Original Article

Research Synthesis Methods

Received 18 February 2014, Revised 07 February 2016, Accepted 14 March 2016 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1211

Neither fixed nor random: weighted least squares meta-regression

T. D. Stanley[a*] and Hristos Doucouliagos[b]

Our study revisits and challenges two core conventional meta-regression estimators: the prevalent use of 'mixed-effects' or random-effects meta-regression analysis and the correction of standard errors that defines fixed-effects meta-regression analysis (FE-MRA). We show how and explain why an unrestricted weighted least squares MRA (WLS-MRA) estimator is superior to conventional random-effects (or mixed-effects) meta-regression when there is publication (or small-sample) bias that is as good as FE-MRA in all cases and better than fixed effects in most practical applications. Simulations and statistical theory show that WLS-MRA provides satisfactory estimates of meta-regression coefficients that are practically equivalent to mixed effects or random effects when there is no publication bias. When there is publication selection bias, WLS-MRA always has smaller bias than mixed effects or random effects. In practical applications, an unrestricted WLS meta-regression is likely to give practically equivalent or superior estimates to fixed-effects, random-effects, and mixed-effects meta-regression approaches. However, random-effects meta-regression remains viable and perhaps somewhat preferable if selection for statistical significance (publication bias) can be ruled out and when random, additive normal heterogeneity is known to directly affect the 'true' regression coefficient.

Keywords: meta-regression; weighted least squares; random effects; fixed effect

1. Introduction

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 MRAs 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 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' MRA (RE-MRA), encompassing mixed effects. The conventional status of RE-MRA 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).

This paper investigates whether an unrestricted WLS approach to meta-regression (WLS-MRA) is comparable with random-effects meta-regression and whether it can successfully correct observational research's routine misspecification and publication biases. Our simulations show that the *unrestricted* WLS-MRA is likely to be as good as and often better than conventional RE-MRA, in actual applications. We also investigate several sources of random heterogeneity in the target regression coefficient and document when random-effects meta-regression is likely to provide adequate estimates and when it is likely to be dominated by WLS-MRA.

In this paper, we confine our attention to the MRA of observational estimates of regression coefficients. Elsewhere, Stanley and Doucouliagos (2015) apply this same unrestricted weighted least squares principle to basic meta-analyses (i.e., simple weighted averages) of standardized mean differences and log odds ratios from randomized controlled trials.

[a]Hendrix College, 1600 Washington St., Conway, AR 72032, USA
[b]Department of Economics, Deakin University, 221 Burwood Highway, Burwood 3125 Victoria, Australia
*Correspondence to: T. D. Stanley, Hendrix College, 1600 Washington St., Conway, AR 72032, USA.
E-mail: Stanley@hendrix.edu

> When there is publication selection bias, the unrestricted weighted least squares approach dominates random effects. . . In practical applications, an unrestricted weighted least squares weighted average will often provide superior estimates to both conventional fixed and random effects (Stanley and Doucouliagos, 2015, p. 2116).

Here, we extend this unrestricted WLS estimation approach to MRA where moderator variables are used to summarize and explain observed heterogeneity among reported effect sizes.

Economists have long applied *unrestricted* WLS-MRA to summarize estimated reported regression coefficients in economics and to explain heterogeneity in reported estimates. It automatically allows for both heteroscedasticity and excess between-study heterogeneity (Stanley and Jarrell, 1989; Stanley and Doucouliagos, 2012). In Stanley and Doucouliagos (2012), we speculate that RE-MRA will be more biased than an unrestricted WLS when the reported research literature contains selection for statistical significance (conventionally called 'publication' or 'small-sample' bias). Unfortunately, the presence of 'publication', 'reporting', or 'small-sample' bias is common in many areas of research (Sterling *et al*., 1995; Gerber *et al*., 2001; Gerber and Malhorta, 2008; Hopewell *et al*., 2009; Doucouliagos and Stanley, 2013). Simulations presented here demonstrate that our conjecture has merit. RE-MRA is indeed more biased than WLS-MRA in the presence of publication selection, reporting, or small-sample bias.

All of these alternative meta-regression approaches—WLS-MRA, fixed-effects meta-regression analysis (FE-MRA), RE-MRA, and mixed-effects MRA—employ WLS.[1] WLS has long been used by meta-analysts: Stanley and Jarrell (1989), 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, FE-MRA, mixed-effect MRA, and RE-MRA restrict the WLS multiplicative constant to be one, whereas the unrestricted WLS does not. To our knowledge, no other meta-analyst has suggested that the *unrestricted* WLS meta-regression should routinely replace RE-MRA, mixed-effect MRA, and FE-MRA.

The central purpose of this paper is to evaluate the relative performance of FE-MRA, RE-MRA, and unrestricted WLS-MRA. We consider various ways in which heterogeneity may unfold (namely, through random omitted-variable bias, direct random additive heterogeneity, and random moderator heterogeneity) with and without publication selection bias. When there are no publication or small-sample biases, our simulations demonstrate how WLS-MRA provides confidence intervals practically equivalent to random effects. With publication selection bias, the unrestricted WLS always has smaller bias than random-effects meta-regression. 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 routine applications.

We do not mean to imply that there are never good *theoretical* reasons to prefer the RE-MRA model when there is excess, additive, and normal between-study heterogeneity or to prefer the FE-MRA model when there is no excess heterogeneity. In fact, all of our simulations impose exactly those conditions that make either the RE-MRA model or the FE-MRA model theoretically valid. Rather, we wish to document the robustness of WLS meta-regression. That is, we demonstrate that the unrestricted WLS-MRA estimation approach has statistical properties (i.e., bias, mean squared error (MSE), and coverage) comparable with or superior to both RE-MRA and FE-MRA, when the theoretical assumptions that underpin these conventional models (RE-MRA and FE-MRA) are entirely true.

In terms of models, we fully accept that the FE-MRA model is true when there is no excess, between-study heterogeneity and that the RE-MRA is true when there is excess, between-study heterogeneity. We offer no new MRA models. However, we document how a different estimation approach, our unrestricted WLS, is as good as or a better way to estimate these conventional meta-regression models. The robustness of unrestricted WLS meta-regression will often make it superior to random-effects or mixed-effects meta-regression in practical applications.

In this investigation of alternative meta-regression estimators, we also show that a meta-regression model with binary dummy variables can adequately correct for misspecification biases routinely found in observational research. Thus, our study demonstrates how a general, unrestricted WLS-MRA can remove or reduce a wide variety of biases routinely contained in social science research and thereby validates Stanley and Jarrell's (1989) MRA model.

2. The 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 $\mathbf{y}$ is an $\mathrm{L} \times 1$ vector of all comparable reported empirical effects in an empirical literature of L estimates. $\mathbf{M}$ is an $\mathrm{L} \times \mathrm{K}$ matrix of explanatory or moderator variables, the first column of which contains all 1s. $\boldsymbol{\beta}$ is a $\mathrm{K} \times 1$ vector of

[1] *In the context of meta-regression, multiple effects are estimated; thus, 'fixed-effects rather than 'fixed-effect' meta-regression is the more appropriate term.*

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 succeeding simulations, M_j is defined in this way. $\boldsymbol{\varepsilon}$ is an $L \times 1$ vector of residuals representing the unexplained errors of the reported empirical effects, and $\boldsymbol{\varepsilon} \sim (0, \mathbf{V})$. That is, **V** is the variance–covariance matrix of **y**, and $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^t) = \mathbf{V}$. **V** has variance of the jth estimated effect on the principal diagonal and zeros elsewhere[2]:

$$\mathbf{V} = \begin{vmatrix} \sigma_1^2 & 0 & . & . & 0 \\ 0 & \sigma_2^2 & & & 0 \\ . & & . & & . \\ . & & & . & . \\ 0 & 0 & . & . & \sigma_L^2 \end{vmatrix}.$$

Equation (1) cannot be adequately estimated by ordinary least squares, 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 an MRA. Thus, $\boldsymbol{\varepsilon}$ from Equation (1) cannot be assumed to be i.i.d., and ordinary least squares is almost never appropriate for MRA. At a minimum, meta-analysts need to adjust for this heteroscedasticity, and WLS are the traditional econometric remedy.

Weighted least squares estimates of Equation (1) are

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

where

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

σ_j^2 is the variance of the jth estimated effect, y_j, and σ^2 is a nonzero constant, which is routinely estimated by the MSEs of the estimated regression residuals by WLS statistical routines (Davidson and MacKinnon, 2004; Wooldridge, 2002).[3] Note that $\mathbf{V}^{-1}$ has $1/\sigma^2$ on the principal diagonal, and zero elsewhere. These inverse variances are conventionally regarded as the 'weights' in WLS routines and statistical packages.

We do not 'assume' that the variance–covariance structure is multiplicative. Aitken (1935), following Gauss (1823), proved that $\widehat{\boldsymbol{\beta}}$ in the WLS formula, Equation (2), is mathematically invariant to any nonzero multiplicative constant (e.g., σ^2 in $\boldsymbol{\Omega}$**).** That is, if **V** is multiplied by some arbitrary constant, say 10, to obtain $\boldsymbol{\Omega}$, then the resulting estimators, $\widehat{\boldsymbol{\beta}}_{\Omega}$, and a second one, which is denoted $\widehat{\boldsymbol{\beta}}_V$ and uses **V** as in Equation (2), will be identical. Thus, they must have exactly the same statistical properties: expected value, bias, consistency, efficiency, and MSE. This invariance to a multiplicative constant is an obvious mathematical property of all WLS estimators, Equation (2), because $\sigma^2/\sigma^2 = 1$, for all $\sigma^2 \neq 0$.

Invariance to a multiplicative constant is also an unavoidable property for RE-MRA. That is, its variance–covariance may also be multiplied by a nonzero constant without affecting the RE-MRA estimates. When the principal diagonal of **V** is replaced by $\sigma_j^2 + \tau^2$, as RE-MRA does, and computed by Equation (2) to obtain $\widehat{\boldsymbol{\beta}}_V$, and then **V** is further multiplied by some nonzero arbitrary constant to obtain $\boldsymbol{\Omega}$ and $\widehat{\boldsymbol{\beta}}_{\Omega}$ the resulting two WLS estimators from Equation (2) would possess all of the same statistical properties: expected value, bias, consistency, efficiency, and MSE as the conventional RE-MRA (i.e., $\widehat{\boldsymbol{\beta}}_{\Omega} = \widehat{\boldsymbol{\beta}}_V$).

The purpose of this paper is not to argue that one model is better than another. Rather, we are interested in which meta-regression estimation approach is likely to provide better statistical properties in realistic applications, even if key dimensions about a given model's structure are not valid. In particular, our simulations always generate excess heterogeneity, randomly and *additively*, as assumed by the RE-MRA model. That is, the individual total

[2] *Following the random-effects model, we also assume that estimates are independent across studies and that there is only one estimate per study. Correlation among estimates poses no theoretical difficulty to meta-regression when they are known. Since Aitken (1935), it has been widely recognized that the Gauss–Markov theorem and its desirable statistical properties will also hold in the case of correlated errors—generalized least squares. In economics, most studies report multiple estimates, and estimates within the same study are likely to be dependent on each other. These meta-regression models are easily adapted to accommodate this dependence using panel or multilevel methods or by calculating cluster-robust standard errors (Stanley and Doucouliagos, 2012).*

[3] *At this point in our discussion of weighted least squares, it is irrelevant which matrix, **Ω** or **V**, is interpreted as containing the 'true' variances. Regardless, the WLS estimates will be exactly the same. Furthermore, standard WLS statistical software will automatically estimate the multiplicative constant and adjust the standard errors accordingly. We discuss the more nuanced differences between **Ω** and **V** and among the FE-MRA, RE-MRA, and WLS-MRA approaches in the following pages.*

variances are forced to have an additive structure ($\sigma_j^2 + \tau^2$), whereas **V** in the WLS-MRA approach only contains estimates of the σ_j^2 terms. This paper investigates how robust WLS-MRA estimation is to such complications and misspecification in its variance–covariance structure, **V**.

We fully acknowledge that fixed-effects and random-effects meta-regression may be preferred in certain circumstances on theoretical grounds. In this paper, we demonstrate that even when there are valid theoretical reasons to prefer fixed effects or to prefer random effects, the unrestricted WLS meta-regression will, nonetheless, possess practically equivalent and sometimes superior statistical properties.

Weighted least squares is a special case of generalized least squares, where the variance–covariance matrix, **V**, has the aforementioned diagonal structure. Aitken (1935) generalized the famous Gauss–Markov theorem, proving that least squares estimates are minimum variance within the class of unbiased linear estimators for all positive semi-definite variance–covariance matrices (Jacquez *et al*., 1968; Stigler, 1986; Greene, 1990; Wooldridge, 2002; Davidson and MacKinnon, 2004). Thus, WLS, which also includes RE-MRA, is the best linear unbiased estimator, when the individual variances are known. This applies to the RE-MRA estimator as well. As long as the variances are known, which is frequently assumed in the meta-analysis literature, then the mathematics of Gauss–Markov theorem will fully follow, regardless of whether the total variances are 'additive' or 'multiplicative'. When the variance–covariances in **V** are known, the Gauss–Markov theorem will hold, and the resulting WLS estimator, $\widehat{\boldsymbol{\beta}}$, will be best linear unbiased estimator and invariant to further multiplication by any arbitrary nonzero constant.

The *unrestricted* WLS-MRA is calculated by substituting squared standard errors for σ_j^2 and allowing the proportional constant, σ^2, to be automatically estimated by the mean squared error, MS_E (Davidson and MacKinnon, 2004). That is, WLS-MRA allows, but does not assume, that the variances differ by a proportional constant, σ^2. In contrast, both fixed-effects and random-effects meta-regression restricts σ^2 to be one, thereby failing to make use of the WLS's remarkable multiplicative invariance property. For example, Hedges and Olkin (1985) explicitly discuss how it is necessary to divide the fixed-effects meta-regression coefficients' standard errors by $\sqrt{MS_E}$. Later, we show that there is never any practical reason to divide by $\sqrt{MS_E}$. That is, the statistical properties (coverage, bias, and MSE) of WLS-MRA are not improved by dividing its standard errors by $\sqrt{MS_E}$ and are very likely to be harmed significantly when FE-MRA is misapplied to general cases.

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., socioeconomic status, age, 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 (personal observation). Thus, a central objective for meta-regression is to explain as much systematic heterogeneity as possible while accommodating any random residual heterogeneity. Since Stanley and Jarrell (1989), multiple unrestricted WLS meta-regression has been the primary way to conduct meta-analyses in economics.

3.1. *Weighted least squares meta-regression*

What is not fully appreciated among meta-analysts is that unrestricted WLS estimates, Equation (2), automatically adjusts for heterogeneity or 'over-dispersion' by estimating σ^2 from WLS's MS_E and that the resilience of WLS to heterogeneity causes the resulting WLS-MRA estimates to rival or best both fixed-effects and random-effects estimates. As discussed earlier, 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 they retain desirable properties even when a bad estimate of σ^2 is used. In contrast, random effects estimates 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; Sidik and Jonkman, 2007; Hoaglin, 2015). It is WLS-MRA's resilience to excess between-study heterogeneity that makes it worthy of further study and application.

Although the ability of WLS to test heterogeneity has been acknowledged, the unrestricted WLS-MRA has thus far been dismissed by other meta-analysts (Thompson and Sharp, 1999; Baker and Jackson, 2013). Because meta-analysts have previously viewed this multiplicative variance structure as a requirement for using WLS-MRA rather than as WLS's resilience to poor heterogeneity estimates, it has not been 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, and we assume further that it is incorrect. However, our simulations demonstrate that their recommendations do not follow even if the random-effects model with its additive variance is true and strictly imposed upon the research record. In the succeeding simulations, we force RE-MRA's model to be true and show that there is still little practical difference between WLS-MRA and RE-MRA, when there is no selection for statistical significance (i.e., publication, reporting, or small-sample bias). When there is publication selection bias, WLS-MRA has consistently smaller bias and often smaller MSE than the corresponding random-effects meta-regression estimator. Because systematic reviewers can never rule out publication or small-sample bias in practice, WLS-MRA is a viable choice even when there is excess additive heterogeneity. Before we turn to the design of our simulations and their findings, we take a short detour to compare and discuss the well-known fixed-effects and random-effects meta-regression models.

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 merely adds a second random term to MRA model (1).

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

where $\mathbf{v}$ is an $L \times 1$ vector of random effects, assumed to be independently distributed as $N(0, \tau^2)$ as well as independent of both $\boldsymbol{\varepsilon}$ and $\mathbf{M}\boldsymbol{\beta}$. τ^2 is the excess, between-study heterogeneity variance. $\boldsymbol{\varepsilon}$ is an $L \times 1$ vector of residuals representing the unexplained sampling errors of the reported empirical effects. Note that this RE-MRA model assumes that any excess random heterogeneity comes from an additive term, $\mathbf{v}$, whereas WLS-MRA's estimates are invariant to any excess multiplicative variance, σ^2, without assuming that between-study variances are additive or multiplicative. We do not regard WLS's multiplicative variance structure as a requirement, but rather as a robustness property. In everything that we do in this paper, we assume that the random-effects model is correct. We are merely arguing, that even when random-effects model is correct, the meta-regression parameters, $\boldsymbol{\beta}$, are estimated as well or better by an entirely different approach to accommodating excess heterogeneity, the unrestricted WLS-MRA.

Random-effects 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 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, $1/\left(SE_j^2 + \hat{\tau}^2\right)$, in a restrictive ($\sigma^2 = 1$) WLS context (Raudenbush, 1994). That is, once $\hat{\tau}^2$ is estimated, $SE_j^2 + \hat{\tau}^2$ becomes the principal diagonal of $\mathbf{V}$ in Equation (2) with σ^2 restricted to be one. Last, WLS equation (2) is used to provide the RE-MRA estimate of $\boldsymbol{\beta}$. Our succeeding simulations are based on Raudenbush's (1994) iterative maximum likelihood algorithm. We recognize that there are other algorithms for calculating RE-MRA coefficients. However, we selected this one because it produces MRA coefficients and standard errors that 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, neither additive nor multiplicative, and thereby constrains WLS's common variance term, σ^2, to be equal to one. FE-MRA's estimates, $\hat{\boldsymbol{\beta}}$, are identical to WLS-MRA—recall Equations (1) and (2). The only difference is that FE-MRA's approach further divides WLS-MRA's standard errors by $\sqrt{MS_E}$. The unrestricted WLS-MRA estimation approach is identical to the FE-MRA approach except that σ^2 is not restricted to be equal to one; hence, WLS-MRA may be regarded as 'unrestricted' WLS, and FE-MRA may be seen as a 're-parameterization' of traditional WLS regression.

Excess heterogeneity is not accommodated by FE-MRA's, because σ^2 is constrained to be one. As a result, the confidence intervals produced by FE-MRA are widely recognized to be too narrow if misapplied to unconditional inference—that is, to inferences about future research results or to cases where populations that might differ in some way from the population sampled (Hedges, 1994; Borenstein *et al.*, 2009). The problem is that fixed effects are often applied to settings for which it was not designed (i.e., unconditional inference) and therefore purport to be more precise than they actually are. If one merely wishes to make an inference about the fixed population mean for a population known to have only a single underlying true effect, then fixed-effects' standard errors and confidence intervals are correct. However, they are not appropriate when making inferences to what might be found in the future research record or to the true underlying value of the effect in question because neither context can guarantee either a fixed population or the absence of genuine heterogeneity.

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 approach (Lipsey and Wilson, 2001; Johnson and Huedo-Medina, 2012). Our simulations demonstrate that there are no circumstances for which fixed-effects estimates have notably better statistical properties than unrestricted WLS-MRA approach. When there is no excess heterogeneity and the population is identical to the one sampled, our simulations show that FE-MRA and WLS-MRA produce identical estimates with *exactly the same* coverage deviation, on average, from the nominal 95% level. Needless to say, when there is excess heterogeneity and fixed effects are nonetheless inappropriately applied, WLS-MRA provides superior coverage.

4. Simulations

Although the Gauss–Markov theorem is nearly two centuries old and WLS have been widely known and well established for many decades, 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 direct applicability of the Gauss–Markov theorem to RE-MRA. Second, the unrestricted WLS-MRA, Equation (2), involves a multiplicative error variance to estimate excess variance and the resulting confidence intervals, whereas our succeeding simulations *add* random unexplained heterogeneity to the data generating process *in all cases*. As we discuss earlier, WLS-MRA's multiplicative variance–covariance structure is not an assumption but rather an unavoidable mathematical property of all WLS meta-regressions, including random effects.

However, in *all* of our simulations, we intentionally misspecify WLS-MRA's variance–covariance structure. In practice, unexplained heterogeneity might well be additive, as assumed by RE-MRA. Furthermore, it is important to be sure that WLS-MRA estimates and their confidence intervals will have adequate statistical properties when heterogeneity and errors are generated in any realistic manner. In these simulations, we rely on WLS's invariance to multiplicative excess variance to estimate the meta-regression coefficients robustly even though additive heterogeneity, exactly as assumed by RE-MRA, is forced into *all* of our simulations. This paper does not question RE-MRA assumptions or the additivity of the heterogeneity variance. We investigate whether WLS-MRA's mathematically invariance to all nonzero multiplicative constants is sufficiently strong to overcome the intentional misspecification of the total variance structure and produce estimates comparable with or better than random effects. The point to these simulations is to demonstrate that WLS-MRA is robust to an additive variance–covariance structure and often performs as well as or better than conventional meta-regression methods in practical applications, even when RE-MRA's theoretical assumptions are imposed upon *all* of the estimates that are meta-analyzed and simulated.

These simulations also generate *systematic heterogeneity* as an 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). Because, the list of independent variables is always reported in research using regressions, reviewers can easily identify whether any given, potentially relevant, variable is omitted or not. Omitting a relevant explanatory variable is an omnipresent threat to the validity of applied econometrics and observational social science, in general. The resulting omitted-variable bias is well known and widely recognized in all econometric texts (Judge *et al.*, 1982; Stanley and Jarrell, 1989; Davidson and MacKinnon, 2004). However, in spite of a previous simulation study (Koetse *et al.*, 2010), what might be in doubt is whether such a crude binary variable adequately corrects this bias within the context of meta-regression. In the simulations presented here, we strive both to be realistic and to advantage of the random-effects approach; thus, these simulations present a genuine challenge to our unrestricted WLS meta-regression estimator.

4.1. Simulation design

Our simulation design begins with past simulation studies of meta-regression methods—Stanley (2008), Stanley *et al.* (2010), and Stanley and Doucouliagos (2014)—which, in turn, were calibrated to mirror several published meta-analyses (Stanley, 2008). Here, we generalize these past simulation designs to allow for systematic heterogeneity, different mechanisms for generating additive random heterogeneity, and a wider range of parameters to ensure that we are challenging and investigating WLS-MRA, fully. Also, we 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 subsequent discussions for more details.

Essentially, data are generated, and a regression coefficient is computed, $\widehat{\alpha}_{1j}$, representing one empirical effect reported in the research literature, y_j. Also, see Figures 1 and 2 for further details. This process of generating data and estimating the target regression coefficient is repeated either 20 or 80 times, producing the MRA sample. A meta-regression sample size of 80 is chosen because it provides sufficient power in most meta-regression applications and there are typically hundreds of reported effects in economics and the other social sciences when regression coefficients are estimated. For example, among 159 meta-analyses in economics, the median number of reported estimates is 192 and the mean is 403 (Ioannidis *et al.*, 2016). Twenty is chosen for the lower limit of the MRA sample size because it is a rather small sample size for any regression estimate. Eighty-eight percent of those meta-analyses that summarize regression estimates in economics have sample sizes greater than 20.

Next, alternative approaches WLS-MRA, RE-MRA, and FE-MRA are employed to estimate the underlying 'true' regression parameter corrected for misspecification biases that are contained in half of the research literature.

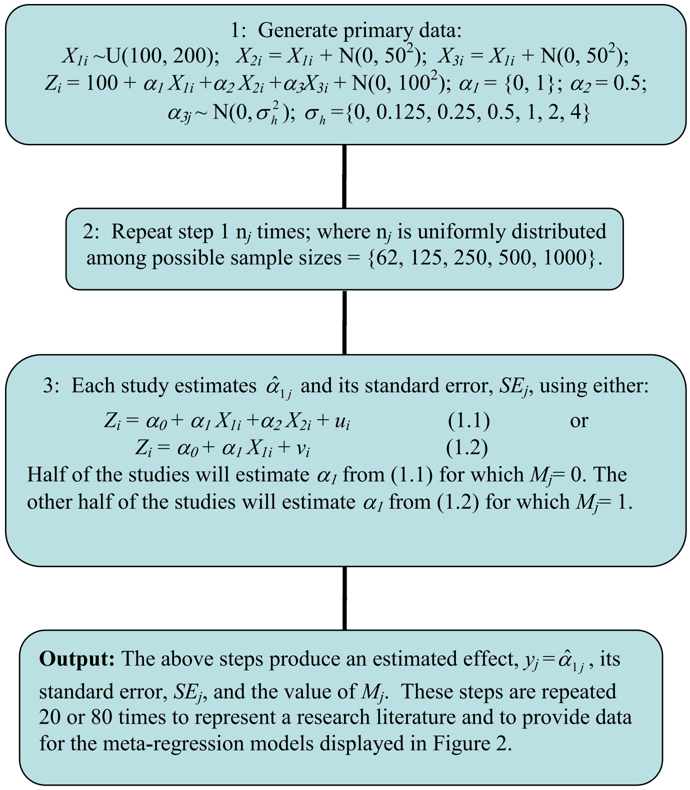

Figure 1. Schema for the simulation of primary studies

In practice, omitted-variable bias is an omnipresent threat to observational studies in the social sciences. Initially, we assume that there are no publication selection biases. Later, we allow selection for statistical significance (publication bias), and, last, we model publication selection bias using the selected estimate's standard error (SE_j) or variance as an additional moderator variable (Egger *et al.*, 1997; Stanley and Doucouliagos, 2014).

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

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

$u_i \sim$ N(0, 100^2), $X_{1i} \sim$ U(100, 200), $\alpha_1 = \{0, 1\}$, $\alpha_2 = 0.5$, and n_j is the number of observations available to the jth primary study.[4] The empirical effect of interest is the estimate of α_1, $\widehat{\alpha}_{1j}$. The partial correlation between Z and X_1 is 0.32, calculated from 100,000 replications of Equation (4), when $\alpha_1 = 1$ and there is no excess heterogeneity. It falls to 0.17 at the highest level of heterogeneity. As routinely observed among systematic reviews of regression coefficients, a wide range of sample sizes is assumed to be used to estimate α_1 in the primary literature; n_j = {62, 125, 250, 500, 1000}. One of two regression models is 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 an 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 ($\widehat{\alpha}_{1j}$) will be biased. This omitted-variable bias will be $\alpha_2 \cdot \gamma_{12}$, where $\gamma_{12} = 1$ 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. Whether or not X_2 is included in any individual simulated study is random with probability 0.5. When X_2 is omitted, $M_j = 1$;

[4] *The statistical properties of the primary regression estimates are not affected by the distribution of the independent variables in Equation (4). That is, if one or more of these X* variables are binary, for example X_2, no substantive change will occur. Regression, primary or meta, takes the independent variables as fixed constants; thus, it does not matter which distribution one choose to generate them in a simulation. If X_2, for example, were binary, then its coefficient, α_2, and its correlation with X_1 would likely need to be adjusted to give the same average results as those reported in the tables of this paper.

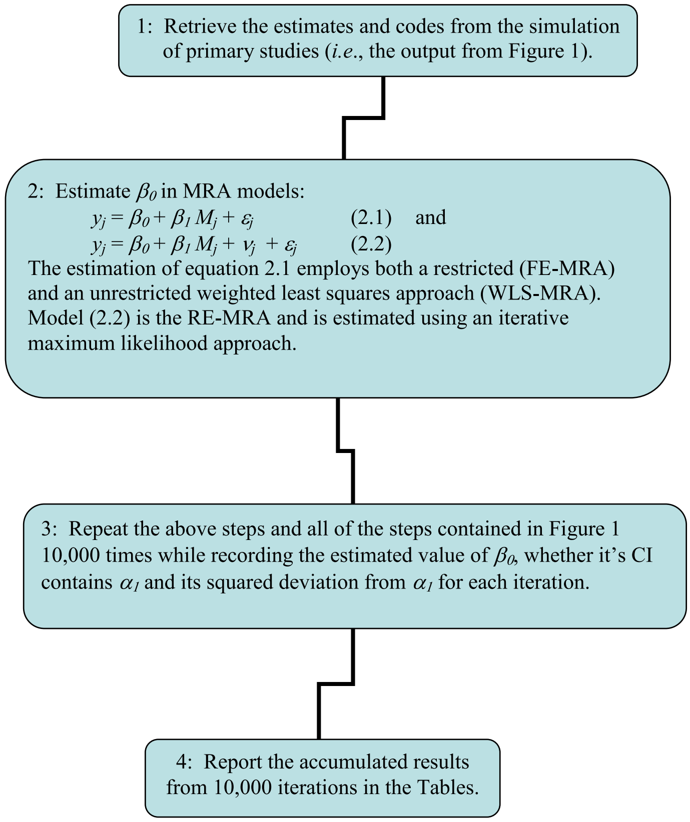

Figure 2. Simulation of alternative meta-regression estimators

$M_j = 0$, otherwise. M_j then becomes an independent, or moderator, variable in the reviewer's meta-regression model:

$$y_j = \beta_0 + \beta_1 M_j + \varepsilon_j. \tag{5}$$

As before, y_j is the jth reported effect, $\hat{\alpha}_{1j}$, and ε_j is the usual random regression disturbance. If M_j correctly accommodates and adjusts for this omitted-variable bias, β_0 will be equal to the mean of the random-effects distribution, α_1. MRA model (5) is then estimated using either fixed-effects, random-effects, or unrestricted WLS approach with $\hat{\alpha}_{1j}$'s squared standard error, SE_j^2, as the estimate of σ_j^2. Needless to say, the RE-MRA adds a second random term, v_j, to Equation (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 Tables 1 and 2. See Figures 1 and 2 for further details about the simulations.

Past simulations have found that the relative size of the unexplained heterogeneity is the most influential dimension of the statistical properties of alternative meta-regression estimators (Stanley, 2008; Stanley *et al.*, 2010; Stanley and Doucouliagos, 2014). In our present simulations, unexplained random heterogeneity is induced in three ways: (i) as random omitted-variable bias; (ii) by adding a random disturbance, $N(0, \tau^2)$, directly to the true regression coefficient, α_1; and (iii) by allowing the true total effect of X_1 on Z, $\alpha_1 + \alpha_{3j}$, to depend on random variations in some moderator variable (age, gender, and income) thought to influence the phenomenon in question.

Our main results focus on random omitted-variable bias generated through a second omitted-variable, X_3, and calibrated by σ_h. Random effects, α_{3j}, are generated for each study from $N(0, \sigma_h^2)$. α_{3j} is fixed for a given primary study but is random across studies. $E(\hat{\alpha}_{1j}) = \alpha_1 + \alpha_{3j}$. Thus, $\sigma_h^2 = \tau^2$ in conventional random-effects terms. Like X_2, X_3 is also generated in a way that makes it correlated with X_1. However, the mean of the sampling distribution, $E(\hat{\alpha}_{1j})$, is now forced to be $\alpha_1 + \alpha_{3j}$, rather than α_1. That is, the true total effect of an increase in X_1 is $\alpha_1 + \alpha_{3j}$.[5] In this context, α_1 is the direct effect of X_1 on Z, and α_3 is the indirect effect of X_1 on Z through X_3. The mean of the

[5] *Omitted-variable bias in estimating the direct effect of* X_1 on Z is $\alpha_3 \cdot \gamma_{13}$, where γ_{13} is the slope coefficient of a regression of X_3 on X_1 and set equal to 1 by our simulations.

Table 1. Coverage percentage of FE-MRA, RE-MRA, and WLS-MRA (nominal level = 0.95).

MRA sample size	σ_h, excess heterogeneity	True effect	I^2	FE-MRA	RE-MRA	WLS-MRA
20	0	0	0.0948	0.9489	0.9544	0.9505
20	0.125	0	0.2433	0.8769	0.9218	0.9350
20	0.25	0	0.6014	0.7067	0.9082	0.9079
20	0.5	0	0.8503	0.4740	0.9191	0.9000
20	1.0	0	0.9465	0.3088	0.9254	0.9110
20	2.0	0	0.9761	0.2277	0.9265	0.9339
20	4.0	0	0.9858	0.1909	0.9233	0.9464
80	0	0	0.0936	0.9495	0.9553	0.9525
80	0.125	0	0.2469	0.8741	0.9429	0.9350
80	0.25	0	0.6011	0.7007	0.9371	0.9058
80	0.5	0	0.8493	0.4769	0.9495	0.9079
80	1.0	0	0.9465	0.3173	0.9433	0.9167
80	2.0	0	0.9761	0.2384	0.9460	0.9440
80	4.0	0	0.9858	0.2047	0.9472	0.9528
20	0	1	0.0593	0.9545	0.9603	0.9531
20	0.125	1	0.3186	0.8738	0.9187	0.9278
20	0.25	1	0.6465	0.7070	0.8996	0.9064
20	0.5	1	0.8687	0.4688	0.9183	0.8996
20	1.0	1	0.9517	0.3125	0.9220	0.9119
20	2.0	1	0.9777	0.2301	0.9227	0.9378
20	4.0	1	0.9863	0.1851	0.9252	0.9455
80	0	1	0.0589	0.9532	0.9568	0.9532
80	0.125	1	0.3179	0.8704	0.9382	0.9282
80	0.25	1	0.6471	0.7040	0.9444	0.9138
80	0.5	1	0.8683	0.4765	0.9460	0.9049
80	1.0	1	0.9517	0.3153	0.9427	0.9240
80	2.0	1	0.9777	0.2364	0.9468	0.9393
80	4.0	1	0.9863	0.1947	0.9436	0.9566
Average				0.5349	0.9352	0.9286

Notes: FE-MRA, RE-MRA, and WLS-MRA refer to the fixed-effects, random-effects, and unrestricted weighted least squares meta-regression estimates, respectively, of β_0 in MRA model (5). Coverage proportions 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 the random excess heterogeneity relative to sampling error. All of these measures are calculated empirically for each replication and averaged across 10,000 replications.

distribution of a given estimated effect is $\alpha_1 + \alpha_{3j}$, and α_{3j} is randomly generated for each study from N(0, σ_h^2). Thus, $\widehat{\alpha}_{1j}$ estimates a heterogeneous effect just as random-effects assumes. Unlike X_2, X_3 is an omitted variable in all of the primary studies that estimate $\widehat{\alpha}_{1j}$, rather than in half of them. Because all studies omit X_3, no meta-regression can correct for its omission in a primary study.

We begin with this case of random omitted-variable bias, because we believe that it is the most realistic in observational social science research where regression is employed. In the social sciences, everyone recognizes that the phenomenon under study might be influenced by a very large number of factors or variables and that it is impossible or impractical to control for all of them in a given observational study. We believe that unmodeled, omitted-variable bias is a main source of both excess unexplained heterogeneity and selection bias in applied econometrics and other areas of observational research. However, Section 4.3 reports the results of simulations where excess heterogeneity is generated directly, and Section 4.4 presents simulations where random additive heterogeneity is generated through random variations in the mean of some moderator variable that influences the 'true' effects in question.

Inducing random heterogeneity through omitted-variable bias adds a random term, $\sigma_{3j} \sim \mathrm{N}\left(0, \sigma_h^2\right)$, to the mean of the sampling distribution of $\widehat{\alpha}_{1j}$, just as assumed in the conventional RE-MRA. Values of random heterogeneity, σ_h^2, were selected to encompass what is found in actual meta-analyses, as measured by I^2 (Higgins and Thompson, 2002). For example, I^2 is 90% among US minimum wage elasticities (Doucouliagos and Stanley, 2009), 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

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

MRA sample size	σ_h, excess heterogeneity	True effect	I^2	RE-MRA |Bias|	WLS-MRA |Bias|	RE-MRA MSE	WLS-MRA MSE
20	0	0	0.0948	0.00059	0.00041	0.00554	0.00549
20	0.125	0	0.2433	0.00105	0.00124	0.00829	0.00845
20	0.25	0	0.6014	0.00091	0.00157	0.01498	0.01687
20	0.5	0	0.8503	0.00085	0.00031	0.03555	0.04661
20	1.0	0	0.9465	0.00087	0.00282	0.11340	0.13435
20	2.0	0	0.9761	0.00157	0.00014	0.40591	0.35193
20	4.0	0	0.9858	0.01341	0.00148	1.62793	0.88102
80	0	0	0.0936	0.00048	0.00051	0.00110	0.00109
80	0.125	0	0.2469	0.00059	0.00040	0.00173	0.00179
80	0.25	0	0.6011	0.00029	0.00021	0.00331	0.00386
80	0.5	0	0.8493	0.00030	0.00077	0.00833	0.01066
80	1.0	0	0.9465	0.00023	0.00031	0.02669	0.02919
80	2.0	0	0.9761	0.00012	0.00099	0.09887	0.06928
80	4.0	0	0.9858	0.00240	0.00203	0.38644	0.15535
20	0	1	0.0593	0.00046	0.00042	0.00564	0.00558
20	0.125	1	0.3186	0.00186	0.00172	0.00825	0.00837
20	0.25	1	0.6465	0.00147	0.00164	0.01487	0.01706
20	0.5	1	0.8687	0.00068	0.00107	0.03607	0.04736
20	1.0	1	0.9517	0.00118	0.00358	0.11352	0.13372
20	2.0	1	0.9777	0.00075	0.00247	0.39659	0.33989
20	4.0	1	0.9863	0.01035	0.01111	1.61637	0.83945
80	0	1	0.0589	0.00067	0.00067	0.00110	0.00109
80	0.125	1	0.3179	0.00013	0.00013	0.00172	0.00177
80	0.25	1	0.6471	0.00068	0.00060	0.00333	0.00389
80	0.5	1	0.8683	0.00009	0.00048	0.00822	0.01035
80	1.0	1	0.9517	0.00012	0.00063	0.02720	0.02953
80	2.0	1	0.9777	0.00163	0.00005	0.09808	0.06986
80	4.0	1	0.9863	0.00195	0.00040	0.38633	0.15414
Average				0.00163	0.00136	0.19483	0.12064

Notes: RE-MRA and WLS-MRA refer to the random-effects and unrestricted weighted least squares 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 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.

correlations of CEO pay and corporate performance (Doucouliagos, *et al.*, 2012), 99.2% among the income elasticities of health care (Costa-Font *et al.*, 2013), and 84% among the partial correlation coefficients of UK's minimum wage increases and employment (De Linde Leonard *et al.*, 2014). 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^2 values of 80% or 90% are the norm.[6]

We calculate I^2 in the succeeding tables by 'empirically' following Higgins and Thompson (2002) and averaging them across 10,000 replications. Empirical estimates of I^2 are biased upward when there is little or no excess heterogeneity (i.e., $\sigma_h = 0$). Like $\hat{\tau}^2$, conventional practice is to truncate I^2 at zero.

Table 1 reports the coverage of WLS-MRA, FE-MRA, and RE-MRA estimates of β_0 in MRA model (5). Needless to say, the RE-MRA version adds a random term to Equation (5), explicitly estimates its variance, and uses this estimate in the weights (and variance–covariance) 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

[6]*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 earlier, we have completed an additional three meta-analyses. Among test of market efficiency 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%.

intervals that are routinely employed by meta-analysts, we generated random datasets in the aforementioned 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.

Last, we also allow publication selection (or reporting) bias in the simulations reported in Table 3. When publication selection is permitted, random values of all the relevant variables are generated in the same way as discussed earlier and sketched in Figure 1 until a statistically significant positive effect, $\widehat{\alpha}_{1j}$, is generated by chance. Unlike simulations of others, we allow selective reporting across several dimensions simultaneously—random sampling error, random heterogeneity (generated through three different pathways), and systematic heterogeneity. Thus, our results are more, likely to be applicable to actual applications than previous studies of selective reporting or publication bias. To conserve space, we assume that such selection for statistical significance occurs in half the reported empirical estimates. For this half, only statistically significant positive effects are retained, becoming output in Figure 1 and the data used for meta-regression (Figure 2). For the other half, the first random estimate is retained and used regardless of whether it is statistically significant or not. In other papers where the focus is on the magnitude of publication bias, how to identify it and how to correct it, we vary the incidence of publication selection from 0% to 100% (Stanley, 2008; Stanley *et al.*, 2010; Stanley and Doucouliagos, 2014). The focus of the current investigation is not on the magnitude of publication bias, per se, but rather on the relative biases and MSEs of RE-MRA and WLS-MRA when publication bias is a genuine possibility. Thus, it is sufficient to show that WLS-MRA has smaller bias and MSE than RE-MRA when there is some

Table 3. Bias and MSE of RE-MRA and WLS-MRA with 50% publication selection bias.

MRA sample size	σ_h, excess heterogeneity	True effect	I^2	RE-MRA |Bias|	WLS-MRA |Bias|	RE-MRA MSE	WLS-MRA MSE
20	0	0	0.1689	0.0348	0.0328	0.0151	0.0147
20	0.125	0	0.3241	0.0581	0.0510	0.0218	0.0209
20	0.25	0	0.5697	0.1140	0.0957	0.0414	0.0397
20	0.5	0	0.8102	0.2367	0.1964	0.1084	0.1035
20	1.0	0	0.9264	0.4510	0.3470	0.3259	0.2677
20	2.0	0	0.9670	0.8138	0.5692	1.0391	0.6824
20	4.0	0	0.9809	1.5212	0.8595	3.6524	1.6393
80	0	0	0.1551	0.0361	0.0345	0.0039	0.0037
80	0.125	0	0.3589	0.0668	0.0593	0.0085	0.0075
80	0.25	0	0.6184	0.1322	0.1148	0.0237	0.0200
80	0.5	0	0.8372	0.2659	0.2250	0.0824	0.0643
80	1.0	0	0.9362	0.4900	0.3891	0.2687	0.1815
80	2.0	0	0.9701	0.8939	0.6092	0.8868	0.4365
80	4.0	0	0.9818	1.6566	0.8880	3.0617	0.9305
20	0	1	0.0825	0.0135	0.0128	0.0056	0.0056
20	0.125	1	0.2358	0.0168	0.0129	0.0083	0.0085
20	0.25	1	0.5325	0.0350	0.0221	0.0155	0.0171
20	0.5	1	0.8083	0.0916	0.0583	0.0412	0.0477
20	1.0	1	0.9255	0.2415	0.1669	0.1567	0.1483
20	2.0	1	0.9666	0.5566	0.3541	0.6540	0.4317
20	4.0	1	0.9806	1.2326	0.6554	2.8299	1.1672
80	0	1	0.0450	0.0101	0.0096	0.0012	0.0012
80	0.125	1	0.2591	0.0158	0.0115	0.0020	0.0019
80	0.25	1	0.5926	0.0314	0.0172	0.0042	0.0041
80	0.5	1	0.8364	0.0940	0.0570	0.0163	0.0133
80	1.0	1	0.9349	0.2564	0.1740	0.0886	0.0559
80	2.0	1	0.9695	0.6142	0.3756	0.4553	0.1978
80	4.0	1	0.9817	1.3571	0.6591	2.1485	0.5624
Average				0.4049	0.2521	0.5703	0.2527
Average for 100% publication selection bias				0.9649	0.6536	2.0600	0.8625

Notes: RE-MRA and WLS-MRA refer to the random-effects and unrestricted weighted least squares 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 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.

moderate amount of publication selection. As requested by an anonymous referee, we also report average bias and MSE for the case of 100% publication bias.

4.2. *Random omitted-variable biases: results*

Table 1 reports the coverage percentages as well as a measure of excess random heterogeneity, I^2, discussed earlier; 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' mean effect {0, 1} is computed, giving the coverage percentages found in the last three columns of Table 1.

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 because their coverages, on average, deviate from the nominal 95% by *exactly* the same amount. 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 with RE-MRA's coverage. On average, RE-MRA coverage is 0.6% closer to the nominal 95% level than WLS-MRA coverage, 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 can produce acceptable confidence intervals, comparable with those of RE-MRA and that FE-MRA's confidence intervals will be unacceptable for most realistic applications.

Last, 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.

Table 2 reports the bias and MSE of these meta-regression approaches 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 and 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% of a small 'true' mean effect, 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 the unrestricted WLS meta-regression is practically equivalent to random-effects (or mixed-effects) meta-regression when there is no publication bias. We do not report bias and MSE for FE-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.

In Sections 4.3 and 4.4, we generate random heterogeneity in other ways to gage how that might affect the relative performance of these alternative approaches to meta-regression, and it does have some effect. However, our central focus is to investigate whether RE-MRA is more biased than WLS-MRA when there is publication bias. Table 3 reports the bias and MSE of these meta-regression estimators when 50% of the estimates are reported only when they are statistically significantly positive. In the columns labeled '|Bias|', the average absolute bias of the MRA estimate of β_0 from Equation (5), or its random-effects equivalent, is reported. These average absolute biases are calculated by the absolute value of the difference between the average of these 10,000 simulations and the true mean effect = {0, 1}.

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 often 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 actual effect. In a recent survey of empirical economics summarized by 159 meta-analyses, Ioannidis *et al.* (2016) find that the median of the medians of reported effects is exaggerated by factor of two or more, using a conservative approach, FE-WLS, to estimate the 'true' effect. Such residual selection biases can be quite large and can have important practical consequences for policy and practice. At least one-third of empirical economics results are exaggerated by a factor of 4 or more (Ioannidis *et al.*, 2016). 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 WLS meta-regression when there is publication selection bias?

Table 3 demonstrates that RE-MRA is more biased and less MSE efficient (higher MSE) than WLS-MRA when there is publication bias. In all cases, WLS-MRA has smaller bias than random-effects (also Figure 3), and it 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

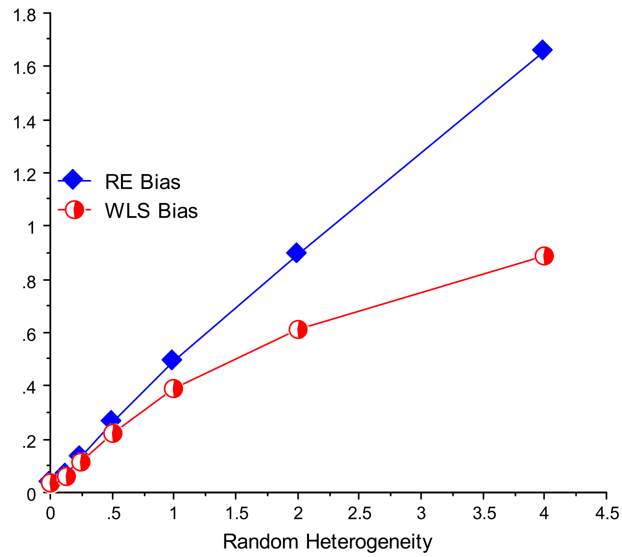

Figure 3. Plot of bias versus σ_h for RE-MRA and WLS-MRA: random omitted-variable bias with 50% publication bias (true effect = 0; MRA $n = 80$)

is 61% larger. Where the bias is largest, WLS-MRA makes its greatest relative improvement over RE-MRA. Although all MRA approaches suffer from notable publication bias if there is selection for statistically significant results and if this selection is left uncorrected, the bias and MSE efficiency of WLS-MRA are much better than RE-WLS's.

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. Although the absolute size of the differences in biases and in MSEs worsen for RE-MRA in this scenario, the relative (or ratio of) bias and MSE are roughly the same as before. Before we turn to the issue of reducing these potentially large publication biases, we report simulations that use different pathways for generating additive excess heterogeneity.

4.3. *Direct random, additive heterogeneity*

Thus far, we have modeled heterogeneity as random omitted-variable bias, because omitted-variable bias is ubiquitous in observational social science research. In the simulations reported in this section, we assume that excess heterogeneity is generated directly and additively to the 'true' mean effect, α_1. Table 4 reports the coverage percentages when excess heterogeneity is direct and additive, exactly as assumed by random-effects theory. In this simulation design, there are no random omitted-variable biases. That is, X_{3i} is not included in the original regression model used to generate Z_i in Equation (4). Instead, random heterogeneity is created by adding $v_j \sim N\left(0, \sigma_h^2\right)$ directly to α_1.

When there is no publication selection bias, Table 4 shows that RE-MRA's confidence intervals are, on average, 2.7% closer to the nominal level (95%) than are WLS-MRA's confidence intervals. Table 5 shows the corresponding bias and MSE, when there is no publication selection bias. WLS-MRA has slightly smaller bias than RE-MRA, but RE-MRA's MSE is a little better. Although these are small differences, either way, in practical application, the edge goes to RE-MRA, when there is no publication bias. But then, this is the pure case of random-effects where heterogeneity is directly added to the true mean effect without any practical complications, so RE-MRA would be expected to perform best here. What is surprising is that the improvements are so small as to be practically inconsequential, even in this unrealistic case.

On the other hand, when there is publication selection, the edge goes to WLS-MRA—see Table 6. In all of these cases, the unrestricted WLS has smaller bias than random-effects (Table 6 and Figure 4). In two-thirds of these, MSE is smaller for WLS-MRA. One might interpret these simulation results as favoring RE-MRA when there is no publication bias and preferring WLS-MRA when there is publication bias. However, because all tests for publication bias have low power, selective reporting can never be ruled out it practice. The advantage of each approach over the other in this simulation design is rather small; thus, we consider these meta-regression approaches to be practically equivalent when excess heterogeneity is directly added to the 'true' mean regression coefficient, in the exact way that RE-MRA's model assumes.

But is it reasonable to expect that the observed heterogeneity of a regression coefficient will be transmitted in such a direct way? How would such 'pure' direct heterogeneity occur in practice? We accept that there is much genuine heterogeneity in social science observational research. However, this

Table 4. Coverage of FE-MRA, RE-MRA, and WLS-MRA from direct random heterogeneity (nominal level = 0.95).

MRA sample size	σ_h, excess heterogeneity	True effect	I^2	FE-MRA	RE-MRA	WLS-MRA
20	0	0	0.0963	0.9504	0.9553	0.9518
20	0.125	0	0.2538	0.8731	0.9172	0.9313
20	0.25	0	0.5521	0.7054	0.9078	0.9073
20	0.5	0	0.8369	0.4461	0.9172	0.8875
20	1.0	0	0.9533	0.2423	0.9209	0.8754
20	2.0	0	0.9880	0.1194	0.9213	0.8765
80	0	0	0.0608	0.9467	0.9518	0.9506
80	0.125	0	0.2817	0.8672	0.9399	0.9288
80	0.25	0	0.6166	0.6940	0.9421	0.9093
80	0.5	0	0.8651	0.4407	0.9430	0.8868
80	1.0	0	0.9624	0.2321	0.9390	0.8850
80	2.0	0	0.9904	0.1200	0.9436	0.8822
20	0	1	0.0965	0.9518	0.9563	0.9541
20	0.125	1	0.2563	0.8719	0.9194	0.9304
20	0.25	1	0.5540	0.7030	0.9061	0.9041
20	0.5	1	0.8366	0.4439	0.9176	0.8882
20	1.0	1	0.9535	0.2387	0.9178	0.8718
20	2.0	1	0.9881	0.1231	0.9197	0.8714
80	0	1	0.0606	0.9486	0.9551	0.9515
80	0.125	1	0.2825	0.8695	0.9381	0.9283
80	0.25	1	0.6168	0.6933	0.9409	0.9057
80	0.5	1	0.8652	0.4385	0.9411	0.8914
80	1.0	1	0.9624	0.2296	0.9408	0.8827
80	2.0	1	0.9904	0.1201	0.9442	0.8809
Average				0.5529	0.9332	0.9055

Notes: FE-MRA, RE-MRA, and WLS-MRA refer to the fixed-effects, random-effects, and unrestricted weighted least squares meta-regression estimates, respectively, of β_0 in MRA model (5). Coverage proportions 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 the random excess heterogeneity relative to sampling error. All of these measures are calculated empirically for each replication and averaged across 10,000 replications.

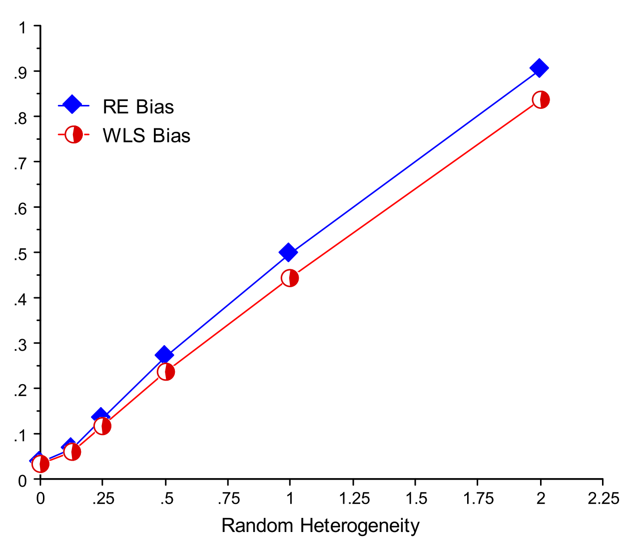

Figure 4. Plot of bias versus σ_h for RE-MRA and WLS-MRA: direct random heterogeneity with 50% publication bias (true effect = 0; MRA n = 80)

Table 5. Bias and MSE of RE-MRA and WLS-MRA from direct random heterogeneity.

MRA sample size	σ_h, excess heterogeneity	True effect	I^2	RE-MRA |Bias|	WLS-MRA |Bias|	RE-MRA MSE	WLS-MRA MSE
20	0	0	0.0948	0.00096	0.00100	0.00536	0.00533
20	0.125	0	0.2433	0.00044	0.00018	0.00844	0.00868
20	0.25	0	0.6014	0.00071	0.00141	0.01519	0.01754
20	0.5	0	0.8503	0.00073	0.00081	0.03636	0.05175
20	1.0	0	0.9465	0.00270	0.00006	0.11932	0.20004
20	2.0	0	0.9761	0.00230	0.00160	0.42989	0.75526
80	0	0	0.0936	0.00041	0.00044	0.00110	0.00110
80	0.125	0	0.2469	0.00070	0.00054	0.00173	0.00179
80	0.25	0	0.6011	0.00010	0.00003	0.00337	0.00402
80	0.5	0	0.8493	0.00035	0.00066	0.00843	0.01235
80	1.0	0	0.9465	0.00191	0.00009	0.02829	0.04747
80	2.0	0	0.9761	0.00171	0.00301	0.10516	0.18479
20	0	1	0.0593	0.00026	0.00021	0.00549	0.00545
20	0.125	1	0.3186	0.00121	0.00077	0.00838	0.00850
20	0.25	1	0.6465	0.00080	0.00112	0.01491	0.01752
20	0.5	1	0.8687	0.00165	0.00084	0.03769	0.05536
20	1.0	1	0.9517	0.00424	0.00679	0.11878	0.19899
20	2.0	1	0.9777	0.00958	0.00001	0.42403	0.73770
80	0	1	0.0589	0.00049	0.00050	0.00110	0.00110
80	0.125	1	0.3179	0.00020	0.00021	0.00175	0.00181
80	0.25	1	0.6471	0.00112	0.00110	0.00333	0.00401
80	0.5	1	0.8683	0.00015	0.00066	0.00862	0.01251
80	1.0	1	0.9517	0.00074	0.00124	0.02815	0.04774
80	2.0	1	0.9777	0.00329	0.00596	0.10536	0.18212
Average				0.00153	0.00122	0.06334	0.10679

Notes: RE-MRA and WLS-MRA refer to the random-effects and unrestricted weighted least squares 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 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.

heterogeneity would typically depend on other factors—regions, industry, gender, age, and so on—systematic differences potentially identifiable and accommodated through a multiple regression in the primary research. Meta-regression should model and thereby filter out these effects as omitted variables, as Tables 1, 2, and 3 have demonstrated.

Alternatively, one might argue that there are so many potential sources of heterogeneity that they can be treated as if they were random, N(0, σ_h^2). But how do these sources of heterogeneity manifest themselves? In the previous Section 4.2, we assumed that there are so many variables potentially affecting the social science phenomenon in question that we can treat their effect on estimated effects as a random, normal omitted-variable bias. But how could random heterogeneity affect the *true* mean regression coefficient directly without intermediates? Systematic heterogeneity or random heterogeneity, which comes from not fully controlling for the effects of moderator variables or through random misspecification biases, might easily affect the expected value of the estimated regression coefficient. But how can random heterogeneity directly affect the true mean population value of the marginal effect of one variable on another without any mediation? Perhaps, at random, some other important population characteristic varies from one dataset used in the primary research to the next? But then, our target effect would likely vary according to how the mean of this important characteristic varies from one study to the next. In the next section, we simulate this transmission channel for excess random heterogeneity.

4.4. *Random moderator heterogeneity*

In our third set of simulation experiments, random deviations of a moderator variable (e.g., age, gender, income, and genome) are assumed to be the source of heterogeneity in true effect, α_1. This simulation design is identical to the direct random heterogeneity case reported in Section 4.3, except that the additive term for 'true' effect, α_1, now depends on the random deviations of the observed mean of a moderator variable. This simulation design is

Table 6. Bias and MSE of RE and WLS from direct random heterogeneity and 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	0.0963	0.0377	0.0358	0.0151	0.0147
20	0.125	0	0.2538	0.0601	0.0536	0.0220	0.0211
20	0.25	0	0.5521	0.1181	0.1000	0.0428	0.0419
20	0.5	0	0.8369	0.2428	0.2040	0.1115	0.1132
20	1.0	0	0.9533	0.4629	0.3975	0.3391	0.3626
20	2.0	0	0.9880	0.8270	0.7323	1.0829	1.2669
80	0	0	0.0608	0.0357	0.0341	0.0039	0.0037
80	0.125	0	0.2817	0.0663	0.0588	0.0083	0.0074
80	0.25	0	0.6166	0.1329	0.1161	0.0238	0.0204
80	0.5	0	0.8651	0.2696	0.2367	0.0844	0.0712
80	1.0	0	0.9624	0.4968	0.4444	0.2753	0.2425
80	2.0	0	0.9904	0.9023	0.8366	0.9043	0.8550
20	0	1	0.0965	0.0112	0.0105	0.0056	0.0055
20	0.125	1	0.2563	0.0168	0.0131	0.0086	0.0087
20	0.25	1	0.5540	0.0308	0.0177	0.0156	0.0177
20	0.5	1	0.8366	0.0902	0.0595	0.0412	0.0535
20	1.0	1	0.9535	0.2470	0.1978	0.1652	0.2169
20	2.0	1	0.9881	0.5937	0.5207	0.7103	0.9505
80	0	1	0.0606	0.0103	0.0098	0.0012	0.0012
80	0.125	1	0.2825	0.0158	0.0116	0.0020	0.0020
80	0.25	1	0.6168	0.0332	0.0190	0.0043	0.0043
80	0.5	1	0.8652	0.0946	0.0616	0.0165	0.0150
80	1.0	1	0.9624	0.2632	0.2155	0.0927	0.0857
80	2.0	1	0.9904	0.6259	0.5679	0.4761	0.4726
Average				0.2369	0.2064	0.1855	0.2023
Average for 100% publication selection bias				0.6098	0.5545	0.7283	0.6410

Notes: RE-MRA and WLS-MRA refer to the random-effects and unrestricted weighted least squares 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.

meant to represent those cases where coincidental variations in some key characteristic of the data used by the primary study cause differences in the underlying 'true' effect.

For example, consider the value of a statistical life (VSL). It is often calculated from the estimated regression coefficient on the risk of fatality in a hedonic wage equation (Stanley *et al.*, 2013). VSL is a key parameter in assessing benefits from health and safety policies and regulations (United States Environmental Protection Agency (EPA), 1997). However, it is widely recognized that VSL is itself dependent on the age and income of the workers investigated (Viscusi and Aldy, 2003; Viscusi, 2011; Doucouliagos *et al.*, 2014). Because data used to estimate the value of a statistical life are often drawn from a given country, region, or group of industries, the underlying average income is likely to vary greatly from one estimate of VSL to the next (Viscusi and Aldy, 2003; Doucouliagos *et al.*, 2014). Government agencies acknowledge the importance of this heterogeneity and adjust VSL values to the relevant average income level with the help of an estimated income elasticity of the VSL (United States Environmental Protection Agency, 2010; U.S. Dept. of Transportation, 2011). It is widely recognized that estimated regression coefficients, VSL in this example, can be influenced by incidental variation of some other important moderator variable.

Tables 7 and 8 contain the results of 10,000 replications of alternative approaches to meta-regression under various conditions when heterogeneity is induced through random variations in the mean of a moderator variable. These simulation results clearly demonstrate that WLS-MRA is superior to RE-MRA. When there is no publication selection, unrestricted WLS produce coverage rates as good as random effects. On average, WLS-MRA's coverage deviates from the nominal 95% level by 0.84% while RE-MRA is off by 0.92%—no practical difference.

When there is publication selection bias, WLS-MRA dominates RE-MRA (Table 8 and Figure 5). Here, WLS-MRA has a smaller bias and MSE than RE-MRA in all cases. On average, RE-MRA's bias is 64% larger than WLS-MRA's bias, and RE-MRA's MSE is 137% bigger. Unlike coverage rates, these differences in bias and MSE will often be of practical import.

Table 7. Coverage of FE-MRA, RE-MRA, and WLS-MRA from random mean heterogeneity (nominal level = 0.95).

MRA sample size	True effect	I^2	FE-MRA	RE-MRA	WLS-MRA
20	0	0.0951	0.9494	0.9541	0.9506
20	0	0.2094	0.9201	0.9400	0.9561
20	0	0.4820	0.8375	0.9192	0.9590
20	0	0.7969	0.6152	0.9350	0.9611
20	0	0.9409	0.3662	0.9381	0.9597
20	0	0.9847	0.1897	0.9331	0.9612
80	0	0.0593	0.9532	0.9575	0.9558
80	0	0.2134	0.9163	0.9410	0.9533
80	0	0.5212	0.8376	0.9465	0.9576
80	0	0.8149	0.6167	0.9504	0.9633
80	0	0.9463	0.3690	0.9488	0.9610
80	0	0.9860	0.1929	0.9444	0.9625
20	1	0.0960	0.9495	0.9542	0.9532
20	1	0.2108	0.9162	0.9338	0.9503
20	1	0.4818	0.8370	0.9258	0.9587
20	1	0.7968	0.6141	0.9330	0.9570
20	1	0.9413	0.3679	0.9366	0.9618
20	1	0.9847	0.1905	0.9370	0.9650
80	1	0.0583	0.9487	0.9548	0.9512
80	1	0.2136	0.9260	0.9497	0.9580
80	1	0.5217	0.8392	0.9455	0.9570
80	1	0.8146	0.6188	0.9485	0.9589
80	1	0.9463	0.3597	0.9473	0.9638
80	1	0.9860	0.1929	0.9474	0.9650
Average			0.6468	0.9584	0.9426

Notes: FE-MRA, RE-MRA, and WLS-MRA refer to the fixed-effects, random-effects, and unrestricted weighted least squares meta-regression estimates, respectively, of β_0 in MRA model (5). Coverage proportions of these estimates are reported in the last three columns. 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 the random excess heterogeneity relative to sampling error. All of these measures are calculated empirically for each replication and averaged across 10,000 replications.

If heterogeneity is induced indirectly through either random variations in some moderator variable or random omitted variables, there is no reason to prefer random-effects meta-regression estimates and often good reason not to.

4.5. Meta-regression accommodation of publication and misspecification biases

Clearly, publication selection biases can be quite large. When there is much heterogeneity, small effects can be manufactured from nothing and reported effects are often double the underlying true effects; see Tables 3, 6, and 8 and Ioannidis *et al.* (2016). Unfortunately, such high levels of heterogeneity are common in the social sciences. How can these potentially large publication biases be reduced and their practical consequences minimized? Table 9 reports the results for both the unrestricted WLS and the random-effects estimators of 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 a WLS-MRA 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 Doucouliagos and Stanley (2009) and Stanley and Doucouliagos (2012 and 2014), Table 9 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 approach. Needless to repeat, the latter adds a random-effects component to Equation (7) and estimates its variance.

Table 8. Bias and MSE of RE and WLS from random mean heterogeneity and 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	0.0962	0.0352	0.0332	0.0149	0.0144
20	0.125	0	0.2088	0.0581	0.0518	0.0211	0.0195
20	0.25	0	0.4833	0.1177	0.0879	0.0423	0.0317
20	0.5	0	0.7963	0.2763	0.1829	0.1434	0.0786
20	1.0	0	0.9413	0.5387	0.3531	0.4816	0.2289
20	2.0	0	0.9846	0.9906	0.6423	1.6828	0.7536
80	0	0	0.9961	0.0358	0.0343	0.0039	0.0038
80	0.125	0	0.2146	0.0577	0.0503	0.0068	0.0057
80	0.25	0	0.5212	0.1277	0.0922	0.0223	0.0131
80	0.5	0	0.8147	0.3009	0.1907	0.1047	0.0447
80	1.0	0	0.9464	0.5703	0.3549	0.3673	0.1466
80	2.0	0	0.9860	1.0401	0.6528	1.2293	0.4924
20	0	1	0.0979	0.0133	0.0126	0.0056	0.0055
20	0.125	1	0.2079	0.0172	0.0146	0.0071	0.0069
20	0.25	1	0.4834	0.0344	0.0225	0.0127	0.0109
20	0.5	1	0.7967	0.1151	0.0546	0.0511	0.0284
20	1.0	1	0.9413	0.2977	0.1458	0.2530	0.1054
20	2.0	1	0.9847	0.7209	0.4135	1.1669	0.4787
80	0	1	0.0593	0.0105	0.0100	0.0012	0.0012
80	0.125	1	0.2144	0.0142	0.0120	0.0016	0.0015
80	0.25	1	0.5213	0.0322	0.0187	0.0035	0.0026
80	0.5	1	0.8146	0.1111	0.0454	0.0204	0.0072
80	1.0	1	0.9463	0.3097	0.1401	0.1289	0.0361
80	2.0	1	0.9860	0.7542	0.3954	0.7068	0.2163
Average				0.2742	0.1671	0.2700	0.1139
Average for 100% publication selection bias				0.6649	0.4455	0.8900	0.3833

Notes: RE-MRA and WLS-MRA refer to the random-effects and unrestricted weighted least squares 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 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.

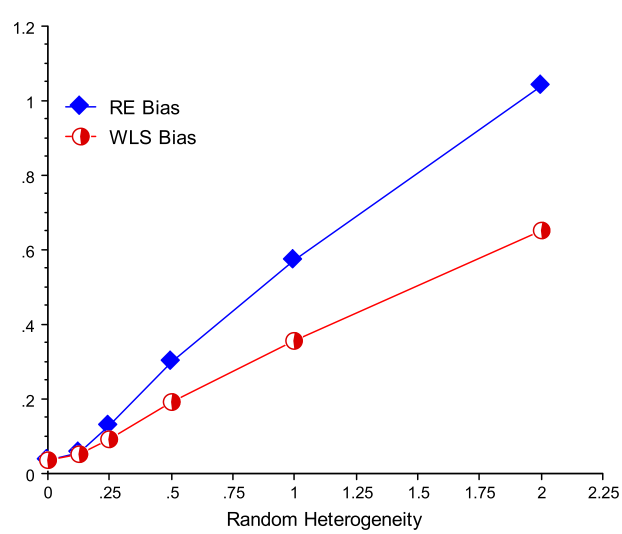

Figure 5. Plot of bias versus σ_h for RE-MRA and WLS-MRA: random moderator heterogeneity with 50% publication bias (true effect = 0; MRA n = 80)

The simulations reported in Tables 9 and 10 induce additive random heterogeneity through random omitted-variable bias, because we believe that this is the most realistic way that excess heterogeneity is generated among observational regression results. This is also the transmission mechanism that produces the largest publication biases and thereby has the greatest need for some correction or accommodation. As before, unrestricted WLS clearly dominates random effects (Table 9). WLS-FAT-PET-MRA has smaller bias than RE-FAT-PET-MRA in 93% of the cases (Table 9), and WLS-FAT-PET-MRA's average bias and MSE are notably smaller than those of RE-FAT-PET-MRA.

However, it is also important to recognize that there are large reductions 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 rather small effect (partial correlations vary from 0.17 to 0.32).

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, 2014). This new approach is a hybrid between the conventional Egger regression, Equation (7), and a meta-regression that uses the estimate's variance as a moderator variable in place of its standard error:

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

See Stanley and Doucouliagos (2014) for the theoretical motivation for this approach and its validation. There, it is shown that MRA model (7) has smaller bias when PET finds no genuine empirical effect (i.e., accept H_0: $\beta_0 = 0$),

Table 9. Bias and MSE of RE-FAT-PET-MRA and WLS-FAT-PET-MRA with 50% publication bias.

Sample size	σ_h, excess heterogeneity	True effect	I^2	RE-MRA |Bias|	WLS-MRA |Bias|	RE-MRA MSE	WLS-MRA MSE
20	0	0	0.0948	0.1658	0.1636	0.0522	0.0516
20	0.125	0	0.2433	0.1545	0.1463	0.0609	0.0596
20	0.25	0	0.6014	0.1052	0.0888	0.0702	0.0746
20	0.5	0	0.8503	0.0105	0.0005	0.1173	0.1475
20	1.0	0	0.9465	0.0898	0.0792	0.3300	0.3565
20	2.0	0	0.9761	0.1157	0.0605	1.0056	0.8423
20	4.0	0	0.9858	0.1997	0.0377	3.0241	1.8469
80	0	0	0.0936	0.1428	0.1414	0.0249	0.0245
80	0.125	0	0.2469	0.1219	0.1108	0.0221	0.0200
80	0.25	0	0.6011	0.0679	0.0476	0.0165	0.0160
80	0.5	0	0.8493	0.0401	0.0513	0.0254	0.0311
80	1.0	0	0.9465	0.1483	0.1383	0.0892	0.0805
80	2.0	0	0.9761	0.2105	0.1372	0.2625	0.1463
80	4.0	0	0.9858	0.3858	0.0600	0.8413	0.2520
20	0	1	0.0593	0.0270	0.0266	0.0172	0.0172
20	0.125	1	0.3186	0.0311	0.0287	0.0255	0.0261
20	0.25	1	0.6465	0.0316	0.0271	0.0459	0.0526
20	0.5	1	0.8687	0.0245	0.0214	0.1085	0.1351
20	1.0	1	0.9517	0.0103	0.0017	0.3081	0.3324
20	2.0	1	0.9777	0.0105	0.0490	0.9767	0.7545
20	4.0	1	0.9863	0.0531	0.1669	2.8764	1.6098
80	0	1	0.0589	0.0251	0.0247	0.0039	0.0039
80	0.125	1	0.3179	0.0293	0.0269	0.0060	0.0061
80	0.25	1	0.6471	0.0340	0.0292	0.0104	0.0117
80	0.5	1	0.8683	0.0247	0.0222	0.0229	0.0275
80	1.0	1	0.9517	0.0162	0.0018	0.0673	0.0638
80	2.0	1	0.9777	0.0448	0.0327	0.2262	0.1224
80	4.0	1	0.9863	0.2100	0.1128	0.7002	0.2309
Average				0.0904	0.0655	0.4049	0.2623

Notes: RE-MRA and WLS-MRA refer to the random-effects and unrestricted weighted least squares 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 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.

Table 10. Bias and MSE of RE-PET-PESSE-MRA and WLS-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	0.0948	0.0665	0.0646	0.0287	0.0285
20	0.125	0	0.2433	0.0598	0.0518	0.0383	0.0384
20	0.25	0	0.6014	0.0437	0.0256	0.0598	0.0670
20	0.5	0	0.8503	0.0018	0.0157	0.1204	0.1548
20	1.0	0	0.9465	0.0785	0.0709	0.3322	0.3790
20	2.0	0	0.9761	0.1020	0.0652	1.0227	0.8587
20	4.0	0	0.9858	0.1365	0.0621	3.0835	1.8331
80	0	0	0.0936	0.0538	0.0522	0.0074	0.0073
80	0.125	0	0.2469	0.0412	0.0305	0.0088	0.0084
80	0.25	0	0.6011	0.0193	0.0061	0.0124	0.0143
80	0.5	0	0.8493	0.0426	0.0655	0.0275	0.0361
80	1.0	0	0.9465	0.1448	0.1330	0.0900	0.0858
80	2.0	0	0.9761	0.2079	0.1360	0.2650	0.1493
80	4.0	0	0.9858	0.3579	0.0545	0.8183	0.2604
20	0	1	0.0593	0.0207	0.0245	0.0175	0.0141
20	0.125	1	0.3186	0.0204	0.0263	0.0258	0.0231
20	0.25	1	0.6465	0.0143	0.0200	0.0436	0.0468
20	0.5	1	0.8687	0.0066	0.0023	0.1043	0.1298
20	1.0	1	0.9517	0.0302	0.0182	0.3158	0.3554
20	2.0	1	0.9777	0.0087	0.0297	0.9750	0.7608
20	4.0	1	0.9863	0.0113	0.1646	2.9422	1.6443
80	0	1	0.0589	0.0200	0.0038	0.0038	0.0015
80	0.125	1	0.3179	0.0195	0.0021	0.0055	0.0024
80	0.25	1	0.6471	0.0198	0.0054	0.0096	0.0071
80	0.5	1	0.8683	0.0060	0.0061	0.0230	0.0246
80	1.0	1	0.9517	0.0496	0.0322	0.0690	0.0644
80	2.0	1	0.9777	0.0567	0.0072	0.2208	0.1226
80	4.0	1	0.9863	0.1672	0.1132	0.7076	0.2314
Average				0.0646	0.0460	0.4064	0.2625

Notes: RE-MRA and WLS-MRA refer to the random-effects and unrestricted weighted least squares 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.

while MRA model (8) has smaller bias when PET finds a genuine empirical effect (i.e., reject H$_0$: $\beta_0 = 0$). Thus, Stanley and Doucouliagos (2014) recommend a conditional estimator, called 'PET-PEESE'. When the conventional *t*-test of H$_0$: $\beta_0 = 0$ from MRA (7) is rejected, MRA model (8) is used to estimate β_0; otherwise, MRA model (7) is used to estimate β_0.

Table 10 reports estimates from this conditional MRA approach over the same conditions as reflected in previous simulations and tables. As before, the unrestricted WLS estimates have much smaller bias and MSE than the corresponding RE estimates, and both PET-PEESE-MRA approaches reduce publication bias beyond the multiple Egger regression model reported in Table 9.

The lesson from all of these simulations taken together is that the unrestricted WLS-MRA *estimator* is practically as good as RE-MRA when there is no selection for statistical significance. When there is publication bias, unrestricted WLS-MRA clearly dominates RE-MRA.

5. Discussion

What explains the success of the unrestricted WLS 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; Stanley, 2008; Stanley *et al.*, 2010: Henmi and Copas, 2010; Stanley and Doucouliagos, 2015), the strong performance of the unrestricted WLS meta-regression 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, $1/\left(SE_i^2 + \hat{\tau}^2\right)$, explains much of the superior statistical performance of WLS-MRA when there is publication selection, reporting, or small-sample bias. Nonetheless, it is surprising to learn that WLS-MRA is often as good as RE-MRA when there is no publication selection bias, because RE-MRA's assumption of random, additive, and normally distributed heterogeneity is forced into all of these simulations.

The unrestricted WLS approach works especially well when heterogeneity is indirect (recall Sections 4.2 and 4.4) because the reported standard errors can be biased or coincidentally proportional. With random omitted-variable bias, the estimated variances of the regression coefficients will be biased (Kmenta, 1971; Davidson and MacKinnon, 2004). Although this bias should not especially advantage WLS, random-effects estimates are sensitive to knowing the within-study variances accurately and thereby correctly separating within-study from between-study variances. The multiplicative variance–covariance structure of WLS, recall Equation (2), is an invariance or robustness property rather than a strict requirement for its application. Allowing for some unknown proportionality offers a flexible approach to excess heterogeneity even if the total variance is additive. Also, proportionality may inadvertently creep into the variance structure when the heterogeneity is indirect. For example, if random variations in the mean of some important moderator variable are the source of heterogeneity, then the variances of the excess heterogeneity and the estimated regression coefficients are both likely to be related to sample size, thereby inducing some proportionality to the estimated variance–covariance matrix.

If there is no publication bias and if indirect random heterogeneity can be ruled out, then random-effects meta-regression also works well. However, even here, WLS-MRA will generally be practically equivalent to random effects. While RE-MRA is often adequate when publication selection bias is known not to be a factor, it is almost never possible to rule out publication bias 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 have publication selection bias. When reviewers fail to do so and there is publication selection or reporting bias, our simulations show that RE-MRA may have much larger biases and MSE than WLS-MRA.

In the worst case for WLS-MRA—direct heterogeneity, no publication selection bias, and high levels of heterogeneity—the confidence intervals for the unrestricted WLS meta-regression are 5–6% further from the nominal level of 95% than RE-MRA's confidence intervals. For this particular combination of circumstances, meta-analysts could justify a preference for RE-MRA over WLS-MRA. But how could reviewers identify whether this combination of conditions applies to a particular area of research?

As discussed earlier, we can never be sure that there is no publication selection (or reporting or small-sample) bias because the power of these tests is low. Likewise, indirect heterogeneity can never be ruled out. In observational studies, we know that omitted-variable biases are pervasive. Hundreds of meta-analyses have confirmed the importance of omitted-variable biases. MAER-Net's reporting guidelines consider it essential to code and accommodate potential omitted-variable biases in all meta-regressions (Stanley *et al.*, 2013). In experimental research, how can a reviewer be confident that observed heterogeneity is induced directly upon the treatment effect? For example, it is common for random variations in subjects' age, gender, health, genome, and so on to affect the observed effect size. When such variations are related to the means of some moderator variable, our simulations show that the unrestricted WLS's statistical properties are much better than random effects'. Stanley and Doucouliagos (2015) show that this same unrestricted WLS approach also works quite well relative to simple conventional meta-analysis (i.e., fixed-effects and random-effects weighted averages), even when there is direct additive heterogeneity in log odds ratios or in the standardized mean differences from randomized controlled trials. In much the same way as demonstrated in this paper, simulations of simple meta-analysis weighted averages across both RCTs and observational studies reveal this same unrestricted WLS estimation approach "clearly dominating random effects when there is publication (or small-sample) bias and . . . confidence intervals are practically equivalent or superior to random effects when there is no publication bias" (Stanley and Doucouliagos, 2015, p. 10).

6. Conclusions

The central contribution of this study is to demonstrate that an unrestricted WLS multiple meta-regression (WLS-MRA) provides a viable and practical alternative to random-effects meta-regression (RE-MRA). When there is no publication selection, reporting, or small-sample biases, these two estimation approaches are practically

[7] *To address an anonymous reviewer's concern, we do not mean to imply that this general WLS-MRA is somehow a 'solution to the publication selection problem.' In a previous paper published by this journal, we recommend MRA models (7) and (8) to 'reduce publication selection bias' (Stanley and Doucouliagos, 2014). RE-MRA and WLS-MRA versions of these PET-PEESE models are simulated in this paper and reported in Tables 9 and 10. In this paper, we show that both approaches succeed in reducing publication bias when it is present with WLS-MRA dominating the RE-MRA version of PET-PEESE-MRA.*

equivalent. More importantly, when there is selection for statistical significance, WLS-MRA always has smaller bias and typically smaller MSE than RE-MRA. Thus, our recommendation is that WLS-MRA should be adopted as the conventional approach for meta-regression of observational research, especially in the social sciences where heterogeneity is routinely indirect and selection for statistical significance is a constant threat.[7]

This study makes further contributions to our understanding of meta-regression. It validates the multiple meta-regression model advanced by Doucouliagos and Stanley (2009) and Stanley and Doucouliagos (2014) and its ability to correct omnipresent misspecification biases and simultaneously accommodate publication bias. Simple binary dummy variables for the presence of possible misspecification biases as moderator variables in an MRA serve as a viable filter for these potential sources of contamination to scientific inference. Suitably constructed MRAs can identify misspecification biases and correct publication selection. The resulting coefficients from the MRA can thus quantify the impact of moderators on treatment effects after correcting for various common biases in the research record. Last, 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 by meta-analysts (Hedges and Olkin, 1985; Lipsey and Wilson, 2001; Konstantopoulos and Hedges, 2004; Johnson and Huedo-Medina, 2012). Even when applied to conditional inferences where all characteristics are exactly the same as current research and it is known that there is no heterogeneity, not dividing by $\sqrt{MS_E}$ (WLS-MRA) produces bias, MSE, and confidence intervals equivalent to those of fixed-effects meta-regression. Although there might be reasons of mathematical aesthetics or theory to prefer FE-MRA over WLS-MRA, there are never practical statistical reasons for doing so. On the other hand, if FE-MRA is misapplied to any other case or should there be unexpected excess heterogeneity, WLS-MRA's confidence intervals are much better. Thus, WLS-MRA is statistically at least as good as fixed-effect meta-regression, *in all cases*, and preferable to fixed effects in the great majority of practical applications.

The statistical properties of unrestricted WLS multiple meta-regression are practically equivalent to both conventional meta-regression approaches (fixed- and random-effects meta-regression) *when these conventional 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, then WLS-MRA has demonstrably better properties. Unfortunately, in practice, neither excess heterogeneity nor publication bias can be ruled out because tests of both excess heterogeneity and publication biases are known to have low power. Thus, caution favors WLS-MRA in practical application. *There is little or nothing of practical consequence to lose but potentially much to gain in employing an unrestricted WLS estimator.*

Nor does the application of this unrestricted WLS estimator limit how a systematic reviewer may interpret her results. The reviewer is free to employ the WLS-MRA estimator, assume that the random-effects model is entirely valid, and then interpret WLS-MRA's estimates in the context of RE-MRA that there is a distribution of 'true' effects. All of our simulations, for which $\sigma_h > 0$, assume that the RE-MRA model is true and that its interpretation is entirely valid. Nonetheless, these same simulations demonstrate that the practical statistical properties of WLS-MRA are as good as or better than RE-MRA when RE-MRA's model is true. Should, instead, the reviewer wish to assume that there is only a single 'true' effect and that the FE-MRA model is correct, she is also free to interpret the estimates from WLS-MRA in this FE-MRA context. Our simulations assume that this FE-MRA model is entirely correct in all of the cases for which $\sigma_h = 0$, and yet WLS-MRA's statistical properties are, on average, identical to FE-MRA. The application of an unrestricted WLS approach does not constrain how a reviewer may choose to interpret her findings. The central purpose of Stanley and Doucouliagos (2015) and the current study is to show that unrestricted WLS estimators are as good as or better than conventional meta-analysis and conventional meta-regression methods, regardless of which conventional model is assumed or might be true. To suggest otherwise is disingenuous.

Nonetheless, a note of further caution is warranted. Although our simulations employ a wide range of the relevant parameters across several pathways of excess heterogeneity, there could be special circumstances that arise for particular areas of research that might alter the relative performance of these meta-regression methods. The real world is likely to contain a wider variety of complications than those we have simulated. The complex interactions of several potential difficulties might affect relative performances. Last, our simulations are based on regression estimates. Although previous research has documented the advantage of this unrestricted WLS estimation approach over random effects in the calculation of simple weighted averages from randomized controlled trials (Stanley and Doucouliagos, 2015), further research is needed to investigate whether a similar finding holds for MRA.

Acknowledgements

We acknowledge support from the Czech Science Foundation (grant 15-02411S).

References

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

Baker RD, Jackson D 2013. Meta-analysis inside and outside of particle physics: two traditions that should converge? *Research Synthesis Methods* **4**: 109–124.

Baker WL, White CM, Cappelleri JC, Kluger J, Coleman CI 2009. Understanding heterogeneity in meta-analysis: the role of meta-regression. *The International Journal of Clinical Practice* **63**: 1426–1434.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR 2009. Introduction to Meta-analysis. Chichester: John Wiley and Sons.

Copas JB, Lozada C 2009. The radial plot in meta-analysis: approximations and applications. *Journal of the Royal Statistical Society, Series C* **58**: 329–344.

Costa-Font J, McGuire A, Stanley TD 2013. Publication selection in health policy research: the winner's curse hypothesis. *Health Policy* **109**: 78–87.

Davidson R, MacKinnon JG 2004. Econometric Theory and Methods. Oxford: Oxford University Press.

De Linde Leonard M, Stanley TD, Doucouliagos H(C) 2014. Does the UK minimum wage reduce employment? A meta-regression analysis. *British Journal of Industrial Relations* **52**: 499–520.

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

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

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

Doucouliagos C(H), Stanley TD, Giles M 2012. Are estimates of the value of a statistical life exaggerated? *Journal of Health Economics* **31**: 197–206.

Doucouliagos C, Stanley TD, Viscusi WK 2014. Publication selection and the income elasticity of the value of a statistical life. *Journal of Health Economics* **33**: 67–75.

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

Gauss CF 1823. Theoria Combinationis Observationum Erroribus Minimis Obnoxiae. G ttingen: Dieterich.

Gerber AS, Green DP, Nickerson D 2001. Testing for publication bias in political science. *Political Analysis* **9**: 385–92.

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

Greene WE 1990. Econometric Analysis. New York: Macmillan.

Hedges LV 1994. Fixed effects models. In Cooper H, Hedges LV (eds.). The Handbook of Research Synthesis(pp. 285–299). New York: Russell Sage.

Hedges LV, Olkin I 1985. Statistical Methods for Meta-analysis. Orlando: Academic Press.

Hedges LV, Vevea JL 1998. Fixed- and random-effects models in meta-analysis. *Psychological Methods* **3**: 486–504.

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

Higgins JPT, Thompson SG 2002. Quantifying heterogeneity in meta-analysis. *Statistics in Medicine* **21**: 1539–1558.

Higgins JPT, Thompson SG 2004. Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine* **23**: 1663–1682.

Hoaglin DC 2015. Misunderstandings of Q and 'Cochrane Q-test' in meta-analysis. *Statistics in Medicine* (wileyonlinelibrary.com) . DOI:10.1002/sim.6632.

Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K 2009. Publication bias in clinical trials due to statistical significance or direction of trial result. *Cochrane Review*. Available at http://www.thecochranelibrary.com .

Ioannidis JPA, Stanley TD, Doucouliagos C 2016. The power of bias in economics research. *The Economic Journal*, forthcoming. Available as SWP, Economics Series 2016/1. Deakin University: Melbourne, Australia. http://www.deakin.edu.au/__data/assets/pdf_file/0007/477763/2016_1.pdf [Accessed February 1, 2016]. .

Jacquez JA, Mather F, Crawford C 1968. Linear regression with non-constant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. *Biometrics* **24**: 607–26.

Johnson BT, Huedo-Medina TB 2012. Depicting estimates using the intercept in meta-regression models: the moving constant technique. *Research Synthesis Methods* **2**: 204–20.

Judge GG, Hill RC, Griffiths W, Lutkepohl H, Lee TC 1982. Introduction to the Theory and Practice of Econometrics. New York: John Wiley and Sons.

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

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

Koetse MJ, Florax RJGM, de Groot HLF 2010. Consequences of effect size heterogeneity on meta-analysis: a Monte Carlo experiment. *Statistical Methods and Applications* **19**: 217–36.

Konstantopoulos S, Hedges LV 2004. Meta-analysis. In Kaplan D (ed.). Quantitative Methodology for the Social Sciences(pp. 281–295). Thousand Oaks: Sage Publications.

Krassoi Peach E, Stanley TD 2009. Efficiency wages, productivity and simultaneity: a meta-regression analysis. *Journal of Labor Research* **30**: 262–8.

Lipsey MW, Wilson DB 2001. Practical Meta-analysis. Thousand Oaks: Sage Publications.

Moreno SG, Sutton AJ, Ades AE, Stanley TD, Abrams KR, Peters JL, Cooper NJ 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.

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

Raudenbush SW 1994. Random effects models. In Cooper H, Hedges LV (eds.). The Handbook of Research Synthesis(pp. 301–321). New York: Russell Sage.

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.

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

Stanley TD 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 TD, Doucouliagos H(C) 2012. Meta-regression Analysis in Economics and Business. Oxford: Routledge.

Stanley TD, Doucouliagos H(C) 2014. Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods* **5**: 60–78.

Stanley TD, Doucouliagos H(C) 2015. Neither fixed nor random: weighted least squares meta-regression analysis. *Statistics in Medicine* **34**: 2116–27.

Stanley TD, Doucouliagos H(C), Giles M, *et al.* 2013. Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys* **27**: 390–94.

Stanley TD, Jarrell SB 1989. Meta-regression analysis: a quantitative method of literature surveys. *Journal of Economic Surveys* **3**: 161–70.

Stanley TD, Jarrell SB, 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 TD, Rosenbaum WL, Weinkam JJ 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 JAC 2009. Meta-Analysis in Stata: An Updated Collection from the *Stata Journal*. College Station, TX: Stata Press.

Stigler SM 1986. The History of Statistics. Cambridge, MA: Belknap Press.

Thompson SG, Sharp SJ 1999. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* **18**: 2693–2708.

Turner RM, Davy J, Clarke MJ, Thompson SG, Higgins JPT 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.

U.S. Dept. of Transportation 2011. Office of the Assistant Secretary for Transportation Policy, Memorandum: Treatment of the Economic Value of Statistical Life in Departmental Analyses—2011 Interim Adjustment. http://www.dot.gov/policy/transportation-policy/treatment-economic-value-statistical-life (accessed 1.05.15).

United States Environmental Protection Agency. 2010. Valuing Mortality Risk Reductions for Environmental Policy: A White Paper. Science Advisory Board—Environmental Economics, Review Draft.

United States Environmental Protection Agency (EPA). 1997. The Benefits and Costs of the Clean Air Act, 1970 to 1990, EPA 410-R-97-002.

Viscusi WK 2011. Policy challenges of the heterogeneity of the value of statistical life. *Foundations and Trends in Microeconomics* **6**: 99–172.

Viscusi WK, Aldy JE 2003. The value of a statistical life: a critical review of market estimates throughout the world. *Journal of Risk and Uncertainty* **27**: 5–76.

White IR 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.

