Statistica Sinica **33** (2023), 2209-2231
doi:https://doi.org/10.5705/ss.202020.0330

DISTRIBUTED EMPIRICAL LIKELIHOOD APPROACH TO INTEGRATING UNBALANCED DATASETS

Ling Zhou[1], Xichen She[2] and Peter X.-K. Song[2]

[1]*Southwestern University of Finance and Economics and* [2]*University of Michigan*

Abstract: This paper proposes a distributed empirical likelihood (DEL) method for performing an integrative analysis of multiple data sources with the flexibility of handling either homogeneous or heterogeneous data. The proposed DEL method does not require pooling individual data sets into a centralized operational platform, so the privacy of subject-level information in individual data sources is protected. The DEL method is shown to be almost surely equal to the centralized empirical likelihood approach that would be adopted if individual data sets were combined and stored at one place. We establish the large-sample properties and algorithm convergence of the DEL method. We also illustrate the numerical performance of the DEL method using simulation studies and a real-data example, in which the DEL method is clearly advantageous over the classical meta-estimation method when analyzing unbalanced data sets.

Key words and phrases: ADMM, data integration, data privacy, divide-and-conquer, meta estimation.

1. Introduction

One of the primary tasks in data integration is to combine data from different sources in order to perform an analysis in a comprehensive and unified manner, overcoming the limitations of separate analyses of individual data sources (Lenzerini (2002); Halevy, Rajaraman and Ordille (2006)). For example, consider a situation in which one data source contains observations collected from students in elementary schools, while another source contains measurements from students in middle schools. These two data sources present different ranges of age, and merging them can yield statistics and conclusions generalizable to a broader population than is possible with an analysis of a single data source. In many applications, appropriately integrating multiple data sources enables practitioners to empower their data analysis to answer a scientific hypothesis of interest (Citro (2014); Lohr and Raghunathan (2017); National Academies of Sciences Engineering and Medicine (2017)). Proposed first by Owen (1988, 1991), the empirical

Corresponding author: Peter X.-K. Song, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA. E-mail: pxsong@umich.edu.

likelihood (EL) method is one of the primary statistical methods for estimation and inference, owing to several methodological advantages over other existing methods. For example, the EL method requires minimal parametric model assumptions for data distributions, allows us to construct data-driven confidence regions, and can organically handle auxiliary information. Such properties are particularly appealing for data integration; see Owen (2001) and the references therein for a comprehensive overview of the EL method.

This paper develops a distributed EL (DEL) method for an integrative data analysis when multiple sources of subject-level data are accessible at local sites, but combining individual data sets is prohibited. Specifically, consider K independently sampled data sets from K study sites, $\boldsymbol{W}^{[k]} = \{\boldsymbol{W}_{ki}\}_{i=1}^{n_k}$, for $k = 1, \ldots, K$, with the kth data set consisting of n_k independent and identically distributed (i.i.d.) observations. To analyze the kth data set alone, we estimate a $(q+1)$-dimensional parameter $\boldsymbol{\theta}_{k0} \in \boldsymbol{\Xi} \subset \mathcal{R}^{q+1}$ using the following set of moment conditions: $E\{g_k(\boldsymbol{W}_{ki}; \boldsymbol{\theta}_{k0})\} = \mathbf{0}$, for $i = 1, \ldots, n_k, k = 1, \ldots, K$, where $g_k(\boldsymbol{W}_{ki}; \boldsymbol{\theta})$ is an m_k-dimensional estimating function with $m_k \geq (q+1)$. We consider two scenarios of integrative analyses of practical importance: **(a)** the homogeneity case: both the parameters and the estimating functions are assumed to be the same for all K data sources, that is, $\boldsymbol{\theta}_{k0} \equiv \boldsymbol{\theta}_0$ and $g_k(\cdot) \equiv g(\cdot)$, for all k; and **(b)** the partial homogeneity case: $\boldsymbol{\theta}_{k0} \equiv \boldsymbol{\theta}_0$, where the moment conditions $g_k(\cdot)$ may be different for k. Scenario **(a)** is typically considered in a classical meta-analysis in that the data distributions in the individual data sources are required to be reasonably balanced, so that reliable site-specific statistics may be obtained prior to a meta-estimation. However, this individual balance may fail to hold in practice. For example, for the motivating US kidney transplant data, three binary covariates on transplant recipients (namely, obesity, previous transplant, and hepatitis C serology in region Guam) take the same value with no variability (i.e., zero variances). In this case, the classical meta-type method fails, while our proposed DEL method works properly, because it aggregates the estimating functions across study sites, whereas the meta-estimation directly aggregates site-specific summary statistics.

Few studies have examined the EL methodology in integrative data analyses. Most published works focus on improving the EL method under a single data set by incorporating auxiliary information. For example, Chen and Kim (2014) developed an EL method for a finite population, incorporating features of the sampling design using suitable constraints. Han and Lawless (2019) studied an EL estimation, using a certain summary of auxiliary information to improve the efficiency. Along this line of research, Huang, Qin and Tsai (2016) proposed a

double EL method for estimating a survival time distribution by synthesizing individual-level data from an external data source; see also Chen and Qin (1993); Qin (2000); Chaudhuri, Handcock and Rendall (2008), and Qin et al. (2014), among others. However, the objective of an integrative data analysis differs from those of the aforementioned approaches, because it pertains to a joint inference with multiple data sources, in which each data set is regarded as a primary information source and treated equally in the distributed estimation.

One straightforward solution to data integration is the so-called *centralized* method; that is, all data sets are combined, stored, and processed in a central computing facility. In this case, the EL method may be applied directly to analyze the combined data. The feasibility of such an approach depends on the availability of the combined data, which may be limited by computing facilities, data use agreements, or data privacy considerations. Data privacy is one of the biggest barriers to sharing and combining data. For example, although sharing patients' health records across hospitals is beneficial for medical research, this process requires significant administrative effort and time. As a result, it is tedious and contingent on information censoring, owing to data privacy concerns and regulatory policies related to data sharing. In addition, the computational burden may become substantial as the volume of merged data increases. For omics data, imaging data, and mobile health data generated by modern high-throughput technologies, storing and processing all the data at a single computing facility may be prohibitive. In such cases, divide-and-conquer strategies using distributed computing are popular. Note that a centralized data analysis may be less efficient, or even invalid, when substantial data heterogeneity exists across data sources.

A *meta-analysis* combines site-specific summary statistics (Simmonds and Higgins (2007); Kovalchik (2012)). It uses a weighted average of individual summary statistics to produce an overall estimator for a common population attribute $\boldsymbol{\theta}_0$ of interest:

$$\hat{\boldsymbol{\theta}}_{meta} = \left(\sum_{k=1}^{K} V_k^{-1} \right)^{-1} \left(\sum_{k=1}^{K} V_k^{-1} \hat{\boldsymbol{\theta}}_k \right), \tag{1.1}$$

where $\hat{\boldsymbol{\theta}}_k$ is, *say*, an EL estimate obtained from the kth data source, and $V_k = \mathrm{Var}(\hat{\boldsymbol{\theta}}_k)$ is the corresponding variance. Having no closed-form expression, an approximate variance V_k is given for a large n_k; see Qin and Lawless (1994). The validity of the meta-estimation in (1.1) is easily justified when all n_k are large by using the means of confidence distribution method first proposed by Efron (1993),

and later advocated by Xie and Singh (2013). However, the large-sample behavior may fail with finite sample sizes. For example, in practice, some data sources have small sample sizes, for which approximate variances are no longer reliable. As a result, the meta estimator in (1.1) may perform poorly. The estimation of V_k may be improved using a resampling method, but this incurs excessive computational costs. Furthermore, some variables may show unbalanced distributions when the sample sizes are small across data sources (e.g., measurements of a covariate with little variability). In this case, the meta-estimation method can fail.

We investigate a new approach to distributed computing and inference, which we call the **d**istributed **e**mpirical **l**ikelihood (DEL) method. It performs an EL estimation and inference without pooling individual data sets or sharing subject-level information across data sets. To address scenarios **(a)** and **(b)**, we present two variants of the DEL method: **(i)** the doubly constrained (DOC) DEL (DEL.DOC) method, which is suitable for the homogeneity scenario **(a)**; and **(ii)** the singly constrained (SIC) DEL (DEL.SIC) method, which applies to the partial homogeneity scenario **(b)**. Compared with the conventional meta-estimation method, the DEL methodology offers four advantages. First, in both scenarios, our DEL method still works in the presence of unbalanced distributions of variables across multiple data sources, where the meta-type method fails. Second, in scenario **(a)**, the DEL.DOC estimator is asymptotically equivalent to that obtained by the centralized method (hereafter, referred to as the centralized EL estimator (CEL)) in the "almost surely" sense (Stout (1974)); that is, the DEL.DOC method produces an EL estimator that is equal, *almost surely*, to the CEL estimator. In contrast, the meta-type estimator in (1.1) is only weakly equivalent to the CEL estimator in the sense of "in distribution" (Liu, Liu and Xie (2015)). Third, the DEL.DOC method provides a valid inference under a fixed K and when K diverges to infinity at any rate. Fourth, neither the DEL.DOC method nor the DEL.SIC method requires a direct estimation of the individual variances V_k in the operation of the DEL method.

The rest of the paper is organized as follows. Section 2 introduces the necessary notation and presents the DEL.DOC method under the homogeneity scenario **(a)**. Section 3 extends the DEL method to the partial homogeneity scenario **(b)**, where we also investigate the theoretical properties of the DEL.SIC method. Simulation studies and a real-data example are used to evaluate the two proposed methods in Sections 4 and 5, respectively. Section 6 concludes the paper. All technical details can be found in the Supplementary Material.

2. DEL Method in the Homogeneity Setting

Under scenario **(a)** of homogeneous parameters $\boldsymbol{\theta}_{k0} \equiv \boldsymbol{\theta}_0$ and estimating functions $g_k(\cdot) \equiv g(\cdot)$, for all k, we present the doubly constrained DEL (DEL.DOC) method after a brief review of the centralized EL (CEL) method.

2.1. The CEL method

Consider the kth data set $\boldsymbol{W}^{[k]}$ of n_k *i.i.d.* observations. The parameter of interest $\boldsymbol{\theta}_0 \in \boldsymbol{\Xi}$ satisfies the mean-zero moment condition: $E_{\boldsymbol{\theta}_0}\{g(\boldsymbol{W}_{ki}; \boldsymbol{\theta}_0)\} = \mathbf{0}$. Here, $\boldsymbol{\Xi} \subset \mathcal{R}^{q+1}$ is a compact set containing $\boldsymbol{\theta}_0$ as an interior point. According to the empirical likelihood theory (Owen (1988); Qin and Lawless (1994)), the parameter $\boldsymbol{\theta}_0$ may be estimated by maximizing the following objective function based on data set $\boldsymbol{W}^{[k]}$: $L_k(\boldsymbol{\theta}) = \sup_{p_{ki}} \prod_{i=1}^{n_k} n_k p_{ki}$, subject to $p_{ki} \geq 0$, for $i = 1, \ldots, n_k, \sum_{i=1}^{n_k} p_{ki} = 1$, and $\sum_{i=1}^{n_k} p_{ki} g(\boldsymbol{W}_{ki}; \boldsymbol{\theta}) = \mathbf{0}$. The EL estimator $\hat{\boldsymbol{\theta}}_k = \operatorname{argmax}_{\boldsymbol{\theta} \in \boldsymbol{\Xi}} L_k(\boldsymbol{\theta})$. For each $\boldsymbol{\theta} \in \boldsymbol{\Xi}$, let $\mathcal{T}_{n_k}(\boldsymbol{\theta}) = \cap_{i=1}^{n_k} \{\boldsymbol{t} : \boldsymbol{t}^T g_{ki}(\boldsymbol{\theta}) < 1\}$, where $g_{ki}(\boldsymbol{\theta}) = g(\boldsymbol{W}_{ki}; \boldsymbol{\theta})$. Denote $h_k(\boldsymbol{\theta}, \boldsymbol{t}) = n_k^{-1} \sum_{i=1}^{n_k} \log\big(1 - \boldsymbol{t}^T g_{ki}(\boldsymbol{\theta})\big)$. Here, $\hat{\boldsymbol{\theta}}_k$ may be obtained by solving the saddle-point problem:

$$\hat{\boldsymbol{\theta}}_k = \operatorname*{argmin}_{\boldsymbol{\theta} \in \boldsymbol{\Xi}} \sup_{\boldsymbol{t} \in \mathcal{T}_{n_k}(\boldsymbol{\theta})} h_k(\boldsymbol{\theta}, \boldsymbol{t}). \tag{2.1}$$

When multiple data sources $\boldsymbol{W}^{[k]}$, for $k = 1, \ldots, K$, are available, as long as the capacity of the computing facility permits it, the CEL method can be performed by first aggregating all data sets on a centralized computing site, and then acquiring an estimator $\hat{\boldsymbol{\theta}}_{cen}$ of the form

$$\hat{\boldsymbol{\theta}}_{cen} = \operatorname*{argmin}_{\boldsymbol{\theta} \in \boldsymbol{\Xi}} \sup_{\boldsymbol{t} \in \mathcal{T}_n(\boldsymbol{\theta})} \left\{ \sum_{k=1}^{K} w_k h_k(\boldsymbol{\theta}, \boldsymbol{t}) \right\}, \tag{2.2}$$

where $w_k = n_k/n$, $n = \sum_{k=1}^{K} n_k$, and $\mathcal{T}_n(\boldsymbol{\theta}) = \cap_{k=1}^{K} \mathcal{T}_{n_k}(\boldsymbol{\theta})$.

This CEL estimator $\hat{\boldsymbol{\theta}}_{cen}$ is the same as the classical EL estimator. Thus, all relevant finite-sample and large-sample properties of the empirical likelihood theory hold and apply for $\hat{\boldsymbol{\theta}}_{cen}$ under suitable regularity conditions. Arguably, in the case **(a)** of homogeneity, the CEL method is the method of choice for performing an EL estimation and inference, as long as an adequate computational facility is available, and no data sharing barriers exist across multiple data sources.

2.2. The DEL.DOC method

As noted in Section 1, obtaining the CEL estimator $\hat{\boldsymbol{\theta}}_{cen}$ in (2.2) can be challenging or even prohibitive. To address this challenge, we propose the DEL.DOC method. Inspired by the divide-and-conquer strategy, our solution leads to a new method of distributed estimation and inference in the context of empirical likelihood. Our proposed DEL.DOC method uses data stored on separate locations, without needing to communicate or share subject-level information across individual data sets.

The divide-and-conquer strategy can be implemented by the following doubly constrained EL optimization problem:

$$\begin{aligned} &\min_{\boldsymbol{\Theta}\in\mathcal{D}_{\boldsymbol{\Theta}}} \sup_{\boldsymbol{T}\in\mathcal{D}_{\mathcal{T}_n(\boldsymbol{\Theta})}} \left\{\sum_{k=1}^{K} w_k h_k(\boldsymbol{\theta}_k, \boldsymbol{t}_k)\right\}, \\ &\text{subject to } \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \cdots = \boldsymbol{\theta}_K, \text{ and } \boldsymbol{t}_1 = \boldsymbol{t}_2 = \cdots = \boldsymbol{t}_K, \end{aligned} \tag{2.3}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K)$, $\boldsymbol{T} = (\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_K)$, $\mathcal{D}_{\boldsymbol{\Theta}} = \{\boldsymbol{\Theta} : \boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$ with $\boldsymbol{\theta}_k \in \boldsymbol{\Xi}\}$, and $\mathcal{D}_{\mathcal{T}_n(\boldsymbol{\Theta})} = \{\boldsymbol{T} : \boldsymbol{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_K)$ with $\boldsymbol{t}_k \in \mathcal{T}_{n_k}(\boldsymbol{\theta}_k)\}$. The double constraints refer to the sets of equality constraints on $\boldsymbol{\theta}_k$ and $\boldsymbol{t}_k$, which guarantee that the resulting EL estimator is equivalent to the centralized EL $\hat{\boldsymbol{\theta}}_{cen}$ estimation. This optimization problem (2.3) is different to the conventional meta estimation given in (1.1), where the combined estimator is based directly on individual EL $\hat{\boldsymbol{\theta}}_k$. In contrast, the proposed DEL.DOC method in (2.3) gives a combined estimator using an aggregated objective function, with the equality constraints imposed directly on the parameters $\boldsymbol{\theta}_k$ and $\boldsymbol{t}_k$. Such global constraints lead to a new combined EL estimator of $\boldsymbol{\theta}_0$. Note that without such constraints, the individual parameters $\boldsymbol{\theta}_k$ are estimated separately in parallel, with no data integration. To solve (2.3), we invoke the mean of the *alternating direction method of multipliers* (ADMM) (Boyd et al. (2011)) using the following optimization problem:

$$\begin{aligned} &\min_{\substack{\boldsymbol{\Theta}\in\mathcal{D}_{\boldsymbol{\theta}} \\ \boldsymbol{a}\in\boldsymbol{\Xi}}} \sup_{\substack{\boldsymbol{T}\in\mathcal{D}_{\mathcal{T}_n} \\ \boldsymbol{b}\in\mathcal{T}_n(\boldsymbol{a})}} \left\{\sum_{k=1}^{K} w_k h_k(\boldsymbol{\theta}_k, \boldsymbol{t}_k)\right\}, \\ &\text{subject to } \boldsymbol{\theta}_k \equiv \boldsymbol{a}, \ \boldsymbol{t}_k \equiv \boldsymbol{b}, \ k = 1, \ldots, K. \end{aligned} \tag{2.4}$$

The constrained optimization in (2.4) can be solved using Algorithm 1, the convergence of which is established in Proposition 1. The final output estimates are denoted by $\hat{\boldsymbol{\theta}}_{del.doc} = \boldsymbol{a}^*$; that is, $\boldsymbol{a}^*$ is the converged value of $\boldsymbol{a}^{(s)}$ at iteration s.

Remark 1. In Step 6 of Algorithm 1, the updated $\boldsymbol{\theta}_k^{(s)}$ is obtained by minimizing a modified EL function that is expanded with a quadratic term consisting of the combined estimate $\boldsymbol{a}^{(s)}$, the step size $\boldsymbol{u}_{1,k}^{(s)}$, and the learning rate matrix $\boldsymbol{\Omega}_{1,k}$. The rationale for adding such an expansion to the EL method stems from the fact that this quadratic term pulls separate local estimates toward a common overall estimate, enabling us to effectively "borrow" information from other data sets during iterations, with no need to access subject-level observations of other data sources. In Theorem 1, we show that the estimator $\hat{\boldsymbol{\theta}}_{del.doc}$ is equal, almost surely, to the CEL $\hat{\boldsymbol{\theta}}_{cen}$ obtained in (2.2). In other words, the proposed DEL.DOC method provides the same solution as that obtained by running analyses with the aggregated data sets once. In addition, with this quadratic term, the proposed DEL method overcomes the problem of unbalanced variables, as shown in both the simulation and the real-data examples.

In a classical meta-analysis, the meta-estimator in (1.1) takes a simple one-step weighted average of data set-specific estimators $\hat{\boldsymbol{\theta}}_k$ using (2.1). With these estimates $\hat{\boldsymbol{\theta}}_k$, proper weights (e.g., estimates of the variances V_k) are chosen to ensure the validity of the meta-estimator. In the proposed DEL.DOC method, all separate estimates $\boldsymbol{\theta}_k^{(s)}$ are calibrated iteratively by the quadratic term in the ADMM algorithm step (ii). Therefore, we can use a relatively liberal choice of the learning rate matrix $\boldsymbol{\Omega}_{1,k}$ in the weighted averaging procedure to combine individual estimates. From this point of view, the meta-estimator (1.1) may be regarded essentially as a one-step approximation to the proposed DEL.DOC estimator. Because the DEL.DOC method does not explicitly calculate the variances V_k, it avoids both the difficulty of estimating the variance V_k, and the potential numerical instability caused by a poor estimate of V_k with a small sample size n_k.

2.3. Theoretical properties

In the case of homogeneity, $\boldsymbol{\theta}_{k0} \equiv \boldsymbol{\theta}_0$ and $g_k(\cdot) \equiv g(\cdot)$, for all k, we let $\boldsymbol{S} \equiv \boldsymbol{S}_1 \equiv \cdots \equiv \boldsymbol{S}_K$ and $\boldsymbol{Q} \equiv \boldsymbol{Q}_1 \equiv \cdots \equiv \boldsymbol{Q}_K$, where $\boldsymbol{S}_k$ and $\boldsymbol{Q}_k$ are the sensitivity and variability matrices defined in conditions (C5) and (C3), respectively. Below, we first present the algorithmic convergence of the ADMM algorithm in Proposition 1, which is proposed to implement the DEL.DOC method. Its proof is given in the Supplementary Material.

Proposition 1. (Algorithmic convergence of the ADMM for the DEL.DOC method). *Denote* $\boldsymbol{U}_r = (\boldsymbol{u}_{r,1}, \ldots, \boldsymbol{u}_{r,K})$, *for* $r = 1, 2$. *Let* $L_{01}(\boldsymbol{\Theta}, \mathbf{T}, \boldsymbol{a}, \boldsymbol{U}_1) = \sum_{k=1}^K w_k h_k(\boldsymbol{\theta}_k, \boldsymbol{t}_k) + \sum_{k=1}^K \boldsymbol{u}_{1,k}^T \boldsymbol{\Omega}_{1,k}(\boldsymbol{\theta}_k - \boldsymbol{a})$, *and* $L_{02}(\boldsymbol{\Theta}, \mathbf{T}, \boldsymbol{b}, \boldsymbol{U}_2) = -\sum_{k=1}^K w_k h_k$

$(\boldsymbol{\theta}_k, \boldsymbol{t}_k) + \sum_{k=1}^{K} \boldsymbol{u}_{2,k}^T \boldsymbol{\Omega}_{2,k}(\boldsymbol{t}_k - \boldsymbol{b})$. *Suppose that (i) there exists* $(\boldsymbol{\Theta}^*, \mathbf{T}^*, \boldsymbol{a}^*, \boldsymbol{b}^*, \boldsymbol{U}_1^*, \boldsymbol{U}_2^*)$, *for which the following inequalities hold for all* $\boldsymbol{\Theta} \in \mathcal{D}_{\boldsymbol{\theta}}, \mathbf{T} \in \mathcal{D}_{\mathcal{T}_n}, \boldsymbol{a} \in \Xi, \boldsymbol{b} \in \mathcal{T}_n(\boldsymbol{a}), \boldsymbol{U}_1 \in \mathcal{R}^{(q+1)\times K}$, *and* $\boldsymbol{U}_2 \in \mathcal{R}^{m_1+\cdots+m_K}$: $L_{01}(\boldsymbol{\Theta}^*, \mathbf{T}, \boldsymbol{a}^*, \boldsymbol{U}_1) \le L_{01}(\boldsymbol{\Theta}^*, \mathbf{T}, \boldsymbol{a}^*, \boldsymbol{U}_1^*) \le L_{01}(\boldsymbol{\Theta}, \mathbf{T}, \boldsymbol{a}, \boldsymbol{U}_1^*)$, *and* $L_{02}(\boldsymbol{\Theta}, \mathbf{T}^*, \boldsymbol{b}^*, \boldsymbol{U}_2) \le L_{02}(\boldsymbol{\Theta}, \mathbf{T}^*, \boldsymbol{b}^*, \boldsymbol{U}_2^*) \le L_{02}(\boldsymbol{\Theta}, \mathbf{T}, \boldsymbol{b}, \boldsymbol{U}_2^*)$; *and (ii) there exist initial values* $(\boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}, \boldsymbol{U}_1^{(0)}, \boldsymbol{U}_2^{(0)})$, *such that*

$$\sum_{k=1}^{K} \left\{ \left(\boldsymbol{u}_{1,k}^{(0)} - \boldsymbol{u}_{1,k}^*\right)^T \boldsymbol{\Omega}_{1,k} \left(\boldsymbol{u}_{1,k}^{(0)} - \boldsymbol{u}_{1,k}^*\right) + \left(\boldsymbol{a}^{(0)} - \boldsymbol{a}^*\right)^T \boldsymbol{\Omega}_{1,k} \left(\boldsymbol{a}^{(0)} - \boldsymbol{a}^*\right) \right\} \le M_1,$$

$$\sum_{k=1}^{K} \left\{ \left(\boldsymbol{u}_{2,k}^{(0)} - \boldsymbol{u}_{2,k}^*\right)^T \boldsymbol{\Omega}_{2,k} \left(\boldsymbol{u}_{2,k}^{(0)} - \boldsymbol{u}_{2,k}^*\right) + \left(\boldsymbol{b}^{(0)} - \boldsymbol{b}^*\right)^T \boldsymbol{\Omega}_{2,k} \left(\boldsymbol{b}^{(0)} - \boldsymbol{b}^*\right) \right\} \le M_2,$$

for some positive constants M_1 *and* M_2. *Then, as the iteration* $s \to \infty$, *we have*

$$\max_k \left\|\boldsymbol{\theta}_k^{(s)} - \boldsymbol{\theta}_k^*\right\|_2 \to 0, \quad \textit{and} \quad \|\boldsymbol{a}^{(s)} - \boldsymbol{a}^*\|_2 \to 0,$$

$$\max_k \left\|\boldsymbol{t}_k^{(s)} - \boldsymbol{t}_k^*\right\|_2 \to 0, \quad \textit{and} \quad \|\boldsymbol{b}^{(s)} - \boldsymbol{b}^*\|_2 \to 0.$$

Proposition 1 indicates that, according to condition (ii), both the initial values and the learning rate matrices $\boldsymbol{\Omega}_{r,k}$ could affect the performance of the DEL.DOC method. When the number of data sources, K, is large, proper initial values are required to ensure the convergence of the algorithm. Theorem 1 gives the almost sure equality between $\hat{\boldsymbol{\theta}}_{del.doc}$ and $\hat{\boldsymbol{\theta}}_{cen}$, as well as the asymptotic properties of $\hat{\boldsymbol{\theta}}_{del.doc}$.

Theorem 1. *Assume Conditions* (C1)–(C5) *in the Appendix hold.*

(a) If the initial values satisfy $\boldsymbol{a}^{(0)} = \boldsymbol{\theta}_0 + O(K^{-1/2})$, $\boldsymbol{b}^{(0)} = O(K^{-1/2})$, $\boldsymbol{U}_1^{(0)} = \mathbf{0}$, *and* $\boldsymbol{U}_2^{(0)} = \mathbf{0}$, *then Prob*$(\hat{\boldsymbol{\theta}}_{del.doc} = \hat{\boldsymbol{\theta}}_{cen}) = 1$.

(b) If the conditions of the initial values given in part (a) hold, as $n \to \infty$,

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\theta}}_{del.doc} - \boldsymbol{\theta}_0 \\ \hat{\boldsymbol{t}}_{del.doc} - \mathbf{0} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{J}_{del.doc}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}^{-1} - \boldsymbol{Q}^{-1}\boldsymbol{S}\boldsymbol{J}_{del.doc}^{-1}\boldsymbol{S}^T\boldsymbol{Q}^{-1} \end{bmatrix}\right),$$

where the Godambe information matrix $\boldsymbol{J}_{del.doc} = \boldsymbol{S}^T\boldsymbol{Q}^{-1}\boldsymbol{S}$.

The proof of Theorem 1 is given in the Supplementary Material. According to Theorem 1, when the number of data sources, K, is fixed, the initial values of $\boldsymbol{a}$ and $\boldsymbol{b}$ may be chosen easily, *say*, as certain reasonable constant vectors. However, if K increases, the initial value of $\boldsymbol{a}^{(0)}$ should be chosen to be close to the true value $\boldsymbol{\theta}_0$ at the rate $O(K^{-1/2})$.

Algorithm 1. Algorithm for DEL.DOC Method.

procedure DEL.DOC (round, ϵ, $\mathbf{\Omega}_{r,k}$, $r = 1, 2$, $k = 1, \ldots, K$) ▷ Input

Determine the initial local estimates $\{\boldsymbol{\theta}_k^{(0)}\}_{k=1}^K$, $\{\boldsymbol{t}_k^{(0)}\}_{k=1}^K$ for each data set (e.g., individual estimates $\hat{\boldsymbol{\theta}}_k$ and corresponding $\hat{\boldsymbol{t}}_k$), initial global estimates $\boldsymbol{a}^{(0)}$ and $\boldsymbol{b}^{(0)}$ (e.g., simple averages of local estimates), and initial step sizes $\{\boldsymbol{u}_{1k}^{(0)}\}_{k=1}^K$, $\{\boldsymbol{u}_{2k}^{(0)}\}_{k=1}^K$ (e.g., all set to $\mathbf{0}$).

$iter \leftarrow 1$

$converge \leftarrow$ FALSE

while !$converge$ & $iter \leq round$ **do**

Update $\boldsymbol{\theta}_k, \boldsymbol{t}_k, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{u}_{1,k}, \boldsymbol{u}_{2,k}$ by

$$\boldsymbol{\theta}_k^{(s+1)} = \underset{\boldsymbol{\theta}_k}{\operatorname{argmin}} \left\{ w_k h_k(\boldsymbol{\theta}_k, \boldsymbol{t}_k^{(s)}) + \frac{1}{2} \left(\boldsymbol{\theta}_k - \boldsymbol{a}^{(s)} + \boldsymbol{u}_{1,k}^{(s)} \right)^T \mathbf{\Omega}_{1,k} \left(\boldsymbol{\theta}_k - \boldsymbol{a}^{(s)} + \boldsymbol{u}_{1,k}^{(s)} \right) \right\};$$

$$\boldsymbol{t}_k^{(s+1)} = \underset{\boldsymbol{t}_k}{\operatorname{argmin}} \left\{ - w_k h_k(\boldsymbol{\theta}_k^{(s+1)}, \boldsymbol{t}_k) + \frac{1}{2} \left(\boldsymbol{t}_k - \boldsymbol{b}^{(s)} + \boldsymbol{u}_{2,k}^{(s)} \right)^T \mathbf{\Omega}_{2,k} \left(\boldsymbol{t}_k - \boldsymbol{b}^{(s)} + \boldsymbol{u}_{2,k}^{(s)} \right) \right\};$$

$$\boldsymbol{a}^{(s+1)} = \left(\sum_{k=1}^K \mathbf{\Omega}_{1,k} \right)^{-1} \left(\sum_{k=1}^K \mathbf{\Omega}_{1,k} \boldsymbol{\theta}_k^{(s+1)} \right), \boldsymbol{b}^{(s+1)} = \left(\sum_{k=1}^K \mathbf{\Omega}_{2,k} \right)^{-1} \left(\sum_{k=1}^K \mathbf{\Omega}_{2,k} \boldsymbol{t}_k^{(s+1)} \right);$$

$$\boldsymbol{u}_{1,k}^{(s+1)} = \boldsymbol{u}_{1,k}^{(s)} + \boldsymbol{\theta}_k^{(s+1)} - \boldsymbol{a}^{(s+1)}, \ \boldsymbol{u}_{2,k}^{(s+1)} = \boldsymbol{u}_{2,k}^{(s)} + \boldsymbol{t}_k^{(s+1)} - \boldsymbol{b}^{(s+1)};$$

where $\mathbf{\Omega}_{r,k}$, $r = 1, 2$, $k = 1, \ldots, K$ are prespecified $(q+1) \times (q+1)$-dimensional positive-definite symmetric matrices pertaining to the algorithmic convergence rate.

if $\max_k \left\| \boldsymbol{\theta}_k^{(s+1)} - \boldsymbol{\theta}_k^{(s)} \right\|_2 < \epsilon$ and $\max_k \left\| \boldsymbol{t}_k^{(s+1)} - \boldsymbol{t}_k^{(s)} \right\|_2 < \epsilon$ for a prespecified threshold ϵ, *say* 10^{-4}, **then** $converge \leftarrow$ TRUE

$iter \leftarrow iter + 1$

return $\mathbf{\Theta}, \boldsymbol{T}, \boldsymbol{a}, \boldsymbol{b}$.

▷ Output

3. DEL Method in the Partial Homogeneity Setting

Now, we consider scenario **(b)** of $\boldsymbol{\theta}_{k0} \equiv \boldsymbol{\theta}_0$, but where $g_k(\cdot)$ varies over k. We propose the DEL.SIC method to perform the meta EL estimation. The asymptotic properties of the DEL.SIC method are thoroughly investigated in this section.

3.1. DEL.SIC method

The DEL.DOC method imposes two sets of equality constraints, $\boldsymbol{\theta}_k \equiv \boldsymbol{a}$ and $\boldsymbol{t}_k \equiv \boldsymbol{b}$, for $k = 1, \ldots, K$, in order to enforce homogeneity on both the parameters $\boldsymbol{\theta}_{k0}$ and the estimation functions $g_k(\cdot)$. This is the typical situation considered routinely in a classical meta-analysis. However, note that in most

practical studies, the second set of constraints on g_k may be rarely satisfied, owing to inter-data heterogeneity. For example, in the setting of instrumental variable models (Imbens (2002); Newey and Smith (2004)), different data sets may have their own instrumental variables, leading to different numbers of moment constraints. Thus, $g_k(\cdot)$ may vary in terms of its form and dimension. In this case, $\boldsymbol{t}_k$ are of different dimensions; hence, it is impossible to require $\boldsymbol{t}_k \equiv \boldsymbol{b}$. A natural modification of the DEL.DOC method is to remove the second set of constraints on $\boldsymbol{t}_k$, resulting in a relaxed DEL method with only a single set of constraints $\boldsymbol{\theta}_k \equiv \boldsymbol{a}$. Interestingly, as shown in Theorem 3, this generalization helps to handle the homogeneity case **(a)**, in which the resulting DEL.SIC estimator appears to have smaller asymptotic variances than those obtained by the DEL.DOC method.

Specifically, the DEL.SIC method relaxes step 6 in Algorithm 1, given as follows:

$$\begin{aligned}
\boldsymbol{\theta}_k^{(s+1)} &= \underset{\boldsymbol{\theta}_k}{\operatorname{argmin}} \left\{ w_k f_k(\boldsymbol{\theta}_k) + \frac{1}{2} \left(\boldsymbol{\theta}_k - \boldsymbol{a}^{(s)} + \boldsymbol{u}_k^{(s)} \right)^T \boldsymbol{\Omega}_k \left(\boldsymbol{\theta}_k - \boldsymbol{a}^{(s)} + \boldsymbol{u}_k^{(s)} \right) \right\}; \\
\boldsymbol{a}^{(s+1)} &= \left(\sum_{k=1}^K \boldsymbol{\Omega}_k \right)^{-1} \left(\sum_{k=1}^K \boldsymbol{\Omega}_k \boldsymbol{\theta}_k^{(s+1)} \right); \ \boldsymbol{u}_k^{(s+1)} = \boldsymbol{u}_k^{(s)} + \boldsymbol{\theta}_k^{(s+1)} - \boldsymbol{a}^{(s+1)}; \quad (3.1)
\end{aligned}$$

where $f_k(\boldsymbol{\theta}) = \sup_{\boldsymbol{t} \in \mathcal{T}_{n_k}(\boldsymbol{\theta})} \left\{ n_k^{-1} \sum_{i=1}^{n_k} \log\left(1 - \boldsymbol{t}^T g_{ki}(\boldsymbol{\theta})\right) \right\}$, and $\boldsymbol{\Omega}_k$ is a prespecified learning rate. Denote the converged value of $\boldsymbol{a}^{(s)}$ from the relaxed ADMM algorithm as $\hat{\boldsymbol{\theta}}_{del.sic}$. The convergence of the ADMM algorithm implemented for the DEL.SIC method via (3.1), stated in Proposition 2 in the Supplementary Material, can be proved using similar arguments to those in Proposition 1. See the Supplementary Material.

Remark 2. The DEL.SIC method provides an interesting interpretation as a new approach to aggregating ELs. Consider an objective function $L(\boldsymbol{\theta})$ that aggregates K individual ELs, each having its own set of p_{ki}: $L(\boldsymbol{\theta}) = \sup_{p_{ki}} \prod_{k=1}^K \prod_{i=1}^{n_k} n_k p_{ki}$, subject to $p_{ki} \geq 0, \sum_{i=1}^{n_k} p_{ki} = 1$, and $\sum_{i=1}^{n_k} p_{ki} g_{ki}(\boldsymbol{\theta}) = 0$, for $k = 1, \ldots, K$. In this formulation, different $g_k(\cdot)$ are allowed in the aggregation. Let $\check{\boldsymbol{\theta}}_{cen}$ be the EL solution to the above objective function obtained using a centralized computation method. After some simple calculations, we have that

$$\check{\boldsymbol{\theta}}_{cen} = \underset{\boldsymbol{\theta} \in \Xi}{\operatorname{argmin}} \sum_{k=1}^K w_k f_k(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta} \in \Xi}{\operatorname{argmin}} \left\{ \sum_{k=1}^K w_k \sup_{\boldsymbol{t}_k \in \mathcal{T}_{n_k}(\boldsymbol{\theta})} h_k(\boldsymbol{\theta}, \boldsymbol{t}_k) \right\}. \quad (3.2)$$

The relaxed ADMM algorithm (3.1) enables us to search for the solution to the reformulated optimality problem given in (3.2), with the following con-

straints: $\min_{\boldsymbol{\Theta}\in\mathcal{D}_{\boldsymbol{\Theta}}}\sup_{\boldsymbol{T}\in\mathcal{D}_{\mathcal{T}_n(\boldsymbol{\Theta})}}\{\sum_{k=1}^K w_k h_k(\boldsymbol{\theta}_k,\boldsymbol{t}_k)\}$, subject to $\boldsymbol{\theta}_1=\cdots=\boldsymbol{\theta}_K$. It coincides with the proposed estimator $\hat{\boldsymbol{\theta}}_{del.sic}$. The next subsection presents the asymptotic properties of the CLE $\check{\boldsymbol{\theta}}_{cen}$ and the almost sure equality between $\check{\boldsymbol{\theta}}_{cen}$ and $\hat{\boldsymbol{\theta}}_{del.sic}$.

3.2. Asymptotic properties

The asymptotic consistency and normality of $\check{\boldsymbol{\theta}}_{cen}$ are established in Theorems 2 and 3, respectively, both of which are new, owing to the relaxation with varying $g_k(\cdot)$ in the EL method. Theorem 5 presents the almost sure equality between $\check{\boldsymbol{\theta}}_{cen}$ and $\hat{\boldsymbol{\theta}}_{del.sic}$. Theorem 6 concerns the asymptotic distribution of the EL ratio statistic used for a distributed inference. All proofs are given in the Supplementary Material. Let $n_{\min}\overset{def}{=}\min\{n_1,\ldots,n_K\}$.

Theorem 2. (Consistency of $\check{\boldsymbol{\theta}}_{cen}$). *If Conditions* (C1)–(C3) *in the Appendix hold, then* $\check{\boldsymbol{\theta}}_{cen}\xrightarrow{p}\boldsymbol{\theta}_0$ *as* $n_{\min}\to\infty$. *Moreover, let* $\check{g}_k=n_k^{-1}\sum_{i=1}^{n_k}g_k(\boldsymbol{W}_{ki};\check{\boldsymbol{\theta}}_{cen})$ *and* $\check{\boldsymbol{t}}_k=\operatorname{argmax}_{\boldsymbol{t}\in\mathcal{T}_{n_k}(\check{\boldsymbol{\theta}}_{cen})}w_k h_k(\check{\boldsymbol{\theta}}_{cen},\boldsymbol{t})$, *We have (i)* $\check{g}_k=O_p(n_k^{-1/2})$; *(ii)* $\check{\boldsymbol{t}}_k$ *exists with probability approaching one; and (iii)* $\check{\boldsymbol{t}}_k=O_p(n_k^{-1/2})$.

The estimation consistency above is well established in the literature.

Theorem 3. (Asymptotic normality of $\check{\boldsymbol{\theta}}_{cen}$). *Under Conditions* (C1)–(C5) *in the Appendix, if* $K=O(n^{1/2-\delta})$, *for some* $0<\delta\le 1/2$ *and* $n_{\min}\to\infty$, *we have*

$$\begin{pmatrix}\sqrt{n}(\check{\boldsymbol{\theta}}_{cen}-\boldsymbol{\theta}_0)\\ \sqrt{n_1}(\check{\boldsymbol{t}}_1-\boldsymbol{0})\\ \vdots\\ \sqrt{n_K}(\check{\boldsymbol{t}}_K-\boldsymbol{0})\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(\boldsymbol{0},\begin{bmatrix}\boldsymbol{J}_{del.sic}^{-1} & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{Q}_1^{-1}-w_1\boldsymbol{P}_{11} & \cdots & -w_1\boldsymbol{P}_{1K}\\ \vdots & \vdots & \cdots & \vdots\\ \boldsymbol{0} & -w_K\boldsymbol{P}_{K1} & \cdots & \boldsymbol{Q}_K^{-1}-w_K\boldsymbol{P}_{KK}\end{bmatrix}\right),$$

where $\boldsymbol{J}_{del.sic}=\lim_{n_{\min}\to\infty}(\sum_{k=1}^K w_k\boldsymbol{S}_k^T\boldsymbol{Q}_k^{-1}\boldsymbol{S}_k)$ *and* $\boldsymbol{P}_{ij}=\boldsymbol{Q}_i^{-1}\boldsymbol{S}_i\boldsymbol{J}_{del.sic}^{-1}\boldsymbol{S}_j^T\boldsymbol{Q}_j^{-1}$.

Theorem 4 compares the estimation efficiency between $\check{\boldsymbol{\theta}}_{cen}$ in (3.2) and $\hat{\boldsymbol{\theta}}_{cen}$ in (2.2).

Theorem 4. *If the dimensions of* $g_k(\cdot)$ *are the same, that is,* $m_1=\cdots=m_K$, *then* $\boldsymbol{J}_{del.sic}\ge\boldsymbol{J}_{cen}$, *where* $\boldsymbol{J}_{cen}=\lim_{n\to\infty}\{(\sum_{k=1}^K w_k\boldsymbol{S}_k^T)(\sum_{k=1}^K w_k\boldsymbol{Q}_k)^{-1}\times(\sum_{k=1}^K w_k\boldsymbol{S}_k)\}$ *is the asymptotic covariance of the centralized EL estimator* $\hat{\boldsymbol{\theta}}_{cen}$ *in* (2.2). *The inequality* $\ge$ *is in the Löwner's sense. Moreover, the equal variance occurs if and only if* $\boldsymbol{S}_k^T\equiv\boldsymbol{Q}_k$, *for* $k=1,\ldots,K$, *or* $\boldsymbol{S}_1^T\boldsymbol{Q}_1^{-1}=\cdots=\boldsymbol{S}_K^T\boldsymbol{Q}_K^{-1}$.

Theorem 4 implies the following interesting results.

Corollary 1. *In the homogeneity case* ***(a)****,* $\hat{\boldsymbol{\theta}}_{cen}$ *(equivalent to* $\hat{\boldsymbol{\theta}}_{del.doc}$ *a.s.) and* $\check{\boldsymbol{\theta}}_{cen}$ *have the same estimation efficiency.*

Theorem 5. *Under the same conditions as those of Theorem* 3*, if the initial values* $\boldsymbol{a}^{(0)} = \boldsymbol{\theta}_0 + O(K^{-1/2})$ *and* $\boldsymbol{u}^{(0)} = \mathbf{0}$*, then Prob*$(\hat{\boldsymbol{\theta}}_{del.sic} = \check{\boldsymbol{\theta}}_{cen}) = 1$*.*

The proof of Theorem 5 is similar to that of Theorem 1 by taking $(\boldsymbol{\Theta}^*, \boldsymbol{a}^*, \boldsymbol{u}^*) = (\check{\boldsymbol{\Theta}}_{cen}, \check{\boldsymbol{\theta}}_{cen}, \mathbf{0})$, and thus is omitted here.

Remark 3. Under Theorem 5, it is easy to see Theorems 2–4 also hold for $\hat{\boldsymbol{\theta}}_{del.sic}$. According to Theorem 4, when the DEL.DOC method is applicable, the asymptotic variance of $\hat{\boldsymbol{\theta}}_{del.sic}$(or $\check{\boldsymbol{\theta}}_{cen}$) is no larger than the variance of $\hat{\boldsymbol{\theta}}_{del.doc}$ (or $\hat{\boldsymbol{\theta}}_{cen}$).

Theorem 3 indicates that the asymptotic normality of $\hat{\boldsymbol{\theta}}_{del.sic}$ holds when the number of data sets K grows at a slower rate than $\sqrt{n}$, and all sample sizes of the data sets should tend to infinity. This order of K, $O(n^{1/2-\delta})$, in Theorem 3 may be relaxed by invoking a high-order bias correction technique. This is because aggregating the EL estimates across K data sets reduces the order of the variance, but does not reduce the order of the estimation bias. Denote $\boldsymbol{\Upsilon} = (\boldsymbol{\theta}^T, \boldsymbol{t}_1^T, \ldots, \boldsymbol{t}_K^T)^T$. Following Firth (1993), we consider a qth-order Taylor expansion of $Q(\boldsymbol{\Upsilon})$, the estimating function for solving the optimization given in (3.2), from which we construct the high-order bias corrected estimator $\tilde{\boldsymbol{\Upsilon}}_q = \check{\boldsymbol{\Upsilon}} + S^{-1}\boldsymbol{r}_q$, where $\check{\boldsymbol{\Upsilon}} = (\check{\boldsymbol{\theta}}_{cen}^T, \check{\boldsymbol{t}}_1^T, \ldots, \check{\boldsymbol{t}}_K^T)^T$, S is a $p \times p$ matrix $E[\nabla Q(\boldsymbol{\Upsilon}_0)]$, and $\boldsymbol{r}_q$ is a certain p-element vector yielded from the q-order Taylor expansion. We can show that the order of the estimation bias for this new estimator $\tilde{\boldsymbol{\Upsilon}}_q$ is $K^{q+2}n^{-(q+1)} + (K/n)^{(q+1)/2}$. In order for this bias to be asymptotically ignorable, it is sufficient to set $K = o(n^{1-1.5/(q+2)})$, for $q \geq 1$, for which the resulting bias is at a higher order than $n^{-1/2}$. The details can be found in the Supplementary Material.

Additionally, Theorem 3 implies that (a) in the DEL.SIC method, $\hat{\boldsymbol{\theta}}_{del.sic}$ and $\hat{\boldsymbol{t}}_k$ are asymptotically independent, (b) $\hat{\boldsymbol{\theta}}_{del.sic}$(or $\check{\boldsymbol{\theta}}_{cen}$) has the same convergence rate as those of $\hat{\boldsymbol{\theta}}_{del.doc}$ and $\hat{\boldsymbol{\theta}}_{cen}$, and (c) from Theorem 1, the asymptotic variance of the individual estimator for $\boldsymbol{t}$ based only on the kth data set is $n_k^{-1}[\boldsymbol{Q}_k^{-1} - \boldsymbol{Q}_k^{-1}\boldsymbol{S}_k\left(\boldsymbol{S}_k^T\boldsymbol{Q}_k^{-1}\boldsymbol{S}_k\right)^{-1}\boldsymbol{S}_k^T\boldsymbol{Q}_k^{-1}]$, which is different from that of $\hat{\boldsymbol{t}}_k$ in the DEL.SIC method. Here, $\hat{\boldsymbol{t}}_k = \operatorname{argmax}_{\boldsymbol{t}\in\mathcal{T}_{n_k}(\hat{\boldsymbol{\theta}}_{del.sic})} w_k h_k(\hat{\boldsymbol{\theta}}_{del.sic}, \boldsymbol{t})$.

As in the EL literature, we consider the following EL ratio (ELR) statistic for testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$: $W_E(\boldsymbol{\theta}_0) = 2n\{\mathcal{L}(\hat{\boldsymbol{\theta}}_{del.sic}, \hat{\mathbf{T}}) - \mathcal{L}(\boldsymbol{\theta}_0, \mathbf{T}_0)\}$, where $\boldsymbol{t}_{k0} = \operatorname{argmax}_{\boldsymbol{t}\in\mathcal{T}_{n_k}(\boldsymbol{\theta}_0)} h_k(\boldsymbol{\theta}_0, \boldsymbol{t})$, for $k = 1, \ldots, K$, and $\mathcal{L}(\boldsymbol{\theta}, \mathbf{T}) = \sum_{k=1}^K w_k h_k(\boldsymbol{\theta}, \boldsymbol{t}_k)$, with $\mathbf{T}_0 = (\boldsymbol{t}_{10}, \boldsymbol{t}_{20}, \ldots, \boldsymbol{t}_{K0})$.

Theorem 6. *Under the conditions of Theorem* 3 *and under the null hypothesis* $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, $W_E(\boldsymbol{\theta}_0) \xrightarrow{d} \chi^2_{q+1}$ *as* $n_{\min} \to \infty$ *and* $\sum_{k=1}^{K} n_k^{-1/2} \to 0$.

Note that when all data sets are of equal size, that is, $n_k \equiv \tilde{n}$, Theorem 6 requires that $\tilde{n} \to \infty$ and $K = O(\tilde{n}^{1/3-\delta})$, for some $0 < \delta \leq 1/3$. To construct a confidence interval, we consider the profile ELR statistic proposed in Qin and Lawless (1994, Corollary 5). Let $\boldsymbol{\theta}^T = (\theta_1, \boldsymbol{\theta}_2^T)^T$, where θ_1 is the parameter of interest and $\boldsymbol{\theta}_2$ is the subvector of the nuisance parameters. For $H_0 : \theta_1 = \theta_{10}$, the profile ELR test statistic is $W_2 = 2n(\mathcal{L}(\hat{\boldsymbol{\theta}}_{del.sic}, \hat{\boldsymbol{T}}) - \mathcal{L}(\theta_{10}, \hat{\boldsymbol{\theta}}_2^{null}, \hat{\boldsymbol{T}}^{null}))$, where $(\hat{\boldsymbol{\theta}}_2^{null}, \hat{\boldsymbol{T}}^{null}) = \arg\min_{\boldsymbol{\theta}_2} \sup_{\boldsymbol{T}} \{\sum_{k=1}^{K} w_k h_k(\boldsymbol{\theta}_k, \boldsymbol{t}_k)\}$, subject to $\theta_{1k} \equiv \theta_{10}$, and $\boldsymbol{\theta}_{2k} \equiv \boldsymbol{\theta}_2$, for $k = 1, \ldots, K$, where $\boldsymbol{\theta}_2$ is a common subvector across K data sets to be estimated. Following Theorem 6, $W_2 \to \chi^2_1$ as $n_{\min} \to \infty$. The details can be found in the Supplementary Material. Thus, a $(1-\alpha)100\%$ confidence interval satisfies $Prob[W_2 \leq \chi^2_1(1-\alpha)] = 1-\alpha$. We numerically find the lower and upper limits of a 95% confidence interval, $[\hat{\theta}_1^l, \hat{\theta}_1^u]$, as follows. First, (i) calculate the 0.05 upper quantile of χ^2_1, denoted by $q_{0.05}$, such that $F(q_{0.05}) = 0.95$, where F is the chi-square distribution function with degree one. Second, solve the equation $W_2(\theta_1) = q_{0.05}$, giving two roots $\hat{\theta}_1^u > \hat{\theta}_1^l$. Then, the interval length is calculated as $|\hat{\theta}_1^u - \hat{\theta}_1^l|$.

4. Examples and Numerical Illustration

Example 1. (Estimation with auxiliary information). In this example, we revisit the two-sample problem considered in Qin and Lawless (1994) from the perspective of a distributed EL estimation and inference. We simulate *i.i.d.* observations of trivariate random variables (X, Y, Z) according to the following models: $X \sim$ Weibull$(2, 1)$, $Y = X + \epsilon_1$, and $Z = X + \epsilon_2$, where $(\epsilon_1, \epsilon_2) \sim$ BVN$(0, 0, 1, 1, -0.6)$ and (ϵ_1, ϵ_2) are independent of X. We consider a scenario of $K = 3$ data sets with unequal sample sizes. The parameter of interest is the tail probability $p = P(X \geq \xi_{0.9})$, with $\xi_{0.9}$ being the 0.9-quantile of Weibull$(2, 1)$. Clearly, the true value $p_0 = 0.1$. Suppose we have additional information from external sources (Y, Z) of, *say*, $E(Y) = \mu_{y0} = 2$ and $E(X) = E(Z)$. Such auxiliary information is relevant to the parameter p, because X is correlated with the two auxiliary variables (Y, Z). To incorporate this information in the estimation of p, we set up a joint estimating function of the form $g(x, y, z; p) = (\mathbf{1}(x \geq \xi_{0.9}) - p, y - \mu_{y0}, x - z)^T$, which is unbiased because $E\{g(X, Y, Z; p_0)\} = \mathbf{0}$. We assess and compare the performance of the DEL.DOC method with that of the following methods: (i) the naive sample proportion estimator with no use of auxiliary information; (ii) the CEL estimator based on the combined data set, and (iii) the classical meta

estimator using the inverse variance weighting. For the meta estimator, the individual variances V_k are separately estimated using the EL asymptotic variance given in Qin and Lawless (1994). Table 1 shows the summary results over 2,000 replicates under different sample sizes. The results include the average bias, empirical standard error (ESE), average asymptotic standard error (ASE), coverage probability (CP), average interval length (AIL) of the 95% confidence interval, and average computational time (Time), measured in seconds. The ASEs are reported only for the naive and meta estimators based on the asymptotic normality. The two methods use the estimated asymptotic variances to construct the confidence intervals, whereas all other methods use the ELR statistic according to Theorem 6. Table 1 shows that the performance of the proposed DEL.DOC method is virtually identical to that of the centralized EL method, with minor differences owing to the numerical implementation. These two top performers give very small estimation biases and adequate coverage probabilities under all cases of the sample sizes considered. The meta method clearly underestimates the standard error, with a noticeably lower coverage probability than 95%, especially when some of the data sets have small sample sizes. The coverage probability of the DEL.SIC method is lower than the nominal 95% level with a small sample size for $n = 60, 120$ and $n_k = 20, 40$. This is not surprising, because the DEL.SIC method requires individual $n_k \to \infty$. The meta method performs worst, with much lower coverage probabilities around 70% and 80%. When all individual data sets are sufficiently large, the results of the meta estimation are close to those of the two top methods. Furthermore, the naive estimation method has the largest ESE and AIL, because it does not use the auxiliary information in the estimation.

Example 2. (Log-linear model with over-dispersion). This example examines violation of the homogeneity in the sense that the second-order moments of $g_k(\cdot)$ are different across the data sets, owing to heterogeneous over-dispersion. We simulate *i.i.d.* observations of (Y, X) from the following Poisson-gamma model: $Y|\theta, X \sim \text{Poi}(\theta\mu)$, with $\log(\mu) = \beta_0 + \beta_1 X$, $X \sim$ Uniform[-1,1], where $(\beta_0, \beta_1) = (0, 1)$, and θ is a multiplicative random effect with $E(\theta) = 1$, the distribution of which is specified differently over $K = 3$ data sets. Specifically, for the first data set, θ is degenerated as $\theta \equiv 1$, so no over-dispersion exists; for the second data set, θ follows a *gamma* distribution, $\theta \sim \text{Gamma}(5, 5)$, and for the third dataset, θ follows a two-point distribution, $\theta \in \{0, 4\}$, with $P(\theta = 4) = 1/4$. Using the correctly specified marginal mean model, the goal is to estimate the slope parameter β_1. The unbiased estimating function is given by

Table 1. Summary of the Bias($\times 10^{-3}$), ESE($\times 10^{-3}$), ASE($\times 10^{-3}$), CP(%), and AIL($\times 10^{-2}$) of the 95% confidence interval and the average computational time (in seconds) over 2,000 replicates under six sample sizes when estimating the tail probability of the Weibull distribution with auxiliary information.

Method	BIAS	ESE	ASE	CP(AIL)	Time	BIAS	ESE	ASE	CP(AIL)	Time
	$n_k = (20, 20, 20)$					$n_k = (80, 80, 80)$				
Naive	0.55	39.06	37.81	93.35(14.82)	0.00	0.37	19.75	19.28	92.75(7.56)	0.00
CEL	2.16	26.97	—	95.88(10.37)	0.05	0.21	13.64	—	95.65(5.29)	0.08
Meta	−13.99	45.18	19.65	69.77 (7.70)	0.32	−1.72	14.64	12.94	91.20(5.07)	0.34
DOC	3.18	26.43	—	96.15(10.39)	4.86	0.21	13.64	—	95.65(5.29)	3.70
SIC	7.27	27.95	—	90.96 (9.30)	2.55	0.28	13.78	—	94.49(5.26)	2.35
	$n_k = (40, 40, 40)$					$n_k = (40, 80, 120)$				
Naive	0.55	27.60	27.12	95.55(10.63)	0.00	0.37	19.75	19.28	92.75(7.56)	0.00
CEL	0.27	19.73	—	94.19 (7.43)	0.06	0.21	13.64	—	95.65(5.29)	0.08
Meta	−4.62	25.02	17.17	83.35 (6.73)	0.31	−2.20	16.51	12.86	89.51(5.04)	0.35
DOC	0.37	19.59	—	94.38 (7.43)	3.91	0.21	13.64	—	95.65(5.29)	3.93
SIC	0.59	20.45	—	91.67 (7.18)	2.35	0.40	13.82	—	94.48(5.24)	2.54
	$n_k = (150, 150, 150)$					$n_k = (80, 150, 220)$				
Naive	0.51	14.34	14.13	93.30 (5.54)	0.00	0.51	14.34	14.13	93.30(5.54)	0.00
CEL	0.29	9.94	—	94.90 (3.88)	0.10	0.29	9.94	—	94.90(3.88)	0.10
Meta	−0.64	10.32	9.69	93.30 (3.80)	0.42	−0.73	10.41	9.69	92.75(3.80)	0.42
DOC	0.30	9.94	—	94.95 (3.88)	3.26	0.29	9.94	—	94.95(3.88)	3.43
SIC	0.37	10.00	—	94.50 (3.86)	2.63	0.30	10.04	—	94.30(3.86)	2.78

$g(x, y; \boldsymbol{\beta}) = (1, x)^T \{y - \exp(\beta_0 + \beta_1 x)\}$, which satisfies $E[g(Y, X; \boldsymbol{\beta})] = \mathbf{0}$. Despite the absence of random effects in the estimating function, the outcome y is over-dispersed, with different dispersion parameters over three data sets. To evaluate the influence of the heterogeneity in the variances, we apply the DEL.DOC and DEL.SIC methods to estimate β_1. Their performance is also compared with that of the standard GLM method under no over-dispersion, the CEL method based on the combined data set, and the meta method with an inverse variance weighting. The results are summarized from 2,000 replicates in Table 2, where the last column shows the relative 95% confidence interval length (RIL) of a method compared with that of the CEL method.

Similarly to the findings in Example 1, the DEL.DOC method and the CEL method exhibit almost identical performance. However, the results of the DEL.SIC method are very close to those of the meta EL method, both of which have smaller ESEs and shorter AILs than those of the DEL.DOC and CEL methods. This numerical evidence confirms the theoretical results in Theorem 3 in the case of heterogeneity, when the second-order moments of the estimating functions g_k are different, owing to varying over-dispersion generated by different distribu-

Table 2. Summary of the BIAS($\times 10^{-2}$), ESE($\times 10^{-2}$), ASE($\times 10^{-2}$), CP(%), and AIL($\times 10^{-1}$) of the 95% confidence interval, relative interval length compared with that of the CEL method (RIL), and average computation time (in seconds) over 2,000 replicates under different sample sizes when estimating the slope parameter β_1 in a log-linear model with over-dispersion.

n_k	method	BIAS	ESE	ASE	CP	AIL	RIL	Time
(50,100,150)	GLM	−0.18	17.34	10.24	74.55	4.01	0.58	0.00
	CEL	−0.18	17.34	—	94.80	6.88	1.00	0.88
	Meta	−0.75	14.36	13.04	93.33	5.11	0.74	0.14
	DOC	−0.19	17.31	—	94.80	6.88	1.00	7.90
	SIC	0.67	14.41	—	93.50	5.20	0.76	6.20
(100,200,300)	GLM	0.14	12.43	7.21	74.00	2.83	0.58	0.00
	CEL	0.14	12.43	—	94.75	4.86	1.00	1.02
	Meta	−0.38	9.77	9.45	94.20	3.70	0.76	0.18
	DOC	0.13	12.39	—	94.75	4.86	1.00	7.57
	SIC	0.01	9.76	—	94.60	3.75	0.77	6.42
(200,400,600)	GLM	0.17	8.91	5.09	73.20	1.99	0.58	0.01
	CEL	0.17	8.91	—	94.50	3.44	1.00	1.55
	Meta	−0.25	6.95	6.76	94.35	2.65	0.77	0.25
	DOC	0.17	8.88	—	94.50	3.44	1.00	10.54
	SIC	−0.43	6.96	—	94.60	2.68	0.78	8.79

tions of the random effects. When naively using the standard GLM method with the over-dispersion ignored, although producing consistent point estimates, it severely underestimates the variance of the estimator. Note that the DEL.SIC method automatically adjusts for heterogeneous over-dispersion, without explicitly modeling it, which is rather appealing in practice.

Example 3. (Log-linear model with unbalanced variables). This example concerns data with unbalanced variables. We simulate *i.i.d.* observations of (Y, X_1, X_2, X_3) from the following Poisson regression model: $Y|X \sim \text{Po}(\mu)$, with $\log(\mu) = \beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3$, where $(\beta_0, \beta_1, \beta_2, \beta_3) = (0, 1, 1, 1)$ and $X_1 \sim \text{U}[-1, 1]$. To create an imbalance, X_{i2} and X_{i3} are generated independently from Bernoulli $B(1, p)$ with varying p. Set the total sample size of $n = 300$ subjects, which are randomly assigned to $K = 10$ data sets satisfying $n_k \equiv 30$. Five different levels of unbalanced variables are considered, with $p = 0.05, 0.1, 0.2, 0.5$. To mimic the real kidney transplant data, we further consider an extremely unbalanced situation (S0) with $X_2 \sim B(1, 0.05)$. One data set is generated by $X_3 \sim B(1, 0.5)$, one has $X_3 \equiv 0$, and the other eight have $X_3 \equiv 1$. We compare the proposed DEL method with the centralized GLM method, centralized EL method, meta GLM method, and meta EL method. The BIAS, ESE, CP, and AIL of the 95% confidence interval, proportion of the algorithmic convergence (PAC), and com-

Table 3. Summary of the results for estimating β_3 in a log-linear model. Different scenarios of imbalances in the distributions of the covariates are given by various probabilities p, and S0 represents an extremely unbalanced situation.

p	0.5	0.2	0.1	0.05	S0	0.5	0.2	0.1	0.05	S0
	Centralized GLM					Centralized EL				
BIAS	0.00	−0.00	−0.00	−0.01	−0.00	0.00	−0.01	−0.00	−0.01	−0.00
ESE	0.07	0.08	0.11	0.16	0.06	0.07	0.08	0.11	0.16	0.06
CP	94.45	94.65	94.75	95.65	94.85	94.60	94.30	92.85	92.33	94.70
AIL	0.26	0.32	0.44	0.62	0.22	0.25	0.32	0.43	0.59	0.22
Time	0.00	0.00	0.00	0.00	0.00	30.12	30.89	30.86	33.35	23.37
PAC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
	Distributed EL									
BIAS	0.00	−0.00	−0.00	−0.01	−0.00					
ESE	0.07	0.08	0.11	0.16	0.06					
CP	94.55	94.30	92.83	92.60	94.67					
AIL	0.25	0.32	0.43	0.59	0.22					
Time	29.47	19.12	14.43	20.66	39.02					
PAC	1.00	1.00	1.00	1.00	0.99					
	Meta GLM					Meta EL				
BIAS	−0.01	0.00	0.02	0.02	–	−0.01	0.01	–	–	–
ESE	0.07	0.08	0.11	0.19	–	0.08	0.09	–	–	–
CP	94.30	94.80	95.05	81.82	–	79.61	78.21	–	–	–
AIL	0.26	0.32	0.43	0.54	–	0.20	0.22	–	–	–
Time	0.02	0.02	0.02	0.02	–	5.83	6.32	–	–	–
PAC	1.00	0.97	0.42	0.01	0.00	0.88	0.04	0.00	0.00	0.00

putation time (Time) from 2,000 replicates are reported in Table 3. It is easy to see that the performance of the DEL method is the same as that of the CEL method, regardless of the imbalance levels. However, the two meta-type methods fail as the imbalance becomes noticeably severe.

5. Real-Data Example

The Scientific Registry of Transplant Recipients (SRTR) provides epidemiological data and statistical analyses related to solid organ transplantation in the United States. Post-transplantation graft survival is the clinical outcome of most importance for patients who receive a kidney transplant, and understanding its associated risk factors is of clinical interest. Graft failure at the fifth year (1 for yes and 0 for no) after organ replacement therapy is analyzed using a logistic model that includes the following eight covariates: donor's and recipient's age, BMI (0 if not obese and 1 if obese), and gender (0 for female and 1 for male), as well as each recipient's previous transplant (prev_tx, 0 for no and 1 for yes) and hepatitis C serology (ree_hcv, 0 for negative and 1 for positive). This illustrative analysis concerns kidney transplant recipients from three regions: Alaska

Table 4. Results of the logistic regression with the SRTR data using the distributed empirical likelihood (DEL) method, Meta method (Meta2), and region-specific maximum likelihood estimations on Alaska (GLM_AK) and Wyoming (GLM_WY).

	DEL		Meta2		MLE_AK		MLE_WY	
	EST	p-val ($\times 10^{-1}$)	EST	p-val ($\times 10^{-1}$)	EST	p-val ($\times 10^{-1}$)	EST	p-val ($\times 10^{-1}$)
Intercept	−1.94	0.01	−1.77	0.01	−1.96	0.22	−1.83	0.09
rec_sex	0.55	0.14	0.42	1.19	0.67	1.17	0.38	2.69
don_sex	−0.57	0.16	−0.48	0.65	−1.13	0.07	−0.20	5.57
rec_age	−0.01	0.18	−0.01	1.31	−0.01	5.02	−0.02	1.46
don_age	0.00	1.37	0.00	9.64	−0.01	7.04	0.01	7.21
rec_bmi	0.19	1.02	0.26	3.45	0.13	7.45	0.49	2.03
don_bmi	−0.52	0.45	−0.30	3.90	−0.94	1.34	−0.10	8.17
prev_tx	0.43	0.86	0.33	3.53	1.00	0.51	−0.18	7.29
rec_hcv	−0.34	1.20	−0.24	7.47	−0.49	6.53	−0.17	8.72

(AK), Guam (GU), and Wyoming (WY), with data sizes of 449, 9, and 401, respectively, during the period 1987 to 2017. Because of the limited data size in GU, the covariates don_bmi, prev_tx, and rec_hcv take the same values with no variability, and thus cannot be included in the regression model as independent variables. Because a standard logistic regression is not feasible for the data in region GU, the meta method fails to combine the results from these three regions. For the purpose of comparison, we apply the method for two regions, namely, AK and WY; see the results in Table 4 under Meta2, and the region-specific logistic regression analyses on AK (GLM_AK) and WY (GLM_WY) using the R package GLM, based on the maximum likelihood estimation. In contrast, our proposed DEL method still combines the three regions, generating several interesting results; see Table 4. From the DEL analysis of the SRTR data from the three regions, we see that (i) male donors have a higher five-year graft survival than that of female donors, but a lower five-year graft survival than that of female recipients; (ii) older recipients tend to have a slightly lower five-year graft failure risk than younger recipients do; and (iii) recipients who receive repeated transplants tend to have the same five-year graft survival as those having a first-time transplant.

6. Discussion

We have developed a DEL methodology for performing EL estimation and inference, with no need to pool individual data sets or share subject-level information across multiple data sets. This is a useful approach for protecting data privacy and overcoming data-sharing barriers in practice. Two forms of

the DEL method are proposed, namely, the DEL.SIC and the DEL.DOC methods, under homogeneity and heterogeneity scenarios, respectively. The former is the setting routinely postulated in classical meta-analyses. Both our analytic and our numerical results show that the DEL method is almost surely equivalent to the centralized EL method, which processes aggregated data at a centralized operation platform.

There is an interesting connection between the meta-estimation and the DEL method. That is, the one-step update of the DEL.SIC method in equation (3.1) gives rise to the meta-estimate (1.1). When the weighting matrix is specified by the variance of the estimator, the resulting meta-estimator is equivalent *in distribution* to the centralized EL estimator. In contrast, our DEL estimator implemented using the ADMM-based iterative weighting is *almost surely* equal to the centralized EL estimator. It is known that the mode of almost sure equivalency is stronger than the mode of equivalency in distribution. One key practical advantage of the DEL method is its ability to handle unbalanced data distributions across multiple data sets, which often occurs in discrete variables. The meta-estimation method may fail when the imbalance is extreme, although this method is computationally faster than the DEL method.

In practice, data heterogeneity appears in various forms, which calls for context-dependent solutions. Our contributions to the DEL method may provide useful techniques for studying data heterogeneity across multiple data sources. Two possible directions of future research are a sparsity regularization of the parameter $\boldsymbol{\theta}$ and a post-fusion inference. The former pertains to the issue of reconciling sparse solutions obtained from different data sets, and the latter includes a debiasing procedure to correct the estimation bias for valid statistical inferences.

Supplementary Material

The online Supplementary Material contains additional notation, simulation results, and technique details, including proofs of the theorems.

Acknowledgments

The authors are grateful to the anonymous reviewers for their constructive comments and suggestions. Zhou's research was partially supported by the Fund of National Natural Science (Nos. 11901470, 11931014 and 11829101) and by the Fundamental Research Funds for the Central Universities (Nos. JBK190904 and JBK 1806002). Song's research was supported in part by the National In-

stitutes of Health grant R01 ES024732 and National Science Foundation grant DMS1811734.

Appendix: Conditions

Here are regularity conditions required to establish key theoretical properties for the proposed DEL.DOC estimators.

(C1) The true parameter $\boldsymbol{\theta}_0$ is an interior point in a compact set $\boldsymbol{\Xi}$ and is the unique solution to $E_{\boldsymbol{\theta}_0}\{g_k(\boldsymbol{W}_{ki};\boldsymbol{\theta})\} = \mathbf{0},\ i = 1,\ldots,n_k,\ k = 1,\ldots,K.$

(C2) Estimating function $g_k(\boldsymbol{W}_{ki};\boldsymbol{\theta})$ is continuous at each $\boldsymbol{\theta} \in \boldsymbol{\Xi}$ with probability 1 for $i = 1,\ldots,n_k,\ k = 1,\ldots,K$, and for some $\alpha > 2$, $E_{\boldsymbol{\theta}_0}(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Xi},1\le k\le K} \|g_k(\boldsymbol{W}_{ki};\boldsymbol{\theta})\|_2^{\alpha}) < \infty$.

(C3) Variability matrices $\boldsymbol{Q}_k = E_{\boldsymbol{\theta}_0}\left\{g_k(\boldsymbol{W}_{ki};\boldsymbol{\theta}_0)g_k^T(\boldsymbol{W}_{ki};\boldsymbol{\theta}_0)\right\}, k = 1,\ldots,K$, are positive-definite.

(C4) $g_k(\boldsymbol{W}_{ki};\boldsymbol{\theta})$ is continuously differentiable in a neighborhood of $\boldsymbol{\theta}_0$, say $\mathcal{D}$, and $E_{\boldsymbol{\theta}_0}\left(\sup_{\boldsymbol{\theta}\in\mathcal{D},1\le k\le K} \|\partial g_k(\boldsymbol{W}_{ki};\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T\|_2\right) < \infty$.

(C5) $\operatorname{rank}(\boldsymbol{S}_k) = q + 1$, where sensitivity matrix $\boldsymbol{S}_k = E_{\boldsymbol{\theta}_0}\{\partial g_k(\boldsymbol{W}_{ki};\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\}$.

All these five conditions (C1)-(C5) are assumed in the seminal work of empirical likelihood (EL) by Qin and Lawless (1994) in the context of estimating functions. Condition (C1) is a mild regularity condition that requires unbiased estimating functions for estimation consistency, while conditions (C3) and (C5) are routinely imposed in order to obtain a valid sandwich covariance matrix in the asymptotic normality. Conditions (C2) and (C4) appear slightly stronger with the uniform upper bounds than the classical situation of one dataset, which is postulated to deal with the case of $K \to \infty$. In other words, these uniform bound conditions may be relaxed when the number of datasets K is fixed. Checking conditions (C2) and (C4) may be done by case by case. Let us look at Example 2, where the mean moment condition g_k used for parameter estimation is the same over the data sources but outcomes are over-dispersed with heterogeneous over-dispersion. It is known that the analytic expressions of the moment condition and its derivative are given as follows: $g_k(\boldsymbol{W};\boldsymbol{\theta}) = (1,x)^T\{y - \exp(\theta_0 + \theta_1 x)\}$, and $\partial g_k(\boldsymbol{W};\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T = -(1,x)^T(1,x)\exp(\theta_0 + \theta_1 x)$, respectively, which are continuous in parameters θ_0 and θ_1, and the same over k. Then, it follows that for some $\alpha > 2$, with a compact set $\boldsymbol{\Xi}$ under Condition (C1), we have

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Xi},1\le k\le K}\|g_k(\boldsymbol{W}_{ki};\boldsymbol{\theta})\|_2^{\alpha}\right)\\
&=\mathbb{E}_{\boldsymbol{\theta}_0}\left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Xi}}\left[(1+X^2)(Y-\exp(\theta_0+\theta_1X))^2\right]^{\alpha/2}\right\}\\
&\le c\sum_{j=0}^{[\alpha+1]}\mathbb{E}_{\boldsymbol{\theta}_0}\left\{(1+X^2)^{\alpha/2}Y^{\alpha-j}\exp(c_1X)^j\right\}\\
&\le\frac{c}{2}\sum_{j=0}^{[\alpha+1]}\left\{\mathbb{E}_{\boldsymbol{\theta}_0}(Y^{2\alpha-2j})+\mathbb{E}_{\boldsymbol{\theta}_0}\left[(1+X^2)^{\alpha}\exp(c_1X)^{2j}\right]\right\}<\infty,
\end{aligned}$$

where $[a]$ denotes the largest integer smaller than a, c and c_1 are two positive constants, and the last inequality holds as long as $\mathbb{E}_{\boldsymbol{\theta}_0}(Y^j)<\infty$, and $\mathbb{E}_{\boldsymbol{\theta}_0}(X^{j_1}\exp(c_1X)^{j_2})<\infty$, for $j,j_1,j_2=1,\ldots,2[\alpha+1]$. In Example 2, these two moment conditions automatically hold because Y follows a Poisson distribution for data source #1, a negative-binomial distribution (resulted from the Poisson-gamma convolution) for data source #2, and a two-component mixture of Poisson distributions for data source #3, as well as $X\sim U[-1,1]$. Similarly, under a compact neighborhood $\mathcal{D}$, we have $\mathbb{E}_{\boldsymbol{\theta}_0}\left(\sup_{\boldsymbol{\theta}\in\mathcal{D},1\le k\le K}\|\partial g_k(\boldsymbol{W}_{ki};\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T\|_2\right)=\mathbb{E}_{\boldsymbol{\theta}_0}(\sup_{\boldsymbol{\theta}\in\mathcal{D}}[(1+2X^2+X^4)\times\exp(\theta_0+\theta_1X)^2]^{1/2})<\infty$, where the last inequality holds because condition (C1) and $X\sim U[-1,1]$.

References

Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* **3**, 1–122.

Chaudhuri, S., Handcock, M. S. and Rendall, M. S. (2008). Generalized linear models incorporating population level information: An empirical-likelihood-based approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 311–328.

Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* **80**, 107–116.

Chen, S. and Kim, J. K. (2014). Population empirical likelihood for nonparametric inference in survey sampling. *Statistica Sinica* **24**, 335–355.

Citro, C. F. (2014). From multiple modes for surveys to multiple data sources for estimates. *Survey Methodology* **40**, 137–161.

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika* **80**, 3–26.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38.

Halevy, A., Rajaraman, A. and Ordille, J. (2006). Data integration: The teenage years. In *Proceedings of the 32nd International Conference on Very large Data Bases*. VLDB Endowment.

Han, P. and Lawless, J. F. (2019). Empirical likelihood estimation using auxiliary summary information with different covariate distributions. *Statistica Sinica* **29**, 1321–1342.

Huang, C.-Y., Qin, J. and Tsai, H.-T. (2016). Efficient estimation of the Cox model with auxiliary subgroup survival information. *Journal of the American Statistical Association* **111**, 787–799.

Imbens, G. W. (2002). Generalized method of moments and empirical likelihood. *Journal of Business & Economic Statistics* **20**, 493–506.

Kovalchik, S. A. (2012). Aggregate-data estimation of an individual patient data linear random effects meta-analysis with a patient covariate-treatment interaction term. *Biostatistics* **14**, 273–283.

Lenzerini, M. (2002). Data integration: A theoretical perspective. In *Proceedings of the Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. ACM.

Liu, D., Liu, R. Y. and Xie, M. (2015). Multivariate meta-analysis of heterogeneous studies using only summary statistics: Efficiency and robustness. *Journal of the American Statistical Association* **110**, 326–340.

Lohr, S. L. and Raghunathan, T. E. (2017). Combining survey data with other data sources. *Statistical Science* **32**, 293–312.

National Academies of Sciences Engineering and Medicine (2017). *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps.* National Academies Press, Washington, D.C..

Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72**, 219–255.

Owen, A. B. (1991). Empirical likelihood for linear models. *The Annals of Statistics* **19**, 1725–1747.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC, London.

Qin, J. (2000). Combining parametric and empirical likelihoods. *Biometrika* **87**, 484–490.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics* **22**, 300–325.

Qin, J., Zhang, H., Li, P., Albanes, D. and Yu, K. (2014). Using covariate-specific disease prevalence information to increase the power of case-control studies. *Biometrika* **102**, 169–180.

Simmonds, M. and Higgins, J. (2007). Covariate heterogeneity in meta-analysis: Criteria for deciding between meta-regression and individual patient data. *Statistics in Medicine* **26**, 2982–2999.

Stout, W. F. (1974). *Almost Sure Convergence*. Academic Press, New York.

Xie, M.-G. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review* **81**, 3–39.

Ling Zhou

Center of Statistical Research, Southwestern University of Finance and Economics, Chengdu 611130, China.

E-mail: zhouling@swufe.edu.cn

Xichen She
Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA.
E-mail: xichens@umich.edu
Peter X.-K. Song
Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA.
E-mail: pxsong@umich.edu

(Received August 2020; accepted December 2021)