DOI: 10.1111/joes.12633

JOURNAL OF ECONOMIC SURVEYS WILEY

ARTICLE

Accounting for sample overlap in economics meta-analyses: The generalized-weights method in practice

Pedro R. D. Bom[1] | **Heiko Rachinger**[2]

[1]Deusto Business School, University of Deusto, Bilbao, Spain

[2]Department of Applied Economics, University of Balearic Islands, Palma, Spain

Correspondence
Pedro R. D. Bom, Deusto Business School, University of Deusto, Bilbao, Spain.
Email: pedro.bom@deusto.es

Funding information
Spanish Ministry of Science and Innovation - State Research Agency, Grant/Award Number: PID2020-114646RB-C43; Basque Government Department of Education, Grant/Award Number: IT1429-22; Jesus de Gangoiti Barrera Foundation

Abstract

Meta-analyses in economics frequently exhibit considerable overlap among primary samples. If not addressed, sample overlap leads to efficiency losses and inflated rates of false positives at the meta-analytical level. In previous work, we proposed a generalized-weights (GW) approach to handle sample overlap. This approach effectively approximates the correlation structure between primary estimates using information on sample sizes and overlap degrees in the primary studies. This paper demonstrates the application of the GW method to economics meta-analyses, addressing practical challenges that are likely to be encountered. We account for variations in data aggregation levels, estimation methods, and effect size metrics, among other issues. We derive explicit covariance formulas for different scenarios, evaluate the accuracy of the approximations, and employ Monte Carlo simulations to demonstrate how the method enhances efficiency and restores the false positive rate to its nominal level.

KEYWORDS
data aggregation, generalized weights, inference, meta-analysis, moderators variables, partial correlation coefficients, sample overlap

JEL CLASSIFICATION
C13

1 | INTRODUCTION

Ideally, meta-analysis would integrate empirical outcomes derived from independent samples. However, in economics, where observational research predominates, empirical studies frequently and imperfectly reuse the same datasets. This is especially evident in research areas reliant on aggregate data, such as economic growth or fiscal policy, where multiple studies can use the same time series or panel data repeatedly. For instance, in the meta-analysis by Bom and Ligthart (2014) on the output elasticity of public capital, approximately half of the analyzed estimates from roughly half of the studies were based on US data. While not all primary samples perfectly align over time, the period from 1970 to 1985 is encompassed in the vast majority of them (we will revisit this example in Section 6). We refer to this characteristic of meta-analysis data as sample overlap. Recent notable meta-analyses that exhibit significant sample overlap include Bajzik et al. (2020), Cazachevici et al. (2020), and Gechert et al. (2022), among numerous other possible examples.[1]

In previous work (Bom & Rachinger, 2020; henceforth BR), we extensively investigated the statistical challenges associated with sample overlap in meta-analysis. Through analytical and quantitative analysis, BR demonstrated that sample overlap can introduce asymptotic biases, substantial efficiency losses, and elevated rates of false positives at the meta-analysis level. The root of the problem lies in the treatment of overlapping observations as distinct realizations, when in reality, they represent the same data. Consequently, this leads to artificially precise meta-estimates, resulting in reduced efficiency and potentially distorted inference. To address this issue, BR proposed a two-step solution. The first step is to estimate the complete variance–covariance matrix to capture the correlation structure induced by sample overlap. BR showed that the elements of this matrix can be feasibly estimated using readily available information in the primary studies, such as basic sample descriptions and standard errors. The second step is to use this matrix to appropriately weight the primary outcomes. This procedure, which BR called the generalized-weights (GW) estimator, effectively solves the sample overlap problem by assigning lower weights to estimates derived from overlapping samples.

Unfortunately, BR's GW solution was derived in a simplified framework and is, thus, not readily applicable to many economics meta-analyses. Several complications can arise in the implementation process. First, primary studies often employ overlapping data aggregated at different levels. For instance, one study might use time-series data for a given country, while another study uses a panel dataset for its regions. Second, primary estimates may be derived using different estimation methods, such as ordinary least squares (OLS) or instrumental variables (IV). Third, in economics meta-analyses, regression coefficients are frequently transformed into partial correlation coefficients (PCCs). The first contribution of this paper is to address these intricacies by deriving covariance expressions specific to each case.

In deriving the covariance expressions, we assume that the primary studies report a homogeneous estimate of the slope coefficient of a simple regression model.[2] Our second contribution is then to show, using simulations, that the GW estimator thus constructed can also be applied in a few important extensions. First, they can be applied to slope coefficients of multiple regressions. Second, we introduce systematic heterogeneity in the primary population model and show that the GW estimator can be used to estimate meta-regressions designed to explain such heterogeneity. Our covariance expressions can be extended in these directions because, having been derived for simple homogeneous cases, they tend to overstate the true covariances in more complex heterogeneous cases. Our conservative approach is, thus, simple yet versatile, giving rise to a few relatively simple formulas that can be used in a wide range of overlapping scenarios.

Finally, as our third contribution, we aim to enhance the accessibility of the GW estimator for the wider meta-analysis research community. To achieve this, we have developed Stata and R codes that automate the entire procedure. The codes, along with comprehensive instructions documents, can be downloaded from https://osf.io/g5t2j. The utilization of the codes merely requires the collection of basic information pertaining to the primary samples during the coding stage. This includes details such as the time period covered by the sample, spatial units considered (e.g., countries or regions), estimation methods employed, and other customary coded information like standard errors and moderator variables. Collecting the necessary information to address sample overlap should only require a minor additional effort during the coding stage.

The paper is organized as follows. In Section 2, we present an overview of the sample overlap problem and the GW solution. Section 3 focuses on the challenges that are likely to arise in practice and provides corresponding solutions. The simulation framework and results are outlined in Section 4. In Sections 5 and 6, we detail the implementation of the code and present an application of the method to the meta-sample of output elasticities of public capital from Bom and Ligthart (2014). Section 7 offers a general discussion, while Section 8 concludes the paper with final remarks. Detailed technical derivations can be found in the Web Appendix.

2 | SAMPLE OVERLAP

2.1 | The problem

Consider a simplified meta-analysis scenario, where the focus is on the relationship between two variables, denoted as x and y. Assume that multiple observational studies employ OLS on a simple regression of y on x using samples of potentially varying sizes. Let the true population slope of this regression be denoted by θ and assume it is common (i.e., homogeneous) across all studies. Suppose the classical linear regression model's standard assumptions hold, ensuring that the OLS estimates are unbiased with valid standard errors. Additionally, assume that all studies report their findings regardless of the direction or magnitude (i.e., no publication selection). Although the assumptions of homogeneity and no publication bias may not hold in practice, they serve to illustrate our main point in a simplified manner.

In this simple setting, the meta-analysis model can be represented as

$$\hat{\theta}_i = \theta + \epsilon_i, \tag{1}$$

where $\hat{\theta}_i$ represents the reported estimate of θ in study i, and ϵ_i denotes the sampling error component. The sampling error is approximately normally distributed, with a zero mean and a

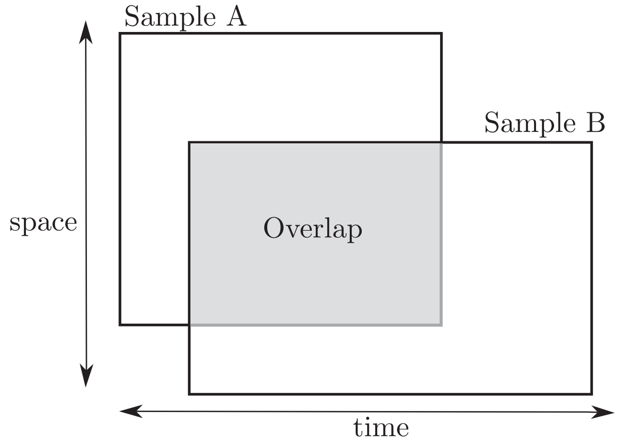

FIGURE 1 Sample overlap over space and time.

heteroskedastic variance $\sigma^2_{\theta_i}$, which is inversely proportional to the primary sample size. When the primary samples do not overlap, the fixed effects (FE) estimator, which is a weighted average of the reported estimates using inverse-variance weights $1/\sigma^2_{\theta_i}$, becomes the optimal meta-analysis estimator for θ. Technically, the sampling error term ϵ_i would be uncorrelated across studies and only its heteroskedastic nature would have to be addressed.[3]

Let us now consider a meta-analysis scenario where certain observations appear repeatedly in multiple primary samples. We refer to this phenomenon as sample overlap. Figure 1 provides an illustration with two primary samples, labeled as A and B, measured across space and time for some relevant variable. Sample overlap is represented by the shaded area, which corresponds to the observations included in both samples. Sample overlap is commonly observed in economics, particularly in research fields that heavily rely on aggregated data, such as time series or panel data at the country level. In such cases, a single dataset may be used in multiple studies, leading to substantial sample overlap. BR provide a nonexhaustive list of examples of meta-analyses where sample overlap is prevalent. Additionally, sample overlap can also occur, albeit potentially in a more balanced manner, when several estimates are reported and collected from the same study using the same dataset, as is often encountered in meta-analysis.

In the presence of sample overlap, the sampling error term ϵ_i is no longer uncorrelated. BR show that this overlap-induced correlation causes meta-estimation inefficiencies and high rates of false positives in standard estimators. This stems from the double-counting of overlapping observations. When sample overlap exists but is disregarded in model (1), two overlapping primary observations, essentially representing the same underlying realization, receive twice the weight of a nonoverlapping observation, despite providing equivalent information. This discrepancy in weights leads to meta-estimation inefficiencies. Furthermore, it decreases the variability of the primary estimates, thereby artificially lowering the standard error of the meta-estimate. As a consequence of this spurious precision, the probability of erroneously rejecting the null hypothesis of no effect increases above its nominal level.

2.2 | The GW solution

BR propose a simple yet flexible solution to sample overlap. They demonstrate that the covariances between overlapping estimates that are induced by sample overlap can be estimated using

TABLE 1 Covariance formulas for the various data aggregation cases and estimation methods.

Type of data aggregation	Both OLS or IV (1)	p = OLS, q = IV (2)	p = IV, q = OLS (3)
(1) None	$C_{qp}\sqrt{\frac{\text{var}(\epsilon_p)}{N_q}}\sqrt{\frac{\text{var}(\epsilon_q)}{N_p}}$	$C_{qp}\frac{\text{var}(\epsilon_p)}{N_q}$	$C_{qp}\frac{\text{var}(\epsilon_q)}{N_p}$
(2) Temporal[a]			
(3) Spatial[a]	$C_{qp}\frac{K}{G}\sqrt{\frac{\text{var}(\epsilon_p)}{N_q}}\sqrt{\frac{\text{var}(\epsilon_q)}{N_p}}$	$C_{qp}\frac{K}{G}\frac{\text{var}(\epsilon_p)}{N_q}$	$C_{qp}\frac{K}{G}\frac{\text{var}(\epsilon_q)}{N_p}$
(4) Double[a]			
(5) Co-aggregation[b]	$C_{qp}\frac{K}{TG}\sqrt{\frac{\text{var}(\epsilon_p)}{N_q}}\sqrt{\frac{\text{var}(\epsilon_q)}{N_p}}$	$C_{qp}\frac{K}{TG}\frac{\text{var}(\epsilon_p)}{N_q}$	$C_{qp}\frac{K}{TG}\frac{\text{var}(\epsilon_q)}{N_p}$

Note: C_{pq} = number of observations in the dataset of p that overlap with the dataset of q. N_p = sample size of p. N_q= sample size of q. $\text{var}(\epsilon_p)$ = estimated variance of estimate p. $\text{var}(\epsilon_q)$ = estimated variance of estimate q. K = number of overlapping spatial units in the spatially disaggregated dataset. G = number of spatial units aggregated (spatial aggregation factor). T = number of time periods aggregated (temporal aggregation factor). For detailed derivations of these formulas, see the Web Appendix.

[a]The subscript p refers to the estimate from the disaggregated dataset and the subscript q refers to the estimate from the aggregated dataset.

[b]The subscript p refers to the estimate from the spatially disaggregated dataset, and the subscript q refers to the estimate from the temporally disaggregated dataset.

information typically reported in primary studies, such as sample sizes, the number of overlapping observations, and standard errors. For instance, let us consider two estimates, denoted as A and B, with corresponding standard errors se(A) and se(B), derived from primary samples of sizes N_A and N_B that share C observations. The covariance between A and B can be estimated as $C \times \frac{\text{se}(A)}{\sqrt{N_B}} \times \frac{\text{se}(B)}{\sqrt{N_A}}$ (see Table 1). Note that this formula corresponds to Proposition 2 in BR, after using the observed standard errors to estimate the unobserved ratio of variances σ_u^2/σ_x^2 (see the Web Appendix for details). By applying this formula to each pair of estimates, one can obtain the variance–covariance matrix of the error term ϵ, which describes the entire correlation structure among primary estimates resulting from sample overlap. The diagonal elements of this matrix represent the variances of the estimates (i.e., the squared reported standard errors), as assumed by the FE/WLS estimator. However, they can also incorporate a heterogeneity component, similar to the Random Effects (RE) estimator (see below).

By using the estimated variance–covariance matrix in a generalized least squares (GLS) manner, we derive the GW estimator. Let $\hat{\theta}$ represent the vector of primary estimates and Ω denote the estimated variance–covariance matrix. The GW estimator of θ is obtained as $\tilde{\theta}_{GW} = (\iota'\Omega^{-1}\iota)^{-1}\iota'\Omega^{-1}\hat{\theta}$, where ι is a vector of ones. The variance of the meta-estimate, $\tilde{\theta}_{GW}$, is given by $\text{Var}(\tilde{\theta}_{GW}) = (\iota'\Omega^{-1}\iota)^{-1}$. Essentially, the GW estimator assigns weights to the primary estimates based on their respective contributions of nonoverlapping sampling information, diminishing the importance of primary observations that appear in multiple samples. Hence, the GW estimator assigns greater weight to estimates derived from larger samples, as they exhibit smaller variances in the main diagonal of Ω; and it assigns more weight to estimates derived from samples with less overlap with other samples, as they have smaller covariances off the main diagonal of Ω.

The GW estimator can handle any type of sample overlap, whether within or between studies, as long as the essential information needed to estimate the covariances between overlapping estimates, such as sample sizes, standard errors, and the number of overlapping observations, can be collected from the primary studies. This information is typically available in most applications. However, it is important to note that the GW estimator, as presented in BR, is a simplified approach and does not address several potential complications that may arise in practical settings.

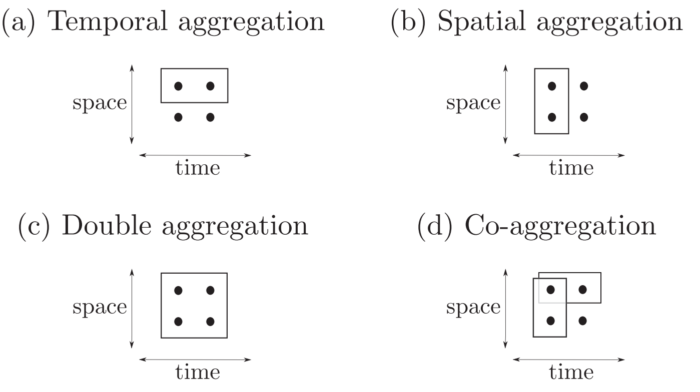

FIGURE 2 The four cases of aggregation overlap. (a) Temporal aggregation. (b) Spatial aggregation. (c) Double aggregation. (d) Co-aggregation.

The subsequent section addresses and provides analytical solutions for three such complications: (1) primary data aggregated at different levels, (2) different estimation methods, and (3) standardized regression coefficients. Furthermore, we examine the performance of the GW estimator in two significant extensions: (1) when the regressions are obtained from multiple regression models, and (2) when the meta-analysis model is expanded to incorporate a meta-regression that accounts for systematic heterogeneity.

3 | ISSUES AND SOLUTIONS

This section addresses three significant issues commonly encountered in economics meta-analysis: data aggregation, utilization of different estimation methods, and conversion to PCCs. We derive precise expressions for the covariance between overlapping estimates in each of these scenarios. In Section 4, we show how these formulas can also be used to handle multiple regressions and meta-regressions.

3.1 | Different levels of data aggregation

Primary studies often use overlapping data aggregated at different levels. For example, one study may employ an annual time-series dataset while another study uses a quarterly time-series dataset, in which case, the annual data aggregate the quarterly data. Similarly, one study may focus on country-level data while another study uses overlapping data from multiple regions within the same country, in which case the country-level data aggregate the regional-level data. In panel data settings, more generally, one dataset may aggregate another in terms of space, time, or both, although it can also occur that two samples aggregate each other in terms of space and time. Section 6 presents specific examples of primary studies included in Bom and Ligthart's (2014) meta-analysis that employ data aggregated at different levels.

We specifically examine four types of data aggregation, as depicted in Figure 2:

1. Temporal aggregation: one observation from one dataset aggregates the observations of two or more time periods from another dataset.
2. Spatial aggregation: one spatial unit in a dataset aggregates two or more spatial units from another dataset.

3. Double aggregation: one observation in one dataset aggregates two or more observations in another dataset across both space and time.
4. Co-aggregation: two datasets mutually aggregate each other along the spatial and temporal dimensions.

Let us first consider the scenario of temporal aggregation, where each observation in one dataset aggregates T observations in the other dataset. We denote T as the temporal aggregation factor. For instance, if one sample comprises annual observations and the other sample consists of quarterly observations, then $T = 4$. The covariance between overlapping estimates for this type of data aggregation is provided in row (2) and column (1) of Table 1 (for a complete derivation, please refer to the Web Appendix). It shares the same expression as the case without aggregation except that the number of overlapping observations between samples (C_{qp}) pertains necessarily to the aggregated sample.

In the case of spatial aggregation, an observation in one dataset aggregates G spatial units in another dataset. We denote G as the spatial aggregation factor. For instance, let us consider two datasets: one is a time-series dataset representing an entire country, and the other is a panel dataset consisting of observations from 10 regions within that country. Clearly, each country-level observation aggregates the data from its ten regions, resulting in $G = 10$. It is important to note that the spatially disaggregated dataset may contain information for only K spatial units, where $K \leq G$. For example, in the given sample, if the regional panel dataset includes data for only five regions, then $K = 5$ (which is less than $G = 10$). The covariance between overlapping estimates for this type of data aggregation is provided in row (3) and column (1) of Table 1.

Spatial and temporal aggregation can occur simultaneously, resulting in what we term as double aggregation. In its simplest form, one sample aggregates the other dataset both spatially and temporally. Consider, for example, a scenario where there are two samples available for the same country: a time series dataset consisting of annual observations and a panel dataset comprising quarterly observations for some regions within the country. In this case, the time series data aggregate the panel data across both space and time. As displayed in Table 1, the covariance expression for the case of double aggregation is the same as that for spatial aggregation, while considering the possibility that $K \leq G$.

Let us now consider the case of co-aggregation, where the samples aggregate each other along both spatial and temporal dimensions. In this scenario, one sample exhibits greater disaggregation along the spatial dimension, while the other sample demonstrates more disaggregation along the temporal dimension. For instance, we can consider a time series dataset consisting of quarterly observations for a specific country and an annual panel dataset representing various regions within that country. Here, the time series dataset is more disaggregated along the temporal dimension, while the panel dataset is more disaggregated along the spatial dimension. The covariance between overlapping estimates in this co-aggregation case is determined by the expression provided in row (4) and column (1) of Table 1.

To evaluate the accuracy of our covariance estimates, we perform a simple Monte Carlo simulation exercise (refer to Appendix A for more details). The simulation involves generating two sets of OLS estimates for the slope coefficient of a simple regression model using simulated overlapping data. This process is repeated one million times to ensure a good approximation of the true covariance between the two estimates. Subsequently, we compare the estimates derived from our covariance formulas with the true covariance obtained from the simulation results.

Figure 3 presents the distribution of the ratio between our covariance estimates and the true covariance for the different cases of data aggregation discussed earlier. The accuracy of the esti-

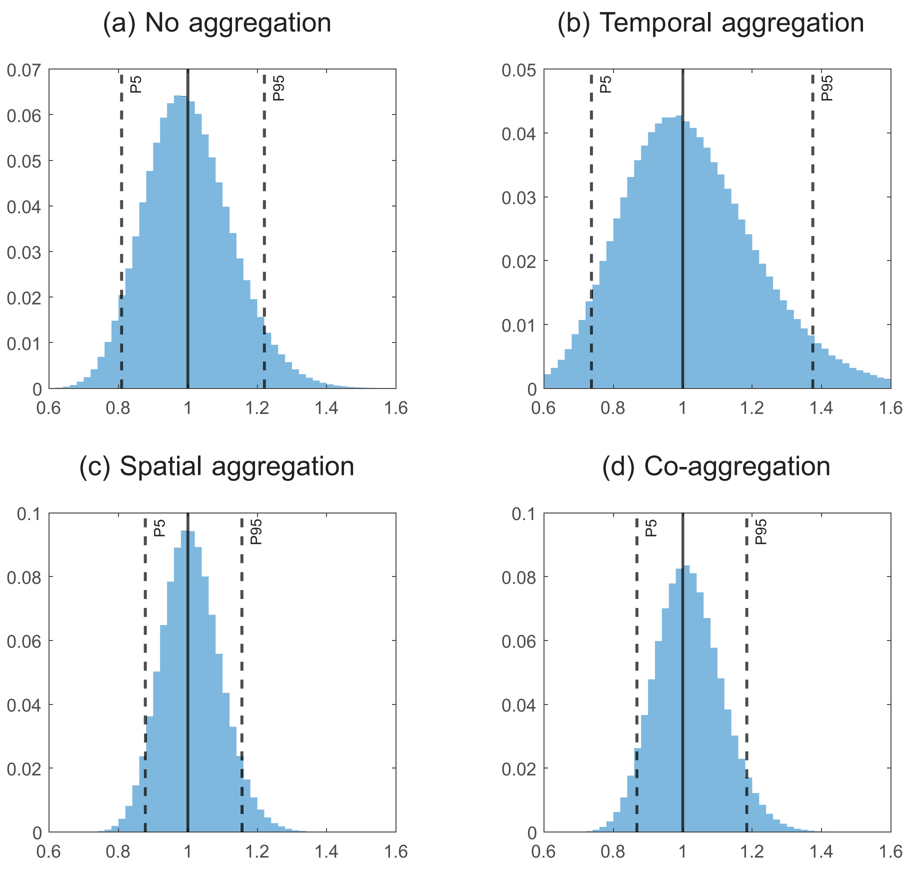

FIGURE 3 Covariance approximations: OLS estimates, various aggregation cases. (a) No aggregation. (b) Temporal aggregation. (c) Spatial aggregation. (d) Co-aggregation. The figures display the empirical distribution of the ratio of estimated covariances (using the formulas displayed in Table 1) to the true covariance. The dashed vertical lines labeled "P5" and "P95" indicate the 5th and 95th percentiles, respectively. For further simulation details, see Appendix A. [Colour figure can be viewed at wileyonlinelibrary.com]

mates is indicated by how close this ratio is to one. Our findings reveal that the distribution of values is roughly symmetric around one, indicating that the covariance estimates are reasonably unbiased to the true covariance. In the absence of aggregation, approximately 90% of the estimates deviate from the true covariance by less than 20%. However, when temporal aggregation is present, resulting in a smaller size for the aggregated sample, the covariance estimates exhibit slightly more dispersion. On the other hand, in cases of spatial or co-aggregation, where both spatial and temporal dimensions are involved, the covariance estimates are naturally more precise.

3.2 | Different estimation methods

The second issue pertains to the utilization of different estimation methods in primary studies. While OLS is the predominant method for estimating linear regression models, IV methods are also commonly employed. The primary motivation behind employing IV methods is to address endogeneity concerns, such as those arising from reverse causality or measurement error. Various approaches exist for implementing IV methods, ranging from two-stage least squares (2SLS) to the generalized method of moments (GMM).[4] The underlying principle of these IV methods is to use the exogenous variation in one or more IV to mitigate the endogeneity problem.

The utilization of IV introduces new sources of variability, which reduces the correlation between overlapping estimates. The extent of this reduction in correlation depends on the number and diversity of instruments employed. Due to the potentially large number of instruments, it becomes impractical to derive covariance formulas encompassing general IV estimates. We adopt instead a conservative approach. We derive a single covariance formula under the assumption that all IV primary studies use the same instrument. This approach provides an upper bound estimate for the covariance, potentially overestimating the covariance when IV studies employ different or multiple instruments. By assigning less weight to such estimates relative to their sample sizes, this procedure ensures that standard errors and *p*-values of significance tests are not artificially small.

Table 1 provides the covariance expressions for overlapping IV estimates in the various data aggregation scenarios discussed in the previous section. As shown in column (1), the covariance between two overlapping IV estimates can be estimated using the same formula as that for two overlapping OLS estimates. However, when combining an OLS estimate and an IV estimate, the covariance is expressed differently, as shown in columns (2) and (3). Notably, these expressions solely involve the estimated variance of the OLS estimate. This is because, as derived in Section A.3.2 of the Web Appendix (specifically, Proposition 4), the overlapping covariance between an OLS estimate and an IV estimate is determined by the overlapping variance of the OLS estimate.

The accuracy of the covariance approximations involving OLS and IV estimates is depicted in Figure 4, using the same simulation design employed in the previous section (see Appendix A for details). The figure displays the ratio of the covariance estimates for two OLS estimates in panel (a), for two IV estimates in panel (b), and for one OLS estimate and one IV estimate in panels (c) and (d), all relative to the true covariance. For brevity, we present here only the results for the case of co-aggregation; the remaining cases of data aggregation can be found in the Web Appendix. As depicted in Figure 4, the distributions of the ratios are approximately symmetrical and centered around one in all cases. This indicates that the covariance estimates are reasonably unbiased with respect to the true covariance. Moreover, the covariance estimates demonstrate considerable accuracy, with approximately 90% of the estimates falling within an error band of around 20% in panels (a)–(c). In panel (d), where the IV estimate represents the spatially disaggregated one, the error band is slightly wider.

While the covariance expression using solely the variance of the OLS estimate provides a roughly unbiased estimator for the covariance between OLS and IV estimates, as demonstrated by our simulations, an important comment is in order here. In certain practical cases, as is the case in Bom and Ligthart's (2014) meta-analysis data (see Section 6), the reported IV standard error may actually be smaller than the OLS standard error, given the respective sample sizes. Several factors can contribute to this situation, such as reporting errors, variations in specifications, or disparities in data definitions. Hence, using only the OLS standard error in the covariance expression raises the possibility that the covariance exceeds the product of the standard errors, a nonsensical outcome that would imply a correlation coefficient larger than one.[5] This could render the variance–covariance matrix nonpositive definite, thereby presenting numerical challenges in the implementation of the method.

Should this issue arise in practical applications, we propose a simple and conservative solution. Instead of using the OLS-IV covariance formulas presented in columns (2) and (3) of Table 1, one

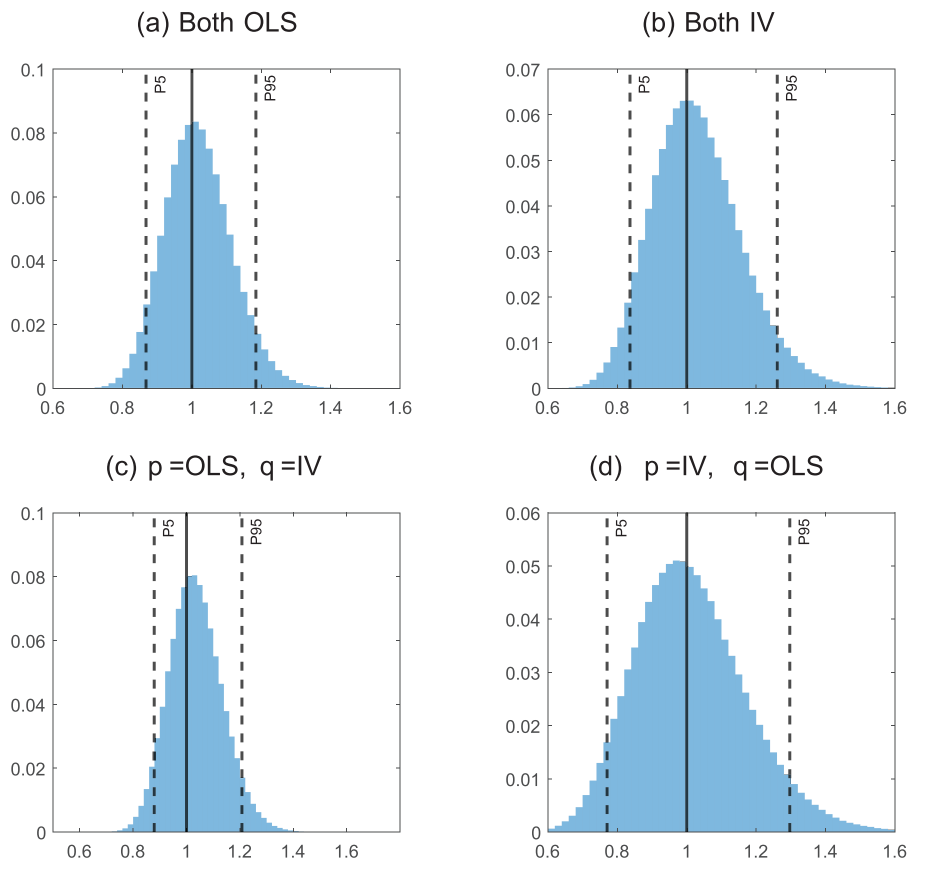

FIGURE 4 Covariance approximations: IV and OLS estimates, co-aggregation. (a) Both OLS. (b) Both IV. (c) p = OLS, q = IV. (d) p = IV, q = OLS. The figures display the empirical distribution of the ratio of estimated covariances (using the formulas displayed in Table 1) to the true covariance. Panel (c) refers to the formula in row (5) and column (2); panel (d) refers to the formula in row (5) and column (3). The dashed vertical lines labeled "P5" and "P95" indicate the 5th and 95th percentiles, respectively. For further simulation details, see Appendix A. [Colour figure can be viewed at wileyonlinelibrary.com]

can employ the covariance formula provided in column (1), assuming two OLS estimates. This approach effectively assumes that the correlation coefficient between an OLS and an IV estimate is equal to that between two OLS estimates. As the former correlation coefficient is smaller than the latter, this solution takes a conservative stance. Moreover, it remains feasible as it guarantees the positive definiteness of the variance–covariance matrix (except when there are cases of perfect overlap, as discussed in see Section 7). While implementing this solution may incur some efficiency cost, it effectively safeguards the rate of false rejections at the nominal level. This solution is implemented in the application presented in Section 6.

3.3 | Standardized effect sizes

Regression coefficients are often transformed into partial correlation coefficients (PCCs) using the formula $PCC = t/\sqrt{t^2 + df}$, where t represents the t-ratio, and df stands for the degrees of freedom of the regression coefficient (cf. Stanley & Doucouliagos, 2012, p.25). This common procedure in economics meta-analyses is often necessary to ensure comparability of regression coefficients across studies (for recent applications, see Gechert & Heimberger, 2022; Heimberger, 2022). In the context of simple regressions, PCCs are equivalent to the so-called beta (or standardized) coefficients, which represent the regression coefficients of the model with standardized variables (i.e., variables previously divided by their standard deviations). For multiple regressions, PCCs can be interpreted as standardized coefficients of a simple regression model after controlling for the effects of all other variables.

To determine the covariance between overlapping PCCs, we adopt a straightforward approach. First, we recognize that PCCs are linear transformations of the underlying regression coefficients. This is evident in simple regressions, and it extends to multiple regressions as well (see Section A.4.9 of the Web Appendix). Second, we take into account that correlation coefficients are invariant to linear transformations of the variables. Consequently, the population correlation coefficient between overlapping PCCs should be equal to that between overlapping regression coefficients. Leveraging this property, we can employ the covariance formulas for regression coefficients to calculate the covariance between overlapping PCCs, after adjusting for the change in the scale of the variables (see the Web Appendix for details).

It is important to acknowledge that the equality between correlation coefficients for overlapping regression coefficients and PCCs does not hold exactly in the samples. The reason is that the ratio of standard deviations used to convert regression coefficients into PCCs must be estimated, leading to variations across studies. The linear transformation of regression coefficients performed by PCCs is, thus, study-specific. Consequently, the correlation coefficient between PCCs is, on average, lower than that between the underlying regression coefficients. Hence, our approach remains conservative, tending to overstate the true covariance between overlapping PCCs.

Figure 5 presents the simulation results regarding the accuracy of covariance estimates between overlapping PCCs converted from OLS estimates (the results for PCCs converted from IV estimates can be found in the Web Appendix). The distributions of estimated covariances, expressed as the ratio to the true covariance, display peaks around values between 1.1 and 1.2. This indicates that the estimated covariances tend to overstate the true covariance by approximately 10%–20%. Notably, the estimated covariances exhibit substantially less dispersion for PCCs than for regression coefficients, reflecting the bounded nature of PCCs.

3.4 | Other relevant issues

In the following sections, we address a few additional issues that are likely to arise in practical applications of the GW estimator. We demonstrate in Section 4, using Monte Carlo simulations, how the previously derived covariance expressions can be used to address these issues too.

3.4.1 | Multiple regression models

In most cases, the coefficients subject to meta-analysis are the slope parameters of multiple regression models, where the effects of multiple variables are controlled. The covariance expressions derived earlier for simple regression models can also be applied to multiple regression models, as demonstrated in Section 4.

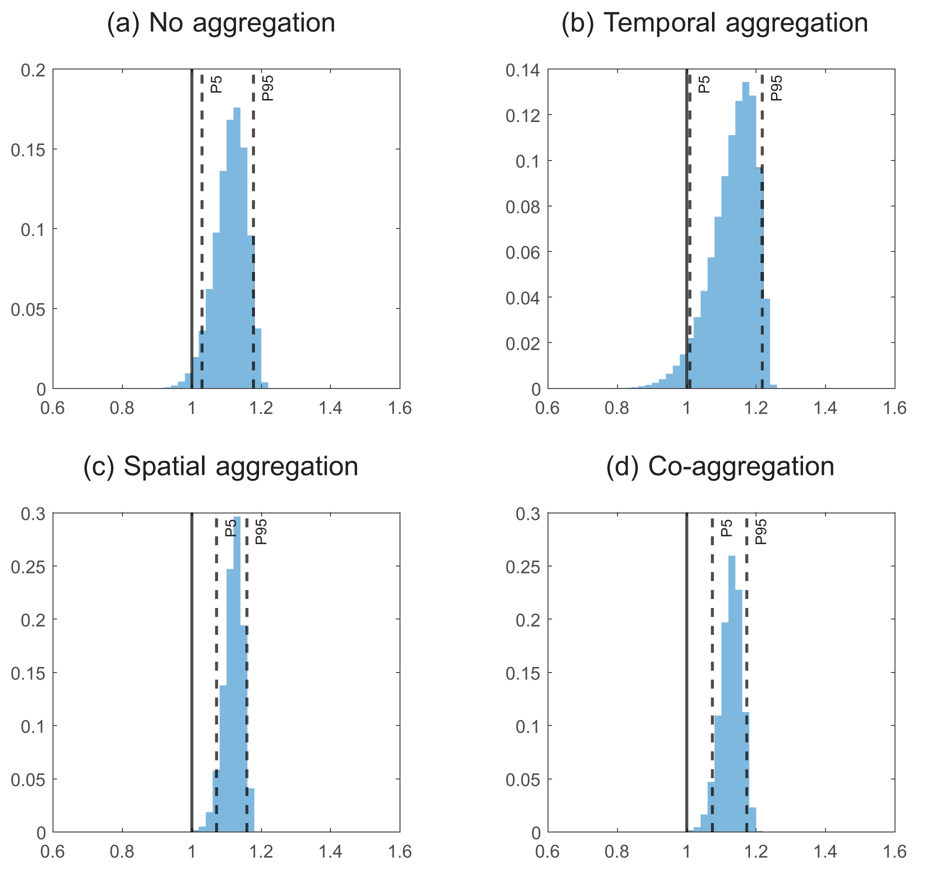

FIGURE 5 Covariance approximations: PCCs, various data aggregation cases. (a) No aggregation. (b) Temporal aggregation. (c) Spatial aggregation. (d) Co-aggregation. The figures display the empirical distribution of the ratio of estimated covariances (using the formulas displayed in Table 1, converted into PCCs as described in the Web Appendix) to the true covariance. The dashed vertical lines labeled "P5" and "P95" indicate the 5th and 95th percentiles, respectively. For further simulation details, see Appendix A. [Colour figure can be viewed at wileyonlinelibrary.com]

3.4.2 | Random heterogeneity

Applied meta-analyses often encounter substantial heterogeneity, meaning that estimates vary significantly beyond what can be attributed to sampling error alone. In some cases, this excess variation is considered random and treated as unobserved, leading to RE models. As shown by BR, the GW method can effectively handle such heterogeneity by incorporating a fixed variance component to the error term's variance–covariance matrix. This component can be estimated in various ways (see Langan et al., 2017) and helps avoid singularity issues in situations with extreme sample overlap.

3.4.3 | Meta-regressions

Systematic heterogeneity across studies are usually handled in meta-analysis using meta-regressions, which are regression models using estimates as dependent variables and study design variables (usually represented as binary moderator variables) as independent variables. The GW method is also suitable for meta-regressions, as shown through simulations in Section 4.

4 | SIMULATIONS

We evaluate the statistical properties of the GW estimator through Monte Carlo simulations (see Appendix B for details). The design of the simulations follow closely BR. Essentially, we simulate a meta-analysis scenario with M primary studies reporting estimates and standard errors of the slope coefficient (θ) in a regression model of y on x (later extended to multiple regression). We introduce random heterogeneity by assuming that θ follows a normal distribution with mean $\bar{\theta} = 0$ and variance $\sigma_\theta^2 = 0.04$. The value of M is set to 32, 128, or 512, representing small, medium, and large meta-analyses, respectively. For the primary sample sizes (N), we consider four values—60, 120, 180, and 240—equally allocated to the M primary studies. To induce sample overlap, we introduce a parameter λ, referred to as the extensive-margin degree of overlap, denoting the fraction of the M primary samples that contain some observations in common. We vary λ with values 1/8, 1/4, and 1/2. In turn, the intensity of sample overlap is governed by parameter ρ, called the intensive-margin degree of overlap. In particular, overlapping samples have ρN observations in common, where ρ takes on the values 0.1, 0.3, and 0.5. Higher values of λ or ρ correspond to a greater degree of sample overlap. BR empirically justify these parameter values based on Bom and Ligthart's (2014) data (pp.821–822).

We compare two estimators, the GW estimator and the RE estimator. The RE estimator would be the natural estimator to use in this environment, if sample overlap were ignored, in view of the heterogeneous nature of θ. The GW estimator extends the RE estimator by incorporating the nonzero off-diagonal entries in the variance–covariance matrix caused by sample overlap. To evaluate the performance of the estimators, we use two statistical metrics: (1) the effective rate of false rejections of a t-test for $\bar{\theta} = 0$, which assesses the size distortions induced by sample overlap by examining if the effective false rejection rate deviates from the nominal 5% rate; and (2) the mean squared error (MSE), which measures the efficiency gains achieved at the meta-analysis level when accounting for sample overlap. We do not consider bias as a statistical metric since we assume the primary estimates are unbiased in our baseline simulations, meaning any weighted average of them, including RE or GW, is also unbiased. In our extensions, we account for cases where some primary estimates may be biased, but these can be addressed in a meta-regression model.

We first consider the issues discussed in the previous section, namely data aggregation, the use of OLS/IV methods, and conversion to PCCs, always assuming a simple regression model of y on x as the data generating process at the primary level. Tables 2 and 3 present the results for the rate of false rejections and MSE, respectively, for the case of overlapping between co-aggregated estimates (panels (a) and (b)), overlapping between OLS and IV estimates (panels (c) and (d)), and conversion to PCCs (panels (e) and (f)). The findings in Table 2 indicate that, in all cases, the RE estimator becomes increasingly oversized as the degree of overlap (larger values of ρ or λ) increases, especially for large meta-samples. However, the GW estimator effectively restores the false rejection rate close to its nominal level. Moreover, Table 3 shows that accounting for sample overlap can reduce the MSE substantially, between twofold and fivefold in scenarios of large meta-samples with considerable overlap.

Next, we examine the performance of the GW estimator in handling sample overlap in two important extensions of our baseline framework. First, we consider a multiple regression model,

TABLE 2 Sizes of a t-test for the RE and GW estimators: data aggregation, different estimation methods, and conversion to PCCs.

	$M = 32$			$M = 128$			$M = 512$		
$\rho \backslash \lambda$	0.125	0.25	0.5	0.125	0.25	0.5	0.125	0.25	0.5
(a) Co-aggregation, RE estimator									
0.1	5.63	5.88	6.61	5.21	5.45	7.33	5.72	7.35	13.70
0.3	6.07	6.18	7.18	5.22	6.93	12.37	6.50	12.22	27.55
0.5	6.16	6.38	8.48	5.68	8.19	15.76	7.73	16.08	36.51
(b) Co-aggregation, GW estimator									
0.1	5.36	5.56	5.79	5.09	4.94	5.28	5.28	5.36	5.64
0.3	5.39	5.75	5.86	4.89	5.68	5.82	5.43	5.39	5.88
0.5	5.81	5.67	6.25	5.08	5.51	5.68	5.23	5.14	5.94
(c) OLS-IV estimates, RE estimator									
0.1	5.97	5.98	6.91	5.67	6.11	11.06	6.65	10.72	24.08
0.3	6.35	6.45	9.55	5.96	9.17	20.16	8.73	19.45	44.06
0.5	6.38	6.99	12.64	6.81	11.87	27.44	11.16	27.13	52.99
(d) OLS-IV estimates, GW estimator									
0.1	5.30	5.15	4.52	4.77	4.47	4.63	4.80	4.72	4.91
0.3	5.61	5.10	4.58	4.62	4.84	4.95	4.54	5.00	6.00
0.5	5.56	5.01	4.90	4.95	5.06	4.59	4.69	5.10	5.91
(e) PCCs, RE estimator									
0.1	5.28	5.33	6.63	4.92	5.54	10.30	5.24	10.45	24.69
0.3	5.34	6.31	9.44	5.88	8.23	20.27	8.81	20.44	45.03
0.5	5.19	6.69	11.49	6.57	11.24	29.01	11.47	27.04	54.96
(f) PCCs, GW estimator									
0.1	5.63	5.43	5.60	5.18	4.90	5.49	4.78	5.47	6.01
0.3	5.58	5.83	6.24	5.68	5.38	6.42	5.86	5.51	5.98
0.5	5.38	5.59	5.67	5.46	5.44	5.97	5.27	5.31	6.44

Note: The reported values refer to the rates of false rejections of a t-test that the average effect is zero at the 5% significance level. M denotes the size of the meta-sample, λ denotes the fraction of overlapping estimates (extensive-margin overlap), and ρ denotes the average correlation between overlapping estimates (intensive-margin overlap). In panels (a) and (b), half of the overlapping estimates are aggregated over space and half are aggregated over time. In panels (c) and (d), half of the overlapping estimates are based on OLS and half are based on IV.

where θ is one of several slope coefficients. Specifically, we investigate multiple regression models with two and six independent variables. The results, presented in panels (a)–(d) of Table 4, are consistent with the case of a simple regression model: sample overlap leads to a significant increase in the rate of false rejections for the RE estimator, while the GW estimator effectively brings it back close to the nominal level of 5%. Additionally, we observe similar improvements in the MSE when using the GW estimator, as shown in panels (a)–(d) of Table 5. To offer intuition for this extension, note that the Frisch–Waugh–Lovell theorem ensures multiple regressions can be reduced to simple regressions by removing the effects of the remaining variables.

Second, we introduce observed heterogeneity in the reported estimates of θ by incorporating biases that are addressed at the meta-analysis level using a meta-regression. Specifically, we assume a multiple regression model with two independent variables: x and an additional control

TABLE 3 Mean squared errors (×1000) for the RE and GW estimators: data aggregation, different estimation methods, and conversion to PCCs.

	$M = 32$			$M = 128$			$M = 512$		
$\rho \setminus \lambda$	**0.125**	**0.25**	**0.5**	**0.125**	**0.25**	**0.5**	**0.125**	**0.25**	**0.5**
(a) Co-aggregation, RE estimator									
0.1	1.450	1.432	1.529	0.362	0.365	0.429	0.094	0.106	0.155
0.3	1.426	1.464	1.598	0.361	0.408	0.560	0.101	0.139	0.294
0.5	1.457	1.475	1.745	0.377	0.441	0.683	0.110	0.177	0.429
(b) Co-aggregation, GW estimator									
0.1	1.450	1.431	1.526	0.361	0.363	0.417	0.093	0.099	0.125
0.3	1.426	1.463	1.585	0.359	0.396	0.491	0.096	0.107	0.151
0.5	1.455	1.471	1.700	0.372	0.411	0.521	0.098	0.111	0.157
(c) OLS-IV estimates, RE estimator									
0.1	1.750	1.787	1.882	0.455	0.476	0.639	0.122	0.157	0.303
0.3	1.820	1.848	2.257	0.460	0.576	1.001	0.140	0.248	0.670
0.5	1.812	1.929	2.708	0.493	0.667	1.374	0.165	0.339	1.064
(d) OLS-IV estimates, GW estimator									
0.1	1.750	1.780	1.860	0.452	0.466	0.586	0.116	0.130	0.184
0.3	1.818	1.822	2.167	0.453	0.514	0.702	0.117	0.144	0.213
0.5	1.809	1.893	2.414	0.473	0.539	0.739	0.123	0.143	0.226
(e) PCCs, RE estimator									
0.1	1.420	1.403	1.558	0.355	0.388	0.529	0.097	0.134	0.273
0.3	1.436	1.530	1.920	0.386	0.477	0.880	0.122	0.221	0.607
0.5	1.426	1.600	2.202	0.406	0.559	1.217	0.142	0.297	0.964
(f) PCCs, GW estimator									
0.1	1.419	1.401	1.549	0.354	0.380	0.473	0.092	0.108	0.154
0.3	1.435	1.520	1.839	0.376	0.416	0.592	0.100	0.114	0.174
0.5	1.422	1.558	1.966	0.381	0.436	0.627	0.099	0.115	0.182

Note: M denotes the size of the meta-sample, λ denotes the fraction of overlapping estimates (extensive-margin overlap), and ρ denotes the average correlation between overlapping estimates (intensive-margin overlap). In panels (a) and (b), half of the overlapping estimates are aggregated over space and half are aggregated over time. In panels (c) and (d), half of the overlapping estimates are based on OLS and half are based on IV.

variable that is correlated with x. We divide the studies into two groups. In equal proportions, one group estimates the correct model, while the other group omits the control variable. Each group thus estimates a fundamentally different parameter. At the meta-analysis level, we employ a meta-regression that includes a moderator variable to identify the omission of the control variable in the second group of studies. We focus on the constant term of the meta-regression as the estimate of interest. Panels (e) and (f) of Tables 4 and 5 present the results for this exercise. We further extend this analysis by considering five moderator variables accounting for five different biases, as shown in panels (g) and (h). The results in both cases exhibit similar characteristics: the GW estimator corrects the size distortions observed with the RE estimator and improves the MSE, consistent with our previous findings.

TABLE 4 Sizes of a t-test for the RE and GW estimators: multiple regressions and meta-regressions.

	$M = 32$			$M = 128$			$M = 512$		
$\rho \setminus \lambda$	0.125	0.25	0.5	0.125	0.25	0.5	0.125	0.25	0.5
(a) Multiple regression with two variables, RE estimator									
0.1	5.36	6.21	6.94	5.22	6.22	10.67	6.47	10.13	24.43
0.3	5.77	6.75	9.32	6.29	9.49	21.13	9.46	20.85	44.88
0.5	5.90	7.18	11.76	6.50	11.14	28.27	11.86	28.47	54.10
(b) Multiple regression with two variables, GW estimator									
0.1	5.11	5.66	5.30	4.83	4.76	4.75	5.05	4.69	4.80
0.3	5.50	5.36	4.68	5.40	5.33	4.52	4.96	4.50	4.79
0.5	5.34	5.16	4.86	4.71	4.62	4.74	4.90	4.88	4.77
(c) Multiple regression with six variables, RE estimator									
0.1	5.77	5.94	6.54	5.45	6.36	9.31	6.19	10.01	20.59
0.3	6.00	6.53	8.57	5.54	8.15	17.27	8.20	16.54	38.77
0.5	5.67	6.52	10.14	6.02	9.92	23.73	9.71	24.67	49.70
(d) Multiple regression with six variables, GW estimator									
0.1	5.40	5.29	4.20	4.79	4.58	3.70	4.65	4.95	4.16
0.3	5.37	4.84	3.82	4.45	4.42	3.72	4.71	4.33	4.11
0.5	5.17	4.55	3.56	4.64	4.17	3.54	4.49	4.63	3.98
(e) Meta-regression with one moderator, RE estimator									
0.1	5.98	6.03	5.74	4.98	5.92	8.24	5.47	7.84	15.97
0.3	5.76	6.39	8.12	5.21	6.68	12.53	7.26	13.00	32.02
0.5	5.93	6.89	9.19	5.88	8.62	18.68	8.60	18.00	41.03
(f) Meta-regression with one moderator, GW estimator									
0.1	6.17	5.94	5.36	4.90	5.31	5.14	4.65	4.98	4.53
0.3	5.91	6.02	5.89	4.87	4.79	4.48	5.48	5.01	5.18
0.5	6.02	6.08	5.26	5.05	5.09	5.02	4.93	5.04	4.97
(g) Meta-regression with five moderators, RE estimator									
0.1				5.87	6.07	7.41	5.76	7.46	13.96
0.3				6.05	6.97	11.82	6.61	11.34	26.71
0.5				6.26	7.75	16.49	8.15	15.33	36.81
(h) Meta-regression with five moderators, GW estimator									
0.1				5.76	5.49	4.62	4.94	4.69	4.65
0.3				5.82	5.22	4.19	5.14	4.86	4.29
0.5				5.46	4.79	4.4	4.99	4.75	4.14

Note: The reported values refer to the rates of false rejections of a t-test that the average effect is zero at the 5% significance level. M denotes the size of the meta-sample, λ denotes the fraction of overlapping estimates (extensive-margin overlap), and ρ denotes the average correlation between overlapping estimates (intensive-margin overlap). Meta-regressions with $M = 32$ and five moderators often give rise to perfect multicollinearity; the results are thus not reported.

5 | IMPLEMENTATION

The GW estimator may present implementation challenges in practical applications, particularly in obtaining the number of overlapping observations between each pair of primary samples

TABLE 5 Mean squared errors (×1000) for the RE and GW estimators: multiple regressions and meta-regressions.

	$M = 32$			**$M = 128$**			**$M = 512$**		
$\rho\backslash\lambda$	**0.125**	**0.25**	**0.5**	**0.125**	**0.25**	**0.5**	**0.125**	**0.25**	**0.5**
(a) Multiple regression with two variables, RE estimator									
0.1	1.589	1.673	1.815	0.414	0.452	0.602	0.114	0.148	0.294
0.3	1.612	1.745	2.151	0.454	0.552	0.997	0.138	0.245	0.669
0.5	1.648	1.811	2.513	0.458	0.629	1.325	0.161	0.335	1.043
(b) Multiple regression with two variables, GW estimator									
0.1	1.589	1.672	1.806	0.413	0.442	0.539	0.109	0.119	0.168
0.3	1.611	1.729	2.029	0.443	0.490	0.646	0.112	0.127	0.185
0.5	1.644	1.765	2.231	0.428	0.482	0.688	0.113	0.131	0.193
(c) Multiple regression with six variables, RE estimator									
0.1	2.085	2.146	2.258	0.528	0.560	0.708	0.143	0.180	0.313
0.3	2.132	2.194	2.577	0.541	0.656	1.052	0.163	0.257	0.668
0.5	2.081	2.269	2.916	0.568	0.731	1.410	0.182	0.363	1.056
(d) Multiple regression with six variables, GW estimator									
0.1	2.086	2.150	2.263	0.528	0.560	0.708	0.141	0.162	0.220
0.3	2.139	2.215	2.625	0.540	0.620	0.831	0.143	0.164	0.245
0.5	2.085	2.264	2.888	0.552	0.633	0.902	0.144	0.174	0.257
(e) Meta-regression with one moderator, RE estimator									
0.1	3.430	3.432	3.491	0.814	0.866	1.025	0.221	0.258	0.384
0.3	3.383	3.541	4.007	0.837	0.941	1.328	0.237	0.350	0.761
0.5	3.355	3.637	4.246	0.872	1.041	1.775	0.266	0.426	1.175
(f) Meta-regression with one moderator, GW estimator									
0.1	3.430	3.431	3.471	0.811	0.854	0.971	0.213	0.229	0.258
0.3	3.381	3.525	3.883	0.827	0.870	1.019	0.213	0.235	0.278
0.5	3.346	3.575	3.896	0.842	0.905	1.111	0.216	0.227	0.290
(g) Meta-regression with five moderators, RE estimator									
0.1				1.091	1.108	1.246	0.273	0.307	0.453
0.3				1.081	1.203	1.587	0.297	0.394	0.797
0.5				1.101	1.250	1.997	0.319	0.489	1.184
(h) Meta-regression with five moderators, GW estimator									
0.1				1.092	1.112	1.247	0.270	0.285	0.361
0.3				1.079	1.166	1.372	0.278	0.296	0.383
0.5				1.097	1.163	1.437	0.279	0.306	0.392

Note: M denotes the size of the meta-sample, λ denotes the fraction of overlapping estimates (extensive-margin overlap), and ρ denotes the average correlation between overlapping estimates (intensive-margin overlap). Meta-regressions with $M = 32$ and five moderators often give rise to perfect multicollinearity; the results are thus not reported.

(i.e., C_{pq}), especially when data aggregation is involved. To facilitate the practical use of the GW estimator in meta-analyses with sample overlap, we have developed a code in Stata and R that automates the entire process. The code is freely available for download at https://osf.io/g5t2j and is accompanied by detailed instructions on its usage.

Implementing the GW estimator using our code requires the user to provide an Excel file containing the meta-analysis data. The essential information to include in this file are the primary estimates, standard errors, indicators for OLS or IV estimates, indicators for unstandardized regression coefficients or PCCs, and basic sample information to infer the degree of overlap between samples. If the user intends to perform a meta-regression, the moderator variables must also be provided. The sample information should specify the time frequency of the data (e.g., quarterly or annual) and the covered period (e.g., 1975–2020), as well as the spatial units considered (e.g., countries or regions) for each primary sample.

With the basic sample size information provided by the user, the code calculates the number of overlapping estimates between any pair of samples, determines the corresponding covariances between primary estimates, and finally estimates the meta-analysis model using the GW estimator. The output is returned in the form of a linear regression model, containing only a constant for a GW weighted average or including moderator variables for a GW meta-regression. This streamlined approach simplifies the implementation process and allows researchers to effectively account for sample overlap in their meta-analyses.

6 | APPLICATION

The application of the GW estimator is illustrated using the meta-analysis data by Bom and Ligthart (2014), which consist of 578 estimates from 68 studies on the output elasticity of public capital. The primary studies in this meta-analysis employ the so-called production function approach, which treats public capital as an input alongside private capital and labor. Most studies assume a Cobb–Douglas production function, which is transformed into a multiple linear regression model by taking logarithms of the variables. Hence, the typical study performs a regression of the log of output on the logs of public capital, private capital, and labor, with some studies also controlling for the effects of other variables. The output elasticity of public capital is represented by the slope coefficient of the log of public capital in this regression.

We focus on Bom and Ligthart's (2014) meta-analysis data because they provide a good illustration of the complications discussed above. First, like many macroeconomics meta-analyses, they exhibit substantial sample overlap within and across studies. In fact, almost half the included estimates (278) from almost half the studies (28) rely on US data. The studies differ in data type and coverage, with some using state-level panel data and others nationwide time-series data for different periods. However, the majority of studies encompass data from the period 1970–1985. Consequently, Bom and Ligthart's (2014) results strongly reflect the specific US experience during that period. Second, aggregation issues feature prominently in their meta-study. While most studies employ country-level time-series or panel data, a considerable number of studies use regional data as well. Finally, their data exhibit substantial heterogeneity, both systematic and random (or unobserved).

We implement the GW estimator using the Stata code described in Section 5. To illustrate the computation of the variance–covariance matrix elements, we select eight estimates from Bom and Ligthart's (2014) dataset, numbered 1 through 8. Table 6 provides relevant information on the samples from which these estimates are derived. The first four estimates are based on annual

TABLE 6 Characteristics of a selection of primary samples in Bom and Ligthart's (2014) meta-analysis.

Estimate	Authors (year)	Data type	Time period	Spatial unit(s)	Sample size	SE
1	Ratner (1983)	Time series	1949–73	US	25	0.099
2	Aschauer (1989)	Time series	1949–67	US	19	0.154
3	Aschauer (1989)	Time series	1968–85	US	18	0.108
4	Aschauer (1989)	Time series	1953–85	US	33	0.027
5	Munnell (1990)	Regional panel	1970–86	9 US states	153	0.030
6	Otto and Voss (1994)	Time series	1966–89	Australia	24	0.229
7	Otto and Voss (1996)	Time series	1966q1–89q4	Australia	136	0.080
8	Kamps (2006)	Country panel	1960–01	22 countries[a]	924	0.060

[a]Including United States and Australia.

TABLE 7 Estimated variance and covariances for a selection of primary samples.

	1	2	3	4	5	6	7	8
1	0.00972							
2	0.01326	0.02380						
3	0.00302	0	0.01170					
4	0.00192	0.00246	0.00213	0.00071				
5	0.00003	0	0.00018	0.00003	0.00093			
6	0	0	0	0	0	0.05240		
7	0	0	0	0	0	0.00765	0.00634	
8	0.00055	0.00056	0.00091	0.00024	0.00001	0.00222	0.00045	0.00361

Note: The numbers of the samples correspond to the studies described in Table 6.

time-series data for the United States, spanning different periods. Estimate 5 comes from a panel of nine US states. Estimates 6 and 7 are obtained from yearly and quarterly Australian time series, spanning the period 1966–1989. Lastly, estimate 8 is derived from a panel of 22 countries, including Australia and the United States, for the period 1960–2001.

Table 7 presents the elements of the estimated variance–covariance matrix corresponding to the primary estimates described in Table 6. The main diagonal contains the estimated variances of the primary estimates, which are given by the square of the reported standard errors in the primary studies (reported in Table 6 for the eight selected studies). The FE/WLS estimator would assume the diagonal matrix thus obtained as the variance–covariance matrix. To obtain the RE estimator—and the GW estimator, in this application, due to existence of perfectly overlapping samples in the meta-analysis—the estimated variance of the heterogeneity error component must be added to this diagonal. We estimate this variance using the DerSimonian and Laird (1986) method, resulting in a value of 0.0059. The elements outside the main diagonal represent the covariances. Notably, these covariances are zero between the United States and Australian estimates as they do not overlap. Similarly, the covariance between estimates 2 and 3, both pertaining to the United States but without overlapping over time, is also zero.

Let us illustrate the computation of several estimated covariances in Table 7 using the relevant formulas from Table 1. First, for samples 1 and 2, aggregated at the same level with an overlap of 19 years, the estimated covariance is $19\sqrt{0.099^2/25}\sqrt{0.154^2/19} = 0.01326$. Second, for samples 1 and 5, with incomplete spatial aggregation and overlapping for 4 years (1970–73) and nine US

TABLE 8 Estimation results for Bom and Ligthart's (2014) data: RE and GW estimators.

	RE estimator		GW estimator	
	Mean only	**Meta-regression**	**Mean only**	**Meta-regression**
$\bar{\theta}$	0.126	0.145	0.134	0.149
	(0.008)	(0.007)	(0.013)	(0.012)
Cointegration		−0.037		−0.095
		(0.028)		(0.035)
Endogeneity		−0.027		−0.054
		(0.024)		(0.029)
Core		0.064		0.017
		(0.020)		(0.030)
Regional data		−0.208		−0.170
		(0.016)		(0.028)
Sample median		−0.002		−0.005
		(0.0011)		(0.0018)

Note: The parameter $\bar{\theta}$ represents the mean output elasticity of public capital. The standard errors are given in parentheses.

states out of 50, the covariance is $4(9/50)\sqrt{0.099^2/25}\sqrt{0.030^2/153} = 0.00003$. Third, for estimates 6 and 7, both concerning Australia over the same time period, but with the former aggregating the latter temporally, we find $24\sqrt{0.229^2/24}\sqrt{0.080^2/136} = 0.00765$. Finally, for estimates 6 and 8, where the latter features Australia among 22 countries in the panel, the covariance is $24\sqrt{0.229^2/24}\sqrt{0.060^2/924} = 0.00222$. We should not be deceived by the apparent small magnitude of these covariances. For instance, the implied correlation coefficient between estimates 1 and 2 is $0.01326/\sqrt{0.00972 \times 0.02380} = 0.87$.

After computing the elements of the variance–covariance matrix, the code proceeds to estimate the specified meta-analysis model using GW (in conjunction with FE and RE), with or without moderators. It is worth noting that the code allows for an unrestricted constant of the variance–covariance matrix, effectively making FE equivalent to the WLS estimator (Stanley and Doucouliagos, 2017). We consider two meta-analysis models: one with only a constant (mean), and another including several moderators. We include five of the most relevant moderator variables considered in the original study: "cointegration" (=1 if the study finds evidence of a cointegration relationship), "endogeneity" (=1 if the study corrects for endogeneity), "core" (=1 if the study defines public capital as core infrastructure capital), "regional data" (=1 if the data are defined at the regional level), and "sample median" (median sample year divided by 1000). We compared the RE and GW estimates of these models. The results are reported in Table 8.

In both models, with and without moderators, the point estimate of the mean output elasticity of public capital is similar for both the RE and GW estimators. However, two important differences between the two sets of estimates are observed. First, the point estimates of the coefficients of the moderator variables differ in certain cases, such as for the variables cointegration, endogeneity, and core. One possible explanation for these differences is the large amount of overlap in Bom and Ligthart's (2014) data, as shown by BR in their Figure 4 (p.822), which leads to very different weights for RE and GW. Another possible explanation is the likely presence of publication bias, better handled by allowing for multiplicative heterogeneity, as shown for WLS (Stanley & Doucouliagos, 2017; Stanley et al., 2022, 2023). Like WLS, GW allows for an unrestricted constant to the variance–covariance matrix; RE does not.

Second, and more importantly, the standard errors of both $\bar{\theta}$ and the coefficients of the moderators are much larger for the GW estimator compared to the RE estimator. For instance, in the mean-only model, the standard error of the estimate of $\bar{\theta}$ is more than 50% larger for the GW estimator; in the meta-regression model, it is almost twice as large. Moreover, the standard errors of the coefficients of the moderators are always larger for the GW estimator. This discrepancy is not attributed to efficiency losses; as recalled from Section 4, GW is more efficient than RE in the presence of sample overlap. The reason lies instead in the fact that RE, by neglecting the repeated occurrence of sampling information in multiple primary studies, achieves artificially high precision. On the other hand, the GW estimator corrects for this spurious precision by appropriately downweighting estimates based on overlapping observations.

7 | DISCUSSION

We have shown how to implement the GW solution to solve the problem of sample overlap in economics meta-analyses, solving several challenges that are likely to be encountered in practical applications. We have provided simple formulas to estimate the covariances between sample-overlapping estimates when primary data are aggregated at different levels, when using OLS and IV methods, and when regression coefficients are converted into PCCs. These formulas were derived for simple regressions, assuming homogeneity across studies, and simulations confirmed their broader applicability to slope coefficients from multiple regressions and meta-regressions addressing systematic heterogeneity.

Our approach is conservative as we assume simple and homogeneous models, leading to covariance formulas that tend to overstate the true covariances in more complex cases. In other words, these formulas provide upper bounds for the true covariances. This conservative nature allows our GW method to work effectively for various scenarios not specifically studied in this paper. For instance, the covariance formula for OLS estimates will still yield a conservative covariance if the overlapping primary estimates are based on different estimation methods, such as GLS (e.g., a RE estimate in a panel setting). Similarly, our formula for IV estimates, assuming the use of one and the same instrument in all IV studies, provides a conservative estimate for the covariance between overlapping estimates obtained from more general IV methods, such as 2SLS or GMM, where multiple and varied instruments are used. This conservative feature also allows IV estimates to be treated, even more conservatively, as OLS estimates, if required, especially when dealing with cases where the variance–covariance matrix is not positive definite (see Section 3.2).

Stated even more generally, any study design differences across studies will lead to lower covariances between overlapping estimates than those indicated by our formulas. This principle applies not only to various estimation methods but also extends to discrepancies in model specifications (e.g., differences in the number and types of control variables included), variations in the definitions of variables (e.g., the use of different price deflators), and even potential changes in primary data due to data revisions or updates. Despite incurring some unavoidable efficiency costs, our conservative approach offers significant advantages in terms of simplicity and versatility. The use of a few simple formulas allows for their application in numerous overlapping cases without compromising the validity of inference at the meta-analysis level.

Most meta-analyses in economics involve excess variation in the reported estimates (Stanley & Doucouliagos, 2019). In such cases, a key objective of meta-analysis is to identify the factors responsible for this systematic heterogeneity. This is typically achieved by extending the meta-analysis model to include moderator variables—so-called meta-regressions (Stanley & Doucouliagos, 2019). Meta-regressions have become a fundamental component of meta-analysis research in economics. As demonstrated, the GW estimator is versatile and applicable not only for simple meta-analysis models of means but also for estimating meta-regressions. It could be applied, therefore, in one important type of meta-regressions including the standard error (or its square) as an additional regressor in order to detect and correct for publication bias (Stanley & Doucouliagos, 2014; Ioannidis et al., 2017; Bom & Rachinger, 2019).

Excess variation in reported estimates may partly arise from random or unobserved factors. Random heterogeneity can be addressed additively by incorporating an extra variance component term on the main diagonal of the variance–covariance matrix (RE estimator). Alternatively, it can be handled multiplicatively by the WLS estimator, which leaves unrestricted the constant multiplying the variance–covariance matrix (Stanley & Doucouliagos, 2017), a more suitable method in the presence of size-dependent heterogeneity and publication bias (Stanley et al., 2022; Stanley et al., 2023). The GW estimator, encompassing both WLS and RE as special cases, can effectively accommodate random heterogeneity in both ways, thus retaining WLS's advantages over RE when dealing with publication bias. However, in cases where there are primary samples with exact overlap, an additive component of the RE type becomes necessary for the variance–covariance matrix to be invertible, even if the constant is left unrestricted.

Meta-studies have addressed estimate dependency in other ways. Two commonly employed approaches are cluster-robust methods and multilevel (hierarchical) models. Cluster-robust methods involve defining disjoint clusters of correlated observations (MacKinnon et al., 2023). In meta-analysis, clusters are typically defined at the study or country level (e.g., Neisser, 2021; Havranek et al., 2022), because samples tend to overlap within studies or countries. These methods adjust the standard errors for within-cluster correlation without altering the point estimates. While potentially effective in balanced cases of sample overlap, cluster-robust methods have two main drawbacks. First, by construction, they do not address efficiency issues. Second, defining clusters can be challenging or altogether infeasible in complex sample overlap structures.

As an example of the latter, consider a meta-analysis scenario where data for one country are reused across several primary studies, either as individual time series or within country panels, as shown in Bom and Ligthart's (2014) application (see Section 6). In such cases, defining clusters of overlapping data poses a challenge. Clustering at the study level is inaccurate as data overlaps not only within but also across studies. Clustering at the country level is infeasible, given that panels include multiple countries. GW, however, can straightforwardly address this and any other types of sample overlap.

Multilevel models assume a hierarchical (nested) structure of correlation among observations (for a general discussion of these models in meta-analysis, see Cheung, 2019; for an application to an economics meta-analysis, see Ugur et al., 2018). In meta-analyses of studies using data from several countries, for instance, one level could represent the country, and another, nested within the former, could pertain to the study itself. This nested correlation structure is then employed in a GLS fashion, akin to the GW method, to adjust both the point estimates and the standard errors. The main advantage of multilevel models over clustering methods is their enhanced efficiency. However, in complex cases of sample overlap, like the example mentioned above, defining the nested structure in these models can be as challenging as identifying clusters in cluster-robust methods. Additionally, these models involve the joint estimation of model parameters and variance components, which can be technically demanding. GW, by contrast, obviates the need for such estimation, relying entirely on overlap information obtained exogenously. In scenarios where the overlap structure fits neatly into a nested hierarchical model, multilevel models and GW are expected to yield similar results.

In summary, GW offers several advantages over cluster methods and multilevel models. First, it can effectively manage any type of overlap-induced dependency, no matter how irregular or complex, within and between studies, provided the number of overlapping observations can be identified. Second, the overlap structure is exogenously determined by information collected from the studies, simplifying the meta-analysis model estimation. Third, it effectively tackles both inference and efficiency concerns. These benefits may be especially significant in complex and unbalanced cases of sample overlap and less so in situations where overlap can be neatly organized into disjoint groups, like studies or countries. Lastly, it is important to note that GW is tailored for addressing sample overlap and does not directly address other types of estimate dependency. However, it can be combined with multilevel or cluster-robust methods for such purposes, making these approaches complementary to GW.

8 CONCLUDING REMARKS

Strong overlap between primary samples is common in economics meta-analyses, especially in research fields relying on aggregate data (e.g., time series or country-level panels). Sample overlap leads to double-counting, causing excessive weight assigned to empirical outcomes based on overlapping observations. Technically, this overlap induces correlation between primary estimates, resulting in efficiency losses and inference distortions at the meta-analysis level. For instance, as we demonstrate in Bom and Rachinger (2020), the likelihood of finding false positives may be greatly exaggerated due to sample overlap.

In Bom and Rachinger (2020), we address sample overlap by estimating the elements of the variance–covariance matrix describing the correlation structure between primary outcomes resulting from sample overlap. Our proposed solution, the GW estimator, properly weights primary empirical outcomes using this matrix. In simple terms, the GW method assigns weights to estimates based on their share of nonoverlapping information.

This paper demonstrates the implementation of the GW methods in economics meta-analyses, resolving several challenges frequently encountered in practice. For instance, primary data are often aggregated at different levels across spatial units (e.g., countries or regions) or over time (e.g., quarters or years). Additionally, different estimation methods (e.g., OLS or IV) are used in obtaining primary estimates, and regression coefficients are often standardized into PCCs. We derive analytical expressions for covariance between overlapping estimates in such cases, and through simulations, we show that these expressions can be applied to estimate not only mean effects but also more general meta-regressions.

To facilitate the implementation of the GW estimator, we provide a code in Stata and R that calculates the degree of overlap between primary samples, computes the variance–covariance matrix describing the correlation structure between primary outcomes, and estimates the meta-analysis model. The user only needs to input basic information on primary samples to calculate the number of overlapping observations. Besides routinely coded sample sizes and standard errors in applied meta-analyses, the user should also specify the time periods and spatial units of observation, requiring minimal extra effort at the coding stage.

Sample overlap is widespread in economics meta-analysis, appearing in different patterns, from more balanced structures where overlap occurs mostly within studies to more irregular structures where it happens within and between studies. Fortunately, the GW method's flexibility allows it to handle any type of overlap-induced correlation structure. Furthermore, the GW estimator can be combined with other models addressing different types of correlation, such as cluster-robust methods. By offering a streamlined and automated tool that facilitates its implementation, we hope to make the GW method an appealing choice for applied researchers conducting meta-analysis with sample overlap.

ACKNOWLEDGMENTS

The authors would like to thank seminar participants at the CEBA talk seminar and the University of Balearic Islands and conference participants at the MAER-net 2023 colloquium for helpful comments and suggestions. We also thank participants of the MAER-net 2021 colloquium for feedback on an earlier version of this paper. The first author gratefully acknowledges financial support from the Jesus de Gangoiti Barrera Foundation and from the Basque Government Department of Education (IT1429-22). The second author gratefully acknowledges financial support from the Spanish Ministry of Science and Innovation—State Research Agency (PID2020-114646RB-C43).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Matlab simulation codes and application data are available at https://osf.io/gcmtw/. Stata and R codes for applying the Generalized-Weights method are provided at https://osf.io/g5t2j/.

ORCID

Pedro R. D. Bom https://orcid.org/0000-0002-5643-2975

ENDNOTES

[1] See tab. 1 in Bom and Rachinger (2020) for further examples.

[2] Our primary focus is on regression coefficients as the object of interest for meta-analysis because linear regression models represent the key analytical tool used in economics research. However, concerns regarding sample overlap in meta-analysis have also been raised in the context of experimental research; see Hussein et al. (2022) and Mathes et al. (2023).

[3] Note that the assumption of homogeneity could be easily relaxed by allowing the true effect θ to be a random distribution or by allowing the reported estimates to show random heterogeneity around the true effect. This would give rise to the random effects (RE) estimator, which would account for the presence of an extra fixed component in the error term. Stanley and Doucouliagos (2017) show that a weighted least squares (WLS) approach, which consists of estimating (1) by OLS after dividing each term by the standard error of $\hat{\theta}_i$, is equivalent to the FE estimator but also accommodates, in a multiplicative way, the presence of random heterogeneity (see below).

[4] For the essentials of IV estimation, see, for example, Angrist and Pischke (2008).

[5] Note that this pathological outcome is prevented when using a geometric average of the two standard errors, as done in the covariance formulas for two OLS or two IV estimates.

REFERENCES

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Aschauer, D. A. (1989). Is public expenditure productive? *Journal of Monetary Economics*, *23*(2), 177–200.

Bajzik, J., Havranek, T., Irsova, Z., & Schwarz, J. (2020). Estimating the Armington elasticity: The importance of study design and publication bias. *Journal of International Economics*, *127*, 103383.

Bom, P. R. D., & Ligthart, J. E. (2014). What have we learned from three decades of research on the productivity of public capital? *Journal of Economic Surveys*, *28*, 889–916.

Bom, P. R. D., & Rachinger, H. (2019). A kinked meta-regression model for publication bias correction. *Research Synthesis Methods*, *10*(4), 497–514.

Bom, P. R. D., & Rachinger, H. (2020). A generalized-weights solution to sample overlap in meta-analysis. *Research Synthesis Methods*, *11*(6), 812–832.

Cazachevici, A., Havranek, T., & Horvath, R. (2020). Remittances and economic growth: A meta-analysis. *World Development*, *134*, 105021.

Cheung, M. W. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, *29*, 387–396.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188.

Gechert, S., Havranek, T., Irsova, Z., & Kolcunova, D. (2022). Measuring capital-labor substitution: The importance of method choices and publication bias. *Review of Economic Dynamics*, *45*, 55–82.

Gechert, S., & Heimberger, P. (2022). Do corporate tax cuts boost economic growth? *European Economic Review*, *147*, 104157.

Havranek, T., Irsova, Z., Laslopova, L., & Zeynalova, O. (2022). Publication and attenuation biases in measuring skill substitution. *The Review of Economics and Statistics*, forthcoming.

Heimberger, P. (2022). Does economic globalisation promote economic growth? A meta-analysis. *World Economy*, *45*(6), 1690–1712.

Hussein, H., Nevill, C. R., Mefen, A., Abrams, K. R., Bujkiewicz, S., Sutton, A. J., & Gray, L. J. (2022). Double-counting of populations in evidence synthesis in public health: A call for awareness and future methodological development. *BMC Public Health*, *22*(1), 1–11.

Ioannidis, J. P., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research. *Economic Journal*, *127*(605), F236–F265.

Kamps, C. (2006). New estimates of government net capital stocks for 22 OECD countries, 1960-2001. *IMF Staff Papers*, *53*(1), 120–150.

Langan, D., Higgins, J. P., & Simmonds, M. (2017). Comparative performance of heterogeneity variance estimators in meta-analysis: A review of simulation studies. *Research Synthesis Methods*, *8*(2), 181–198.

MacKinnon, J. G., Ørregaard Nielsen, M., & Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, *232*, 272–299.

Mathes, T., Zhang, Z., Pachanov, A., & Pieper, D. (2023). Systematic reviews and meta-analyses that include registry-based studies: Methodological challenges and areas for future research. *Journal of Clinical Epidemiology*, *156*, 119–122.

Munnell, A. H. (1990). Why has productivity growth declined? Productivity and public investment. *New England Economic Review*, *January/February*, 3–22.

Neisser, C. (2021). The elasticity of taxable income: A meta-regression analysis. *Economic Journal*, *131*, 3365–3391.

Otto, G. D., & Voss, G. M. (1994). Public capital and private sector productivity. *Economic Record*, *70*(209), 121–132.

Otto, G. D., & Voss, G. M. (1996). Public capital and private production in Australia. *Southern Economic Journal*, *62*(3), 723–738.

Ratner, J. B. (1983). Government capital and the production function for U.S. private output. *Economics Letters*, *13*, 213–217.

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

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

Stanley, T. D., & Doucouliagos, H. (2017). Neither fixed nor random: Weighted least squares meta-regression. *Research Synthesis Methods*, *8*(1), 19–42.

Stanley, T. D., & Doucouliagos, H. (2019). *Practical significance, meta-analysis and the credibility of economics* [Discussion paper series, 12458]. IZA.

Stanley, T. D., Doucouliagos, H., & Ioannidis, J. P. (2022). Beyond random effects: When small-study findings are more heterogeneous. *Advances in Methods and Practices in Psychological Science*, *5*, 1–11.

Stanley, T. D., Ioannidis, J. P., Maier, M., Doucouliagos, H., Otte, W. M., & Bartoš, F. (2023). Unrestricted weighted least squares represent medical research better than random effects in 67,308 cochrane meta-analyses. *Journal of Clinical Epidemiology*, *157*, 53–58.

Ugur, M., Awaworyi Churchill, S., & Solomon, E. (2018). Technological innovation and employment in derived labour demand models: A hierarchical meta-regression analysis. *Journal of Economic Surveys*, *32*(1), 50–82.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bom, P. R. D., & Rachinger, H. (2025). Accounting for sample overlap in economics meta-analyses: The generalized-weights method in practice. *Journal of Economic Surveys*, *39*, 904–931. https://doi.org/10.1111/joes.12633

APPENDIX A

This appendix describes the setup of the Monte Carlo simulations used to evaluate the covariance estimates. We generate two overlapping datasets for variables y and x with the data generating process $y = x + u$, where $x \sim U(0,1)$ and $u \sim N(0,1)$. For each dataset, we compute OLS and IV estimates for a simple regression of y on x, along with their corresponding standard errors. The IV estimate uses a third variable z as an instrument for x, determined as $z = \delta x + (1-\delta)\epsilon$, where $\delta = 0.8$ and $\epsilon \sim N(0,1)$. This δ represents the correlation coefficient between z and x. We then convert the estimated regression coefficients to partial correlation coefficients. These steps are repeated one million times, and the true covariance is calculated as the empirical covariance between estimates across these repetitions.

We consider various data aggregation cases as discussed in Section 3.1, except for double aggregation. The sample sizes vary depending on the data aggregation case. In all cases, the time dimension consists of 160 basic time periods, with 120 periods common to both samples and 40 unique to each sample. For temporal aggregation, we use an aggregation factor of $T = 4$, where basic quarters are aggregated into years.

In cases of no aggregation and temporal aggregation, the samples are time series without a spatial dimension ($G = K = 1$). However, for spatial aggregation and co-aggregation, we introduce a spatial dimension with 20 spatial units (interpreted as regions), where the aggregation factor is $G = 20$, and only $K = 10$ units are present in the spatially disaggregated sample. Table A.1 presents the sample sizes and number of overlapping observations for the different data aggregation cases.

TABLE A.1 Sample sizes, overlap, and covariances in the various data aggregation cases.

	No aggregation $T=1, G=1, K=1$	**Temporal aggregation** $T=4, G=1, K=1$	**Spatial aggregation** $T=1, G=20, K=10$	**Co-aggregation** $T=4, G=20, K=10$
N_p	160	160	1600	400
N_q	160	40	160	160
C_{pq}	120	120	1200	300
C_{qp}	120	30	120	120

Note: The subscript p refers to the disaggregated sample and the subscript q refers to the aggregated one. In the case of co-aggregation, p refers to the spatially disaggregated estimate whereas q refers to the temporally disaggregated one.

APPENDIX B

In this appendix, we describe the design of the Monte Carlo simulations used to assess the statistical performance of the GW estimator in a meta-analysis setting with sample overlap. We consider a meta-analysis scenario with $M \in \{32, 128, 512\}$ primary studies. Each primary study provides an estimate and a standard error of the slope parameter θ, which quantifies the (partial) effect of variable x on variable y, based on samples of size $N \in \{60, 120, 180, 240\}$, evenly distributed among the M studies.

In general, the data generating process for y is represented by the multiple regression equation:

$$y = \alpha + \theta x + \phi_1 w_1 + \cdots + \phi_p w_p + u, \qquad \text{(B.1)}$$

where $w_1, \ldots, w_p$ are a set of control variables, and u is an independently and identically distributed error term. For simplicity, and without any loss of generality, we set α to zero. The parameter of interest, θ, is assumed to be randomly heterogeneous. We draw the values of θ from a normal distribution with mean $\bar{\theta} = 0$ and variance $\sigma_\theta^2 = 0.04$. This introduces a degree of parameter heterogeneity of approximately 85% at the meta-analysis level, as measured by the I^2 statistic. Note that the parameter of interest for meta-analysis is $\bar{\theta}$, which we set to zero to study the rate of false rejections. In the case of a simple regression, we set $\phi_1 = \cdots = \phi_p = 0$. For multiple regressions, we examine two scenarios: (1) single additional control variable, where we set $\phi_1 = 1$ and all other ϕ coefficients are zero; and (2) five additional control variables, where $\phi_1 = \cdots = \phi_5 = 0.2$.

We randomly draw x and u from a standard normal distribution. Regarding the control variables, we examine two scenarios. In the first case, the relevant control variables (ws) are also drawn from standard normal distributions, and the primary studies estimate the correctly specified model by including those relevant control variables. This scenario leads to a simple meta-analysis model, which involves a weighted average of the (unbiased) estimates of θ.

The second case involves extending the meta-analysis model to a meta-regression. In this scenario, we assume a multiple regression with either one (w_1) or five control variables ($w_1, \ldots, w_5$), where each control variable w_j is correlated with x. Specifically, $w_j = \psi x + (1 - \psi)N(0, 1)$, where $\psi = 0.4$ represents the correlation coefficient between w_j and x. As a result, primary studies obtain biased estimates of θ when they omit one relevant control variable. To model this situation, we assume that primary studies may either omit no relevant variables or omit some relevant w_j, both with equal probability. If a relevant variable is omitted, each possible w_j is omitted with the same probability. At the meta-analysis stage, the omission of a specific w_j is represented by a moderator variable D_j.

In the case where the primary model features only one additional control variable (w_1), the meta-analysis model is given by a meta-regression with a single moderator identifying estimates from models that exclude this control variable:

$$\hat{\theta}_i = \theta_0 + \theta_1 D_1 + \epsilon_i, \qquad \text{(B.2)}$$

where each θ_1 measures the bias relative to the correctly specified group of estimates. In the case with five additional control variables ($w_1, .., w_5$), the meta-analysis model is then a meta-regression with five moderator variables, each identifying the exclusion of a particular control variable in the primary regression:

$$\hat{\theta}_i = \theta_0 + \theta_1 D_1 + \cdots + \theta_5 D_5 + \epsilon_i, \qquad \text{(B.3)}$$

where each θ_j, $j = 1, \ldots, 5$, measures the bias relative to the correctly-specified group of estimates.

We model sample overlap as in BR. In particular, there are $M_1 = (1 - \lambda)M$ independent samples and $M_2 = \lambda M$ overlapping samples. The parameter $\lambda \in [0, 1]$ thus measures the extensive-margin degree of overlap. Overlapping samples contain ρN overlapping (i.e., repeated) observations. The parameter $\rho \in [0, 1]$ thus measures the intensive-margin degree of overlap. We consider three possible values of λ, 1/8, 1/4, and 1/2; and three possible values of ρ, 0.1, 0.3, and 0.5. As for the type of overlap, we consider Case A of BR, in which overlapping estimates represent a relatively small fraction of the total (i.e., λ is relatively small), but are concentrated on one single underlying sample. This would be the case, for instance, if one single sample (say, of US data) is used repeatedly in many studies.

Our simulations are based on $R = 10,000$ repetitions. We compute two main statistics:

- The rate of false rejections (i.e., the size) of a t-test of $H_0 : \bar{\theta} = 0$; that is, the percentage of repetitions for which H_0 is (incorrectly) rejected at the 5% level.
- The mean squared error: $MSE = (1/R) \sum_{i=1}^{R} (\tilde{\theta}_i - \bar{\theta})^2$, where $\tilde{\theta}_i$ is the ith meta-estimate.