DEAKIN
UNIVERSITY AUSTRALIA

Faculty of Business and Law
School of Accounting, Economics and Finance

School Working Paper

ECONOMICS SERIES

SWP 2013/2

Updated July 2014

Better than Random: Weighted Least Squares Meta-Regression Analysis

T.D. Stanley and Hristos Doucouliagos

Better than Random: Weighted Least Squares Meta-Regression Analysis

by

T.D. Stanley* and Hristos Doucouliagos**

Abstract

Our study revisits and challenges two core conventional meta-regression models: the prevalent use of 'mixed-effects' or random-effects meta-regression analysis (RE-MRA) and the correction of standard errors that defines fixed-effects meta-regression analysis (FE-MRA). We show how and explain why the traditional, unrestricted weighted least squares estimator (WLS-MRA) is superior to conventional random-effects (or mixed-effects) meta-regression when there is publication (or small-sample) bias and as good as FE-MRA in all cases and better in most practical applications. Simulations and statistical theory show that WLS-MRA provides satisfactory estimates of meta-regression coefficients with confidence intervals that are comparable to mixed-or random-effects when there is no publication bias. When there is publication selection bias, WLS-MRA dominates mixed- and random-effects, especially when there is large additive heterogeneity as assumed by the random-effects meta-regression model.

Keywords: meta-regression, weighted least squares, random-effects, fixed-effects.

* Julia Mobley Professor of Economics, Hendrix College, 1600 Washington St., Conway, AR, 72032 USA. Email: Stanley@hendrix.edu. Phone: 1-501-450-1276; Fax: 1-501-450-1400.

** Professor of Economics, School of Accounting, Economics and Finance and Alfred Deakin Research Institute, Deakin University, 221 Burwood Highway, Burwood, 3125, Victoria, Australia. Email: douc@deakin.edu.au. Phone: 61 03 9244 6531.

Better than Random: Weighted Least Squares Meta-Regression Analysis

1. INTRODUCTION

Multiple meta-regression analysis (MRA) is widely used by systematic reviewers to explain the excess systematic variation often observed across research studies, whether experimental, quasi-experimental or observational. Hundreds of meta-regression analyses are conducted each year. The conventional approach to the estimation of multiple meta-regression coefficients and their standard errors is 'random' or 'mixed-effects' MRA (Sharp 1998; Knapp and Hartung, 2003; Higgins and Thompson, 2004; Borenstein *et al*., 2009; Moreno *et al*., 2009; Sterne, 2009; White, 2011). To focus on the essential difference between traditional, unrestricted weighted least squares (WLS), fixed-effects and random/mixed-effects meta-regression, we designate any meta-regression that adds a second independent, random term as a 'random-effects' meta-regression analysis (RE-MRA), encompassing mixed-effects. The conventional status of random-effects meta-regression analysis is most clearly seen by the fact that only RE-MRAs are estimated in STATA's meta-regression routines (Sharp 1998; Sterne, 2009; White, 2011).

In this paper, we investigate how well the traditional, unrestricted weighted least squares approach to meta-regression (WLS-MRA) summarizes estimated regression coefficients from observational research, explains systematic heterogeneity among independent research results, compares to conventional random-effects meta-regression, and how it can successfully correct observational research's routine misspecification biases. Our simulations show that the traditional, 'unrestricted' weighted least squares MRA is as good as or better than conventional random-effects MRA in summarizing and correcting regression estimates. Although we report simulation results only for observational estimates of regression coefficients in this paper, we have reason to believe that our findings generalize to the meta-regression of experimental results as well (Stanley and Doucouliagos, 2013b).

Conventional practice among economists has long been to use the traditional, *unrestricted* weighted least squares meta-regression analysis (WLS-MRA) to explain and summarize estimated regression coefficients because it allows for both heteroscedasticity and excess heterogeneity (Stanley and Jarrell, 1989; Stanley and Doucouliagos, 2012).

One of the weaknesses of random/mixed-effects meta-regression analysis is that it assumes that the random-effects are independent of the model's moderator variables. Otherwise, the regression estimates are known to be biased (Wooldridge, 2002; Davidson and MacKinnon, 2004). For this and other reasons, Stanley and Doucouliagos (2012) speculate that random-effects meta-regression analysis will be more biased than traditional weighted least squares when the reported research literature contains selection for statistical significance (conventionally called, 'publication' or 'small-sample' bias). Unfortunately, the presence of 'publication' or 'small-sample' bias is common in most areas of research, not only economics (Sterling *et al.,* 1995; Gerber *et al*., 2001; Gerber and Malhorta, 2008; Hopewell *et al*., 2009; Doucouliagos and Stanley, 2013). We show in this paper that our conjecture is correct: random-effects MRA is indeed more biased than WLS-MRA in the presence of publication or small-sample bias.

We fully acknowledge that all of these alternative meta-regression approaches: WLS-MRA, FE-MRA, RE-MRA and mixed-effects MRA employ weighted least squares and that WLS has long been used by meta-analysts: Stanley and Jarrell (1989), DuMouchel (1990), Raudenbush (1994), Thompson and Sharp (1999), Higgins and Thompson (2002), Steel and Kammeyer-Mueller (2002), Baker *et al.* (2009), Copas and Lozada (2009), and Moreno *et al*. (2009), to cite a few. However, fixed-, mixed- and random-effects MRA restrict the traditional WLS multiplicative constant to be one; whereas traditional WLS does not. To our knowledge, no meta-analyst has suggested that the traditional, *unrestricted* weighted least squares meta-regression should routinely replace random-, mixed-, and fixed-effects meta-regression analysis.

The central purpose of this paper is to investigate the bias of random-effects meta-regression analysis relative to traditional WLS meta-regression when there is publication selection bias. When there are no publication or small-sample biases, our simulations demonstrate how WLS-MRA provides adequate estimates of meta-regression coefficients and their confidence intervals, comparable to what RE-MRA produces. Unfortunately, systematic reviewers can never be confident that there is no publication bias in any given area of research, because tests for publication and small-sample biases are known to have low power (Egger *et al*., 1997; Stanley, 2008). Thus, systematic reviewers have reason to prefer WLS-MRA over RE-MRA in practical applications. What is especially interesting is that WLS-MRA's improvement over RE-MRA is greatest in those exact circumstances for which RE-MRA is designed—large additive heterogeneity.

In the process of documenting how the traditional unrestricted weighted least squares estimator is superior to conventional random-effects meta-regression, we show that a simple meta-regression model that uses binary dummy variables corrects for misspecification biases routinely found in observational research. Thus, our study demonstrates how a general, unrestricted weighted least squares framework can remove or reduce a wide variety of biases routinely found in social science research.

2. GAUSS-MARKOV THEOREM AND WLS-MRA

Suppose that the reviewer wishes to summarize and explain some reported empirical effect, y_j. The basic form of the meta-regression model needed to explain variation among these reported effects is:

$$\mathbf{y} = \mathbf{M}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where:

- **y** is a Lx1 vector of all comparable reported empirical effects in an empirical literature of L estimates.
- **M** is a LxK matrix of explanatory or moderator variables, the first column of which contains all 1s.
- $\boldsymbol{\beta}$ is a Kx1 vector of MRA coefficients, the first of which represents the 'true' underlying empirical effect investigated. For this interpretation to be true, the moderator variables, **M**, need to be defined in a manner such that M_j= 1 represents the presence of some potential bias and M_j= 0 its absence. In the below simulations, M_j is defined in this way.
- $\boldsymbol{\varepsilon}$ is a Lx1 vector of residuals representing the unexplained errors of the reported empirical effects, and $\boldsymbol{\varepsilon} \sim (0, \mathbf{V})$.

Equation (1) cannot be adequately estimated by ordinary least squares (OLS), because systematic reviewers almost always find large variation among the standard errors of the reported effects. This means that reviewers directly observe large *heteroscedasticity* among reported estimates of effects, which define the dependent variable in a meta-regression analysis. Thus, $\boldsymbol{\varepsilon}$ from equation (1) cannot be assumed to be i.i.d., and OLS is almost never appropriate for MRA. At a minimum, meta-analysts need to adjust for this heteroscedasticity, and weighted least squares are the traditional remedy.

Weighted least squares estimates of MRA (1) are:

$$\hat{\boldsymbol{\beta}} = (\mathbf{M}^t\boldsymbol{\Omega}^{-1}\mathbf{M})^{-1}\mathbf{M}^t\boldsymbol{\Omega}^{-1}\mathbf{y}, \tag{2}$$

where:

$$\boldsymbol{\Omega} = \sigma^2\mathbf{V} = \sigma^2\begin{vmatrix} \sigma_1^2 & 0 & \bullet & \bullet & 0 \\ 0 & \sigma_2^2 & & & 0 \\ \bullet & & \bullet & & \bullet \\ \bullet & & & \bullet & \bullet \\ 0 & 0 & \bullet & \bullet & \sigma_L^2 \end{vmatrix},$$

σ_j^2 is the variance of the j^{th} estimated effect, y_j, and σ^2 is a nonzero constant, which is routinely estimated by the MS_E when replacing $\hat{\boldsymbol{\beta}}$ in equation (1) (Greene, 1990; Davidson and MacKinnon, 2004). Note that $\boldsymbol{\Omega}^{-1}$ has $1/\sigma_j^2$ on the principal diagonal, zero elsewhere. These inverse variances are conventionally regarded as the weights in WLS routines and statistical packages.

WLS is a special case of generalized least squares (GLS) where the variance-covariance matrix, $\boldsymbol{\Omega}$, has the above diagonal structure. Aitken (1935) generalized the famous Gauss-Markov theorem that proves least squares estimates are minimum variance within the class of unbiased linear estimators for all positive semi-definite variance-covariance matrices (Jacquez *et al*., 1969; Stigler, 1986; Greene, 1990; Davidson and MacKinnon, 2004). Thus, WLS will also possess these Gauss-Markov statistical properties when σ_j^2s are known. Note that the WLS estimator, $\hat{\boldsymbol{\beta}}$, is invariant to any nonzero σ^2. This invariance is an obvious mathematical property of the WLS formula, equation (2), because $\sigma^2/\sigma^2=1$, for all $\sigma^2 \neq 0$. In practical application, the traditional *unrestricted* WLS is calculated by substituting consistent estimates (squared standard errors in meta-regressions) for σ_j^2, and σ^2 is automatically estimated by $MS_E = \mathbf{e}^t\mathbf{e}/(L-K)$, for $\mathbf{e} = \mathbf{y}-\mathbf{M}\hat{\boldsymbol{\beta}}$ (Kmenta, 1971; Judge *et al*., 1982; Greene, 1990; Wooldridge, 2002; Davidson and MacKinnon, 2004). In contrast, fixed-effects meta-regression restricts σ^2 to be one by dividing the meta-regression coefficients' standard errors by $\sqrt{MS_E}$ (Hedges and Olkin, 1985), thereby failing to make use of the WLS's multiplicative invariance property. Below, we show that there is never any statistical reason to divide by $\sqrt{MS_E}$.

3. ACCOMMODATING HETEROGENEITY

Meta-analysts in economics and the social sciences routinely observe excess heterogeneity. Heterogeneity is also quite common in medical research (Turner *et al*., 2012). Because individual and social behaviors are often unique, yet conditional upon a legion of factors (*e.g.,* socio-economic status, institutions, culture, framing, experience and history) that can rarely be fully controlled, experimentally or observationally, excess heterogeneity is the norm in economics and social scientific research. For example, among hundreds of meta-analyses conducted in economics, none have reported a Cochran's Q-test which allows the meta-analyst to accept homogeneity. Thus, the central objective for the meta-regression is to explain as much of systematic heterogeneity as possible and to accommodate any remaining heterogeneity.

3.1 Weighted Least Squares Meta-Regression

What is not fully appreciated among meta-analysts is that traditional unrestricted weighted-least squares estimates, MRA model (2), automatically adjusts for heterogeneity or 'over-dispersion' by calculating σ^2 from WLS's MS_E and that the resilience of least squares causes the resulting WLS-MRA estimates to rival or best both fixed- and random-effects estimates. As discussed above, the Gauss-Markov theorem proves that WLS provide unbiased and efficient estimators regardless of the amount of multiplicative over-dispersion. WLS-MRA estimates are invariant to the magnitude of known or unknown heterogeneity, and it retains all of these desirable properties even when a bad estimate of σ^2 is used. In contrast, random effects are highly sensitive to the accuracy of the estimate of the between-study variance, τ^2, and conventional estimates of τ^2 are biased (Raudenbush, 1994; Hedges and Vevea, 1998).

Although the ability of traditional WLS to test heterogeneity has been acknowledged, the traditional WLS-MRA has thus far been dismissed by meta-analysts (Thompson and Sharp, 1999). Because meta-analysts view this multiplicative variance structure as a requirement for using WLS-MRA rather than as WLS's resilience to heterogeneity, it is not highly regarded. "The rationale for using a multiplicative factor for variance inflation is weak. The idea that the variance of the estimated effect within each study should be multiplied by some constant has little intuitive appeal, . . . we do not recommend them in practice" (Thompson and Sharp, 1999, p. 2705).

We fully accept Thompson and Sharp's (1999) premise that the rationale for a multiplicative, rather than an additive, variance structure may be weak or even incorrect; however, our simulations demonstrate that their recommendations do not follow even if the random-effects model with its additive variance is true. In the simulations below, we assume that RE-MRA's model is true and show that the traditional WLS-MRA estimator is unbiased and provides acceptable confidence intervals, comparable to RE-MRA. These simulations find that there is little practical difference between WLS- and RE-MRA, when there is no selection for statistical significance (or publication bias). When there is publication selection bias, WLS-MRA dominates the corresponding random-effects meta-regression. That is, WLS-MRA is less biased and more efficient (smaller MSE) than RE-MRA on average and in the vast majority of cases. Because systematic reviewers can never rule out publication bias in practice, WLS-MRA should be routinely employed. Before we turn to the design of these simulations and their findings, we take a short detour to compare and discuss fixed- and random-effects meta-regressions.

3.2 Conventional Fixed-Effects and Random-Effects Meta-Regression

Random-effects (or mixed-effects) MRA is the conventional meta-regression model for excess heterogeneity. RE-MRA adds a second random term to MRA model (1).

$$\mathbf{y} = \mathbf{M\beta} + \nu + \boldsymbol{\varepsilon}, \tag{3}$$

where ν is a Lx1 vector of random effects, assumed to be independently distributed as $N(0,\tau^2)$ as well as independent of both $\boldsymbol{\varepsilon}$ and $\mathbf{M\beta}$. τ^2 is the random-effects (or heterogeneity) variance. Note that this RE-MRA model assumes that any excess random heterogeneity comes from an additive term, ν; whereas, WLS-MRA is invariant to any excess multiplicative variance, σ^2.

RE-MRA estimates of $\boldsymbol{\beta}$ are derived from either the method of moments or a maximum likelihood approach (Raudenbush, 1994). There are several related algorithms that provide these RE-MRA estimates, but typically they involve a two-step process. In the first step, τ^2 is estimated, and the second-step uses this estimate of τ^2, $\hat{\tau}^2$, to provide weights, $1/(SE_i^2+\hat{\tau}^2)$, in a restrictive (σ^2=1) weighted least squares context (Raudenbush, 1994). Our below simulations are based on Raudenbush's (1994) iterative maximum likelihood algorithm. The resulting RE-MRA estimates and confidence intervals calculated by our simulations are identical to five or more significant digits as those produced by STATA's random-effects meta-regression routine (Sharp, 1998).

In contrast to both RE-MRA and WLS-MRA, fixed-effects meta-regression assumes that there is no excess heterogeneity and constrains WLS's common variance, σ^2, to be equal to one. Otherwise, FE-MRA's model and estimated coefficients are identical to WLS-MRA—recall equations (1) and (2). The only difference is that FE-MRA further divides WLS-MRA's standard errors by $\sqrt{MS_E}$. WLS-MRA is identical to FE-MRA except that σ^2 is not restricted to be equal to one; hence, WLS-MRA may be regarded as 'unrestricted' weighted least squares.

Not only is excess heterogeneity not accommodated by FE-MRA's, it is forbidden. As a result, the confidence intervals produced by FE-MRA are widely recognized to be too narrow if misapplied to unconditional inference—that is, inferences for populations that might have some characteristic(s) that differs from the population sampled (Hedges, 1994; Borenstein *et al*., 2009). Fixed effects are, "in the strictest sense, limited to the factor levels represented in the sample. . . , (but) the generalizations made from them by researchers are typically not constrained precisely to factor levels in the study" (Hedges and Vevea, 1998, p. 488). The problem is that fixed-effects are often applied to settings for which they were not designed (*i.e.*, unconditional inference) and therefore purport to be more precise than they actually are. If one merely wishes to make inferences to a population identical to the one sampled, then fixed-effects' standard errors and confidence intervals are correct. However, they are not appropriate when making inferences to what might be found in future research or to the underlying value of the empirical effect in question because these might concern different conditions.

In recognition of the severe limitation of fixed-effects, some meta-analysts recommend against its use (Borenstein *et al*., 2009), while others view it as a viable meta-regression model (Lipsey and Wilson, 2001; Johnson and Huedo-Medina, 2012). Our simulations demonstrate that there are no circumstances for which fixed-effects have notably better statistic properties than traditional unrestricted WLS-MRA. When there is no excess heterogeneity and the population is identical to the one sampled, FE-MRA and WLS-MRA produce identical estimates and the confidence intervals are practically the same. Needless to say, when there is excess heterogeneity and fixed-effects are nonetheless inappropriately applied, WLS-MRA has clear superior coverage.

4. SIMULATIONS

Although the Gauss-Markov theorem is nearly two centuries old and weighted least squares are well established, WLS-MRA's relative performance under realistic research conditions requires further investigation. For one thing, RE-MRA is not a linear estimator, thereby negating the applicability of the Gauss-Markov theorem. Secondly, the unrestricted WLS-MRA, equation (2), involves a multiplicative error variance; whereas our below simulations *add* random unexplained heterogeneity to the data generating process. The WLS-MRA model is intentionally mis-specified because, in practice, unexplained heterogeneity might well be additive. Furthermore, it is important to be sure that WLS-MRA estimated coefficients will have desirable statistical properties even when the heterogeneity and errors are generated in any realistic manner. This paper does not concern which assumption or model about the structure of heterogeneity variance (multiplicative or additive) is the more correct. Rather, the point to these simulations is to demonstrate that WLS-MRA is preferred in practical applications over random-effects, even when random-effects MRA is the correct model.

This simulation generates *systematic heterogeneity* as omitted-variable bias, and all simulated MRAs model it with a binary dummy variable (0 if the relevant explanatory variable is included in the primary study's regression model; 1 if it is omitted). Omitting a relevant explanatory variable is an omnipresent threat to the validity of applied econometrics and observational research, in general. The resulting omitted-variable bias is well known and widely recognized (Judge *et al*., 1982; Stanley and Jarrell, 1989; Davidson and MacKinnon, 2004). However, what might be in doubt is whether such a crude binary variable adequately corrects this bias within meta-regression. In these simulations, we strive to be both realistic and challenging to the unrestricted WLS-MRA.

4.1 Simulation Design

Our simulation design closely follows past simulation studies of meta-regression models—Stanley (2008), Stanley *et al*., (2010); Stanley and Doucouliagos (2013), which, in turn, were calibrated to mirror several published meta-analyses (Stanley, 2008). Here, we generalize the past simulation design to allow for systematic heterogeneity and a wider range of parameters, because we wanted to be sure that we are challenging our preferred approach, fully. In particular, we fill in and widen the range of random heterogeneity to represent the values observed in recently published meta-analyses for which we can calculate I^2 (Higgins and Thompson, 2002)—see below for more detail.

Essentially, data are generated, and a regression coefficient is computed, representing one empirical effect reported in the research literature. Although estimated regression coefficients are the empirical effects meta-analyzed in the simulations reported here, we have reason to believe that the same general patterns would be found if standardized mean differences from RCTs are used (Stanley and Doucouliagos, 2013b). This process of generating data and estimating the target regression coefficient is repeated either 20 or 80 times (the MRA sample size).

Next, WLS-MRA, RE-MRA and FE-MRA are applied to estimate the underlying 'true' regression parameter corrected for misspecification biases that are contained in half of the research literature. In practice, omitted-variable bias is an omnipresent threat to observational studies. Initially, we assume that there are no publication selection biases. Later, we allow selection for statistical significance (publication bias) and model it using the selected estimate's standard error (SE_j) or variance as an additional moderator variable (Egger *et al*., 1997; Stanley and Doucouliagos, 2013a).

To be more detailed, the dependent variable, Z_i, for the regression model employed by the primary researchers is generated by:

$$Z_i = 100 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i \qquad . \tag{4}$$

$u_i \sim N(0, 100^2)$, $\alpha_1 = \{0, 1\}$ and $\alpha_2 = 0.5$. The empirical effect of interest is the estimate of α_1, $\hat{\alpha}_{1j}$. The correlation between Z and X_1 is either 0.27 (calculated and averaged from 10,000 replications) when $\alpha_1 = 1$ or zero when $\alpha_1 = 0$. Recall that 0.27 represents a small effect size by conventional guidelines (Cohen, 1988). These parameters were selected to make it a genuine challenge to identify this small effect ($\rho = 0.27$) amongst large random and systematic heterogeneity. Furthermore, such small effects are often the norm among regression studies (Doucouliagos, 2011). As routinely observed among systematic reviews of regression coefficients, a wide range of sample sizes are assumed to be used to estimate α_1 in the primary literature—n = {62, 125, 250, 500, 1000}. One of two regression models are used to estimate α_1 in the primary literature: a simple regression with only X_1 as the independent variable and Z as the dependent variable and a second model that employs two independent variables, X_1 and X_2.

X_2 is generated in a manner that makes it correlated with X_1. X_2 is set equal to X_1 plus a N(0, 50^2) random disturbance. When a relevant variable, like X_2, is omitted from a regression but is correlated with the included independent variable, like X_1, the estimated regression coefficient ($\hat{\alpha}_{1j}$) will be biased. This omitted-variable bias will be $\alpha_2 \cdot \gamma_{12}$; where γ_{12}=1 is the slope coefficient of a regression of X_{2i} on X_{1i}. For these simulations, we assume only that the reviewer can identify whether or not X_2 is included in the primary study's estimation model. When X_2 is omitted, M_j =1; M_j =0, otherwise. M then becomes an independent, or moderator, variable in the reviewer's meta-regression model:

$$y_j = \beta_0 + \beta_1 M_j + \varepsilon_j \, . \qquad (5)$$

As before, y_j is the ith reported effect, $\hat{\alpha}_{1j}$, and ε is the usual random regression disturbance. β_0 should be equal to the true regression coefficient, α_1, if M correctly detects and corrects this omitted-variable bias. MRA model (5) is then estimated using either fixed-effects, random-effects or traditional, unrestricted WLS with $\hat{\alpha}_{1j}$'s squared standard error, SE_j^2, as the estimate of σ_j^2 . Needless to say, the random-effects MRA adds a second random term, ν_j, to (5) as in equation (3). The only difference between WLS-MRA and FE-MRA is that FE-MRA further divides WLS-MRA's standard errors by $\sqrt{MS_E}$. Simulation results for these alternative meta-regression estimation approaches are reported in Table 1and 2.

Past simulations have found that the relative size of the unexplained heterogeneity is the most influential dimension (Stanley, 2008). In these simulations, unexplained random heterogeneity is induced by a second omitted-variable bias through X_3 and calibrated by σ_h (Stanley, 2008; Stanley *et al*., 2010; Stanley and Doucouliagos, 2013a). σ_h^2 is the variance of the regression coefficient α_3 in equation (4), and the magnitude of α_3 is generated by N(0, σ_h^2). α_3 is fixed for a given primary study but is random across studies. Thus, $\sigma_h^2 = \tau^2$ in the conventional random-effects context. Like X_2, X_3 is also generated in a way that makes it correlated with X_1, and α_3 is the bias in omitting X_3; however, X_3 is an omitted variable in all of the primary studies that estimate $\hat{\alpha}_{1j}$, rather than half of them. Because all studies omit X_3, our MRA does not and cannot correct this second source of omitted-variable bias.

We choose to induce random heterogeneity through omitted-variable bias, because this adds a random term, α_3, to the reported effects, just as modeled in the conventional 'random-effects' MRA, and allows us to easily calculate σ_h. Also, we believe that un-modeled, omitted-variable bias is the main source of excess unexplained heterogeneity and selection bias in econometrics and other areas of observational research. Values of random heterogeneity, σ_h, were selected to encompass what is found in actual meta-analyses, as measured by I^2, (Higgins and Thompson, 2002). For example, among US minimum wage elasticities, I^2 is 90% (Doucouliagos and Stanley, 2009), I^2 is 87% for efficiency wage elasticities (Krassoi Peach and Stanley (2009), 93% among estimates of the value of statistical life (Doucouliagos, Stanley and Giles, 2012), 97% among the partial correlations of CEO pay and corporate performance (Doucouliagos, Haman and Stanley, 2012), 99.2% among the income elasticities of health care (Costa-Font *et al*., 2013), and I^2 is 84% among the partial correlation coefficients of UK minimum wage increases and employment decreases (De Linde Leonard *et al*., 2013). Needless to say, smaller values of I^2 can also be found throughout the social and medical sciences. However, it is our experience that I^2s of 80 or 90 % are the norm.[1]

[1] An anonymous referee asked how common large values of heterogeneity are. Thus, we calculated I^2 where we could: our published meta-analyses over the last 5 years and works in progress. In addition to the six reported above, we have completed an additional three meta-analyses. Among test of market efficiency

We report two measures of I^2 in the below tables. 'I^2' measures the excess heterogeneity among estimates, $\hat{\alpha}_{1j}$, relative to estimated sampling error before either systematic heterogeneity or publication selection bias is included. I^2 is calculated 'empirically' from Higgins and Thompson (2002) and averaged across 10,000 replications. Empirical estimates of I^2 are biased upward when there is little or no excess heterogeneity (*i.e.,* σ_h =0). Like $\hat{\tau}^2$, conventional practice is to truncate I^2 at zero. Our tables also report a second measure of relative heterogeneity, 'residual I^2', which calculates the amount of excess heterogeneity that remains after systematic variation is first added then explicitly filtered by meta-regressions, equations (5) or (7).

Table 1 reports the coverage of WLS-, FE- and RE-MRA estimates of β_o in MRA model (5). Needless to say, the RE-MRA version adds a random term to equation (5), explicitly estimates its variance, and uses it in the weights matrix. In our simulations, RE-MRA estimates are computed using Raudenbush's (1994) iterative maximum likelihood algorithm. As Raudenbush (1994) observes, estimates converge after only a few iterations. To verify that our maximum likelihood algorithm produces the same RE-MRA estimates and confidence intervals that are routinely employed by meta-analysts, we generated random datasets in the above manner and compare the RE-MRA estimates and their confidence intervals from our maximum likelihood algorithm to those calculated by STATA. Because this process always produces the exact same values of both the estimates and their standard errors to 5 or more significant digits, we are confident that our simulations accurately represent RE-MRA as applied in the field.

Lastly, we also allow publication selection bias in the simulations reported in Tables 3-5. When publication selection is permitted, random values of all the relevant variables are generated in the same way as discussed above until a statistically significant positive effect, $\hat{\alpha}_{1j}$, is generated by chance. To conserve space, we assume that such selection for statistical significance occurs in half the reported empirical estimates. For the other half, the first random estimate is used regardless of whether it is statistically significant or not. In other papers where the focus is on the magnitude of publication

in Asian-Australasian stock markets, I^2 = 95%; across reported estimates of the effect of telecom investment on economic growth, I^2 = 92%; and it is 97% among the price elasticities of alcohol demand. The average I^2 across these nine meta-analyses is 93%.

bias, we vary the incidence of publication from 0 to 100% (Stanley, 2008; Stanley *et al*., 2010; Stanley and Doucouliagos, 2013a). The focus of the current investigation is not on the magnitude of publication bias *per se*, but rather the relative biases and mean square errors (MSE) of RE-MRA and WLS-MRA when publication bias is a genuine possibility.[2] Thus, it is sufficient to show that WLS-MRA has smaller bias and MSE than RE-MRA when there is some moderate amount of publication selection.

4.2 Simulation Results

Table 1 reports the coverage percentages as well as the two relative measures of unexplained random heterogeneity discussed above. 95% confidence intervals are constructed for each replication around the MRA estimates of β_0 from (5) or its random-effects equivalent. The proportion of the 10,000 confidence intervals randomly generated by these simulations that actually contain the 'true' value $\{0,1\}$ is computed, giving the coverage percentages found in the last three columns of Table 1.

Insert Table 1 about here

First, it is clear that dividing WLS-MRA's standard errors by $\sqrt{MS_E}$ is not a good idea—see the FE-MRA column in Table 1. When there is no excess heterogeneity, WLS-MRA is as good as FE-MRA. When there is excess heterogeneity, the coverage of the 'fixed-effects' MRA is unacceptably thin. Unfortunately, such excess heterogeneity is common in the social and medical sciences (Turner *et al*., 2012), and all tests for it are underpowered (Sidik and Jonkman, 2007).

Second, WLS-MRA produces coverage rates that are comparable to RE-MRA's coverage. On average, RE-MRA coverage is 0.6% closer to the nominal 95% than is WLS-MRA, and this difference increases to 1.4% if the Knapp-Hartung corrections for RE-MRA's confidence intervals are used (Knapp and Hartung, 2003). However, ironically, the coverage rates for WLS-MRA are better than RE-MRA's when there is large *additive* heterogeneity, the exact circumstances for which RE-MRA is designed. The message here is that WLS-MRA produce acceptable confidence intervals,

[2] We do, however, report the averages for the case of 100% publication bias.

comparable to RE-MRA, regardless of the amount of heterogeneity, and that FE-MRA's confidence intervals will be unacceptable for most realistic applications.

Lastly, the MRA dummy variable, *M*, succeeds in correcting omitted-variable bias. The average estimate of β_0 from MRA model (5) does not differ from its true value by more than rounding errors. This result can be seen in Table 1 by the closeness of the coverage proportions to their nominal 95% level when $\sigma_h = 0$ and also by the biases and MSEs reported in Table 2.

Insert Table 2 about here

Table 2 reports the bias and MSE (mean square error) of these meta-regression methods when there is no publication selection for statistical significance, the same conditions reported in Table 1. When these 10,000 replications are repeated ten times, the mean absolute deviation from one individual bias reported in Table 2 to another is approximately 0.0004—0.0001 for the MSE. Coverage proportions vary by 0.0006 from one simulation of 10,000 replications to another. The biases reported in Table 2 are practically nil, a bit larger than 0.1%, on average, confirming the viability of using dummy moderator variables, *M*, to remove misspecification biases. Surprisingly, the MSE of WLS-MRA is, on average, 38% smaller than RE-MRA's MSE. Taken together, Tables 1 and 2 demonstrate that traditional, unrestricted weighted least squares meta-regression is at least as good as than random-effects (or mixed-effects) meta-regression. This finding is unexpected, because these simulations involve the exact circumstances—random, additive heterogeneity—for which random-effects is designed to accommodate. Our expectations were that WLS-MRA might be superior to RE-MRA in the presence of publication selection bias, but not when there is no publication bias.

The advantage of WLS-MRA over RE-MRA displayed in Table 2 is attributable to the highest levels of heterogeneity. If the highest level of heterogeneity, $\sigma_h = 4.0$, is removed, both WLS- and RE-MRA have average bias = 0.001 and average MSE = 0.06, with RE-MRA having slightly lower bias (by 0.0002) and WLS-MRA having slightly lower MSE (by 0.003). We wish to make no claim that WLS-MRA is superior to RE-MRA when there is no publication bias. Rather our point is that these meta-regression approaches are practically equivalent when there is no publication selection and that both have negligible bias when misspecification bias is modeled by a dummy moderator variable. Recall that ± 1.0 is a small empirical effect (ρ=0.27); thus, ± 0.001 or even ± 0.01 are negligible from a practice or policy perspective. Even when we look at the most unfavorable conditions for WLS-MRA (moderate to high heterogeneity), RE-MRA offers only a small, practically insignificant, improvement. In practice, these conditions would be very difficult to identify because tests and measures of heterogeneity are not very reliable (Sidik and Jonkman, 2007; Sutton and Higgins, 2007).

Nonetheless, an interesting puzzle remains. What explains the superior performance of WLS-MRA at high levels of heterogeneity? Isn't large additive heterogeneity the exact condition for which RE-MRA is designed? To explain this puzzle, consider the expected value and variance of the estimated effects in the presence of omitted variable bias. In our simulations, when X_3 is an omitted variable, the bias in estimating α_1 is α_3, and these simulations generate α_3 from $N(0, \sigma_h^2)$. As a result, the estimate's variance (SE_j^2) will contain a second additive term that depends directly on the amount of this random omitted variable bias, α_3^2 (Kementa, 1971). That is, the reported variances, SE_j^2, will be the sum of two terms: the usual regression variance (S_j^2) plus α_3^2, which varies directly with σ_h^2. Or, $SE_j^2 = S_j^2 + \alpha_3^2$. As random heterogeneity gets larger and larger, it will gradually dominate the conventional regression variance term, S_j^2, because the latter does not increase with greater random heterogeneity. For the highest levels of heterogeneity, variations in SE_j^2 will be roughly proportional to this heterogeneity, α_3^2, because S_j^2 and its sample variations will be relatively small. Thus, for large levels of heterogeneity, variances are approximately multiplicative—equation (2).

Insert Figures 1 and 2 about here

This rough proportionality of an estimate's variance and random heterogeneity is more clearly seen in Figure 1, which graphs 1,000 random primary study standard errors squared, against the square of the random heterogeneity. In our simulations, we can directly observe the randomly generated heterogeneity. Figure 1 plots these variances from our simulations' largest heterogeneity condition, (σ_h=4.0; I^2=98.6%). At such high levels of heterogeneity, excess heterogeneity dominates conventional sampling errors, and the estimate's reported variance ('SE-squared') will be correlated with the random heterogeneity (ρ=0.5). Figure 1 reveals a fan-shaped scatter and an approximate proportionality between these two variances. Thus, for such large heterogeneity, SE_j^2 is roughly proportional to $\tau^2 = \sigma_h^2$, and WLS's multiplicative variance-covariance structure will be approximately correct. On the other hand, for small heterogeneity (σ_h=0.125; I^2=24.7%), there is virtually no correlation between SE_j^2 and random heterogeneity (ρ = 0.03) because variations in conventional sampling error will dominate random heterogeneity– see Figure 2. Nor is this phenomenon some eccentricity of estimated regression coefficients alone. Stanley and Doucouliagos (2013b) demonstrate that standardized mean differences (Cohen's d) from RCTs exhibit this same proportionality between SE_j^2 and random heterogeneity for large levels of heterogeneity.

Though interesting, the above discussion of the puzzle of high heterogeneity is a mere distraction to our central point and its practical consequences—that WLS-MRA dominates RE-MRA when there is publication bias. Table 3 reports the bias and MSE (mean square error) of these meta-regression estimators when 50% of the estimates are reported only if they are statistically significant and when this publication bias is not explicitly accommodated by either of these meta-regression models. In the columns labeled 'Bias,' the average MRA estimate of β_0 from equation (5), or its random-effects equivalent, is reported. The average biases reported in Tables 2, 3 and 4 are the absolute value of the difference between the average of these 10,000 simulations and the true effect={0,1}.

Insert Table 3 about here

Table 3 clearly reveals how publication bias can be quite large, potentially dominating the actual empirical effect. As theory would suggest, this bias is especially large when there is large heterogeneity. Unfortunately, such large values of I^2 are found in economics and social science research. When there is no genuine empirical effect, the appearance of empirical effects can be manufactured. When there is a small genuine empirical effect, publication selection in half the studies combined with large heterogeneity can double the apparent effect. This publication bias can be quite large and can have important practical consequences for policy and practice. However, the importance of publication bias and its effects on policy are widely reported and well documented throughout the literature. Here, these biases merely serve as a baseline for relative comparison. Next, we turn to our central question: will random-effects meta-regression be more or less biased than weighted least squares meta-regression when there is publication selection bias?

Table 3 demonstrates that RE-MRA is more biased and less efficient (higher MSE) than WLS-MRA when there is publication bias. *In all cases,* WLS-MRA has smaller bias than random-effects, and it also has a smaller MSE in 89% of these cases. On average, RE-MRA's MSE is more than twice that of WLS-MRA, and its bias is 61% larger. Where the bias is largest, WLS-MRA makes its greatest relative improvement over RE-MRA. Although all MRA approaches exhibit notable publication bias if there is selection for statistical significant results and this selection goes uncorrected, the bias and efficiency of WLS-MRA is much better than RE-WLS.

Note that the relative performance of RE-MRA and WLS-MRA does not depend on the incidence of publication selection. The last row of Table 3 reports the average values when there is 100% publication selection for statistical significance. Tables 2-5 do not report the bias or MSE for fixed-effects MRA because these will be identical to WLS-MRA. Recall that FE-MRA and WLS-MRA differ only in how their standard errors are calculated.

How can these potentially large publication biases be reduced and their practical consequences minimized? Table 4 reports the results for both the traditional, unrestricted weighted least squares and the random-effects approaches to a multiple Egger regression (Egger *et al.*, 1997). Recall that an Egger meta-regression uses empirical effects as the dependent variable and their standard errors as the independent or moderator variable.

$$y_j = \beta_0 + \beta_1 SE_j + u_j \,. \tag{6}$$

Egger *et al*. (1997) employ WLS and the conventional t-test of β_1 as a test for the presence of publication bias (sometimes called the 'funnel-asymmetry test' or FAT), while Stanley (2008) uses the WLS-MRA version of equation (6) and the conventional t-test of β_0 as a test for the presence of a genuine empirical effect beyond publication bias (called the 'precision-effect test' or PET). Following Stanley and Doucouliagos (2012; 2013a), Table 4 estimates β_0 from the multiple FAT-PET-MRA:

$$y_j = \beta_0 + \beta_1 SE_j + \beta_2 M_j + u_j \; , \tag{7}$$

using either an unrestricted WLS or random-effects. Needless to repeat, the latter adds a random-effects component to equation (7) and estimates its variance.

Insert Table 4 about here

As before, where publication selection is not accommodated, traditional, unrestricted weighted least squares clearly dominates random effects, see Table 4. It is important to recognize the large reduction of publication bias for both approaches from what is reported in Table 3. On average, bias is reduced by 78% for RE-FAT-PET-MRA and 74% for WLS-FAT-PET-MRA. The amount of bias remaining is practically negligible for WLS-FAT-PET-MRA. Recall that a true effect of 1.0 is a practically small effect (ρ=.27). When compared against one another, WLS-FAT-PET-MRA has smaller bias than RE-FAT-PET-MRA in 93% of the cases (see Table 4), and WLS-FAT-PET-MRA's average bias and MSE are notably smaller than RE-FAT-PET-MRA. As before, the relative advantage of WLS-MRA over RE-MRA is not a function of the incidence of publication selection—see the last row of Table 4.

But can publication bias be reduced further? Recently, a somewhat more complicated, conditional meta-regression approach has been shown to reduce publication selection bias (Stanley and Doucouliagos, 2013a). This new approach is a hybrid between the conventional Egger regression and a meta-regression that uses the estimate's variance as a moderator variable in place of its standard error.

$$y_j = \beta_0 + \beta_1 SE^2_j + \beta_2 M_j + u_j \tag{8}$$

See Stanley and Doucouliagos (2013a) for the theoretical motivation for this approach and its validation. There, it is shown that MRA model (7) has the smaller bias when PET finds no genuine empirical effect (*i.e.,* accept H_0: β_0=0), while MRA model (8) has the smaller bias when PET finds a genuine empirical effect (*i.e.,* reject H_0: β_0=0). Thus, Stanley and Doucouliagos (2013a) recommend a conditional estimator, called 'PET-

PEESE.' When the conventional t-test of H_0: β_0=0 from MRA model (7) is rejected, MRA model (8) is used to estimate β_0; otherwise, MRA model (7) is used to estimate β_0.

Table 5 reports the estimates from this conditional MRA model over the same conditions as reflected in previous simulations and tables. As before, the WLS approach has much smaller bias and MSE, and both PET-PEESE-MRAs reduce publication bias beyond the multiple Egger regression reported in Table 4. Taken together, WLS-MRA is as good as or better than RE-MRA when there is no selection for statistical significance. When there is publication bias, WLS-MRA clearly dominates RE-MRA.

Insert Table 5 about here

5. DISCUSSION

What explains the success of the unrestricted weighted least squares meta-regression approach? Although past research has shown that random-effects *weighted averages* are more biased than fixed-effects when there is publication bias (Poole and Greenland, 1999; Sutton *et al*., 2000; Stanley, 2008; Stanley *et al*., 2010: Henmi and Copas, 2010), the high performance of the unrestricted WLS meta-regression both with and without selection for statistical significance is a surprise. Certainly, the fact that the unrestricted WLS's weights, $1/SE_i^2$, gives relatively more (less) weight to the most (least) precise estimates than does RE-MRA's weights, $1/(SE_i^2+\hat{\tau}^2)$, explains much of the superior statistical performance of WLS-MRA when there is publication selection bias. Nonetheless, it is surprising to learn that WLS-MRA can outperform RE-MRA when there is no publication selection bias, because RE-MRA's assumption of random, additive heterogeneity is forced into our simulations.

We explain this puzzle, in part, as the approximate proportionality of reported estimate's variance and random heterogeneity when there are the highest levels of heterogeneity, making WLS's multiplicative variance structure approximately correct. Even when there is low or moderate heterogeneity, the Gauss-Markov theorem's multiplicative invariance property allows WLS to accommodate heterogeneity in a way that is practically equivalent, in the properties of the resulting estimates, to random-effects. RE-MRA suffers for a further disadvantage over WLS when there exists publication selection. RE-MRA assumes that the 'random-effects', v_j , are independent of the moderator variables. Yet Figure 1 reveals notable correlation between v_j^2 and SE_j^2, a moderator variable in FAT-PET-PEESE, when there is large heterogeneity.

6. CONCLUSIONS

The central contribution of this study is the demonstration that traditional unrestricted weighted least squares multiple meta-regression (WLS-MRA) provides a viable practical alternative to random-effects meta-regression (RE-MRA), one that dominates RE-MRA in those exact circumstances for which RE-MRA is designed (large additive heterogeneity). When there is no publication selection bias, these two approaches are practically equivalent with the edge going to WLS-MRA if the excess heterogeneity is especially large. More importantly, when there is selection for statistical significance, WLS-MRA clearly dominates RE-MRA. Whether or not publication selection bias is explicitly modeled by meta-regression, the bias and MSE of unrestricted weighted least squares meta-regression is notably smaller than random-effects meta-regression. Here too, ironically, the relative performance of WLS-MRA is best when there is large additive heterogeneity. Thus, our recommendation is that WLS-MRA should be adopted as the conventional approach for meta-regression, especially in the social sciences where high levels of heterogeneity are the norm.

Granted that RE-MRA is often adequate when publication selection bias can be ruled out; this is almost never possible in actual meta-analysis practice. Publication selection bias has been found to be quite common and tests for its presence have low power (e.g., Egger *et al*., 1997; Stanley, 2008). Thus, prudence requires that systematic reviewers treat all areas of research as if they had publication bias. When reviewers fail to do so and there is publication selection bias, our simulations show that RE-MRA has much larger biases than WLS-MRA and more than double the MSE.

This study makes further contributions to our understanding of meta-regression. It validates the multiple meta-regression model advanced by Stanley and Doucouliagos (2013a, eq. 10) and its ability to correct omnipresent misspecification biases and simultaneously accommodate publication bias. Simple binary dummy variables for the presence of possible biases as moderator variables in a MRA serves as a viable filter for these potential sources of contamination to scientific inference.

Lastly, this study reveals that it is *never* advisable to divide the standard errors of meta-regression coefficients produced by conventional WLS statistical packages by the square root of MSE, as has often been recommended (Hedges and Olkin, 1985; Lipsey and Wilson, 2001; Konstantopoulos and Hedges, 2004; Johnson and Huedo-Medina, 2012). Even when applied to conditional inferences and all characteristics are exactly the same as current research, not dividing by $\sqrt{MS_E}$ (WLS-MRA) produces CIs as good as fixed-effects (FE-WLS). If FE-MRA is misapplied to any other case or should there be unexpected heterogeneity, WLS-MRA's CIs are much better. Thus, in practical application, WLS-MRA is preferable to both fixed- and random-effects meta-regression. Its statistical properties are practically equivalent to these conventional meta-regression models *when these models are true*. However, if there is excess heterogeneity and a fixed-effects meta-regression is used or if there is publication bias and random-effects is used, WLS-MRA is demonstrably better. Unfortunately in practice, neither excess heterogeneity nor publication bias can be ruled out because tests of excess heterogeneity and publication biases are both known to have low power. Thus, caution favors WLS-MRA in all practical applications. There is nothing of practical consequence to lose but much to gain in employing WLS-MRA.

Nonetheless, a note of caution is warranted. Although our simulations employ a wide range of the relevant parameters, there might be special circumstances that arise for particular areas of research that might alter the relative performance of these meta-regression methods. Also, the real world is likely to contain more complications than those few that we have simulated. The complex interactions of several such complications might also affect relative performances. Lastly, our simulations are based on regression estimates. Although we have reason to believe that our findings will apply to RCTs (Stanley and Doucouliagos, 2013b), further research is needed.

REFERENCES

Aitken, A.C. 1935. On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh* **55**: 42–48.

Baker, W.L., White, C.M., Cappelleri, J.C., Kluger, J., and Coleman, C.I. 2009. Understanding heterogeneity in meta-analysis: the role of meta-regression. *The International Journal of Clinical Practice* **63**: 1426-1434.

Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H.R. 2009. Introduction to Meta-Analysis, Chichester: John Wiley and Sons.

Cohen, J. 1988. *Statistical Power Analysis in the Behavioral Sciences,* 2nd ed. Hillsdale: Erlbaum.

Copas, J.B. and Lozada, C. 2009. The radial plot in meta analysis: Approximations and applications. *Journal of the Royal Statistical Society*: *Series C* (*Applied Statistics*), **58**: 329-344.

Costa-Font, J., McGuire, A. and Stanley, T.D. 2013. Publication selection in health policy research: The winner's curse hypothesis. *Health Policy*, 109: 78– 87.

Davidson, R. and MacKinnon, J.G. 2004. *Econometric theory and methods.* Oxford: Oxford University Press.

De Linde Leonard, M., Stanley, T. D and Doucouliagos, H(C). 2013. Does the UK minimum wage reduce employment? A meta-regression analysis. *British Journal of Industrial Relations*. DOI: 10.1111/bjir.12031.

Doucouliagos H(C). 2011. How Large is Large? Preliminary and relative guidelines for interpreting partial correlations in economics. *SWP*, Economics Series 2011-5, Deakin University.

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

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

Doucouliagos, C(H) and Stanley, T.D. 2013. Theory competition and selectivity: Are all economic facts greatly exaggerated? *Journal of Economic Surveys* **27**: 316-39.

Doucouliagos, C(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.

DuMouchel, W. 1990. Bayesian meta-analysis, in Berry, D. (ed), *Statistical Methodology in the Pharmaceutical Sciences*, Marcel Dekker, New York, pp. 509-529.

Egger, M., Smith, GD., Scheider, M., and Minder, C. 1997. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* **316**: 629-34.

Gerber, A.S., Green, D.P. and Nickerson, D. 2001. Testing for publication bias in political science, *Political Analysis,* **9**: 385-92.

Gerber, A.S. and Malhorta, N. 2008. Publication bias in empirical sociological research, *Sociological Methods and Research,* **25**: 1-28.

Greene, W.E. 1990. *Econometric Analysis,* New York: Macmillan.

Hedges, L.V and Olkin, I. 1985. *Statistical Methods for Meta-Analysis.* Orlando: Academic Press.

Hedges, L.V. 1994. Fixed effects models in H. Cooper and L.V. Hedges (eds.) *The Handbook of Research Synthesis*. New York: Russell Sage, 285-299.

Hedges, L.V. and Vevea, J.L. 1998. Fixed- and random-effects models in meta-analysis. *Psychological Methods* **3**: 486-504.

Henmi, M. and Copas, J.B. Confidence intervals for random effects meta-analysis and robustness to publication bias. *Statistics in Medicine* 2010; **29**: 2969-2983.

Higgins, J.P.T and Thompson, S.G. 2002. Quantifying heterogeneity in meta-analysis. *Statistics in Medicine* **21**: 1539-1558.

Higgins, J.P.T, and Thompson, S.G. 2004. Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine* **23**: 1663-1682.

Hopewell, S., Loudon, K., Clarke, M.J, Oxman, A.D. and Dickersin, K. 2009. Publication bias in clinical trials due to statistical significance or direction of trial result. *Cochrane Review*, Issue 1. Available at http://www.thecochranelibrary.com.

Jacquez, J.A., Mather, F. J., and. Crawford, C. R. 1968. Linear regression with non-constant, unknown error variances: Sampling experiments with least squares, weighted least squares and maximum likelihood estimators. *Biometrics* **24:** 607-626.

Johnson, B.T. and Huedo-Medina, T.B. 2012. Depicting estimates using the intercept in meta-regression models: The moving constant technique. *Research Methods Synthesis*, **2**: 204-20.

Judge, G.G., Hill, R.C., Griffiths, W., Lutkepohl, H. and Lee, T.C. 1982. *Introduction to the Theory and Practice of Econometrics.* John Wiley and Sons: New York.

Kmenta, J. 1971. *Elements of Econometrics,* New York: Macmillan.

Knapp, G., and J. Hartung. 2003. Improved tests for a random effects meta-regression with a single covariate. Statistics in Medicine **22**: 2693-2710

Konstantopoulos, S. and Hedges, L.V. 2004. Meta-analysis in D. Kaplan (ed.) *Quantitative methodology for the social sciences.* Sage Publications: Thousand Oaks: pp. 281-295.

Krassoi Peach, E. and T. D. Stanley. 2009. Efficiency wages, productivity and simultaneity: A meta-regression analysis," *Journal of Labor Research,* 30: 262-8.

Lipsey, M.W. and Wilson, D.B. 2001. *Practical Meta-Analysis,* Sage Publications: Thousand Oaks.

Moreno, S.G, Sutton, A.J, Ades, A.E, Stanley, T.D, Abrams, K.R, Peters, J.L, Cooper, N.J. 2009. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Medical Research Methodology* **9**:2. http://www.biomedcentral.com/1471-2288/9/2.

Raudenbush, S.W. 1994. Random effects models. In H. Cooper and L.V. Hedges (eds.) *The Handbook of Research Synthesis*. New York: Russell Sage, pp. 301-321.

Sharp, S. 1998. sbe23: Meta-analysis regression. *Stata Technical Bulletin* 42: 16-22. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, pp. 148-155. College Station, TX: Stata Press.

Stanley, T.D. 2008. Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection. *Oxford Bulletin of Economics and Statistics* **70**:103-27.

Stanley, T.D. and Doucouliagos, H(C). 2012. *Meta-Regression Analysis in Economics and Business*, Oxford: Routledge, 2012.

Stanley, T.D., Doucouliagos H(C). 2013a. Meta-regression approximations to reduce publication selection bias. *Review Synthesis Methods*. http://onlinelibrary.wiley.com/doi/10.1002/jrsm.1095/pdf

Stanley, T.D., Doucouliagos H(C). 2013b. Neither fixed nor random: Weighted least squares meta-regression analysis. *SWP*, Economics Series 2013-1, Deakin University.

Stanley, T.D. and Jarrell, S.B. 1989. Meta-regression analysis: A quantitative method of literature surveys, *Journal of Economic Surveys* **3**: 161-70.

Stanley T.D., Jarrell, S.B. and Doucouliagos, H(C). 2010. Could it be better to discard 90% of the data? A statistical paradox. *American Statistician* **64**:70-77.

Steel, P. & Kammeyer-Mueller, J. 2002. Comparing meta-analytic moderator search techniques under realistic conditions. *Journal of Applied Psychology*, **87**: 96-111.

Sterling, T.D., Rosenbaum, W.L. and Weinkam, J.J. 1995. Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa, *American Statistician,* **49**: 108-12.

Sterne, J.A. C. 2009. Meta-Analysis in Stata: An Updated Collection from the *Stata Journal*. College Station, TX: Stata Press.

Stigler, S.M. 1986. *The History of Statistics*. Cambridge, MA: Belknap Press.

Sutton, A.J., Song, F., Gilbody, S.M., Abrams, K.R. 2000. Modelling publication bias in meta-analysis: a review. *Statistical Methods in Medical Research* **9**:421-445.

Poole C., Greenland S. 1999. Random-effects meta-analyses are not always conservative. *American Journal of Epidemiology,* **150**:469-475.

Sidik, K. and Jonkman, J.N. 2007. A comparison of heterogeneity variance estimators in combining results of studies, *Statistics in Medicine,* **26**: 1964–81.

Sutton, A.J and Higgins, J.P.T. 2007. Recent developments in meta-analysis, *Statistics in Medicine,* 27:625–50.

Thompson, S.G. and Sharp, S.J. 1999. Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine* **18**: 2693-2708.

Turner, R.M., Davy, J., Clarke, M.J., Thompson, S.G. and Higgins, J.P.T 2012. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* **41**: 818-27.

White, I.R. 2011. Multivariate random-effects meta-regression: updates to *mvmeta*. *Stata Journal*, **11**: 255-270.

Wooldridge, JM. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge: MIT Press

Table 1: Coverage Percentage of FE-, RE-, and WLS-MRA (nominal level = .95)

MRA Sample Size	σ_h, Excess Heterogeneity	True Effect	I^2	Residual I^2	FE-MRA	RE-MRA	WLS-MRA
20	0	0	.0948	.0981	.9489	.9544	.9505
20	0.125	0	.2433	.2540	.8769	.9218	.9350
20	0.25	0	.6014	.5495	.7067	.9082	.9079
20	0.5	0	.8503	.8201	.4740	.9191	.9000
20	1.0	0	.9465	.9347	.3088	.9254	.9110
20	2.0	0	.9761	.9716	.2277	.9265	.9339
20	4.0	0	.9858	.9833	.1909	.9233	.9464
80	0	0	.0936	.0595	.9495	.9553	.9525
80	0.125	0	.2469	.2814	.8741	.9429	.9350
80	0.25	0	.6011	.6067	.7007	.9371	.9058
80	0.5	0	.8493	.8470	.4769	.9495	.9079
80	1.0	0	.9465	.9437	.3173	.9433	.9167
80	2.0	0	.9761	.9740	.2384	.9460	.9440
80	4.0	0	.9858	.9842	.2047	.9472	.9528
20	0	1	.0593	.0969	.9545	.9603	.9531
20	0.125	1	.3186	.2486	.8738	.9187	.9278
20	0.25	1	.6465	.5447	.7070	.8996	.9064
20	0.5	1	.8687	.8207	.4688	.9183	.8996
20	1.0	1	.9517	.9356	.3125	.9220	.9119
20	2.0	1	.9777	.9715	.2301	.9227	.9378
20	4.0	1	.9863	.9832	.1851	.9252	.9455
80	0	1	.0589	.0594	.9532	.9568	.9532
80	0.125	1	.3179	.2808	.8704	.9382	.9282
80	0.25	1	.6471	.6055	.7040	.9444	.9138
80	0.5	1	.8683	.8479	.4765	.9460	.9049
80	1.0	1	.9517	.9436	.3153	.9427	.9240
80	2.0	1	.9777	.9740	.2364	.9468	.9393
80	4.0	1	.9863	.9842	.1947	.9436	.9566
Average					.5349	.9352	.9286

Notes: FE-MRA, RE-MRA and WLS-MRA refer to the fixed-effects, random-effects and unrestricted weighted least square meta-regression estimates, respectively, of β_0 in MRA model (5). Coverage rates of these estimates are reported in the last three columns. σ_h is the standard deviation of random excess additive heterogeneity, υ_i, in equation (3). I^2 is the percent of the total variation among the empirical effects that is attributable to heterogeneity when there is no publication bias or systematic heterogeneity; that is, I^2 measures only the random excess heterogeneity relative to sampling error. Residual I^2 measures the excess heterogeneity that remains after MRA model (5) accounts for systematic heterogeneity. All of these measures are calculated empirically for each replication and averaged across 10,000 replications.

Table 2: Bias and MSE of RE- and WLS-MRA

MRA Sample Size	σ_h, Excess Heterogeneity	True Effect	I^2	Residual I^2	RE-MRA |Bias|	WLS-MRA |Bias|	RE-MRA MSE	WLS-MRA MSE
20	0	0	.0948	.0981	.00059	.00041	.00554	.00549
20	0.125	0	.2433	.2540	.00105	.00124	.00829	.00845
20	0.25	0	.6014	.5495	.00091	.00157	.01498	.01687
20	0.5	0	.8503	.8201	.00085	.00031	.03555	.04661
20	1.0	0	.9465	.9347	.00087	.00282	.11340	.13435
20	2.0	0	.9761	.9716	.00157	.00014	.40591	.35193
20	4.0	0	.9858	.9833	.01341	.00148	1.62793	.88102
80	0	0	.0936	.0595	.00048	.00051	.00110	.00109
80	0.125	0	.2469	.2814	.00059	.00040	.00173	.00179
80	0.25	0	.6011	.6067	.00029	.00021	.00331	.00386
80	0.5	0	.8493	.8470	.00030	.00077	.00833	.01066
80	1.0	0	.9465	.9437	.00023	.00031	.02669	.02919
80	2.0	0	.9761	.9740	.00012	.00099	.09887	.06928
80	4.0	0	.9858	.9842	.00240	.00203	.38644	.15535
20	0	1	.0593	.0969	.00046	.00042	.00564	.00558
20	0.125	1	.3186	.2486	.00186	.00172	.00825	.00837
20	0.25	1	.6465	.5447	.00147	.00164	.01487	.01706
20	0.5	1	.8687	.8207	.00068	.00107	.03607	.04736
20	1.0	1	.9517	.9356	.00118	.00358	.11352	.13372
20	2.0	1	.9777	.9715	.00075	.00247	.39659	.33989
20	4.0	1	.9863	.9832	.01035	.01111	1.61637	.83945
80	0	1	.0589	.0594	.00067	.00067	.00110	.00109
80	0.125	1	.3179	.2808	.00013	.00013	.00172	.00177
80	0.25	1	.6471	.6055	.00068	.00060	.00333	.00389
80	0.5	1	.8683	.8479	.00009	.00048	.00822	.01035
80	1.0	1	.9517	.9436	.00012	.00063	.02720	.02953
80	2.0	1	.9777	.9740	.00163	.00005	.09808	.06986
80	4.0	1	.9863	.9842	.00195	.00040	.38633	.15414
Average					.00163	.00136	.19483	.12064

Notes: RE-MRA and WLS-MRA refer to the random-effects and unrestricted weighted least square meta-regression estimates, respectively, of β_0 in MRA model (5). Bias and MSE of these estimates are reported in the last four columns. σ_h is the standard deviation of random excess additive heterogeneity, υ_i. I^2 is the percent of the total variation among the empirical effects that is attributable to heterogeneity when there is no publication bias or systematic heterogeneity; that is, I^2 measures only the random excess heterogeneity relative to sampling error. Residual I^2 measures the excess heterogeneity that remains after MRA model (5) accounts for systematic heterogeneity. All of these statistical measures are calculated empirically for each replication and averaged across 10,000 replications.

Table 3: Bias and MSE of RE- and WLS-MRA with 50% Publication Selection Bias

MRA Sample Size	σ_h, Excess Heterogeneity	True Effect	I^2	Residual I^2	RE-MRA |Bias|	WLS-MRA |Bias|	RE-MRA MSE	WLS-MRA MSE
20	0	0	.1689	.0893	.0348	.0328	.0151	.0147
20	0.125	0	.3241	.1515	.0581	.0510	.0218	.0209
20	0.25	0	.5697	.5036	.1140	.0957	.0414	.0397
20	0.5	0	.8102	.7944	.2367	.1964	.1084	.1035
20	1.0	0	.9264	.9217	.4510	.3470	.3259	.2677
20	2.0	0	.9670	.9644	.8138	.5692	1.0391	.6824
20	4.0	0	.9809	.9786	1.5212	.8595	3.6524	1.6393
80	0	0	.1551	.0649	.0361	.0345	.0039	.0037
80	0.125	0	.3589	.2815	.0668	.0593	.0085	.0075
80	0.25	0	.6184	.5924	.1322	.1148	.0237	.0200
80	0.5	0	.8372	.8315	.2659	.2250	.0824	.0643
80	1.0	0	.9362	.9345	.4900	.3891	.2687	.1815
80	2.0	0	.9701	.9689	.8939	.6092	.8868	.4365
80	4.0	0	.9818	.9806	1.6566	.8880	3.0617	.9305
20	0	1	.0825	.0662	.0135	.0128	.0056	.0056
20	0.125	1	.2358	.1096	.0168	.0129	.0083	.0085
20	0.25	1	.5325	.4969	.0350	.0221	.0155	.0171
20	0.5	1	.8083	.7968	.0916	.0583	.0412	.0477
20	1.0	1	.9255	.9216	.2415	.1669	.1567	.1483
20	2.0	1	.9666	.9642	.5566	.3541	.6540	.4317
20	4.0	1	.9806	.9785	1.2326	.6554	2.8299	1.1672
80	0	1	.0450	.0374	.0101	.0096	.0012	.0012
80	0.125	1	.2591	.2422	.0158	.0115	.0020	.0019
80	0.25	1	.5926	.5847	.0314	.0172	.0042	.0041
80	0.5	1	.8364	.8338	.0940	.0570	.0163	.0133
80	1.0	1	.9349	.9337	.2564	.1740	.0886	.0559
80	2.0	1	.9695	.9685	.6142	.3756	.4553	.1978
80	4.0	1	.9817	.9805	1.3571	.6591	2.1485	.5624
Average					.4049	.2521	.5703	.2527
Average for 100 % Publication Selection Bias					.9649	.6536	2.0600	.8625

Notes: RE-MRA and WLS-MRA refer to the random-effects and unrestricted weighted least square meta-regression estimates, respectively, of β_0 in MRA model (5). Bias and MSE of these estimates are reported in the last four columns. σ_h is the standard deviation of random excess additive heterogeneity, υ_i. I^2 is the percent of the total variation among the empirical effects that is attributable to heterogeneity when there is no publication bias or systematic heterogeneity; that is, I^2 measures only the random excess heterogeneity relative to sampling error. Residual I^2 measures the excess heterogeneity that remains after MRA model (5) accounts for systematic heterogeneity. All of these statistical measures are calculated empirically for each replication and averaged across 10,000 replications.

Table 4: Bias and MSE of RE- and WLS-FAT-PET-MRA with 50% Publication Bias

Sample Size	σ_h, Excess Heterogeneity	True Effect	I^2	Residual I^2	RE-MRA |Bias|	WLS-MRA |Bias|	RE-MRA MSE	WLS-MRA MSE
20	0	0	.0948	.0893	.16575	.16364	.05218	.05157
20	0.125	0	.2433	.1515	.15454	.14634	.06086	.05960
20	0.25	0	.6014	.5036	.10523	.08881	.07024	.07462
20	0.5	0	.8503	.7944	.01049	.00049	.11731	.14752
20	1.0	0	.9465	.9217	.08977	.07918	.32996	.35647
20	2.0	0	.9761	.9644	.11570	.06047	1.00564	.84226
20	4.0	0	.9858	.9786	.19965	.03766	3.02407	1.84687
80	0	0	.0936	.0649	.14279	.14135	.02492	.02454
80	0.125	0	.2469	.2815	.12187	.11082	.02212	.01996
80	0.25	0	.6011	.5924	.06794	.04757	.01648	.01601
80	0.5	0	.8493	.8315	.04014	.05133	.02536	.03112
80	1.0	0	.9465	.9345	.14829	.13834	.08918	.08054
80	2.0	0	.9761	.9689	.21051	.13721	.26245	.14630
80	4.0	0	.9858	.9806	.38580	.06001	.84128	.25204
20	0	1	.0593	.0662	.02702	.02664	.01722	.01720
20	0.125	1	.3186	.1096	.03110	.02868	.02553	.02607
20	0.25	1	.6465	.4969	.03162	.02711	.04590	.05264
20	0.5	1	.8687	.7968	.02446	.02144	.10850	.13508
20	1.0	1	.9517	.9216	.01027	.00171	.30811	.33235
20	2.0	1	.9777	.9642	.01052	.04896	.97668	.75452
20	4.0	1	.9863	.9785	.05311	.16689	2.87638	1.60977
80	0	1	.0589	.0374	.02505	.02471	.00391	.00390
80	0.125	1	.3179	.2422	.02933	.02690	.00602	.00611
80	0.25	1	.6471	.5847	.03398	.02921	.01039	.01168
80	0.5	1	.8683	.8338	.02473	.02216	.02287	.02753
80	1.0	1	.9517	.9337	.01618	.00175	.06726	.06377
80	2.0	1	.9777	.9685	.04481	.03268	.22621	.12240
80	4.0	1	.9863	.9805	.21000	.11280	.70017	.23085
Average					.09038	.06553	.40490	.26226
Average for 100 % Publication Selection Bias					.3078	.1714	.3108	.2030

Notes: RE-MRA and WLS-MRA refer to the random-effects and unrestricted weighted least square meta-regression estimates, respectively, of β_0 in the multiple MRA model (7). Bias and MSE of these estimates are reported in the last four columns. σ_h is the standard deviation of random excess additive heterogeneity, υ_i. I^2 is the percent of the total variation among the empirical effects that is attributable to heterogeneity when there is no publication bias or systematic heterogeneity; that is, I^2 measures only the random excess heterogeneity relative to sampling error. Residual I^2 measures the excess heterogeneity that remains after MRA model (7) accounts for publication bias and systematic heterogeneity. All of these statistical measures are calculated empirically for each replication and averaged across 10,000 replications.

Table 5: Bias and MSE of RE- and PET-PESSE-MRA with 50% Publication Bias

Sample Size	Excess Heterogeneity (σ_h)	True Effect	I^2	RE-MRA |Bias|	WLS-MRA |Bias|	RE-MRA MSE	WLS-MRA MSE
20	0	0	.0948	.0665	.0646	.0287	.0285
20	0.125	0	.2433	.0598	.0518	.0383	.0384
20	0.25	0	.6014	.0437	.0256	.0598	.0670
20	0.5	0	.8503	.0018	.0157	.1204	.1548
20	1.0	0	.9465	.0785	.0709	.3322	.3790
20	2.0	0	.9761	.1020	.0652	1.0227	.8587
20	4.0	0	.9858	.1365	.0621	3.0835	1.8331
80	0	0	.0936	.0538	.0522	.0074	.0073
80	0.125	0	.2469	.0412	.0305	.0088	.0084
80	0.25	0	.6011	.0193	.0061	.0124	.0143
80	0.5	0	.8493	.0426	.0655	.0275	.0361
80	1.0	0	.9465	.1448	.1330	.0900	.0858
80	2.0	0	.9761	.2079	.1360	.2650	.1493
80	4.0	0	.9858	.3579	.0545	.8183	.2604
20	0	1	.0593	.0207	.0245	.0175	.0141
20	0.125	1	.3186	.0204	.0263	.0258	.0231
20	0.25	1	.6465	.0143	.0200	.0436	.0468
20	0.5	1	.8687	.0066	.0023	.1043	.1298
20	1.0	1	.9517	.0302	.0182	.3158	.3554
20	2.0	1	.9777	.0087	.0297	.9750	.7608
20	4.0	1	.9863	.0113	.1646	2.9422	1.6443
80	0	1	.0589	.0200	.0038	.0038	.0015
80	0.125	1	.3179	.0195	.0021	.0055	.0024
80	0.25	1	.6471	.0198	.0054	.0096	.0071
80	0.5	1	.8683	.0060	.0061	.0230	.0246
80	1.0	1	.9517	.0496	.0322	.0690	.0644
80	2.0	1	.9777	.0567	.0072	.2208	.1226
80	4.0	1	.9863	.1672	.1132	.7076	.2314
Average				.0646	.0460	.4064	.2625

Notes: RE-MRA and WLS-MRA refer to the random-effects and unrestricted weighted least square meta-regression estimates, respectively, of β_0 in multiple MRA model (7) or multiple MRA model (8), conditional on whether H_0: $\beta_0=0$ is rejected. Bias and MSE of these estimates are reported in the last four columns. σ_h is the standard deviation of random additive heterogeneity, υ_i. I^2 is the percent of the total variation among the empirical effects that is attributable to heterogeneity when there is no publication bias or systematic heterogeneity; that is, I^2 measures only the random excess heterogeneity relative to sampling error. All of these statistical measures are calculated empirically for each replication and averaged across 10,000 replications.

Figure 1: Plot of Estimated Variances (SE_i^2) vs. Heterogeneity Variances; σ_h=4; I^2=98.6%

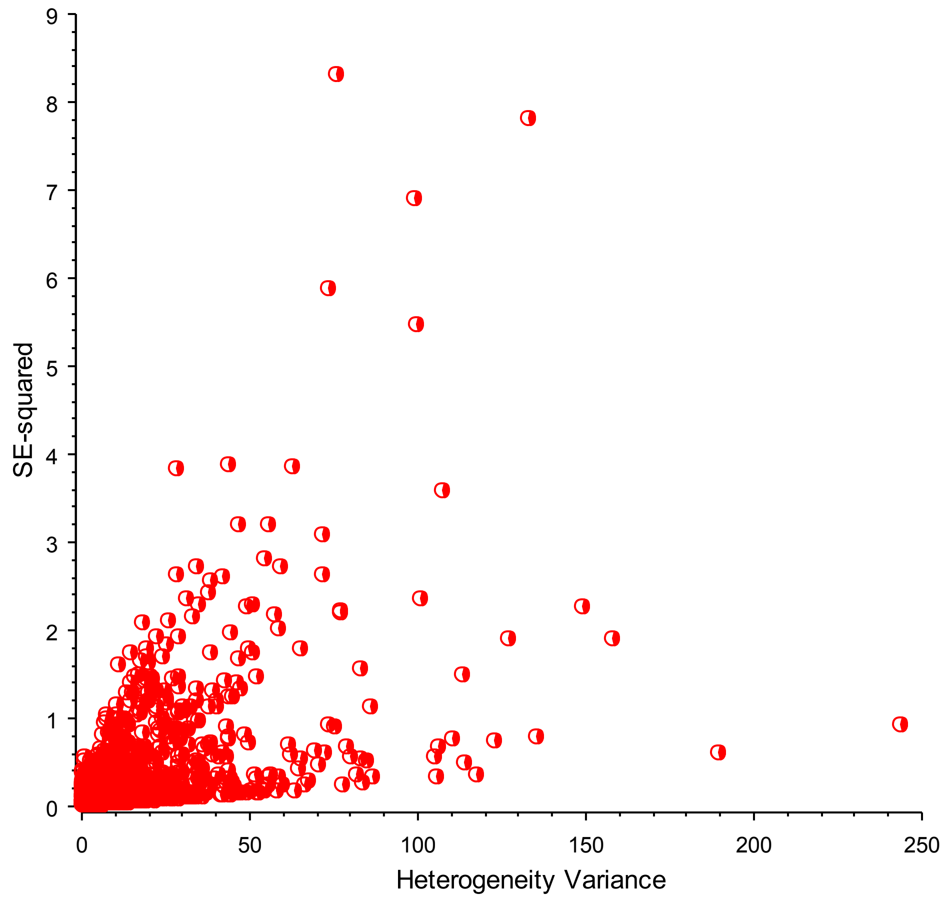

Figure 2: Plot of Estimated Variances (SE_i^2) vs. Heterogeneity Variances; σ_h=.125; I^2=24.7%