data environments, where there is heterogeneity in true effects both across and within studies, the FPP procedure's performance is generally poor. It is unreliable for the first two objectives, and less efficient than some other estimators that are not particularly designed to correct for publication bias. Further, hypothesis tests about the overall mean effect cannot be trusted.

The main conclusions I draw from thie research in this chapter are as follows. First, meta-analyses should routinely report measures of heterogeneity such as I^2. This is not standard practice in the economics and business literatures and should be. Second, future research should more intensively explore the conditions under which FPP performs well. As noted elsewhere (Stanley, 2008; Moreno et al., 2009; Stanley and Doucouliagos, 2014a), publication bias is a serious problem and the FPP procedure has shown great promise in mitigating its deleterious consequences in some cases. Having a better understanding of where the FPP procedure can be successfully applied is an important topic for future research.

1.5. References

Borenstein, M., Hedges, L. V., Higgins, J., and Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97-111.

Chliova, M., Brinckmann, J., and Rosenbusch, N. (2015). Is microcredit a blessing for the poor? A meta-analysis examining development outcomes and contextual considerations. *Journal of Business Venturing*, *30*(3), 467-487.

Costa-Font, J., Gemmill, M., and Rubert, G. (2011). Biases in the healthcare luxury good hypothesis? A meta-regression analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(1), 95-107.

Doucouliagos, H., Haman, J., and Stanley, T.D. (2012). Pay for performance and corporate governance reform. *Industrial Relations*, *51*, 670- 703.

Doucouliagos, H., and Paldam, M. (2013). The robust result in meta-analysis of aid effectiveness: A response to Mekasha and Tarp. *The Journal of Development Studies*, *49*(4), 584-587.

Doucouliagos, H. and Stanley, T.D. (2009a). Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations, 47*, 406-29.

Doucouliagos, H., and Paldam, M. (2009b). The aid effectiveness literature: the sad results of 40 years of research. *Journal of Economic Surveys*, *23*(3), 433-461.

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

Doucouliagos, H., Stanley, T. D., and Viscusi, W. K. (2014). Publication selection and the income elasticity of the value of a statistical life. *Journal of Health Economics*, *33*, 67-75.

Efendic, A., Pugh, G., and Adnett, N. (2011). Institutions and economic performance: A meta-regression analysis. *European Journal of Political Economy*, *27*(3), 586-599.

Eisend, M. (2015). Have we progressed marketing knowledge? A meta-analysis of effect sizes in marketing research. *Journal of Marketing, 79*(3), 23-40.

Gunby, P., Jin, Y., and Reed, W.R. (2017). Did FDI really cause Chinese economic growth? A meta-analysis. *World Development, 90*, 242-255.

Haelermans, C., and Borghans, L. (2012). Wage effects of on-the-job training: A meta-analysis. *British Journal of Industrial Relations*, *50*(3), 502-528.

Havránek, T. (2010). Rose effect and the euro: is the magic gone? *Review of World Economics*, *146*, 241-261.

Higgins, J.P. and Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539-58.

Hughes, A.M., Gregory, M.E., Joseph, D.L., Sonesh, S.C., Marlow, S.L., Lacerenza, C.N., Benishek, L.E., King, H.B., and Salas, E. (2016). Saving lives: A meta-analysis of team training in healthcare. *Journal of Applied Psychology*, *101*(9), 1266-1304.

Iwasaki, I., and Tokunaga, M. (2014). Macroeconomic impacts of FDI in transition economies: a meta-analysis. *World Development*, *61*, 53-69.

Khechine, H., Ndjambou, P., and Lakhal, S. (2016). A meta-analysis of the UTAUT model: Eleven years later. *Canadian Journal of Administrative Sciences*, *33*, 138-152.

Laroche, P. (2016). A meta-analysis of the union–job satisfaction relationship. *British Journal of Industrial Relations*, *54*(4), 709-741.

Lazzaroni, S., and van Bergeijk, P. A. (2014). Natural disasters' impact, factors of resilience and development: A meta-analysis of the macroeconomic literature. *Ecological Economics*, *107*, 333-346.

Linde Leonard, M., Stanley, T. D., and Doucouliagos, H. (2014). Does the UK minimum wage reduce employment? A meta-regression analysis. *British Journal of Industrial Relations*, *52*(3), 499-520.

Moayeri, F., Hsueh, Y.S., Clarke, P., and Dunt, D. (2016). Do model-based studies in chronic obstructive pulmonary disease measure correct values of utility? A meta-analysis. *Value in Health*, *19*(4), 363-373.

Nelson, J. P. (2013). Meta-analysis of alcohol price and income elasticities – with corrections for publication bias. *Health Economics Review*, *3*(1), 1-10.

Reed, W. R. (2015). A Monte Carlo analysis of alternative meta-analysis estimators in the presence of publication bias. *Economics: The Open-Access, Open-Assessment E-Journal*, 9 (2015-30): 1—40. http://dx.doi.org/10.5018/economics-ejournal.ja.2015-30

Reed, W. R., Florax, R. J. G. M., and Poot, J. (2015). A Monte Carlo analysis of alternative meta-analysis estimators in the presence of publication bias. Economics Discussion Papers, No 2015-9, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2015-9

Ringquist, E. (2013). *Meta-analysis for public management and policy*. San Franciso, CA : John Willey & Sons.

Rodríguez, J. and Muro, J. (2015). On the size of sheepskin effects: A meta-analysis. *Economics: The Open-Access, Open-Assessment E-Journal*, *9,* (2015-37): 1—18. http:// dx.doi.org/10.5018/economics-ejournal.ja.2015-37

Sarooghi, H., Libaers, D., and Burkemper, A. (2015). Examining the relationship between creativity and environmental factors. *Journal of Business Venturing*, *30*(5), 714-731.

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

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

Stanley, T. D., Jarrell, S. B., and Doucouliagos, H. (2012). Could it be better to discard 90% of the data? A statistical paradox. *The American Statistician*, *64*(1), 70-77.

Stanley, T.D., and Doucouliagos, H. (2014a). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*(1), 60-78.

Stanley, T.D., and Doucouliagos, H. (2014b). Better than random: Weighted least squares meta-regression analysis. Working paper, Deakin University, School of Accounting, Economics, and Finance, Economics Series, SWP 2013/2, updated February 2014.

Stanley, T. D., and Doucouliagos, H. (2015). Neither fixed nor random: weighted least squares meta-analysis. *Statistics in Medicine*, *34*(13), 2116-2127.

Triana, M. D. C., Jayasinghe, M., and Pieper, J. (2015). Perceived workplace racial discrimination and its correlates: A meta-analysis. *Journal of Organizational Behavior, 36*(4), 491-513.