Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Meta-analysis of independent datasets using constrained generalised method of moments

Menghao Xu & Jun Shao

To cite this article: Menghao Xu & Jun Shao (2020) Meta-analysis of independent datasets using constrained generalised method of moments, Statistical Theory and Related Fields, 4:1, 109-116, DOI: 10.1080/24754269.2019.1630545

To link to this article: https://doi.org/10.1080/24754269.2019.1630545

Published online: 18 Jun 2019.

STATISTICAL THEORY AND RELATED FIELDS
2020, VOL. 4, NO. 1, 109–116
https://doi.org/10.1080/24754269.2019.1630545

Taylor & Francis
Taylor & Francis Group

Meta-analysis of independent datasets using constrained generalised method of moments

Menghao Xu[a] and Jun Shao[a,b]

[a]KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai, People's Republic of China; [b]Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA

ABSTRACT
We propose a constrained generalised method of moments (CGMM) for enhancing the efficiency of estimators in meta-analysis in which some studies do not measure all covariates associated with the response or outcome. Under some assumptions, we show that the proposed CGMM estimators have good asymptotic properties. We also demonstrate the effectiveness of the proposed method through simulation studies with fixed sample sizes.

ARTICLE HISTORY
Received 3 February 2019
Revised 22 April 2019
Accepted 7 June 2019

KEYWORDS
Meta-analysis; generalised method of moments; efficiency

1. Introduction

Because of the availability of multiple datasets, not just summary statistics, from different studies in modern applications, meta-analysis has become an important tool to gain efficiency in estimating a common structural parameter vector of interest from all studies by appropriately using multiple datasets (Hartung, Knapp, & Sinha, 2008; Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003; Schmidt & Hunter, 2014; Simonian & Laird, 1986). There exists a rich literature on how to form optimal calibration equations for improving the efficiency of parameter estimates within various classes of unbiased estimators (Chen & Chen, 2000; Deville & Sarndal, 1992; Lumley, Shaw, & Dai, 2011; Robins, Rotnitzky, & Zhao, 1994; Slud & DeMissie, 2011; Wu, 2003; Wu & Sitter, 2001). The methodology for 'model-based' maximum likelihood estimation has also been studied previously in some special cases of this problem (Chatterjee, Chen, Maas, & Carroll, 2016). A number of researchers have proposed semiparametric maximum likelihood methods for various types of regression models, while accounting for complex sampling designs (Breslow & Holubkov, 1997; Lawless, Wild, & Kalbfleisch, 1999; Qin, Zhang, Li, Albanes, & Yu, 2015; Rao & Molina, 2015; Scott & Wild, 1997).

One issue that has to be addressed with multiple studies is that some studies may not measure all covariates although all studies have the same responses (Chatterjee et al., 2016). Specifically, a past study only measured q of the $p+q$ covariates measured in the current study. Although unobserved covariate values in the past study can be treated as missing covariate values, better statistical procedure may be derived because in each study, a covariate is either observed or missing entirely, which is referred to as systematic missing covariates.

To illustrate the idea, let us consider the special case of two studies. Let Y be a response or outcome of interest, $\boldsymbol{U}$ and $\boldsymbol{X}$ be p- and q-dimensional vectors of associated covariates measured in study 1, and $\boldsymbol{X}$ be the covariate vector measured only in study 2. We focus on the situation where whether $\boldsymbol{U}$ is observed does not affect the conditional means, i.e.,

$$E(Y \mid \boldsymbol{U}, \boldsymbol{X}, \delta) = E(Y \mid \boldsymbol{U}, \boldsymbol{X}) \quad \text{and} \quad E(\boldsymbol{U} \mid \boldsymbol{X}, \delta) = E(\boldsymbol{U} \mid \boldsymbol{X}), \tag{1}$$

where $\delta = k$ for study $k = 1,2$. In the missing data literature, the 'missingness' of $\boldsymbol{U}$ with property (1) is referred to as missing at random, but not missing completely at random.

Suppose that we are interested in the parameters in the conditional mean $E(Y \mid \boldsymbol{U}, \boldsymbol{X})$, which can be called structural parameters. From the first equation in (1), estimation can actually be done using data from study 1. However, we want to make use of data from study 2, which is the purpose of meta-analysis. The second equation, which will be referred to as bridge equation, may enable us to obtain estimators based on data from all studies that are more efficient than those using only data from study 1.

In this article, we assume that the conditional means in (1) follow linear models for both observation and bridge equation. Although more complicated models may be encountered in applications, the discussion with linear models is a good start to this problem.

CONTACT Menghao Xu jason_xmh@163.com

In Section 2, we propose a constraint generalised method of moments for estimation in the case of two studies. Asymptotic distributions of the proposed estimators are established, with which we illustrate when more asymptotically efficient estimators can be obtained. Simulation studies support our asymptotic results and illustrate the magnitude of efficiency gain. Our method can be extended to the case of more than two studies. As an example of extension, in Section 3, we consider the situation of three studies and supplemented with simulation results. The last section contains some technical details.

2. Results for two independent studies

In this section, we consider two studies, indicated by $\delta \in \{1,2\}$, with independent datasets. Following Section 1, we use $Y, \boldsymbol{X}$, and $\boldsymbol{U}$ as the response of interest, the covariate vector measured in both studies, and the covariate vector measured only in study 1, respectively.

2.1. Constrained generalised method of moments

For illustration, we first consider a univariate $\boldsymbol{U} = U$. Assume (1) and linear models for two independent studies as follows:

$$\delta = 1 \quad Y = \beta_u U + {\boldsymbol{\beta}_x}^{\mathrm{T}}\boldsymbol{X} + \varepsilon_1 \tag{2}$$

$$\delta = 2 \quad Y = {\boldsymbol{\eta}_x}^{\mathrm{T}}\boldsymbol{X} + \varepsilon_2 \tag{3}$$

$$\text{bridge} \quad U = \boldsymbol{\gamma}_x^{\mathrm{T}}\boldsymbol{X} + \varepsilon_b, \quad \boldsymbol{\eta}_x = \beta_u\boldsymbol{\gamma}_x + \boldsymbol{\beta}_x \tag{4}$$

where ε_1, ε_2, and ε_b are independent with mean 0 and variances σ_1^2, σ_2^2, and σ_b^2, respectively, β_u, $\boldsymbol{\beta}_x$, $\boldsymbol{\eta}_x$, and $\boldsymbol{\gamma}_x$ are parameter vectors with appropriate dimensions, and the superscript T denotes vector transpose. Models (2)–(4) assume that the structure parameters β_u, $\boldsymbol{\beta}_x$, and $\boldsymbol{\gamma}_x$ are the same for all studies, while the distributions of ε's can vary with studies, exhibiting the heteroscedasticity of the data among different studies.

We are mainly interested in estimating β_u and $\boldsymbol{\beta}_x$ in (2). Instead of using data from study 1 only, we try to make use of data from study 2 to gain estimation efficiency. Condition (4) is for bridging data in two studies to gain efficiency by using the additional data from study 2. It is not necessary. See our discussion in Section 4.

Assume that we have two independent random samples with sizes n_1 and n_2 from studies 1 and 2, respectively. We denote the total sample size from all studies as $n = n_1 + n_2$. From (2)–(4), we construct estimating equations $E\boldsymbol{g}(\boldsymbol{Z},\boldsymbol{\theta}) = \boldsymbol{0}$ and a constraint $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{0}$, where $\boldsymbol{0}$ is the vector of zeros, $\boldsymbol{\theta} = (\beta_u, \boldsymbol{\beta}_x^{\mathrm{T}}, \boldsymbol{\gamma}_x^{\mathrm{T}}, \boldsymbol{\eta}_x^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{\eta}_x - \beta_u\boldsymbol{\gamma}_x - \boldsymbol{\beta}_x$, $\boldsymbol{Z} = (Y, U, \boldsymbol{X}^{\mathrm{T}}, \delta)^{\mathrm{T}}$, $\boldsymbol{g}(\boldsymbol{Z},\boldsymbol{\theta})$ is a column vector with elements of

$$I(\delta = 1)(n/n_1)(\beta_u U + {\boldsymbol{\beta}_x}^{\mathrm{T}}\boldsymbol{X} - Y)U,$$

$$I(\delta = 1)(n/n_1)(\beta_u U + {\boldsymbol{\beta}_x}^{\mathrm{T}}\boldsymbol{X} - Y)\boldsymbol{X},$$

$$I(\delta = 1)(n/n_1)({\boldsymbol{\gamma}_x}^{\mathrm{T}}\boldsymbol{X} - U)\boldsymbol{X},$$

$$I(\delta = 2)(n/n_2)({\boldsymbol{\eta}_x}^{\mathrm{T}}\boldsymbol{X} - Y)\boldsymbol{X},$$

where $I(A)$ is the indicator function of the event A.

Let $\boldsymbol{z}_i = (y_i, u_i, \boldsymbol{x}_i^{\mathrm{T}}, \delta_i)^{\mathrm{T}}$, $i = 1,\dots,n$, be observed data from samples, where $(y_i, u_i, \boldsymbol{x}_i^{\mathrm{T}}, \delta_i)^{\mathrm{T}}$ with $\delta_i = k$ are identically distributed as $(Y, U, \boldsymbol{X}^{\mathrm{T}}, \delta)^{\mathrm{T}}$ with $\delta = k$, $k = 1,2$, and let $\bar{\boldsymbol{g}}(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n \boldsymbol{g}(\boldsymbol{z}_i,\boldsymbol{\theta})$. The two step constrained generalised method of moments (CGMM) is applied as follows.

(1) Compute $\tilde{\boldsymbol{\theta}}_c = \arg\min[\bar{\boldsymbol{g}}(\boldsymbol{\theta})^{\mathrm{T}}\bar{\boldsymbol{g}}(\boldsymbol{\theta})]$ over $\boldsymbol{\theta}$ with constraint $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{0}$.
(2) Compute the weight matrix $\widehat{W} = n[\sum_{i=1}^n \boldsymbol{g}(\boldsymbol{z}_i, \tilde{\boldsymbol{\theta}}_c)\boldsymbol{g}(\boldsymbol{z}_i, \tilde{\boldsymbol{\theta}}_c)^{\mathrm{T}}]^{-1}$.
(3) Compute the two step CGMM estimator $\hat{\boldsymbol{\theta}}_c = \arg\min[\bar{\boldsymbol{g}}(\boldsymbol{\theta})^{\mathrm{T}}\widehat{W}\bar{\boldsymbol{g}}(\boldsymbol{\theta})]$ over $\boldsymbol{\theta}$ with constraint $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{0}$.

We now extend our idea to a multivariate $\boldsymbol{U}$ that is observed in study 1 but not in study 2. Let $\boldsymbol{U}$ be p-dimensional and U_j be its jth component. Then, the previous procedure can still be applied with U, β_u, $\beta_u U$, $\boldsymbol{\gamma}_x$, ε_3, $(n/n_1)I(\delta = 1)(\boldsymbol{\gamma}_x\boldsymbol{X}^{\mathrm{T}} - U)\boldsymbol{X}$, and $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{\eta}_x - \beta_u\boldsymbol{\gamma}_x - \boldsymbol{\beta}_x$ replaced by $\boldsymbol{U}$, $\boldsymbol{\beta}_u = (\beta_{u1},\dots,\beta_{up})^{\mathrm{T}}$, $\boldsymbol{\beta}_u^{\mathrm{T}}\boldsymbol{U}$, $(\boldsymbol{\gamma}_{x1},\dots,\boldsymbol{\gamma}_{xp})$, $\boldsymbol{\varepsilon}_3 = (\varepsilon_{31},\dots,\varepsilon_{3p})^{\mathrm{T}}$, $(n/n_1)I(\delta = 1)((\boldsymbol{\gamma}_{x1}^{\mathrm{T}}\boldsymbol{X} - U_1)\boldsymbol{X}^{\mathrm{T}},\dots,(\boldsymbol{\gamma}_{xp}^{\mathrm{T}}\boldsymbol{X} - U_p)\boldsymbol{X}^{\mathrm{T}})$, and $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{\eta}_x - \boldsymbol{\beta}_u^{\mathrm{T}}(\boldsymbol{\gamma}_{x1},\dots,\boldsymbol{\gamma}_{xp})^{\mathrm{T}} - \boldsymbol{\beta}_x)$, respectively.

2.2. Asymptotic properties

The general theory for the generalised method of moment (GMM) is given in Hansen (1982). The CGMM we proposed in Section 2.1 adds a constraint to the GMM. For the purpose of testing hypotheses, Engle and McFadden (1994) considered the CGMM. We now establish an asymptotic result in a similar manner. For simplicity, we consider only a univariate U.

Let $\boldsymbol{\theta}_0$ denote the true value of the parameter vector $\boldsymbol{\theta}$. For the CGMM estimator $\hat{\boldsymbol{\theta}}_c$ defined in Section 2.1 with the constraint $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{0}$, we have the following result.

Theorem 2.1: *Assume that models* (2)–(4) *hold;* $\boldsymbol{\theta}_0$ *is the unique root of* $E\boldsymbol{g}(\boldsymbol{Z},\boldsymbol{\theta}) = \boldsymbol{0}$; *both* n_1 *and* n_2 *diverge to* ∞ *and* $n_1/n \to h$ *with* $0 < h < 1$; *and the matrices* $\boldsymbol{\Omega} = \lim_{n\to\infty} E\boldsymbol{g}(\boldsymbol{Z},\boldsymbol{\theta}_0)\boldsymbol{g}(\boldsymbol{Z},\boldsymbol{\theta}_0)^{\mathrm{T}}$, $\boldsymbol{G} = \lim_{n\to\infty} E[\partial\boldsymbol{g}(\boldsymbol{Z},\boldsymbol{\theta})/\partial\boldsymbol{\theta}^{\mathrm{T}}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, $\boldsymbol{\Sigma}_x = E(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})$, *and* $\boldsymbol{A} = [\partial\boldsymbol{c}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^{\mathrm{T}}]|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = (\boldsymbol{\gamma}_x, \boldsymbol{I}_q, \beta_u\boldsymbol{I}_q, -\boldsymbol{I}_q)$ *are all of full rank, where* $\boldsymbol{I}_q$ *is the identity matrix of order q. Then,*

$$n^{1/2}(\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}_0) \to_d N\big(\boldsymbol{0}, \boldsymbol{B} - \boldsymbol{B}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A}\boldsymbol{B}\big), \tag{5}$$

where $\boldsymbol{B} = (\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{G})^{-1}$ *and* $\to_d$ *denotes convergence in distribution as* $n \to \infty$.

If we do not use the constraint $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{0}$, then the unconstraint GMM (UGMM) estimator in our problem described in Section 2.1 is the vector of the least square estimators of β_u, $\boldsymbol{\beta}_u$, and $\boldsymbol{\gamma}_x$ based on data in study 1 only and the least squares estimator of $\boldsymbol{\eta}_x$ based on data in study 2 only. Let $\hat{\boldsymbol{\theta}}_0$ be the UGMM estimator. Then

$$n^{1/2}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0) \rightarrow_d N(\mathbf{0}, \boldsymbol{B}), \tag{6}$$

which can be derived in the same manner as deriving (5) but with $\boldsymbol{c}(\boldsymbol{\theta}) = \mathbf{0}$.

Is the CGMM estimator $\hat{\boldsymbol{\theta}}_c$ asymptotically more efficient than the UGMM estimator $\hat{\boldsymbol{\theta}}_0$ because of utilising two data sets? It follows from results (5) and (6) that a component of $\hat{\boldsymbol{\theta}}_c$ is asymptotically more efficient if and only if the corresponding diagonal element of the matrix $\boldsymbol{B}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A}\boldsymbol{B}$ is positive.

To find out the magnitude of efficiency gains in using CGMM, we need to address the issue that the limit h of sample size ratio may be different from 1/2, and need to derive more explicitly the asymptotic covariance matrices in (5) and (6).

Note that the first 1+2q components of $\hat{\boldsymbol{\theta}}_c$, denoted by $\hat{\boldsymbol{\zeta}}_c$, estimates $\boldsymbol{\zeta} = (\beta_u, \boldsymbol{\beta}_x^{\mathrm{T}}, \boldsymbol{\gamma}_x^{\mathrm{T}})^{\mathrm{T}}$ based on data from study 1 with size n_1, whereas the last q components of $\hat{\boldsymbol{\theta}}_c$, denoted by $\hat{\boldsymbol{\eta}}_{xc}$, estimates $\boldsymbol{\eta}_x$ based on data from study 2 with size n_2. From the technical details in Section 5, we obtain from (5) that

$$\begin{pmatrix} n_1^{1/2}(\hat{\boldsymbol{\zeta}}_c - \boldsymbol{\zeta}) \\ n_2^{1/2}(\hat{\boldsymbol{\eta}}_{xc} - \boldsymbol{\eta}_x) \end{pmatrix} \rightarrow_d N \times \big(\mathbf{0}, \boldsymbol{HBH} - \boldsymbol{HBA}^{\mathrm{T}}(\boldsymbol{ABA}^{\mathrm{T}})^{-1}\boldsymbol{ABH}\big), \tag{7}$$

where $\boldsymbol{H}$ is a diagonal matrix whose first 1+2q diagonal elements are $h^{1/2}$ and last q diagonal elements are $(1-h)^{1/2}$. For the special case where δ and $(U, \boldsymbol{X})$ in (1) are independent (so that missing U is completely at random), it is further shown in Section 5 that

$$\boldsymbol{HBH} = \begin{bmatrix} \sigma_1^2/\sigma_b^2 & -\sigma_1^2\boldsymbol{\gamma}_x{}^{\mathrm{T}}/\sigma_b^2 & \mathbf{0} & \mathbf{0} \\ -\sigma_1^2\boldsymbol{\gamma}_x/\sigma_b^2 & \sigma_1^2(\boldsymbol{\Sigma}_x^{-1} + \boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathrm{T}}/\sigma_b^2) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_b^2\boldsymbol{\Sigma}_x^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_2^2\boldsymbol{\Sigma}_x^{-1} \end{bmatrix}, \tag{8}$$

where $\mathbf{0}$ denotes a column or row vector of zeros or a matrix of zeros with an appropriate dimension, and that

$$\boldsymbol{HBA}^{\mathrm{T}}(\boldsymbol{ABA}^{\mathrm{T}})^{-1}\boldsymbol{ABH} = \frac{1}{\Delta}\begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D} \end{bmatrix}, \tag{9}$$

where $\Delta = (1-h)(\sigma_1^2 + \sigma_b^2\beta_u^2) + h\sigma_2^2$ and

$$\boldsymbol{D} = \begin{bmatrix} \sigma_1^4(1-h)\boldsymbol{\Sigma}_x^{-1} & \sigma_1^2\sigma_b^2\beta_u(1-h)\boldsymbol{\Sigma}_x^{-1} & -\sigma_1^2\sigma_2^2[(1-h)h]^{1/2}\boldsymbol{\Sigma}_x^{-1} \\ \sigma_1^2\sigma_b^2\beta_u(1-h)\boldsymbol{\Sigma}_x^{-1} & \sigma_b^4\beta_u^2(1-h)\boldsymbol{\Sigma}_x^{-1} & -\sigma_1^2\sigma_b^2\beta_u[(1-h)h]^{1/2}\boldsymbol{\Sigma}_x^{-1} \\ -\sigma_1^2\sigma_2^2[(1-h)h]^{1/2}\boldsymbol{\Sigma}_x^{-1} & -\sigma_1^2\sigma_b^2\beta_u[(1-h)h]^{1/2}\boldsymbol{\Sigma}_x^{-1} & \sigma_2^4 h\boldsymbol{\Sigma}_x^{-1} \end{bmatrix}. \tag{10}$$

Similarly, if $\hat{\boldsymbol{\zeta}}_0$ and $\hat{\boldsymbol{\eta}}_{x0}$ denote the UGMM estimators of $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}_x$, then

$$\begin{pmatrix} n_1^{1/2}(\hat{\boldsymbol{\zeta}}_0 - \boldsymbol{\zeta}) \\ n_2^{1/2}(\hat{\boldsymbol{\eta}}_{x0} - \boldsymbol{\eta}_x) \end{pmatrix} \rightarrow_d N\big(\mathbf{0}, \boldsymbol{HBH}\big). \tag{11}$$

We define the asymptotic relative efficiency gain in using CGMM estimator $\hat{\theta}_{cj}$, the jth component of $\hat{\boldsymbol{\theta}}_c$, with respect to the unconstraint GMM estimator $\hat{\theta}_{0j}$, the jth component of $\hat{\boldsymbol{\theta}}_0$, to be

$$R_j = \frac{\begin{array}{c}\text{the asymptotic variance of } \hat{\theta}_{0j} \\ -\text{the asymptotic variance of } \hat{\theta}_{cj}\end{array}}{\text{the asymptotic variance of } \hat{\theta}_{0j}}$$

$j = 1, \ldots, 1+3q$. From (7) and (11), we derive R_j's as follows. First, $R_1 = 0$, i.e., there is no gain in estimating β_u. Intuitively, this is because the data set in study 2 does not have information on U. Second, for estimating q components of $\boldsymbol{\beta}_x$,

$$R_j = \frac{(1-h)\sigma_b^2\beta_u^2\sigma^{(j-1)}}{\Delta\sigma_1^2(\sigma_b^2\sigma^{(j-1)} + \gamma_{x(j-1)}^2)}, \quad j = 2, \ldots, q+1,$$

where $\sigma^{(t)}$ is the tth diagonal element of the matrix $\boldsymbol{\Sigma}_x^{-1}$ and γ_{xt} is the tth component of $\boldsymbol{\gamma}_x$. Third, for estimating q components of $\boldsymbol{\gamma}_x$,

$$R_j = \frac{(1-h)\sigma_b^2\beta_u^2}{\Delta}, \quad j = q+2, \ldots, 2q+1.$$

Note that $\Delta^{-1}(1-h)$ is a decreasing function of h. Hence the CGMM estimators of components of $\boldsymbol{\beta}_x$ and $\boldsymbol{\gamma}_x$ are increasingly more efficient when h decreases, i.e., n_2/n_1 increases, which means more information can be borrowed from study 2. Finally, for estimating q components of $\boldsymbol{\eta}_x$,

$$R_j = \frac{h\sigma_2^2}{\Delta}, \quad j = 2q+2, \ldots, 3q+1,$$

which increases when h increases.

Table 1. Simulation variances of CGMM and UGMM estimators ($p=q=1$).

n_1	n_2	Method	Parameter and its true value $\beta_u=1$	$\beta_x=1$	$\gamma_x=1$	$\eta_x=2$
100	100	UGMM	0.010158	0.020502	0.010518	0.010397
		CGMM	0.010217	0.017443	0.007185	0.006881
100	400	UGMM	0.010787	0.019185	0.009858	0.002449
		CGMM	0.010856	0.015174	0.005433	0.002112
400	100	UGMM	0.002446	0.005086	0.002434	0.010239
		CGMM	0.002451	0.004723	0.002061	0.003258

Table 2. Simulation variances of CGMM and UGMM estimators ($p=1, q=2$).

ρ	Method	Parameter and its true value $\beta_u=1$	$\beta_{x1}=1$	$\beta_{x2}=1$	$\gamma_{x1}=1$	$\gamma_{x2}=1$
0	UGMM	0.010563	0.021218	0.020028	0.010662	0.010307
	CGMM	0.010788	0.018011	0.017978	0.007072	0.006917
0.3	UGMM	0.010333	0.021848	0.022418	0.011316	0.011551
	CGMM	0.010543	0.018177	0.019034	0.007819	0.007521
0.6	UGMM	0.010249	0.026688	0.027351	0.016678	0.017572
	CGMM	0.010482	0.021604	0.022424	0.011010	0.011839

2.3. Simulation study

Two simulation studies are carried out to check the empirical performance of the CGMM and UGMM estimators with finite fixed sample sizes. In the first simulation, we consider univariate U and X, i.e., $p=q=1$. The covariate is generated from the standard normal distribution. The covariate U and response Y are generated according to (2)–(4) with ε_1, ε_2, and ε_b independently distributed as standard normal.

Based on 2000 simulations, Table 1 gives the simulation variances of estimators of univariate parameters β_u, β_x, γ_x, and η_x, for both CGMM and UGMM. All simulation biases are less than 0.006 and thus not reported. True values of parameters and different sample sizes are included in Table 1.

A few conclusions can be made from Table 1.

(1) When $n_1=n_2=100$, the simulation relative efficiency gain of CGMM over UGMM is $(-0.58\%, 14.92\%, 31.69\%, 33.82\%)$ for estimating $(\beta_u, \beta_x, \gamma_x, \eta_x)$. This indicates that there is almost no improvement in estimating β_u, but there are substantial gains in estimating other 3 parameters, which supports our asymptotic result discussed in Section 2.2. In fact, the vector of asymptotic relative efficiency gains in theory defined in Section 2.2 is $(0, 1/6, 1/3, 1/3)$, which is very close to the simulation relative gains.

(2) When $n_1=100$ and $n_2=400$, more information from study 2 can be borrowed to estimate parameters in study 1. The simulation relative efficiency gain vector is $(-0.63\%, 20.91\%, 44.99\%, 13.76\%)$. We have more gains in estimating β_x and γ_x, but less gain in estimating η_x. The vector of asymptotic relative efficiency gains in theory defined in Section 2.2 is $(0, 2/9, 4/9, 1/9)$, which is very close to the simulation relative gain.

(3) When $n_1=400$ and $n_2=100$, the simulation relative efficiency gain vector is $(-0.2\%, 7.14\%, 15.32\%, 68.18\%)$. We have less gains in estimating β_x and γ_x, but more gain in estimating η_x. the vector of asymptotic relative efficiency gains in theory defined in Section 2.2 is $(0, 1/12, 1/6, 2/3)$, which is very close to the simulation relative gain.

Our second simulation considers a $q=2$ dimensional $\boldsymbol{X}=(X_1,X_2)^{\mathrm{T}}$, while U is still univariate. Data are generated according to (2) – (4) with $\boldsymbol{X}$ being a two-dimensional normal with zero marginal means, unit marginal variances, and a correlation ρ.

Note that $\boldsymbol{\beta}_x=(\beta_{x1},\beta_{x2})^{\mathrm{T}}$, $\boldsymbol{\gamma}_x=(\gamma_{x1},\gamma_{x2})^{\mathrm{T}}$, and $\boldsymbol{\eta}_x=(\eta_{x1},\eta_{x2})^{\mathrm{T}}$ are all 2-dimensional. Based on 2000 simulations, Table 2 gives the simulation variances of estimators of β_u, β_{x1}, β_{x2}, γ_{x1}, and γ_{x2} for both CGMM and UGMM. Results for the variances of estimators of η_{x1} and η_{x2} are omitted. Again, all simulation biases are less than 0.004 and thus not reported. True values of parameters are included in Table 2. Sample size $n_1=n_2=100$ and $\rho=0$, 0.3, and 0.6 are considered.

Table 2 shows similar results to those in Table 1. In estimating $\boldsymbol{\beta}_x$, the simulation relative gain of CGMM over UGMM ranges from 10% to 20%, while there is no gain in estimating β_u. Increasing the value of ρ, the correlation between two components of $\boldsymbol{X}$ increases the relative efficiency gain, but not substantially.

3. Results for three independent studies

The method and results in Section 2 can be extended to various situations where the number of independent studies is more than 2 and different covariates are observed in different studies. We consider in this section the case of three studies where the response Y and covariates $\boldsymbol{U}$, $\boldsymbol{V}$, and $\boldsymbol{X}$ are observed according to the following with sample sizes in three studies:

Study	Observed				Sample size
$\delta=1$	Y	$\boldsymbol{U}$	$\boldsymbol{V}$	$\boldsymbol{X}$	n_1
$\delta=2$	Y	$\boldsymbol{U}$		$\boldsymbol{X}$	n_2
$\delta=3$	Y		$\boldsymbol{V}$	$\boldsymbol{X}$	n_3

The total sample size from all studies is $n=n_1+n_2+n_3$.

3.1. CGMM

Similar to (1) and (2)–(4), we assume that

$$E(Y\mid \boldsymbol{U},\boldsymbol{V},\boldsymbol{X},\delta)=E(Y\mid \boldsymbol{U},\boldsymbol{V},\boldsymbol{X}), \tag{12}$$

$$E(\boldsymbol{V}\mid \boldsymbol{U},\boldsymbol{X},\delta)=E(\boldsymbol{V}\mid \boldsymbol{U},\boldsymbol{X}), \tag{13}$$

$$E(\boldsymbol{U}\mid \boldsymbol{V},\boldsymbol{X},\delta)=E(\boldsymbol{U}\mid \boldsymbol{V},\boldsymbol{X}), \tag{14}$$

$\delta = 1, 2, 3$, and that

$$\delta = 1 \quad Y = \boldsymbol{\beta}_u^{\mathrm{T}} \boldsymbol{U} + \boldsymbol{\beta}_v^{\mathrm{T}} \boldsymbol{V} + \boldsymbol{\beta}_x^{\mathrm{T}} \boldsymbol{X} + \varepsilon_1 \tag{15}$$

$$\delta = 2 \quad Y = \boldsymbol{\eta}_u^{\mathrm{T}} \boldsymbol{U} + \boldsymbol{\eta}_x^{\mathrm{T}} \boldsymbol{X} + \varepsilon_2 \tag{16}$$

$$\delta = 3 \quad Y = \boldsymbol{\tau}_v^{\mathrm{T}} \boldsymbol{V} + \boldsymbol{\tau}_x^{\mathrm{T}} \boldsymbol{X} + \varepsilon_3 \tag{17}$$

$$\text{bridge} \quad \boldsymbol{U} = \boldsymbol{\gamma}_{uv} \boldsymbol{V} + \boldsymbol{\gamma}_{ux} \boldsymbol{X} + \boldsymbol{\varepsilon}_b, \quad \boldsymbol{V} = \boldsymbol{\gamma}_{vu} \boldsymbol{U} + \boldsymbol{\gamma}_{vx} \boldsymbol{X} - \boldsymbol{\varepsilon}_b$$
$$\text{with } \boldsymbol{\gamma}_{uv} \boldsymbol{\gamma}_{vu} = \boldsymbol{I} \quad \text{and} \quad \boldsymbol{\gamma}_{uv} \boldsymbol{\gamma}_{vx} + \boldsymbol{\gamma}_{ux} = \boldsymbol{0} \tag{18}$$

where $p \times q$ matrix $\boldsymbol{\gamma}_{ux} = (\boldsymbol{\gamma}_{ux1}, \ldots, \boldsymbol{\gamma}_{uxp})^{\mathrm{T}}$, $p \times l$ matrix $\boldsymbol{\gamma}_{uv} = (\boldsymbol{\gamma}_{uv1}, \ldots, \boldsymbol{\gamma}_{uvp})^{\mathrm{T}}$, $l \times q$ matrix $\boldsymbol{\gamma}_{vx} = (\boldsymbol{\gamma}_{vx1}, \ldots, \boldsymbol{\gamma}_{vxl})^{\mathrm{T}}$, and $l \times p$ matrix $\boldsymbol{\gamma}_{vu} = (\boldsymbol{\gamma}_{vu1}, \ldots, \boldsymbol{\gamma}_{vul})^{\mathrm{T}}$. Assume samples are independent and identically distributed within each study and independent among studies and ε's are independent with mean zero. By assumptions (12), (14), (15), (17), and the expression of $\boldsymbol{U}$ in (18), we have the following constraint conditions:

$$\boldsymbol{\beta}_u^{\mathrm{T}} \boldsymbol{\gamma}_{uv} + \boldsymbol{\beta}_v^{\mathrm{T}} = \boldsymbol{\tau}_v^{\mathrm{T}} \quad \text{and} \quad \boldsymbol{\beta}_u^{\mathrm{T}} \boldsymbol{\gamma}_{ux} + \boldsymbol{\beta}_x^{\mathrm{T}} = \boldsymbol{\tau}_x^{\mathrm{T}}. \tag{19}$$

By assumptions (12), (13), (15), (16), and the expression of $\boldsymbol{V}$ in (18), we have the following constraint conditions:

$$\boldsymbol{\beta}_u^{\mathrm{T}} + \boldsymbol{\beta}_v^{\mathrm{T}} \boldsymbol{\gamma}_{vu} = \boldsymbol{\eta}_u^{\mathrm{T}} \quad \text{and} \quad \boldsymbol{\beta}_v^{\mathrm{T}} \boldsymbol{\gamma}_{vx} + \boldsymbol{\beta}_x^{\mathrm{T}} = \boldsymbol{\eta}_x^{\mathrm{T}}. \tag{20}$$

Denote $(\boldsymbol{\beta}_u^{\mathrm{T}}, \boldsymbol{\beta}_v^{\mathrm{T}}, \boldsymbol{\beta}_x^{\mathrm{T}})^{\mathrm{T}}$, $(\boldsymbol{\eta}_u^{\mathrm{T}}, \boldsymbol{\eta}_x^{\mathrm{T}})^{\mathrm{T}}$, and $(\boldsymbol{\tau}_v^{\mathrm{T}}, \boldsymbol{\tau}_x^{\mathrm{T}})^{\mathrm{T}}$ by $\boldsymbol{\beta}$, $\boldsymbol{\eta}$, and $\boldsymbol{\tau}$, respectively. Models (15)–(18) assume that the structure parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}_{vu}$, $\boldsymbol{\gamma}_{vx}$, $\boldsymbol{\gamma}_{uv}$, and $\boldsymbol{\gamma}_{ux}$ are the same for all studies, while the distributions of ε's can vary with studies. We are mainly interested in estimating $\boldsymbol{\beta}$ in (15). Instead of using data from study 1 only, we try to make use of data from studies 2 and 3 to gain estimation efficiency. Condition (18) is needed for bridging data among three studies; without this condition, it is hard to gain any efficiency by using the additional data from study 2 and 3.

Denote vec($\boldsymbol{M}$) a row vector contains all rows in a matrix $\boldsymbol{M}$. From (15)–(20), we construct estimating equations $E\boldsymbol{g}(\boldsymbol{Z}, \boldsymbol{\theta}) = 0$ and a constraint $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{0}$, where $\boldsymbol{Z} = (Y, \boldsymbol{U}^{\mathrm{T}}, \boldsymbol{V}^{\mathrm{T}}, \boldsymbol{X}^{\mathrm{T}}, \delta)^{\mathrm{T}}$, $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\eta}^{\mathrm{T}}, \boldsymbol{\tau}^{\mathrm{T}}, \mathrm{vec}(\boldsymbol{\gamma}_{uv}), \mathrm{vec}(\boldsymbol{\gamma}_{ux}), \mathrm{vec}(\boldsymbol{\gamma}_{vu}), \mathrm{vec}(\boldsymbol{\gamma}_{vx}))^{\mathrm{T}}$, $\boldsymbol{c}(\boldsymbol{\theta}) = (\boldsymbol{\beta}_u^{\mathrm{T}} \boldsymbol{\gamma}_{uv} + \boldsymbol{\beta}_v^{\mathrm{T}} - \boldsymbol{\tau}_v^{\mathrm{T}}, \boldsymbol{\beta}_u^{\mathrm{T}} \boldsymbol{\gamma}_{ux} + \boldsymbol{\beta}_x^{\mathrm{T}} - \boldsymbol{\tau}_x^{\mathrm{T}}, \boldsymbol{\beta}_u^{\mathrm{T}} + \boldsymbol{\beta}_v^{\mathrm{T}} \boldsymbol{\gamma}_{vu} - \boldsymbol{\eta}_u^{\mathrm{T}}, \boldsymbol{\beta}_v^{\mathrm{T}} \boldsymbol{\gamma}_{vx} + \boldsymbol{\beta}_x^{\mathrm{T}} - \boldsymbol{\eta}_x^{\mathrm{T}}, \mathrm{vec}(\boldsymbol{\gamma}_{uv} \boldsymbol{\gamma}_{vu} - \boldsymbol{I}), \mathrm{vec}(\boldsymbol{\gamma}_{uv} \boldsymbol{\gamma}_{vx} + \boldsymbol{\gamma}_{ux}))^{\mathrm{T}}$, $\boldsymbol{g}(\boldsymbol{Z}, \boldsymbol{\theta})$ is a column vector with elements of

$$I(\delta = 1)(n/n_1)[\boldsymbol{\beta}_u^{\mathrm{T}} \boldsymbol{U} + \boldsymbol{\beta}_v^{\mathrm{T}} \boldsymbol{V} + \boldsymbol{\beta}_x^{\mathrm{T}} \boldsymbol{X} - Y] \times (\boldsymbol{U}^{\mathrm{T}}, \boldsymbol{V}^{\mathrm{T}}, \boldsymbol{X}^{\mathrm{T}})^{\mathrm{T}},$$
$$I(\delta = 2)(n/n_2)[\boldsymbol{\eta}_u^{\mathrm{T}} \boldsymbol{U} + \boldsymbol{\eta}_x^{\mathrm{T}} \boldsymbol{X} - Y](\boldsymbol{U}^{\mathrm{T}}, \boldsymbol{X}^{\mathrm{T}})^{\mathrm{T}},$$
$$I(\delta = 3)(n/n_3)[\boldsymbol{\tau}_v^{\mathrm{T}} \boldsymbol{V} + \boldsymbol{\tau}_x^{\mathrm{T}} \boldsymbol{X} - Y](\boldsymbol{V}^{\mathrm{T}}, \boldsymbol{X}^{\mathrm{T}})^{\mathrm{T}},$$
$$I(\delta = 1)(n/n_1)((\boldsymbol{\gamma}_{uv1}^{\mathrm{T}} \boldsymbol{V} + \boldsymbol{\gamma}_{ux1}^{\mathrm{T}} \boldsymbol{X} - U_1)(\boldsymbol{V}^{\mathrm{T}}, \boldsymbol{X}^{\mathrm{T}}), \ldots, \times (\boldsymbol{\gamma}_{uvp}^{\mathrm{T}} \boldsymbol{V} + \boldsymbol{\gamma}_{uxp}^{\mathrm{T}} \boldsymbol{X} - U_p)(\boldsymbol{V}^{\mathrm{T}}, \boldsymbol{X}^{\mathrm{T}}))^{\mathrm{T}},$$
$$I(\delta = 1)(n/n_1)((\boldsymbol{\gamma}_{vu1}^{\mathrm{T}} \boldsymbol{U} + \boldsymbol{\gamma}_{vx1}^{\mathrm{T}} \boldsymbol{X} - V_1)(\boldsymbol{U}^{\mathrm{T}}, \boldsymbol{X}^{\mathrm{T}}), \ldots, \times (\boldsymbol{\gamma}_{vul}^{\mathrm{T}} \boldsymbol{U} + \boldsymbol{\gamma}_{vxl}^{\mathrm{T}} \boldsymbol{X} - V_l)(\boldsymbol{U}^{\mathrm{T}}, \boldsymbol{X}^{\mathrm{T}}))^{\mathrm{T}}.$$

Let $\boldsymbol{z}_i = (y_i, \boldsymbol{u}_i^{\mathrm{T}}, \boldsymbol{v}_i^{\mathrm{T}}, \boldsymbol{x}_i^{\mathrm{T}}, \delta_i)^{\mathrm{T}}$, $i = 1, \ldots, n$, be independent samples, where $(y_i, \boldsymbol{u}_i^{\mathrm{T}}, \boldsymbol{v}_i^{\mathrm{T}}, \boldsymbol{x}_i^{\mathrm{T}}, \delta_i)^{\mathrm{T}}$ with $\delta_i = k$ are identically generated from the distribution of $(Y, \boldsymbol{U}^{\mathrm{T}}, \boldsymbol{V}^{\mathrm{T}}, \boldsymbol{X}^{\mathrm{T}}, \delta)^{\mathrm{T}}$ with $\delta = k$, $k = 1,2,3$. Define $\bar{\boldsymbol{g}}(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \boldsymbol{g}(\boldsymbol{Z}_i, \boldsymbol{\theta})$. The two step CGMM is applied as follows.

(1) Compute $\tilde{\boldsymbol{\theta}}_c = \arg\min[\bar{\boldsymbol{g}}(\boldsymbol{\theta})^{\mathrm{T}} \bar{\boldsymbol{g}}(\boldsymbol{\theta})]$ over $\boldsymbol{\theta}$ with constraint $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{0}$.

(2) Compute the weight matrix $\widehat{W} = n[\sum_{i=1}^n \boldsymbol{g}(\boldsymbol{z}_i, \tilde{\boldsymbol{\theta}}_c) \boldsymbol{g}(\boldsymbol{z}_i, \tilde{\boldsymbol{\theta}}_c)^{\mathrm{T}}]^{-1}$.

(3) Compute the two step CGMM estimator $\hat{\boldsymbol{\theta}}_c = \arg\min[\bar{\boldsymbol{g}}(\boldsymbol{\theta})^{\mathrm{T}} \widehat{W} \bar{\boldsymbol{g}}(\boldsymbol{\theta})]$ over $\boldsymbol{\theta}$ with constraint $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{0}$.

Asymptotic property of CGMM estimator $\hat{\boldsymbol{\theta}}_c$ can be established similarly to Theorem 2.1.

3.2. Simulation study

In this section, we consider univariate U and V, i.e., $p = l = 1$. Then $\boldsymbol{\beta}_u$, $\boldsymbol{\beta}_v$, $\boldsymbol{\eta}_u$, $\boldsymbol{\tau}_v$, $\boldsymbol{\gamma}_{vu}$, $\boldsymbol{\gamma}_{uv}$, and $\boldsymbol{\varepsilon}_b$ reduce to scalars β_u, β_v, η_u, τ_v, γ_{vu}, γ_{uv}, and ε_b, respectively. Two simulation studies are carried out to check the empirical performance of the CGMM and UGMM estimators with finite fixed sample sizes. In the first simulation, we consider univariate X, i.e., $q = 1$. Then $\boldsymbol{\beta}_x$, $\boldsymbol{\eta}_x$, $\boldsymbol{\tau}_x$, $\boldsymbol{\gamma}_{ux}$, and $\boldsymbol{\gamma}_{vx}$ reduce to scalars β_x, η_x, τ_x, γ_{ux}, and γ_{vx}, respectively. The covariate X and V are independently generated from the standard normal distribution. The covariate U and response Y are generated according to (15)–(18) with ε_1, ε_2, ε_3, and ε_b independently distributed as standard normal.

Based on 2000 simulations, Table 3 gives the simulation variances of estimators of parameters β_u, β_v, and β_x for both CGMM and UGMM. Results for the variances of estimators of other parameters are omitted. All simulation biases are less than 0.007 and thus not reported. True values of parameters and different sample sizes are included in Table 3.

It can be seen that messages provided by Table 3 are very similar to those from Table 1. When $n_1 = n_2 = n_3 = 100$, the simulation relative efficiency gain of CGMM over UGMM is (83.04%, 62.05%, 54.97%) for estimating $(\beta_u, \beta_v, \beta_x)$. When $n_1 = 100$, $n_2 = 100$ and $n_2 = 400$, more information from study 3 can be borrowed to estimate parameters in study 1, and the simulation relative efficiency gain vector is (85.11%, 65.35%, 63.08%) for estimating $(\beta_u, \beta_v, \beta_x)$.

Table 3. Simulation variances of CGMM and UGMM estimators ($p=l=q=1$).

				Parameter and its true value		
n_1	n_2	n_2	Method	$\beta_u=1$	$\beta_v=1$	$\beta_x=2$
100	100	100	UGMM	0.010464	0.020405	0.020696
			CGMM	0.001775	0.007744	0.009319
100	100	400	UGMM	0.010181	0.020822	0.021421
			CGMM	0.001516	0.007215	0.007908
100	400	100	UGMM	0.009943	0.020654	0.019776
			CGMM	0.001147	0.006557	0.008078

$\boldsymbol{\eta}=(2,3)^{\mathrm{T}}, \boldsymbol{\tau}=(2,1)^{\mathrm{T}}, (\gamma_{ux},\gamma_{vx},\gamma_{uv},\gamma_{vu})=(-1,1,1,1)$

Table 4. Simulation variances of CGMM and UGMM estimators ($p=l=1, q=2$).

		Parameter and its true value			
ρ	Method	$\beta_u=1$	$\beta_v=1$	$\beta_{x1}=2$	$\beta_{x2}=2$
0	UGMM	0.010922	0.021232	0.021002	0.021755
	CGMM	0.001465	0.005598	0.008043	0.008337
0.3	UGMM	0.011040	0.022697	0.023520	0.023184
	CGMM	0.001534	0.007043	0.008896	0.009251
0.6	UGMM	0.010087	0.020480	0.027241	0.025991
	CGMM	0.001797	0.008067	0.011927	0.011524

$\boldsymbol{\eta}=(2,3,1)^{\mathrm{T}}, \boldsymbol{\tau}=(2,1,3)^{\mathrm{T}}, \gamma_{uv}=\gamma_{vu}=1$
$\boldsymbol{\gamma}_{ux}=(-1,1)^{\mathrm{T}}, \boldsymbol{\gamma}_{vx}=(1,-1)^{\mathrm{T}}$

When $n_1=100$, $n_2=400$ and $n_2=100$, more information from study 2 can be borrowed to estimate parameters in study 1, and the simulation relative efficiency gain vector is (88.46%, 68.25%, 59.15%) for estimating $(\beta_u,\beta_v,\beta_x)$.

Our second simulation considers a $q=2$ dimensional $\boldsymbol{X}=(X_1,X_2)^{\mathrm{T}}$, while U and V are still univariate. Data are generated according to (15)–(18) with $\boldsymbol{X}$ being two dimensional normal with zero marginal means, unit marginal variances, and a correlation ρ.

Based on 2000 simulations, Table 4 gives the simulation variances of estimators of β_u, β_v, β_{x1}, and β_{x2} for both CGMM and UGMM. Results for the variances of estimators of other parameters are omitted. Again, all simulation biases are less than 0.004 and thus not reported. True values of parameters are included in Table 4. Sample size $n_1=n_2=n_3=100$ and $\rho=0$, 0.3, and 0.6 are considered. The results show substantial improvement of CGMM over UGMM, and the effect of ρ is not substantial.

Different from Tables 1 and 2, the simulation relative efficiency gain in estimating β_u is not zero in Tables 3 and 4. This is because that the additional independent study $\delta=2$ provides more information for CGMM in estimating β_u. The same conclusion can be made for estimating β_v.

4. Discussion

We have proposed a CGMM estimator for using information from datasets in different studies. An asymptotic theorem is established in the case of two studies to illustrate that the CGMM estimator is more efficient than the UGMM estimator using data from one study only. Our simulation studies show that the CGMM estimators can achieve major efficiency gains over the UGMM estimators in cases with two or three studies.

Comparing results for three studies with those for two studies, we conclude that the conclusions are similar, but the CGMM procedure is more complicated with three studies. This is still true if we encounter more studies. The improvement of the CGMM over the UGMM (which basically uses within-study data) increases as the number of studies increases, since more datasets are involved when more studies are considered. However, the derivation of CGMM may be messy when there are many studies and datasets.

We consider linear models for data in both observation and bridge patterns. This is not necessary and can be extended. For example, assumptions (3) and (4) may be replaced by a more general assumption on $E(U \mid \boldsymbol{X})$, either parametric or nonparametric. More research is needed to extend the framework and to explore methods that can handle more general model assumptions.

5. Technical details

Proof of Theorem 2.1: Note that $\bar{\boldsymbol{g}}(\boldsymbol{\theta}_0)$ is a sample average of i.i.d. random vectors with mean zero and finite covariance matrix $\boldsymbol{\Omega}$. Then the Lindeberg–Levy central limit theorem implies

$$\boldsymbol{T}_n=\boldsymbol{\Omega}^{-1/2}n^{1/2}\bar{\boldsymbol{g}}(\boldsymbol{\theta}_0)\rightarrow_d N(\boldsymbol{0},\boldsymbol{I}_{1+3q}). \tag{21}$$

Define a Lagrangian for $\hat{\boldsymbol{\theta}}_c$: $L_n(\boldsymbol{\theta},\boldsymbol{\lambda})=Q_n(\boldsymbol{\theta})-\boldsymbol{c}(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{\lambda}$, where $Q_n(\boldsymbol{\theta})=\bar{\boldsymbol{g}}(\boldsymbol{\theta})^{\mathrm{T}}\widehat{W}\bar{\boldsymbol{g}}(\boldsymbol{\theta})$. In this expression, $\boldsymbol{\lambda}$ is a column vector of undetermined Lagrangian multipliers; these are non-zero when the constraints are binding. The first-order conditions for solution of the constrained optimisation problem are

$$\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}=\begin{bmatrix} n^{1/2}\nabla_{\boldsymbol{\theta}}Q_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_c}-\nabla_{\boldsymbol{\theta}}\boldsymbol{c}(\boldsymbol{\theta})^{\mathrm{T}}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_c}n^{1/2}\boldsymbol{\lambda}\\ -n^{1/2}\boldsymbol{c}(\hat{\boldsymbol{\theta}}_c)\end{bmatrix}. \tag{22}$$

Let $\boldsymbol{G}_n(\boldsymbol{\theta})=n^{-1}\sum_{i=1}^{n}\nabla_{\boldsymbol{\theta}}\,\boldsymbol{g}(\boldsymbol{z}_i,\boldsymbol{\theta})$. Since $\tilde{\boldsymbol{\theta}}_c$ is a consistent estimator of $\boldsymbol{\theta}_0$, $\boldsymbol{G}_n(\tilde{\boldsymbol{\theta}}_c)-\boldsymbol{G}=o_p(1)$ and $\hat{\boldsymbol{W}}-\boldsymbol{\Omega}^{-1}=o_p(1)$, where $o_p(1)$ denotes a sequence of random vectors converging to zero in probability. Using these results and Taylor expansions, we have

$$\begin{aligned}n^{1/2}\bar{\boldsymbol{g}}(\hat{\boldsymbol{\theta}}_c)&=n^{1/2}\bar{\boldsymbol{g}}(\boldsymbol{\theta}_0)-\boldsymbol{G}_n(\hat{\boldsymbol{\theta}}_c)n^{1/2}(\hat{\boldsymbol{\theta}}_c-\boldsymbol{\theta}_0)\\ &=\boldsymbol{\Omega}^{1/2}\boldsymbol{T}_n-\boldsymbol{G}n^{1/2}(\hat{\boldsymbol{\theta}}_c-\boldsymbol{\theta}_0)+o_p(1),\\ n^{1/2}\boldsymbol{c}(\hat{\boldsymbol{\theta}}_c)&=n^{1/2}\boldsymbol{c}(\boldsymbol{\theta}_0)+\boldsymbol{A}n^{1/2}(\hat{\boldsymbol{\theta}}_c-\boldsymbol{\theta}_0)+o_p(1)\\ &=\boldsymbol{A}n^{1/2}(\hat{\boldsymbol{\theta}}_c-\boldsymbol{\theta}_0)+o_p(1)\end{aligned}$$

and

$$n^{1/2}\nabla_{\boldsymbol{\theta}}Q_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_c}=\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}n^{1/2}\bar{\boldsymbol{g}}(\hat{\boldsymbol{\theta}}_c)+o_p(1).$$

Substituting these into the first-order conditions in (22) yields

$$\begin{bmatrix} n^{1/2}(\hat{\boldsymbol\theta}_c-\boldsymbol\theta_0) \\ n^{1/2}\boldsymbol\lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol G^{\mathrm T}\boldsymbol\Omega^{-1}\boldsymbol G & \boldsymbol A^{\mathrm T} \\ \boldsymbol A & \boldsymbol 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol G^{\mathrm T}\boldsymbol\Omega^{-1/2}\boldsymbol T_n \\ \boldsymbol 0 \end{bmatrix} + o_p(1). \tag{23}$$

Applying the formula for the inverse of partitioned matrix (Lu & Shiou, 2002) to (23) and the fact that $\boldsymbol B = (\boldsymbol G^{\mathrm T}\boldsymbol\Omega^{-1}\boldsymbol G)^{-1}$ yields

$$\begin{bmatrix} n^{1/2}(\hat{\boldsymbol\theta}_c-\boldsymbol\theta_0) \\ n^{1/2}\boldsymbol\lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol B - \boldsymbol B\boldsymbol A^{\mathrm T}(\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A\boldsymbol B \\ (\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A\boldsymbol B \end{bmatrix} \times \boldsymbol G^{\mathrm T}\boldsymbol\Omega^{-1/2}\boldsymbol T_n + o_p(1). \tag{24}$$

Note that $\boldsymbol B = (\boldsymbol G^{\mathrm T}\boldsymbol\Omega^{-1}\boldsymbol G)^{-1}$ and $\boldsymbol B = \boldsymbol B^{\mathrm T}$ yield

$$\begin{aligned}
&\{[\boldsymbol B - \boldsymbol B\boldsymbol A^{\mathrm T}(\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A\boldsymbol B]\boldsymbol G^{\mathrm T}\boldsymbol\Omega^{-1/2}\} \\
&\quad\times \{[\boldsymbol B - \boldsymbol B\boldsymbol A^{\mathrm T}(\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A\boldsymbol B]\boldsymbol G^{\mathrm T}\boldsymbol\Omega^{-1/2}\}^{\mathrm T} \\
&\quad= [\boldsymbol B - \boldsymbol B\boldsymbol A^{\mathrm T}(\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A\boldsymbol B]\boldsymbol B^{-1} \\
&\quad\times [\boldsymbol B - \boldsymbol B\boldsymbol A^{\mathrm T}(\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A\boldsymbol B]^{\mathrm T} \\
&\quad= [\boldsymbol I - \boldsymbol B\boldsymbol A^{\mathrm T}(\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A][\boldsymbol B - \boldsymbol B\boldsymbol A^{\mathrm T}(\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A\boldsymbol B] \\
&\quad= \boldsymbol B - 2\boldsymbol B\boldsymbol A^{\mathrm T}(\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A\boldsymbol B \\
&\quad+ \boldsymbol B\boldsymbol A^{\mathrm T}(\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T}(\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A\boldsymbol B \\
&\quad= \boldsymbol B - \boldsymbol B\boldsymbol A^{\mathrm T}(\boldsymbol A\boldsymbol B\boldsymbol A^{\mathrm T})^{-1}\boldsymbol A\boldsymbol B.
\end{aligned}$$

Then, result (5) follows from (21) and (24). ■

Proofs of (7)–(10). Let $\boldsymbol H_n$ be a diagonal matrix whose first $1+2q$ diagonal elements to be $h_n^{1/2}$ and last q diagonal elements to be $(1-h_n)^{1/2}$. Then $\boldsymbol H_n \to \boldsymbol H$, which together with (5) imply result (7).

To complete the proof, we now give derivations of (8) – (10). Write

$$\boldsymbol\Omega = \begin{bmatrix} \boldsymbol\Omega_{11} & \boldsymbol\Omega_{12} & \boldsymbol\Omega_{13} \\ \boldsymbol\Omega^{\mathrm T}{}_{12} & \boldsymbol\Omega_{22} & \boldsymbol\Omega_{23} \\ \boldsymbol\Omega^{\mathrm T}{}_{13} & \boldsymbol\Omega^{\mathrm T}{}_{23} & \boldsymbol\Omega_{33} \end{bmatrix},$$

$$\boldsymbol G = \begin{bmatrix} \boldsymbol G_{11} & \boldsymbol G_{12} & \boldsymbol G_{13} \\ \boldsymbol G^{\mathrm T}{}_{12} & \boldsymbol G_{22} & \boldsymbol G_{23} \\ \boldsymbol G^{\mathrm T}{}_{13} & \boldsymbol G^{\mathrm T}{}_{23} & \boldsymbol G_{33} \end{bmatrix},$$

where $\boldsymbol\Omega_{11}$ is $(1+q)\times(1+q)$, $\boldsymbol\Omega_{22}$ and $\boldsymbol\Omega_{33}$ are both $q\times q$, and the dimension of $\boldsymbol G_{ij}$ is the same as that of $\boldsymbol\Omega_{ij}$. By $h_n \to h \neq 0$, and the definitions of $\boldsymbol\Omega$ and $\boldsymbol g$,

$$\begin{aligned}
\boldsymbol\Omega_{11} &= \lim_{n\to\infty} E\Big[\frac{n^2}{n_1^2} I(\delta=1)(\beta_u U \\
&\quad + \boldsymbol\beta_x{}^{\mathrm T}\boldsymbol X - Y)^2 (U,\boldsymbol X^{\mathrm T})^{\mathrm T}(U,\boldsymbol X^{\mathrm T})\Big] \\
&= h^{-2}E[(\varepsilon_1)^2]E[I(\delta=1)]E[(U,\boldsymbol X^{\mathrm T})^{\mathrm T}(U,\boldsymbol X^{\mathrm T})] \\
&= \frac{\sigma_1^2}{h}\begin{bmatrix} E(U^2) & E(U\boldsymbol X^{\mathrm T}) \\ E(U\boldsymbol X^{\mathrm T})^{\mathrm T} & E(\boldsymbol X\boldsymbol X^{\mathrm T}) \end{bmatrix} \\
&= \frac{\sigma_1^2}{h}\boldsymbol K \quad \text{with } \boldsymbol K = \begin{bmatrix} \boldsymbol\gamma_x^{\mathrm T}\boldsymbol\Sigma_x\boldsymbol\gamma_x + \sigma_b^2 & \boldsymbol\gamma_x^{\mathrm T}\boldsymbol\Sigma_x \\ \boldsymbol\Sigma_x\boldsymbol\gamma_x & \boldsymbol\Sigma_x \end{bmatrix},
\end{aligned}$$

where the second equation follows from the assumption that δ and $(U,\boldsymbol X)$ are independent and ε_1 is independent of $(\delta,U,\boldsymbol X)$, the third equation follows from $E[I(\delta=1)]=h$ and ε_1 has variance σ_1^2, and the last equation follows from (4) and the assumption that ε_b with mean zero is independent of $\boldsymbol X$ so that $E(U\boldsymbol X^{\mathrm T}) = E(\boldsymbol\gamma_x^{\mathrm T}\boldsymbol X\boldsymbol X^{\mathrm T} + \varepsilon_b\boldsymbol X^{\mathrm T}) = \boldsymbol\gamma_x^{\mathrm T}\boldsymbol\Sigma_x$ and $E(U^2) = E((\boldsymbol\gamma_x^{\mathrm T}\boldsymbol X + \varepsilon_b)^2) = \boldsymbol\gamma_x^{\mathrm T}\boldsymbol\Sigma_x\boldsymbol\gamma_x + \sigma_b^2$. Similarly, by $E[I(\delta=1)] = h$ and the assumption that δ and $(U,\boldsymbol X)$ are independent, we have

$$\begin{aligned}
\boldsymbol\Omega_{22} &= \lim_{n\to\infty} E\left[\frac{n^2}{n_1^2} I(\delta=1)(\boldsymbol\gamma_x{}^{\mathrm T}\boldsymbol X - U)^2\boldsymbol X\boldsymbol X^{\mathrm T}\right] \\
&= \frac{\sigma_b^2}{h}\boldsymbol\Sigma_x, \\
\boldsymbol\Omega_{33} &= \lim_{n\to\infty} E\left[\frac{n^2}{n_2^2} I(\delta=2)(\boldsymbol\eta_x{}^{\mathrm T}\boldsymbol X - Y)^2\boldsymbol X\boldsymbol X^{\mathrm T}\right] \\
&= \frac{\sigma_2^2}{1-h}\boldsymbol\Sigma_x, \\
\boldsymbol\Omega_{12} &= \lim_{n\to\infty} E\Big[\frac{n^2}{n_1^2} I(\delta=1)(\beta_u U + \boldsymbol\beta_x{}^{\mathrm T}\boldsymbol X - Y) \\
&\quad \times(\boldsymbol\gamma_x{}^{\mathrm T}\boldsymbol X - U)(U,\boldsymbol X^{\mathrm T})^{\mathrm T}\boldsymbol X^{\mathrm T}\Big] \\
&= h^{-1}E\varepsilon_1 E[\varepsilon_b(U,\boldsymbol X^{\mathrm T})^{\mathrm T}\boldsymbol X^{\mathrm T}] = \boldsymbol 0,
\end{aligned}$$

where the last equation is guaranteed by the assumption that $E\varepsilon_1 = 0$. Since $I(\delta=1)I(\delta=2) = 0$, $\boldsymbol\Omega_{13}$ and $\boldsymbol\Omega_{23}$ are $\boldsymbol 0$. Thus,

$$\boldsymbol\Omega^{-1} = \begin{bmatrix} \dfrac{h}{\sigma_1^2}\boldsymbol K^{-1} & \boldsymbol 0 & \boldsymbol 0 \\ \boldsymbol 0 & \dfrac{h}{\sigma_b^2}\boldsymbol\Sigma_x^{-1} & \boldsymbol 0 \\ \boldsymbol 0 & \boldsymbol 0 & \dfrac{1-h}{\sigma_2^2}\boldsymbol\Sigma_x^{-1} \end{bmatrix}$$

$$\text{with } \boldsymbol K^{-1} = \frac{1}{\sigma_b^2}\begin{bmatrix} 1 & -\boldsymbol\gamma_x^{\mathrm T} \\ -\boldsymbol\gamma_x & \sigma_b^2\boldsymbol\Sigma_x^{-1} + \boldsymbol\gamma_x\boldsymbol\gamma_x^{\mathrm T} \end{bmatrix}.$$

By the definitions of $\boldsymbol G$ and $\boldsymbol g$, partial derivatives corresponding to blocks other than diagonal blocks in $\boldsymbol G$ are zero, i.e., $\boldsymbol G_{12}$, $\boldsymbol G_{13}$, and $\boldsymbol G_{23}$ are $\boldsymbol 0$. By $E[I(\delta=1)] = h$, the assumptions that $h_n \to h \neq 0$, and δ is independent of $(U, \boldsymbol{X})$, we have

$$\boldsymbol{G}_{11} = \lim_{n\to\infty} E\left[\frac{n}{n_1} I(\delta = 1)(U, \boldsymbol{X}^{\mathrm{T}})^{\mathrm{T}}(U, \boldsymbol{X}^{\mathrm{T}})\right] = \boldsymbol{K},$$

$$\boldsymbol{G}_{22} = \lim_{n\to\infty} E\left[\frac{n}{n_1} I(\delta = 1)\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\right] = \boldsymbol{\Sigma}_x,$$

$$\boldsymbol{G}_{33} = \lim_{n\to\infty} E\left[\frac{n}{n_2} I(\delta = 2)\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\right] = \boldsymbol{\Sigma}_x.$$

Combining these results, we obtain that

$$\boldsymbol{B} = (\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{G})^{-1} = \begin{bmatrix} \frac{\sigma_1^2}{h}\boldsymbol{K}^{-1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{\sigma_b^2}{h}\boldsymbol{\Sigma}_x^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \frac{\sigma_2^2}{1-h}\boldsymbol{\Sigma}_x^{-1} \end{bmatrix}. \quad (25)$$

By (25) and the definition of $\boldsymbol{H}$, we have the explicit form of $\boldsymbol{HBH}$ in (8).

Note that $\boldsymbol{A} = [\boldsymbol{\gamma}_x, \boldsymbol{I}_q, \beta_u\boldsymbol{I}_q, -\boldsymbol{I}_q]$. the explicit form of $\boldsymbol{HBA}^{\mathrm{T}}(\boldsymbol{ABA}^{\mathrm{T}})^{-1}\boldsymbol{ABH}$ in (9) – (10) follows from (25), the definition of $\boldsymbol{H}$, and

$$\begin{aligned} \boldsymbol{HBA}^{\mathrm{T}} &= (\boldsymbol{0}, h^{-1/2}\sigma_1^2\boldsymbol{\Sigma}_x^{-1}, h^{-1/2}\sigma_b^2\beta_u\boldsymbol{\Sigma}_x^{-1}, \\ &\quad -(1-h)^{-1/2}\sigma_2^2\boldsymbol{\Sigma}_x^{-1})^{\mathrm{T}}, \\ \boldsymbol{ABA}^{\mathrm{T}} &= [h^{-1}\sigma_1^2 + h^{-1}(\sigma_b\beta_u)^2 + (1-h)^{-1}\sigma_2^2]\boldsymbol{\Sigma}_x^{-1}. \end{aligned}$$

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Research supported by the National Natural Science Foundation of China [grant 11831008], and the U.S. National Science Foundation – Division of Mathematical Sciences (US) [grant DMS-1612873].

Notes on contributors

Menghao Xu is a doctoral candidate in college of statistics, East China Normal University. His main research direction is variable selection, missing data and survival analysis.

Jun Shao is a professor in department of University of Wisconsin-Madison and in college of statistics, East China Normal University. His research covers a wide range of fields, s.t. the jackknife, bootstrap and other resampling methods; variable selection and inference with high dimensional data; sample surveys (variance estimation, imputation for nonrespondents); missing data (nonignorable missing, dropout, semi-parametric methods); longitudinal data analysis with missing data and/or measurement error; medical statistics (clinical trials, personalized medicine, bioequivalence). He is the author of Mathematical Statistics, which is a wildly used graduate textbook covering topics in statistical theory essential for graduate students preparing for work on a Ph.D. degree in statistics.

References

Breslow, N. E., & Holubkov, R. (1997). Maximum likelihood estimation of logistic regression parameters under two-phase, outcome-dependent sampling. *Journal of the Royal Statistical Society, Series B*, *59*(2), 447–461.

Chatterjee, N., Chen, Y.-H., Maas, P., & Carroll, R. J. (2016). Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *Journal of the American Statistical Association*, *111*, 107–117.

Chen, Y.-H., & Chen, H. (2000). A unified approach to regression analysis under double sampling design. *Journal of the Royal Statistical Society, Series B*, *62*, 449–460.

Deville, J. C., & Sarndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, *87*, 376–382.

Engle, R. F., & McFadden, D. L. (1994). *Handbook of econometrics* (Vol. 4). Amsterdam: Elsevier Science, North Holland.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*, 1029–1054.

Hartung, J., Knapp, G., & Sinha, K. B. (2008). *Statistical meta-analysis with applications*. New York, NY: Wiley.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560.

Lawless, J. F., Wild, C. J., & Kalbfleisch, J. D. (1999). Semi-parametric methods for response-selective and missing data problems in regression. *Journal of the Royal Statistical Society, Series B*, *61*, 413–438.

Lu, T. T., & Shiou, S. H. (2002). Inverses of 2 × 2 block matrices. *Computers and Mathematics with Applications*, *43*, 119–129.

Lumley, T., Shaw, P. A., & Dai, J. Y. (2011). Connections between survey calibration estimators and semiparametric models for incomplete data. *International Statistical Review*, *79*, 200–220.

Qin, J., Zhang, H., Li, P., Albanes, D., & Yu, K. (2015). Using covariate-specific disease prevalence information to increase the power of case- control studies. *Biometrika*, *102*, 169–180.

Rao, J. N. K., & Molina, I. (2015). *Small area estimation*. New York, NY: Wiley.

Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, *89*, 846–866.

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage Publications.

Scott, A. J., & Wild, C. J.. (1997). Fitting regression models to case-control data by maximum likelihood. *Biometrika*, *84*, 57–71.

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

Slud, E., & DeMissie, D. (2011). Validity of regression meta-analyses versus pooled analyses of mixed linear models. *Mathematics in Engineering, Science and Aerospace*, *2*, 251–265.

Wu, C. (2003). Optimal calibration estimators in survey sampling. *Biometrika*, *90*, 937–951.

Wu, C., & Sitter, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, *96*, 185–193.