



Research Article

George Luta, Mihai Giurcanu*, Kirstine Amris, Gary G Koch, and Pranab K Sen

Weighted Empirical and Euclidean Likelihood Covariate Adjustment

<https://doi.org/10.1515/stat-2020-0107>

Received Mar 10, 2021; accepted Dec 09, 2021

Abstract: Covariate adjustment is often used in statistical analysis of randomized experiments to increase efficiency of estimators of treatment effects. In this paper, we study covariate adjustment based on the empirical and Euclidean likelihoods, and propose weighted versions that arise as natural alternatives. The weighted methods incorporate the auxiliary information that the covariates have equal means among the treatment groups due to randomization. We show that the empirical and the Euclidean likelihoods and their weighted versions are first order equivalent to Koch's nonparametric covariance adjustment. Allowing the weights to be negative, the resulting pseudo Euclidean likelihood is equivalent to Koch's method, and its weighted version can be viewed as a weighted version of Koch's method. In a simulation study, we assess the finite sample properties of the proposed methods. The analysis of a clinical trial data set illustrates an application of these methods to a practical situation.

Keywords: Covariate adjustment, Empirical likelihood, Euclidean likelihood, Weighted methods

2020 Mathematics Subject Classification: Primary 62F30, Secondary 62J10

1 Introduction

Analysis of covariance is used in statistical analysis of randomized experiments to increase efficiency of estimators of treatment effects and to induce equivalence (with respect to the covariates) of the treatment groups generated by randomization (Snedecor and Cochran, 1980). Nonparametric alternatives include the rank analysis of covariance (Quade, 1967), the analysis of covariance based on general rank scores (Puri and Sen, 1971), and the nonparametric covariance adjustment for categorical data (Koch et al., 1982). The presence of covariate information in randomized experiments is similar to the presence of auxiliary information in survey sampling (Fienberg and Tanur, 1987). Auxiliary information, mostly in the form of known finite population means of auxiliary variables, is frequently used in survey analysis to increase the precision of estimators (Deville and Sarndal, 1992).

Nonparametric covariance adjustment strategies are described by Koch et al. (1998) for continuous, dichotomous, and ordinal data in randomized clinical trials with two groups. The authors described extensions of these methods for stratified studies, for multivariate responses, for missing data problems, and for estimation

***Corresponding Author: Mihai Giurcanu:** Department of Public Health Sciences, University of Chicago, 5841 S Maryland Ave, Chicago, IL, 60637, United States of America; Email: giurcanu@uchicago.edu

George Luta: Department of Biostatistics, Bioinformatics, and Biomathematics, Georgetown University, Washington, DC 20057, United States of America

Kirstine Amris: Department of Rheumatology, The Parker Institute, Copenhagen University Hospital, Bispebjerg and Frederiksberg, Frederiksberg, DK-2000, Denmark, EU

Gary G Koch: Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, United States of America

Pranab K Sen: Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, United States of America



of nonparametric measures of treatment effects. Tangen and Koch (2001) further extended the methodology for randomized clinical trials with multiple groups. The power of non-inferiority and superiority tests based on Koch's estimator were evaluated by Lesaffre *et al.* (2002) and Lesaffre and Senn (2003) proposed correction factors to better maintain the nominal levels of the tests on small samples.

The empirical likelihood (EL) was introduced by Owen (1988) as a semiparametric method of inference for general parameters defined by statistical functionals. An advantage of the EL is that it defines an empirical likelihood ratio test (ELRT) which has similar asymptotic properties as the likelihood ratio test (LRT) in the parametric case. The confidence regions obtained by inverting the ELRT enjoy desirable properties: have a data-determined shape, respect the range of parameters, are transformation respecting, and do not require estimation of scale or skewness (Hall and La Scala, 1990). The ELRT statistic was shown to be Bartlett-correctable by DiCiccio *et al.* (1991). Applications of EL for randomized experiments were investigated by Adimari (1995), Jing (1995), Liu *et al.* (2008), and Owen (1991). Weighted EL methods have been proposed by Chen and Sitter (1999), Ren (2001), Glenn and Zhao (2006), and Tsao and Wu (2006). A comprehensive review on applications of EL is provided by Owen (2001) and on EL methods for sample surveys by Wu (2002).

In this paper, we propose covariate adjustment methods for randomized clinical trials that incorporate the auxiliary information of equal means of the covariates among the treatment groups due to randomization similarly to Koch's approach. These methods include the EL, the Euclidean likelihood (UL), and our proposed weighted versions that arise as natural alternatives: the weighted empirical likelihood (WEL) and the weighted Euclidean likelihood (WUL). The UL uses the Euclidean distance instead of the Kullback-Leibler divergence in the optimization problem corresponding to the EL. Allowing the weights to be negative in the case of Euclidean distance, we show that the resulting pseudo UL is equivalent to Koch's method. The weights used for the WEL, WUL, and pseudo WUL incorporate the available auxiliary information. The pseudo Euclidean likelihood (weighted or unweighted) methods have the computational advantage of providing closed-form expressions. We prove that the EL, the UL, the pseudo UL, and their weighted versions are first order equivalent and that the pseudo WUL can be viewed as a weighted version of Koch's method. Our approach is similar to the pseudo EL for sample surveys of Chen and Sitter (1999), Wu and Rao (2006), and Zhong and Rao (2000). Specifically, both our WEL approach and pseudo EL use weights on the components of the EL function. However, while the pseudo EL uses survey weights, we use the weights derived from the auxiliary information. We will show an application of our weighted approach for sample surveys which can be viewed as a weighted version of the pseudo EL of Chen and Sitter (1999) and Wu and Rao (2006).

Although previous work has described the use of Koch's nonparametric covariance adjustment method for multivariate data structures, the focus has not been on multivariate test statistics (Saville *et al.*, 2011; Kawaguchi *et al.*, 2011; Saville and Koch, 2013). As such, the consideration of multivariate test statistics in their own right and the construction of related confidence regions are also novel contributions provided by the current work. Alternative approaches to covariate adjustment have been proposed to estimate treatment effects in clinical trials such as: matching, stratification, inverse probability weighting, and using propensity score as a covariate. For a comparison between these methods, see Elze *et al.* (2017). For a survey on medical papers that use either covariate adjustment, or subgroup analysis, or baseline comparisons in clinical trials reporting as well as statistical ramifications and recommendations for future practice, see Pocock *et al.* (2002). For some unexpected results about the efficiency of treatment effect estimates adjusting for covariates for binary outcomes in logistic regression, see Robinson and Jewell (1991). For an application of covariate adjustment for cost-effectiveness data, see Willan *et al.* (2004). For a principled, practically-feasible approach for covariate adjustment that yields the desired gain of efficiency using semiparametric theory, see Tsiatis *et al.* (2008). Note that these references consider the case of univariate responses while in our paper we consider multivariate responses.

The rest of the article is organized as follows. We first describe the EL, the UL, and the pseudo UL for a multivariate mean with auxiliary information and introduce the weighted methods as natural alternatives. We then extend the methods for the common multivariate mean and for the adjusted treatment effect models. We describe alternative methods of inference for the treatment effect in presence of covariates such as the Koch's method and multivariate analysis of covariance. In a simulation study, we compare, in terms of the root mean squared errors of the estimators and size distortions of the tests, the proposed methods with alternative

methods of inference for treatment effects. We conduct a data analysis of a clinical trial data set to illustrate an application of the methods in practice. We summarize the results developed in this paper in a conclusion and the proofs of theoretical results are provided in the appendix.

2 Multivariate mean with auxiliary information

2.1 Empirical/Euclidean likelihood for the mean

Let $X_{1:n} = \{X_1, \dots, X_n\}$ be an i.i.d. sample from a distribution P on \mathbb{R}^q , and let $V_X = \text{Var}(X)$. The parameter of interest is $\mu_X = E(X)$, where $X \sim P$. The EL function of inference for μ_X is defined as:

$$\hat{\ell}_X^E(\mu) = \max_{p \in \mathbb{P}^n} \left\{ \sum_{i=1}^n \log(np_i) : \sum_{i=1}^n p_i X_i = \mu \right\}, \quad (1)$$

where $\mu \in \mathbb{R}^q$ and $\mathbb{P}^n = \{p = (p_i) \in \mathbb{R}^n : \sum_{i=1}^n p_i = 1, p_i > 0\}$. Strictly speaking, $\hat{\ell}_X^E(\mu)$ is the log-EL function, but for convenience, we refer to it here and elsewhere as the EL function. Since the objective function is strictly concave, then (1) has a unique solution if μ is in the relative interior of the convex hull of $X_{1:n}$. Let $\hat{p}^E(\mu) = (\hat{p}_i^E(\mu)) \in \mathbb{P}^n$ be the solution of (1) assuming that it exists. Owen (1988) showed that

$$\hat{p}_i^E(\mu) = \frac{1}{n(1 + \hat{\lambda}^E(\mu)^T(X_i - \mu))}, \quad 1 \leq i \leq n, \quad (2)$$

where $\hat{\lambda}^E(\mu)$ satisfies $\sum_{i=1}^n \hat{p}_i^E(\mu) X_i = \mu$. Owen (1988) showed that, similarly to Wilks's (1938) theorem for the LRT in parametric models, $-2\hat{\ell}_X^E(\mu_X) \xrightarrow{d} \chi_q^2$, where \xrightarrow{d} denotes convergence in distribution. The ELRT rejects $H_0 : \mu_X = \mu_X^0$ at nominal type I error level α if $-2\hat{\ell}_X^E(\mu_X^0) > \chi_{q;1-\alpha}^2$, where $\chi_{q;1-\alpha}^2$ denotes the $(1 - \alpha)$ quantile of the χ_q^2 distribution. Inverting the ELRT, a $(1 - \alpha) \times 100\%$ confidence region for μ_X is given by $\{\mu \in \mathbb{R}^q : -2\hat{\ell}_X^E(\mu) \leq \chi_{q;1-\alpha}^2\}$.

If instead of the (forward) Kullback-Leibler divergence we use the Euclidean distance in (1), we obtain the Euclidean likelihood (UL) function:

$$\hat{\ell}_X^U(\mu) = \max_{p \in \mathbb{P}^n} \left\{ -(1/2) \sum_{i=1}^n (np_i - 1)^2 : \sum_{i=1}^n p_i X_i = \mu \right\}. \quad (3)$$

Note that (3) is a quadratic optimization problem with inequality constraints which does not have closed form solutions (for a graduate level textbook on convex optimization, see Boyd and Vandenberghe, 2004). If we allow the weights be negative in (3) and maximize over \mathbb{Q}^n , where $\mathbb{Q}^n = \{q = (q_i) \in \mathbb{R}^n : \sum_{i=1}^n q_i = 1\}$, we obtain the pseudo UL function:

$$\hat{\ell}_X^P(\mu) = \max_{q \in \mathbb{Q}^n} \left\{ -(1/2) \sum_{i=1}^n (nq_i - 1)^2 : \sum_{i=1}^n q_i X_i = \mu \right\}. \quad (4)$$

Here and in other places, we use q to either denote a vector $q = (q_i) \in \mathbb{Q}^n$ or the dimension of X ; the notation should be obvious from the context. Strictly speaking, the UL was defined by Owen as the maximizer of (3) over $q \in \mathbb{Q}^n$ (Owen, 2001, section 3.15). However, to be consistent with the definition of EL, we define the UL in (3) as a maximization over $p \in \mathbb{P}^n$ and we call $\hat{\ell}_X^P(\mu)$ the pseudo UL function. Using the method of Lagrange multipliers, Owen (2001) derived the solution for q_i of (4) in terms of μ (see the appendix):

$$\hat{q}_i(\mu) = n^{-1} (1 - (\bar{X} - \mu)^T \hat{V}_X^{-1} (X_i - \bar{X})), \quad 1 \leq i \leq n, \quad (5)$$

where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ and $\hat{V}_X = n^{-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$. Substituting $\hat{q}_i(\mu)$ from (5) in (4), yields:

$$-2\hat{\ell}_X^P(\mu) = n(\bar{X} - \mu)^T \hat{V}_X^{-1} (\bar{X} - \mu). \quad (6)$$

By (5), $\Pr(\hat{q}(\mu_X) \in \mathbb{P}^n) \rightarrow 1$, where $\hat{q}(\mu) = (\hat{q}_i(\mu)) \in \mathbb{Q}^n$. Hence $-2\hat{\ell}_X^U(\mu_X) \xrightarrow{d} \chi_q^2$, so that, the EL, the UL, and the pseudo UL are first order equivalent. Note that the pseudo UL has a computational advantage compared to the EL and UL since the weights have closed-form expressions. Another advantage of pseudo UL is that if μ is not in the convex hull of the observations, the EL and UL are undefined (since the feasible set is the empty set), while the pseudo UL and the corresponding weights are always defined.

2.2 Empirical/Euclidean likelihood for the mean with auxiliary information

Let $Z_{1:n} = \{Z_1, \dots, Z_n\}$ be an i.i.d. sample from a distribution P on \mathbb{R}^{p+q} , where $Z_i = (X_i, Y_i) \in \mathbb{R}^{p+q}$, $X_i \in \mathbb{R}^q$, $Y_i \in \mathbb{R}^p$, $E(X) = \mu_X \in \mathbb{R}^q$ is known, $E(Z) = \mu_Z \in \mathbb{R}^{p+q}$, and $Z = (X, Y) \sim P$. Here and in other places, we use p to either denote a probability vector $p = (p_i) \in \mathbb{P}^n$ or the dimension of Y . Let $V_Z = \text{Var}(Z)$, $V_Y = \text{Var}(Y)$, and $V_{XY} = V_{YX}^T = \text{Cov}(X, Y)$. The parameter of interest is $\mu_Y = E(Y) \in \mathbb{R}^p$. The EL function incorporates the auxiliary information that μ_X is known and is defined as:

$$\hat{\ell}_{Y|X}^E(\mu|\mu_X) = \hat{\ell}_Z^E(\mu_X, \mu) - \hat{\ell}_X^E(\mu_X). \quad (7)$$

Owen (1991) showed that $-2\hat{\ell}_{Y|X}^E(\mu_Y|\mu_X) \xrightarrow{d} \chi_p^2$. Write $\hat{\ell}_{Y|X}^E(\mu|\mu_X)$ equivalently as

$$\hat{\ell}_{Y|X}^E(\mu|\mu_X) = \max_{p \in \mathbb{P}^n} \left\{ \sum_{i=1}^n \log(p_i/\hat{p}_i^E) : \sum_{i=1}^n p_i X_i = \mu_X, \sum_{i=1}^n p_i Y_i = \mu \right\},$$

where $\hat{p}_i^E = \hat{p}_i^E(\mu_X)$ and $\hat{p}_i^E(\mu)$ are given by (2). The maximum empirical likelihood estimator (MELE) of μ_Y was introduced by Qin and Lawless (1994) and is defined as:

$$\hat{\mu}_Y^E = \operatorname{argmax}_{\mu \in \mathbb{R}^p} \hat{\ell}_{Y|X}^E(\mu|\mu_X).$$

Since $\sum_{i=1}^n \hat{p}_i^E X_i = \mu_X$ by definition in (1), then

$$\max_{p \in \mathbb{P}^n} \left\{ \sum_{i=1}^n \log(p_i/\hat{p}_i^E) : \sum_{i=1}^n p_i X_i = \mu_X, \sum_{i=1}^n p_i Y_i = \mu \right\} \leq \max_{p \in \mathbb{P}^n} \left\{ \sum_{i=1}^n \log(p_i/\hat{p}_i^E) : \sum_{i=1}^n p_i X_i = \mu_X \right\},$$

with equality when $\mu = \sum_{i=1}^n \hat{p}_i^E Y_i$. Hence $\hat{\mu}_Y^E = \sum_{i=1}^n \hat{p}_i^E Y_i$. By Theorem 1 of Qin and Lawless (1994),

$$n^{1/2}(\hat{\mu}_Y^E - \mu_Y) \xrightarrow{d} N(0, \Sigma_Y) \quad \text{and} \quad -2\hat{\ell}_{Y|X}^E(\mu_Y|\mu_X) \xrightarrow{d} \chi_p^2,$$

where $\Sigma_Y = V_Y - V_{YX}V_X^{-1}V_{XY}$. Note that both \bar{Y} and $\hat{\mu}_Y^E$ are consistent estimators of μ_Y . Since $\Sigma_Y \preceq V_Y$, where \preceq denotes the Lowener partial ordering, $\hat{\mu}_Y^E$ is first order more efficient than \bar{Y} . If, in particular, $\text{rank}(V_{XY}) \geq p$, then $\Sigma_Y \prec V_Y$ and $\hat{\mu}_Y^E$ is strictly more efficient than \bar{Y} .

The UL function is obtained using the Euclidean distance in (7):

$$\hat{\ell}_{Y|X}^U(\mu|\mu_X) = \hat{\ell}_Z^U(\mu_X, \mu) - \hat{\ell}_X^U(\mu_X).$$

Let $\hat{\mu}_Y^U = \operatorname{argmax}_{\mu \in \mathbb{R}^p} \hat{\ell}_{Y|X}^U(\mu|\mu_X)$ be the maximum UL estimator (MULE) of μ_Y . Since $\sum_{i=1}^n \hat{p}_i^U X_i = \mu_X$, where $\hat{p}^U = (\hat{p}_i^U) = \hat{p}^U(\mu_X) \in \mathbb{P}^n$ and $\hat{p}^U(\mu)$ is the solution (in $p \in \mathbb{P}^n$) of (3), then using a similar argument as above, $\hat{\mu}_Y^U = \sum_{i=1}^n \hat{p}_i^U Y_i$. Maximization over $q \in \mathbb{Q}^n$, by (6), yields the pseudo UL function:

$$\hat{\ell}_{Y|X}^P(\mu|\mu_X) = -(n/2)(\bar{Z} - v)^T \hat{V}_Z^{-1}(\bar{Z} - v) + (n/2)(\bar{X} - \mu_X)^T \hat{V}_X^{-1}(\bar{X} - \mu_X), \quad (8)$$

where $v = (\mu_X, \mu) \in \mathbb{R}^{p+q}$, $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$, and $\hat{V}_Z = n^{-1} \sum_{i=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})^T$. Let $\hat{q}_i = \hat{q}_i(\mu_X)$, where $\hat{q}_i(\mu)$ is given by (5). Let $\hat{\mu}_Y^P = \operatorname{argmax}_{\mu \in \mathbb{R}^p} \hat{\ell}_{Y|X}^P(\mu|\mu_X)$ be the pseudo MULE of μ_Y . Since $\sum_{i=1}^n \hat{q}_i X_i = \mu_X$ by definition in (4), then

$$\hat{\mu}_Y^P = \sum_{i=1}^n \hat{q}_i Y_i = \bar{Y} - \hat{V}_{YX} \hat{V}_X^{-1}(\bar{X} - \mu_X). \quad (9)$$

Simple algebra shows that (see the appendix):

$$-2\hat{\ell}_{Y|X}^P(\mu|\mu_X) = n(\hat{\mu}_Y^P - \mu)^T \hat{\Sigma}_Y^{-1}(\hat{\mu}_Y^P - \mu), \quad (10)$$

where $\hat{\Sigma}_Y = \hat{V}_Y - \hat{V}_{YX}\hat{V}_X^{-1}\hat{V}_{XY}$. Since $\Pr(\hat{q} \in \mathbb{P}^n) \rightarrow 1$, where $\hat{q} = (\hat{q}_i) \in \mathbb{Q}^n$, then $n^{1/2}(\hat{\mu}_Y^U - \mu_Y) \xrightarrow{d} N(0, \Sigma_Y)$ and $-2\hat{\ell}_{Y|X}^U(\mu_Y|\mu_X) \xrightarrow{d} \chi_p^2$. Hence, the EL, the UL, and the pseudo UL are first order asymptotically equivalent.

2.3 Weighted empirical/Euclidean likelihood

Our weighted empirical likelihood (WEL) is a natural alternative to the EL for the mean with auxiliary information. The method can be readily extended for more general parameters similarly to Owen (1988). Instead of minimizing the divergence between $p \in \mathbb{P}^n$ and the vector of uniform weights $n^{-1}\mathbf{1}_n \in \mathbb{P}^n$, where $\mathbf{1}_n \in \mathbb{R}^n$ is the n -vector of ones, the WEL minimizes the divergence between p and $\hat{p}^E = (\hat{p}_i^E) \in \mathbb{P}^n$ while imposing the same constraints as the EL; specifically,

$$\hat{\ell}_{Y|X}^{\text{wE}}(\mu|\mu_X) = \max_{p \in \mathbb{P}^n} \left\{ \sum_{i=1}^n (n\hat{p}_i^E) \log(p_i/\hat{p}_i^E) : \sum_{i=1}^n p_i X_i = \mu_X, \sum_{i=1}^n p_i Y_i = \mu \right\}. \quad (11)$$

If μ is in the relative interior of the convex hull of $Y_{1:n}$, the solution for p_i of (11) in terms of μ has the following expression (see the appendix):

$$\hat{p}_i^{\text{wE}}(\mu) = \frac{\hat{p}_i^E}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - v)}, \quad 1 \leq i \leq n, \quad (12)$$

where $v = (\mu_X, \mu) \in \mathbb{R}^{p+q}$ and $\hat{\lambda}^{\text{wE}}(\mu) \in \mathbb{R}^{p+q}$ satisfies:

$$\sum_{i=1}^n \frac{\hat{p}_i^E(Z_i - v)}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - v)} = 0. \quad (13)$$

By Jensen's inequality,

$$\sum_{i=1}^n (n\hat{p}_i^E) \log(p_i/\hat{p}_i^E) \leq n \log \left(\sum_{i=1}^n \hat{p}_i^E p_i / \hat{p}_i^E \right) = 0,$$

and the maximum is attained for $p_i = \hat{p}_i^E$, $1 \leq i \leq n$. Hence $\hat{\mu}_Y^E = \operatorname{argmax}_{\mu \in \mathbb{R}^p} \hat{\ell}_{Y|X}^{\text{wE}}(\mu|\mu_X)$, and thus, the maximum WEL estimator and the MELE of μ_Y are the same. Theorem 1 shows the large sample behavior of the Lagrange multipliers and of the WEL function in local neighborhoods of μ_Y , as well as the asymptotic null distribution of the weighted ELRT.

Theorem 1. *Let $Z_{1:n} = \{Z_1, \dots, Z_n\}$ be an i.i.d. sample from a distribution P on \mathbb{R}^{p+q} , where $Z_i = (X_i, Y_i) \in \mathbb{R}^{p+q}$, $X_i \in \mathbb{R}^p$, $Y_i \in \mathbb{R}^q$, $E(X) = \mu_X$ is known, $V_Z = \operatorname{Var}(Z)$ is nonsingular, and $Z = (X, Y) \sim P$. Let $\mu_Y = E(Y)$, $V_Y = \operatorname{Var}(Y)$, $V_{XY} = V_{YX}^T = \operatorname{Cov}(X, Y)$, and $\Sigma_Y = V_Y - V_{YX}V_X^{-1}V_{XY}$. Then, for every $K > 0$,*

$$\max_{\mu \in \mathbb{K}_n} \|\hat{\lambda}^{\text{wE}}(\mu) - V_Z^{-1}(\hat{\mu}_Z^E - v)\| = o_P(n^{-1/2}), \quad (14)$$

$$\max_{\mu \in \mathbb{K}_n} |-2\hat{\ell}_{Y|X}^{\text{wE}}(\mu|\mu_X) - n(\hat{\mu}_Y^E - \mu)^T \Sigma_Y^{-1}(\hat{\mu}_Y^E - \mu)| = o_P(1), \quad (15)$$

and

$$\max_{\mu \in \mathbb{K}_n} \max_{1 \leq i \leq n} |\hat{p}_i^{\text{wE}}(\mu) - \hat{p}_i^E (1 - (\hat{\mu}_Z^E - v)^T V_Z^{-1}(Z_i - v))| = o_P(n^{-1}), \quad (16)$$

where $\mathbb{K}_n = \{\mu : \|\mu - \mu_Y\| \leq Kn^{-1/2}\}$, $v = (\mu_X, \mu) \in \mathbb{R}^{p+q}$, and $\hat{\mu}_Z^E = (\mu_X, \hat{\mu}_Y^E) \in \mathbb{R}^{p+q}$. Furthermore,

$$-2\hat{\ell}_{Y|X}^{\text{wE}}(\mu_Y|\mu_X) \xrightarrow{d} \chi_p^2. \quad (17)$$

If we use the Euclidean distance in (11), we obtain the weighted UL (WUL) function:

$$\hat{\ell}_{Y|X}^{\text{WU}}(\mu|\mu_X) = \max_{p \in \mathbb{P}^n} \left\{ -(1/2) \sum_{i=1}^n (n\hat{p}_i^{\text{U}}) (p_i/\hat{p}_i^{\text{U}} - 1)^2 : \sum_{i=1}^n p_i X_i = \mu_X, \sum_{i=1}^n p_i Y_i = \mu \right\}, \quad (18)$$

\hat{p}_i^{U} are defined before (8). Maximizing over $q \in \mathbb{Q}^n$ in (18) and using \hat{q}_i instead of \hat{p}_i^{U} , where \hat{q}_i are defined before (9), we obtain the weighted pseudo UL likelihood function given by:

$$\hat{\ell}_{Y|X}^{\text{WP}}(\mu|\mu_X) = \max_{q \in \mathbb{Q}^n} \left\{ -(1/2) \sum_{i=1}^n (n\hat{q}_i) (q_i/\hat{q}_i - 1)^2 : \sum_{i=1}^n q_i X_i = \mu_X, \sum_{i=1}^n q_i Y_i = \mu \right\}. \quad (19)$$

The solution $\hat{q}^{\text{W}}(\mu) = (\hat{q}_i^{\text{W}}(\mu)) \in \mathbb{Q}^n$ (in $q \in \mathbb{Q}^n$) of (19) is given by (see the appendix):

$$\hat{q}_i^{\text{W}}(\mu) = \hat{q}_i (1 - (\hat{\mu}_Z^{\text{P}} - \nu)^T \tilde{V}_Z^{-1} (Z_i - \hat{\mu}_Z^{\text{P}})), \quad (20)$$

where $\tilde{V}_Z = \sum_{i=1}^n \hat{q}_i (Z_i - \hat{\mu}_Z^{\text{P}})(Z_i - \hat{\mu}_Z^{\text{P}})^T$, $\hat{\mu}_Z^{\text{P}} = (\mu_X, \hat{\mu}_Y^{\text{P}}) \in \mathbb{R}^{p+q}$, and $\nu = (\mu_X, \mu) \in \mathbb{R}^{p+q}$. Substituting $\hat{q}_i^{\text{W}}(\mu)$ for q_i in (19), yields (see the appendix):

$$-2\hat{\ell}_{Y|X}^{\text{WP}}(\mu|\mu_X) = n(\hat{\mu}_Y^{\text{P}} - \mu)^T \tilde{\Sigma}_Y^{-1} (\hat{\mu}_Y^{\text{P}} - \mu), \quad (21)$$

where $\tilde{\Sigma}_Y = \tilde{V}_Y - \tilde{V}_{YX} \tilde{V}_X^{-1} \tilde{V}_{XY}$. By (20), $\Pr(\hat{q}^{\text{W}}(\mu_Y) \in \mathbb{P}^n) \rightarrow 1$, thus $-2\hat{\ell}_{Y|X}^{\text{WU}}(\mu_Y|\mu_X) \xrightarrow{d} \chi_p^2$. Hence, the WEL, WUL, and pseudo WUL are first order equivalent.

2.4 Application of WEL in sample surveys

Note that our approach is similar to the pseudo EL for sample surveys of Chen and Sitter (1999), Wu and Rao (2006), and Zhong and Rao (2000). Specifically, both our WEL and the pseudo EL use weights on the components of the EL function. However, while the pseudo EL uses the normalized survey weights, we use the weights derived from the auxiliary information. In this section, we show an application of our method to sample surveys, and in this case, our WEL method can be viewed as a weighted version of the pseudo EL of Chen and Sitter (1999) and Wu and Rao (2006). The weighted EL for sample surveys can be derived in two steps. In the first step, we derive the weights that minimize the divergence to the normalized survey weights (given by the normalized inverse survey selection probabilities) using the auxiliary information; these weights contain both information about the survey weights and the auxiliary information. In the second step, we use these weights (instead of the normalized survey weights) when we define the WEL function for survey data.

Specifically, let \tilde{d}_i denote the normalized survey weights corresponding to the sample such that $\sum_{i=1}^n \tilde{d}_i = 1$ (for more details, see Wu and Rao, 2006). Then, the pseudo EL function for μ_X is given by:

$$\hat{\ell}_X^{\text{E}}(\mu) = \max_{p \in \mathbb{P}^n} \left\{ \sum_{i=1}^n (n\tilde{d}_i) \log(p_i/\tilde{d}_i) : \sum_{i=1}^n p_i X_i = \mu \right\}. \quad (22)$$

Let $\hat{p}^{\text{E}}(\mu)$ denote the solution (in $p \in \mathbb{P}^n$) of (22). Under regularity conditions (see, e.g., Wu and Rao, 2006), $-2\hat{\ell}_X^{\text{E}}(\mu_X)$ is asymptotically distributed as a chi-squared variable scaled by the design effect. In the second step, consider the WEL function for μ_Y under the auxiliary information that μ_X is known:

$$\hat{\ell}_{Y|X}^{\text{WE}}(\mu|\mu_X) = \max_{p \in \mathbb{P}^n} \left\{ \sum_{i=1}^n (n\hat{p}_i^{\text{E}}) \log(p_i/\hat{p}_i^{\text{E}}) : \sum_{i=1}^n p_i X_i = \mu_X, \sum_{i=1}^n p_i Y_i = \mu \right\}, \quad (23)$$

where $\hat{p}^{\text{E}} = (\hat{p}_i^{\text{E}}) \in \mathbb{P}^n$ and $\hat{p}^{\text{E}} = \hat{p}^{\text{E}}(\mu_X)$. Note that our WEL approach in sample surveys with auxiliary information is different from the pseudo EL approach of Wu and Rao (2006), Chen and Sitter (1999), and Zhong and Rao (2000) and can be viewed as an alternative to the pseudo EL. Notice that our WEL approach to sample surveys blends naturally the weights from auxiliary information and survey weights in statistical models with auxiliary information. The asymptotic and finite sample properties of the corresponding WEL methods in sample surveys for the models considered in this paper is a topic of future research.

3 Common multivariate mean

3.1 Empirical/Euclidean likelihood

Let $X_{1:n_k}^{(k)} = \{X_{k,1}, \dots, X_{k,n_k}\}$ be two independent samples of i.i.d. observations from the distributions P_k on \mathbb{R}^q , where $E(X_1) = E(X_2) = \mu_X$ is the parameter of interest, $V_{X_k} = \text{Var}(X_k)$, $X_k \sim P_k$, $k = 1, 2$. The EL function is given by:

$$\hat{\ell}_{X_1, X_2}^E(\mu) = \max_{p_k \in \mathbb{P}^{n_k}} \left\{ \sum_{k=1}^2 \sum_{i=1}^{n_k} \log(n_k p_{k,i}) : \sum_{i=1}^{n_k} p_{k,i} X_{k,i} = \mu \right\}, \quad (24)$$

where $p_k = (p_{k,i}) \in \mathbb{P}^{n_k}$, $k = 1, 2$. Using Lagrange multipliers (see the appendix), we obtain the solutions for $p_{k,i}$ of (24) in terms of μ :

$$\hat{p}_{k,i}^E(\mu) = \frac{1}{n_k (1 + \hat{\lambda}_k^E(\mu)^T (X_{k,i} - \mu))}, \quad 1 \leq i \leq n_k, \quad k = 1, 2, \quad (25)$$

where $\hat{\lambda}_k^E(\mu)$ satisfy

$$\sum_{i=1}^{n_k} \frac{X_{k,i} - \mu}{n_k (1 + \hat{\lambda}_k^E(\mu)^T (X_{k,i} - \mu))} = 0, \quad k = 1, 2.$$

Let $\hat{\mu}_X^E$ be the MELE of μ_X . Since

$$\hat{\ell}_{X_1, X_2}^E(\mu) \leq \max_{p_k \in \mathbb{P}^{n_k}} \left\{ \sum_{k=1}^2 \sum_{i=1}^{n_k} \log(n_k p_{k,i}) : \sum_{i=1}^{n_1} p_{1,i} X_{1,i} = \sum_{i=1}^{n_2} p_{2,i} X_{2,i} \right\}$$

with equality when $\mu = \sum_{i=1}^{n_k} \hat{p}_{k,i}^E X_{k,i}$, then $\hat{\mu}_X^E = \sum_{i=1}^{n_k} \hat{p}_{k,i}^E X_{k,i}$, where $\hat{p}_k^E = (\hat{p}_{k,i}^E) \in \mathbb{R}^{n_k}$ satisfy

$$(\hat{p}_1^E, \hat{p}_2^E) = \operatorname{argmax}_{p_k \in \mathbb{P}^{n_k}} \left\{ \sum_{k=1}^2 \sum_{i=1}^{n_k} \log(n_k p_{k,i}) : \sum_{i=1}^{n_1} p_{1,i} X_{1,i} = \sum_{i=1}^{n_2} p_{2,i} X_{2,i} \right\}. \quad (26)$$

Assume that $n_k/n \rightarrow \pi_k > 0$, where $n = n_1 + n_2$. By Theorems 1 and 2 of Qin and Lawless (1994),

$$n^{1/2}(\hat{\mu}_X^E - \mu_X) \xrightarrow{d} N(0, \Sigma_X) \quad \text{and} \quad -2\hat{\ell}_{X_1, X_2}^E(\mu_X) + 2\hat{\ell}_{X_1, X_2}^E(\hat{\mu}_X^E) \xrightarrow{d} \chi_q^2,$$

where, with a slight abuse of notation, $\Sigma_X = (\pi_1 V_{X_1}^{-1} + \pi_2 V_{X_2}^{-1})^{-1}$. Note that $\hat{\mu}_X^E$ and \bar{X}_k are consistent estimators of μ_X ; since $\Sigma_X \prec \pi_k^{-1} V_{X_k}$ for $k = 1, 2$, then $\hat{\mu}_X^E$ is strictly more efficient than both \bar{X}_1 and \bar{X}_2 .

If we use the Euclidean distance in (24), we obtain the UL function:

$$\hat{\ell}_{X_1, X_2}^U(\mu) = \max_{p_k \in \mathbb{P}^{n_k}} \left\{ -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k p_{k,i} - 1)^2 : \sum_{i=1}^{n_k} p_{k,i} X_{k,i} = \mu \right\}. \quad (27)$$

Maximizing (27) over $q_k \in \mathbb{Q}^{n_k}$, we obtain the pseudo UL function:

$$\hat{\ell}_{X_1, X_2}^P(\mu) = \max_{q_k \in \mathbb{Q}^{n_k}} \left\{ -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k q_{k,i} - 1)^2 : \sum_{i=1}^{n_k} q_{k,i} X_{k,i} = \mu \right\}. \quad (28)$$

The solutions $\hat{q}_k(\mu) = (\hat{q}_{k,i}(\mu)) \in \mathbb{Q}^{n_k}$ of (28) are given by (see the appendix):

$$\hat{q}_{k,i}(\mu) = n_k^{-1} (1 - (\bar{X}_k - \mu)^T \hat{V}_{X_k}^{-1} (X_{k,i} - \bar{X}_k)), \quad 1 \leq i \leq n_k, \quad k = 1, 2, \quad (29)$$

where $\bar{X}_k = n_k^{-1} \sum_{i=1}^{n_k} X_{k,i}$ and $\hat{V}_{X_k} = n_k^{-1} \sum_{i=1}^{n_k} (X_{k,i} - \bar{X}_k)(X_{k,i} - \bar{X}_k)^T$. Substituting $\hat{q}_{k,i}(\mu)$ from (29) for $q_{k,i}$ in (28), yields:

$$-2\hat{\ell}_{X_1, X_2}^P(\mu) = \sum_{k=1}^2 n_k (\bar{X}_k - \mu)^T \hat{V}_{X_k}^{-1} (\bar{X}_k - \mu). \quad (30)$$

Let $\hat{\mu}_X^P$ be the pseudo MULE of μ_X . By (30),

$$\hat{\mu}_X^P = (n_1 \hat{V}_{X_1}^{-1} + n_2 \hat{V}_{X_2}^{-1})^{-1} \sum_{k=1}^2 n_k \hat{V}_{X_k}^{-1} \bar{X}_k. \quad (31)$$

Since $\Pr(\hat{q}_k \in \mathbb{P}^{n_k}) \rightarrow 1$, where $\hat{q}_k = (\hat{q}_{k,i}(\hat{\mu}_X^P)) \in \mathbb{Q}^{n_k}$, then $n^{1/2}(\hat{\mu}_X^U - \mu_X) \xrightarrow{d} N(0, \Sigma_X)$ and $-2\hat{\ell}_{X_1, X_2}^U(\mu_X) + 2\hat{\ell}_{X_1, X_2}^U(\hat{\mu}_X^U) \xrightarrow{d} \chi_q^2$, where $\hat{\mu}_X^U$ is the MULE of μ_X .

3.2 Weighted empirical/Euclidean likelihood

The WEL function is obtained by minimizing the divergence between $p_k \in \mathbb{P}^{n_k}$ and $\hat{p}_k^E \in \mathbb{P}^{n_k}$ in (24), i.e.,

$$\hat{\ell}_{X_1, X_2}^{WE}(\mu) = \max_{p_k \in \mathbb{P}^{n_k}} \left\{ \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k \hat{p}_{k,i}^E) \log(p_{k,i} / \hat{p}_{k,i}^E) : \sum_{i=1}^{n_k} p_{k,i} X_{k,i} = \mu \right\}, \quad (32)$$

where $\hat{p}_k^E = (\hat{p}_{k,i}^E) \in \mathbb{P}^{n_k}$ are given by (26). Using Lagrange multipliers (see the appendix), we obtain the solutions for $p_{k,i}$ of (32) in terms of μ :

$$\hat{p}_{k,i}^{WE}(\mu) = \frac{\hat{p}_{k,i}^E}{1 + \hat{\lambda}_k^{WE}(\mu)^T (X_{k,i} - \mu)}, \quad 1 \leq i \leq n_k, \quad k = 1, 2, \quad (33)$$

where $\hat{\lambda}_k^{WE}(\mu)$ satisfies

$$\sum_{i=1}^{n_k} \frac{\hat{p}_{k,i}^E (X_{k,i} - \mu)}{1 + \hat{\lambda}_k^{WE}(\mu)^T (X_{k,i} - \mu)} = 0.$$

Similarly as before, the maximum WEL estimator and the MELE of μ_X are the same. Theorem 2 shows the large sample behavior of the Lagrange multipliers and of the WEL function in local neighborhoods of μ_X , as well as the asymptotic null distribution of the weighted ELRT.

Theorem 2. Let $X_{1:n_k}^{(k)} = \{X_{k,1}, \dots, X_{k,n_k}\}$ be two independent i.i.d. samples from the distributions P_k on \mathbb{R}^q , with $E(X_1) = E(X_2) = \mu_X$, where $X_k \sim P_k$, $k = 1, 2$. Let $n = n_1 + n_2$ and assume that $n_k/n \rightarrow \pi_k > 0$ and $V_{X_k} = \text{Var}(X_k)$ are nonsingular. Then, for every $K > 0$,

$$\max_{\mu \in \mathbb{K}_n} \|\hat{\lambda}_k^{WE}(\mu) - V_{X_k}^{-1}(\hat{\mu}_X^E - \mu)\| = o_P(n^{-1/2}), \quad (34)$$

$$\max_{\mu \in \mathbb{K}_n} \left| -2\hat{\ell}_{X_1, X_2}^{WE}(\mu) - \sum_{k=1}^2 n_k (\hat{\mu}_X^E - \mu)^T V_{X_k}^{-1} (\hat{\mu}_X^E - \mu) \right| = o_P(1), \quad (35)$$

and

$$\max_{\mu \in \mathbb{K}_n} \max_{1 \leq i \leq n_k} |\hat{p}_{k,i}^{WE}(\mu) - \hat{p}_{k,i}^E (1 - (\hat{\mu}_X^E - \mu)^T V_{X_k}^{-1} (X_{k,i} - \hat{\mu}_X^E))| = o_P(n^{-1}), \quad (36)$$

where $\mathbb{K}_n = \{\mu : \|\mu - \mu_X\| \leq Kn^{-1/2}\}$. Furthermore,

$$-2\hat{\ell}_{X_1, X_2}^{WE}(\mu_X) \xrightarrow{d} \chi_p^2. \quad (37)$$

If we use the Euclidean distance in (32), we obtain the WUL function:

$$\hat{\ell}_{X_1, X_2}^U(\mu) = \max_{p_k \in \mathbb{P}^{n_k}} \left\{ -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k \hat{p}_{k,i}^U) (p_{k,i} / \hat{p}_{k,i}^U - 1)^2 : \sum_{i=1}^{n_k} p_{k,i} X_{k,i} = \mu \right\}, \quad (38)$$

where $\hat{p}_{k,i}^U = \hat{p}_{k,i}^U(\hat{\mu}_X^U)$ and $\hat{p}_k^U(\mu) \in \mathbb{P}^{n_k}$ is the solution of (27). Maximizing (38) over \mathbb{Q}^{n_k} , we obtain the pseudo WUL function:

$$\hat{\ell}_{X_1, X_2}^{UP}(\mu) = \max_{q_k \in \mathbb{Q}^{n_k}} \left\{ -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k \hat{q}_{k,i}) (q_{k,i} / \hat{q}_{k,i} - 1)^2 : \sum_{i=1}^{n_k} q_{k,i} X_{k,i} = \mu \right\}, \quad (39)$$

where $\hat{q}_{k,i} = \hat{q}_{k,i}(\hat{\mu}_X^P)$ and $\hat{q}_{k,i}(\mu)$ are given by (29). The solutions $\hat{q}_k^W(\mu) = (\hat{q}_{k,i}^W(\mu)) \in \mathbb{Q}^{n_k}$ of (39) are given by (see the appendix):

$$\hat{q}_{k,i}^W(\mu) = \hat{q}_{k,i} (1 - (\hat{\mu}_X^P - \mu)^T \tilde{V}_{X_k}^{-1} (X_{k,i} - \hat{\mu}_X^P)), \quad (40)$$

where $\tilde{V}_{X_k} = \sum_{i=1}^{n_k} \hat{q}_{k,i} (X_{k,i} - \hat{\mu}_X^P)(X_{k,i} - \hat{\mu}_X^P)^T$. Substituting $\hat{q}_{k,i}^W(\mu)$ for q_i in (39), yields:

$$-2\hat{\ell}_{X_1, X_2}^{WP}(\mu) = \sum_{k=1}^2 n_k (\hat{\mu}_X^P - \mu)^T \tilde{V}_{X_k}^{-1} (\hat{\mu}_X^P - \mu). \quad (41)$$

Since $\Pr(\hat{q}_k^W(\mu_X) \in \mathbb{P}^{n_k}) \rightarrow 1$, then $-2\hat{\ell}_{X_1, X_2}^{WU}(\mu_X) \xrightarrow{d} \chi_p^2$, and thus, the WEL, the WUL, and the pseudo WUL are first order equivalent.

3.3 Computational aspects

We can re-express the two-sample EL problem (24) as a one-sample EL for the mean with auxiliary information. To this end, let $p_i = (n_1/n)p_{1,i}$, $Z_i = n/n_1(X_{1,i}, X_{1,i}, 1)$, $p_{n_1+j} = (n_2/n)p_{2,j}$, and $Z_{n_1+j} = n/n_2(0, -X_{2,j}, 0)$, where $1 \leq i \leq n_1$, $1 \leq j \leq n_2$, and $n = n_1 + n_2$. Then (24) is equivalent to

$$\hat{\ell}_{X_1, X_2}^E(\mu) = \max_{p \in \mathbb{P}^n} \left\{ \sum_{i=1}^n \log(np_i) : \sum_{i=1}^n p_i Z_i = \nu \right\}, \quad (42)$$

where $p = (p_i) \in \mathbb{P}^n$, $Z_i \in \mathbb{R}^{2q+1}$, and $\nu = (\mu, 0, 1) \in \mathbb{R}^{2q+1}$. Note that $\hat{\mu}_X^E = \sum_{i=1}^{n_1} \hat{p}_i^E (n/n_1) X_{1,i}$, where $\hat{p}^E = (\hat{p}_i^E) \in \mathbb{P}^n$ is the solution of (42).

Similarly, we can re-express the two-sample WEL problem (32) as a one-sample WEL for the mean. To this end, let $p_i = (n_1/n)p_{1,i}$, $\hat{p}_i^E = (n_1/n)\hat{p}_{1,i}^E$, $Z_i = n/n_1(X_{1,i}, X_{1,i}, 1)$, $p_{n_1+j} = (n_2/n)p_{2,j}$, $\hat{p}_{n_1+j}^E = (n_2/n)\hat{p}_{2,j}^E$, $Z_{n_1+j} = n/n_2(0, -X_{2,j}, 0)$, where $1 \leq i \leq n_1$ and $1 \leq j \leq n_2$. Then (32) is equivalent to:

$$\hat{\ell}_{X_1, X_2}^{WE}(\mu) = \max_{p \in \mathbb{P}^n} \left\{ \sum_{i=1}^n (n\hat{p}_i^E) \log(p_i/\hat{p}_i^E) : \sum_{i=1}^n p_i Z_i = \nu \right\}, \quad (43)$$

where $p = (p_i) \in \mathbb{P}^n$, $Z_i \in \mathbb{R}^{2q+1}$, $\hat{p}^E = (\hat{p}_i^E) \in \mathbb{P}^n$, and $\nu = (\mu, 0, 1) \in \mathbb{R}^{2q+1}$. These reformulations of the two-sample problems as one-sample problems are similar to the reformulation of the stratified case as an unstratified case used by Wu (2005).

4 Treatment effect adjusting for covariates

4.1 Empirical/Euclidean likelihood

Let $Z_{1:n_k}^{(k)} = \{Z_{k,1}, \dots, Z_{k,n_k}\}$ be two independent i.i.d. samples from the distributions P_k on \mathbb{R}^{p+q} , where $Z_{k,i} = (X_{k,i}, Y_{k,i}) \in \mathbb{R}^{p+q}$, $Y_{k,i} \in \mathbb{R}^p$ is the response and $X_{k,i} \in \mathbb{R}^q$ is the covariate corresponding to the i th subject in the k th group, $1 \leq i \leq n_k$, $k = 1, 2$. The parameter of interest is the treatment effect $\Delta_Y = E(Y_1) - E(Y_2)$, where $Z_k = (X_k, Y_k) \sim P_k$. Let $V_{Z_k} = \text{Var}(Z_k)$, $V_{Y_k} = \text{Var}(Y_k)$, $V_{X_k} = \text{Var}(X_k)$, and $V_{X_k Y_k} = V_{Y_k X_k}^T = \text{Cov}(X_k, Y_k)$. The randomization implies that the covariates $X_{k,i}$ have equal means between the treatment groups, i.e., $E(X_1) = E(X_2)$. The randomization also implies that $V_{X_1} = V_{X_2}$; for notational purposes, we do not make this distinction here.

The EL function for (μ_X, Δ_Y) is given by:

$$\hat{\ell}_{Z_1, Z_2}^E(\mu, \Delta) = \max_{p_k \in \mathbb{P}^{n_k}} \left\{ \sum_{k=1}^2 \sum_{i=1}^{n_k} \log(n_k p_{k,i}) : \sum_{i=1}^{n_k} p_{k,i} X_{k,i} = \mu, \sum_{i=1}^{n_1} p_{1,i} Y_{1,i} - \sum_{i=1}^{n_2} p_{2,i} Y_{2,i} = \Delta \right\}. \quad (44)$$

The profile EL function for Δ_Y is $\hat{\phi}_{Z_1, Z_2}^E(\Delta) = \max_{\mu \in \mathbb{R}^q} \hat{\ell}_{Z_1, Z_2}^E(\mu, \Delta)$. Equivalently,

$$\hat{\phi}_{Z_1, Z_2}^E(\Delta) = \max_{p_k \in \mathbb{P}^{n_k}} \left\{ \sum_{k=1}^2 \sum_{i=1}^{n_k} \log(n_k p_{k,i}) : \sum_{i=1}^{n_1} p_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} p_{2,i} Z_{2,i} = v \right\}, \quad (45)$$

where, with a slight abuse of notation, $v = (0, \Delta) \in \mathbb{R}^{p+q}$. Using Lagrange multipliers (see the appendix), we obtain:

$$\hat{p}_{k,i}^E(\Delta) = \frac{1}{-\hat{\eta}_k^E(\Delta) + (-1)^{k+1} n \hat{\lambda}^E(\Delta)^T Z_{k,i}}, \quad (46)$$

where $\hat{\eta}_k^E(\Delta)$ and $\hat{\lambda}^E(\Delta)$ satisfy (b) and (c) of (76). Since $\hat{\phi}_{Z_1, Z_2}^E(\Delta) \leq \hat{\ell}_{X_1, X_2}^E(\hat{\mu}_X^E)$ with equality when $\Delta = \hat{\Delta}_Y^E$, where $\hat{\Delta}_Y^E = \sum_{i=1}^{n_1} \hat{p}_{1,i}^E Y_{1,i} - \sum_{i=1}^{n_2} \hat{p}_{2,i}^E Y_{2,i}$, $\hat{\ell}_{X_1, X_2}^E(\mu)$ is defined in (24), and \hat{p}_k^E are given by (26), then $\hat{\Delta}_Y^E$ is the MELE of Δ_Y . By Theorems 1 and 2 of Qin and Lawless (1994),

$$n^{1/2}(\hat{\Delta}_Y^E - \Delta_Y) \xrightarrow{d} N(0, \Sigma_Y) \quad \text{and} \quad -2\hat{\phi}_{Z_1, Z_2}^E(\Delta_Y) + 2\hat{\phi}_{Z_1, Z_2}^E(\hat{\Delta}_Y^E) \xrightarrow{d} \chi_p^2,$$

where $\Sigma_Y = V_Y - V_{YX} V_X^{-1} V_{XY}$, $V_Y = \pi_1^{-1} V_{Y_1} + \pi_2^{-1} V_{Y_2}$, $V_X = \pi_1^{-1} V_{X_1} + \pi_2^{-1} V_{X_2}$, and $V_{YX} = V_{XY}^T = \pi_1^{-1} V_{Y_1 X_1} + \pi_2^{-1} V_{Y_2 X_2}$.

Both $\hat{\Delta}_Y^E$ and $\bar{\Delta}_Y = \bar{Y}_1 - \bar{Y}_2$ are consistent estimators of Δ_Y . Since $\Sigma_Y \preceq V_Y$, then $\hat{\Delta}_Y^E$ is more efficient than $\bar{\Delta}_Y$. If, in particular, $\text{rank}(V_{YX}) \geq p$, then $\hat{\Delta}_Y^E$ is strictly more efficient than $\bar{\Delta}_Y$. Note that, we use here the auxiliary information that the covariate means are equal due to randomization. However, randomization also implies that higher order moments of the covariates are equal between the treatment groups. Our results apply similarly if we consider equality of higher order moments by redefining X to include higher order moments of covariates, and thus, without loss of generality, we only consider the auxiliary information of equal means.

If we use the Euclidean distance, we obtain the profile UL function $\hat{\phi}_{Z_1, Z_2}^U(\Delta) = \max_{\mu \in \mathbb{R}^q} \hat{\ell}_{Z_1, Z_2}^U(\Delta, \mu)$, where

$$\hat{\ell}_{Z_1, Z_2}^U(\Delta, \mu) = \max_{p_k \in \mathbb{P}^{n_k}} \left\{ -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k p_{k,i} - 1)^2 : \sum_{i=1}^{n_1} p_{1,i} Y_{1,i} - \sum_{i=1}^{n_2} p_{2,i} Y_{2,i} = \Delta, \sum_{i=1}^{n_k} p_{k,i} X_{k,i} = \mu \right\}.$$

Similarly to (45),

$$\hat{\phi}_{Z_1, Z_2}^U(\Delta) = \max_{p_k \in \mathbb{P}^{n_k}} \left\{ -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k p_{k,i} - 1)^2 : \sum_{i=1}^{n_1} p_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} p_{2,i} Z_{2,i} = v \right\}. \quad (47)$$

Let $\hat{\Delta}_Y^U = \text{argmax}_{\Delta \in \mathbb{R}^p} \hat{\phi}_{Z_1, Z_2}^U(\Delta)$ denote the MULE of Δ_Y . Similarly as above, we can write $\hat{\Delta}_Y^U = \sum_{i=1}^{n_1} \hat{p}_{1,i}^U Y_{1,i} - \sum_{i=1}^{n_2} \hat{p}_{2,i}^U Y_{2,i}$. Maximizing (47) over $q_k \in \mathbb{Q}^{n_k}$, yields the profile pseudo-UL function:

$$\hat{\phi}_{Z_1, Z_2}^P(\Delta) = \max_{q_k \in \mathbb{Q}^{n_k}} \left\{ -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k q_{k,i} - 1)^2 : \sum_{i=1}^{n_1} q_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} q_{2,i} Z_{2,i} = v \right\}. \quad (48)$$

The solutions $\hat{q}_k(\Delta) = (\hat{q}_{k,i}(\Delta)) \in \mathbb{Q}^{n_k}$ of (48) are given by (see the appendix):

$$\hat{q}_{k,i}(\Delta) = n_k^{-1} (1 + (-1)^k (n/n_k) (\bar{\Delta}_Z - v)^T \hat{V}_Z^{-1} (Z_{k,i} - \bar{Z}_k)), \quad (49)$$

where $\hat{V}_Z = (n/n_1) \hat{V}_{Z_1} + (n/n_2) \hat{V}_{Z_2}$, $\bar{\Delta}_Z = (\bar{\Delta}_X, \bar{\Delta}_Y) \in \mathbb{R}^{p+q}$, $\bar{\Delta}_X = \bar{X}_1 - \bar{X}_2$, and $\bar{\Delta}_Y = \bar{Y}_1 - \bar{Y}_2$. By (49), we obtain (see the appendix):

$$-2\hat{\phi}_{Z_1, Z_2}^P(\Delta) = n(\bar{\Delta}_Z - v)^T \hat{V}_Z^{-1} (\bar{\Delta}_Z - v). \quad (50)$$

Let $\hat{\Delta}_Y^P$ be the maximum pseudo UL estimator of Δ_Y . By (50), we obtain (see the appendix):

$$\hat{\Delta}_Y^P = \sum_{i=1}^{n_1} \hat{q}_{1,i} Y_{1,i} - \sum_{i=1}^{n_2} \hat{q}_{2,i} Y_{2,i} = \bar{\Delta}_Y - \hat{V}_{YX} \hat{V}_X^{-1} \bar{\Delta}_X, \quad (51)$$

where $\hat{q}_{k,i} = \hat{q}_{k,i}(\hat{\mu}_X^P)$ and $\hat{q}_{k,i}(\mu)$ are given by (29). Notice that Koch's estimator has the same expression as (51) (see section 4.4). By (49), then $\text{Pr}(\hat{q}_k(\Delta_Y) \in \mathbb{P}^{n_k}) \rightarrow 1$, so that $n^{1/2}(\hat{\Delta}_Y^U - \Delta_Y) \xrightarrow{d} N(0, \Sigma_Y)$ and $-2\hat{\phi}_{Z_1, Z_2}^U(\Delta_Y) + 2\hat{\phi}_{Z_1, Z_2}^U(\hat{\Delta}_Y^U) \xrightarrow{d} \chi_p^2$.

4.2 Weighted empirical/Euclidean likelihood

The profile WEL function is given by:

$$\hat{\phi}_{Z_1, Z_2}^{\text{WE}}(\Delta) = \max_{p_k \in \mathbb{P}^{n_k}} \left\{ \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k \hat{p}_{k,i}^{\text{E}}) \log(p_{k,i} / \hat{p}_{k,i}^{\text{E}}) : \sum_{i=1}^{n_1} p_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} p_{2,i} Z_{2,i} = v \right\}, \quad (52)$$

where $\hat{p}_k^{\text{E}} = (\hat{p}_{k,i}^{\text{E}}) \in \mathbb{P}^{n_k}$ satisfy (26). Solving (52), yields (see the appendix):

$$\hat{p}_{k,i}^{\text{WE}}(\Delta) = \frac{n_k \hat{p}_{k,i}^{\text{E}}}{-\hat{\eta}_k^{\text{WE}}(\Delta) + (-1)^{k+1} n \hat{\lambda}^{\text{WE}}(\Delta)^T Z_{k,i}}, \quad (53)$$

where $\hat{\eta}_k^{\text{WE}}(\Delta)$ and $\hat{\lambda}^{\text{WE}}(\Delta) \in \mathbb{R}^{p+q}$ satisfy (b) and (c) of (78). Similarly as before, the maximum WEL estimator and the MELE of Δ_Y are the same. Theorem 3 shows the large sample behavior of the Lagrange multipliers and of the WEL function in local neighborhoods of Δ_Y , and the asymptotic null distribution of the weighted ELRT.

Theorem 3. Let $Z_{1:n_k}^{(k)} = \{Z_{k,1}, \dots, Z_{k,n_k}\}$ be two independent i.i.d. samples from the distributions P_k on \mathbb{R}^{p+q} , where $Z_{k,i} = (X_{k,i}, Y_{k,i}) \in \mathbb{R}^{p+q}$, $X_{k,i} \in \mathbb{R}^q$, $Y_{k,i} \in \mathbb{R}^p$, $E(X_1) = E(X_2)$, $V_{Z_k} = \text{Var}(Z_k)$ are non-singular, and $Z_k = (X_k, Y_k) \sim P_k$. Let $\Delta_Y = E(Y_1) - E(Y_2)$. Assume that $n_k/n \rightarrow \pi_k > 0$, where $n = n_1 + n_2$. Then for every $K > 0$,

$$\max_{\Delta \in \mathbb{K}_n} \|\hat{\lambda}^{\text{WE}}(\Delta) - V_Z^{-1}(\hat{\Delta}_Z^{\text{E}} - v)\| = o_P(n^{-1/2}), \quad (54)$$

$$\max_{\Delta \in \mathbb{K}_n} |-2\hat{\phi}_{Z_1, Z_2}^{\text{WE}}(\Delta) - n(\hat{\Delta}_Y^{\text{E}} - \Delta)^T \Sigma_Y^{-1}(\hat{\Delta}_Y^{\text{E}} - \Delta)| = o_P(1), \quad (55)$$

and

$$\max_{\Delta \in \mathbb{K}_n} \max_{1 \leq i \leq n_k} |\hat{p}_{k,i}^{\text{WE}}(\Delta) - \hat{p}_{k,i}^{\text{E}}(1 + (-1)^k (n/n_k)(\hat{\Delta}_Z^{\text{E}} - v)^T V_Z^{-1}(Z_{k,i} - \tilde{Z}_k^{\text{E}}))| = o_P(n^{-1}), \quad (56)$$

where $\mathbb{K}_n = \{\Delta : \|\Delta - \Delta_Y\| \leq Kn^{-1/2}\}$, $v = (0, \Delta) \in \mathbb{R}^{p+q}$, $\hat{\Delta}_Z^{\text{E}} = \tilde{Z}_1^{\text{E}} - \tilde{Z}_2^{\text{E}}$, $\tilde{Z}_k^{\text{E}} = \sum_{i=1}^{n_k} \hat{p}_{k,i}^{\text{E}} Z_{k,i}$, and $\hat{p}_k^{\text{E}} = (\hat{p}_{k,i}^{\text{E}}) \in \mathbb{P}^{n_k}$ satisfy (26). Furthermore,

$$-2\hat{\phi}_{Z_1, Z_2}^{\text{WE}}(\Delta_Y) \xrightarrow{d} \chi_p^2. \quad (57)$$

The profile WUL function is given by:

$$\hat{\phi}_{Z_1, Z_2}^{\text{WU}}(\Delta) = \max_{p_k \in \mathbb{P}^{n_k}} \left\{ -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k \hat{p}_{k,i}^{\text{U}}) (p_{k,i} / \hat{p}_{k,i}^{\text{U}} - 1)^2 : \sum_{i=1}^{n_1} p_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} p_{2,i} Z_{2,i} = v \right\}, \quad (58)$$

where $\hat{p}_{k,i}^{\text{U}} = \hat{p}_{k,i}^{\text{U}}(\hat{\mu}_X^{\text{U}})$, $\hat{p}_k^{\text{U}}(\mu)$ maximize (27), and $\hat{\mu}_X^{\text{U}}$ is the MULE of μ_X . Maximizing (58) over $q_k \in \mathbb{Q}^{n_k}$, yields the profile pseudo WUL function:

$$\hat{\phi}_{Z_1, Z_2}^{\text{WP}}(\Delta) = \max_{q_k \in \mathbb{Q}^{n_k}} \left\{ -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k \hat{q}_{k,i}) (q_{k,i} / \hat{q}_{k,i} - 1)^2 : \sum_{i=1}^{n_1} q_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} q_{2,i} Z_{2,i} = v \right\}, \quad (59)$$

where $\hat{q}_{k,i} = \hat{q}_{k,i}(\hat{\mu}_X^{\text{P}})$, $\hat{q}_{k,i}(\mu)$ are given by (29), and $\hat{\mu}_X^{\text{P}}$ is given by (31). The solutions $\hat{q}_k^{\text{W}}(\Delta) = (\hat{q}_{k,i}^{\text{W}}(\Delta)) \in \mathbb{Q}^{n_k}$ of (59) in terms of Δ are given by (see the appendix):

$$\hat{q}_{k,i}^{\text{W}}(\Delta) = \hat{q}_{k,i} (1 + (-1)^k (n/n_k)(\hat{\Delta}_Z^{\text{P}} - v)^T \tilde{V}_Z^{-1}(Z_{k,i} - \tilde{Z}_k)), \quad (60)$$

where $\tilde{Z}_k = n_k^{-1} \sum_{i=1}^{n_k} \hat{q}_{k,i} Z_{k,i}$, $\hat{\Delta}_Z^{\text{P}} = \tilde{Z}_1 - \tilde{Z}_2$, $\tilde{V}_Z = \sum_{i=1}^{n_k} \hat{q}_{k,i} (Z_{k,i} - \tilde{Z}_k)(Z_{k,i} - \tilde{Z}_k)^T$, $\tilde{V}_Z = (n/n_1)\tilde{V}_{Z_1} + (n/n_2)\tilde{V}_{Z_2}$, $\hat{q}_{k,i} = \hat{q}_{k,i}(\hat{\mu}_X^{\text{P}})$, and $\hat{q}_{k,i}(\mu)$ are given by (29). Substituting $\hat{q}_{k,i}^{\text{W}}(\Delta)$ for $q_{k,i}$ in (59), yields the following closed-form expression for the profile pseudo WUL function (see the appendix):

$$-2\hat{\phi}_{Z_1, Z_2}^{\text{WP}}(\Delta) = n(\hat{\Delta}_Y^{\text{P}} - \Delta)^T \tilde{\Sigma}_Y^{-1}(\hat{\Delta}_Y^{\text{P}} - \Delta), \quad (61)$$

where $\tilde{\Sigma}_Y = \tilde{V}_Y - \tilde{V}_{YX} \tilde{V}_X^{-1} \tilde{V}_{XY}$ and $\hat{\Delta}_Y^{\text{P}} = (0, \hat{\Delta}_Y^{\text{P}}) \in \mathbb{R}^{p+q}$. Since $\Pr(\hat{q}_k^{\text{W}}(\Delta_Y) \in \mathbb{P}^{n_k}) \rightarrow 1$, then $-2\hat{\phi}_{Z_1, Z_2}^{\text{WU}}(\Delta_Y) \xrightarrow{d} \chi_p^2$. Since the pseudo UL is equivalent to Koch's nonparametric covariance adjustment, the pseudo WUL can be viewed as a weighted version of Koch's method.

4.3 Computational aspects

We can re-express the two-sample EL problem (45) as a one-sample EL. To this end, let $p_i = (n_1/n)p_{1,i}$, $Z_i = n/n_1(Z_{1,i}, 1)$, $p_{n_1+j} = (n_2/n)p_{2,j}$, and $Z_{n_1+j} = -n/n_2(Z_{2,j}, 0)$, where $1 \leq i \leq n_1$, $1 \leq j \leq n_2$, and $n = n_1 + n_2$. Then (45) is equivalent to

$$\hat{\phi}_{Z_1, Z_2}^E(\Delta) = \max_{p \in \mathbb{P}^n} \left\{ \sum_{i=1}^n \log(np_i) : \sum_{i=1}^n p_i Z_i = \nu \right\}, \quad (62)$$

where $p = (p_i) \in \mathbb{P}^n$, $Z_i \in \mathbb{R}^{p+q+1}$, and $\nu = (0, \Delta, 1) \in \mathbb{R}^{p+q+1}$. Note that $\hat{\Delta}_Y^E = \sum_{i=1}^{n_1} (n/n_1) \hat{p}_i^E Y_{1,i} - \sum_{i=1}^{n_2} (n/n_2) \hat{p}_{n_1+i}^E Y_{2,i}$, where $\hat{p}^E = (\hat{p}_i^E) \in \mathbb{P}^n$ is the solution of (62).

Similarly, we can re-express the two-sample WEL problem (52) as a one-sample WEL. To this end, let $p_i = (n_1/n)p_{1,i}$, $\hat{p}_i^E = (n_1/n)\hat{p}_{1,i}^E$, $Z_i = n/n_1(Z_{1,i}, 1)$, $p_{n_1+j} = (n_2/n)p_{2,j}$, $\hat{p}_{n_1+j}^E = (n_2/n)\hat{p}_{2,j}^E$, $Z_{n_1+j} = -n/n_2(Z_{2,j}, 0)$, where $1 \leq i \leq n_1$ and $1 \leq j \leq n_2$. Then (52) is equivalent to:

$$\hat{\phi}_{Z_1, Z_2}^{wE}(\Delta) = \max_{p \in \mathbb{P}^n} \left\{ \sum_{i=1}^n (n\hat{p}_i^E) \log(p_i/\hat{p}_i^E) : \sum_{i=1}^n p_i Z_i = \nu \right\}, \quad (63)$$

where $p = (p_i) \in \mathbb{P}^n$, $\hat{p}^E = (\hat{p}_i^E) \in \mathbb{P}^n$, and $\nu = (0, \Delta, 1) \in \mathbb{R}^{p+q+1}$.

4.4 Alternative estimators for the treatment effect

4.4.1 Unadjusted treatment effect estimation

If the covariate information is not used, the statistic usually employed to estimate Δ_Y is the (unadjusted) treatment effect estimate $\bar{\Delta}_Y = \bar{Y}_1 - \bar{Y}_2$. The Wald test statistic for testing $H_0 : \Delta_Y = \Delta_Y^0$ is $\hat{T}_Y = n(\bar{\Delta}_Y - \Delta_Y^0)^T \hat{V}_Y^{-1} (\bar{\Delta}_Y - \Delta_Y^0)$. Then $n^{1/2}(\bar{\Delta}_Y - \Delta_Y) \xrightarrow{d} N(0, V_Y)$ and $\hat{T}_Y \xrightarrow{d} \chi_p^2$ under H_0 , where V_Y is defined in section 4.1.

4.4.2 Koch's nonparametric covariance adjustment

The randomization implies that the covariates have equal population means, i.e., $E(X_1) = E(X_2)$, and this constitutes the auxiliary information used by Koch's method. The Koch's estimator of Δ_Y is defined as $\hat{\Delta}_Y^K = \operatorname{argmin}_{\Delta \in \mathbb{R}^p} \hat{Q}_Z(\Delta)$, where $\hat{Q}_Z(\Delta) = (\bar{\Delta}_Z - \nu)^T \hat{V}_Z^{-1} (\bar{\Delta}_Z - \nu)$ and $\nu = (0, \Delta) \in \mathbb{R}^{p+q}$. Note that $\hat{Q}_Z(\Delta)$ corresponds to the pseudo UL function (50). Hence $\hat{\Delta}_Y^K = \bar{\Delta}_Y - \hat{V}_{YX} \hat{V}_X^{-1} \bar{\Delta}_X$. Let $\hat{T}_K = n(\hat{\Delta}_Y^K - \Delta_Y^0)^T \hat{\Sigma}_Y^{-1} (\hat{\Delta}_Y^K - \Delta_Y^0)$ denote the Wald statistic; then $n^{1/2}(\hat{\Delta}_Y^K - \Delta_Y) \xrightarrow{d} N(0, \Sigma_Y)$ and $\hat{T}_K \xrightarrow{d} \chi_p^2$ under H_0 , where Σ_Y is defined in section 4.1.

4.4.3 Multivariate analysis of covariance

Consider the multivariate analysis of covariance (MANCOVA) model to estimate Δ_Y :

$$Y_{k,i} = \mu_k + B_k \hat{X}_{k,i} + \epsilon_{k,i} = \mu_k + \begin{pmatrix} \hat{X}_{k,i}^T \beta_{k,1} \\ \vdots \\ \hat{X}_{k,i}^T \beta_{k,p} \end{pmatrix} + \epsilon_{k,i}, \quad (64)$$

where $Y_{k,i} \in \mathbb{R}^p$ and $\hat{X}_{k,i} \in \mathbb{R}^q$ denote the response and the (centered) covariate corresponding to the i th subject in the k th group, $\mu_k \in \mathbb{R}^p$ is the k th group mean, $\beta_{k,j} \in \mathbb{R}^q$ is the linear effect of the covariates on the j th component of the response in the k th group, $B_k = (\beta_{k,j})$ is the $p \times q$ matrix of regression coefficients corresponding to the k th group, n_k is the number of subjects in the k th group, and $\epsilon_{k,i} \sim$ i.i.d. $(0, V_{\epsilon_k})$ are the random errors, with $1 \leq i \leq n_k$, $1 \leq j \leq p$, and $k = 1, 2$. The covariates $\hat{X}_{k,i}$ are centered at 0 by subtracting the (overall) mean, i.e., $\hat{X}_{k,i} = X_{k,i} - \bar{X}$, where $\bar{X} = n^{-1} \sum_{k=1}^2 \sum_{i=1}^{n_k} X_{k,i}$ and $n = n_1 + n_2$. In this parametrization

of the model, μ_k is the k th group mean response obtained when the covariates are set equal to the (overall) population mean. The parameter of interest is the treatment effect $\Delta_Y = \mu_1 - \mu_2$.

We rewrite (64) in an equivalent form

$$Y_{k,i} = \mu_k + (I_p \otimes \hat{X}_{k,i}^T) \beta_k + \epsilon_{k,i},$$

where $\beta_k = \text{vec}(B_k^T) \in \mathbb{R}^{pq}$, $\text{vec}(\cdot)$ is the vectorization operator, and \otimes is the Kronecker product. The least squares estimates (LSE) of μ_k and β_k are given by (see appendix):

$$\hat{\mu}_k = \bar{Y}_k - (I_p \otimes \hat{X}_k^T) \hat{\beta}_k, \quad \hat{\beta}_k = \text{vec}(\hat{V}_{X_k}^{-1} \hat{V}_{X_k Y_k}), \quad (65)$$

where $\hat{X}_k = \bar{X}_k - \bar{X}$. The LSE of Δ_Y is thus given by:

$$\begin{aligned} \hat{\Delta}_Y^M &= \hat{\mu}_1 - \hat{\mu}_2 = \bar{Y}_1 - (I_p \otimes \hat{X}_1^T) \text{vec}(\hat{V}_{X_1}^{-1} \hat{V}_{X_1 Y_1}) - \bar{Y}_2 + (I_p \otimes \hat{X}_2^T) \text{vec}(\hat{V}_{X_2}^{-1} \hat{V}_{X_2 Y_2}) \\ &= \bar{Y}_1 - \hat{V}_{Y_1 X_1} \hat{V}_{X_1}^{-1} \hat{X}_1 - \bar{Y}_2 + \hat{V}_{Y_2 X_2} \hat{V}_{X_2}^{-1} \hat{X}_2. \end{aligned}$$

Since $\hat{X}_1 = \bar{X}_1 - \bar{X} = (n_2/n) \bar{\Delta}_X$ and $\hat{X}_2 = \bar{X}_2 - \bar{X} = -(n_1/n) \bar{\Delta}_X$, then

$$\hat{\Delta}_Y^M = \bar{\Delta}_Y - ((n_2/n) \hat{V}_{Y_1 X_1} \hat{V}_{X_1}^{-1} + (n_1/n) \hat{V}_{Y_2 X_2} \hat{V}_{X_2}^{-1}) \bar{\Delta}_X.$$

Under regularity conditions $n^{1/2}(\hat{\Delta}_Y^M - \Delta_Y) \xrightarrow{d} N(0, \Sigma_M)$, where

$$\Sigma_M = V_Y + (\pi_2 V_{Y_1 X_1} V_{X_1}^{-1} + \pi_1 V_{Y_2 X_2} V_{X_2}^{-1}) V_X (\pi_2 V_{Y_1 X_1} V_{X_1}^{-1} + \pi_1 V_{Y_2 X_2} V_{X_2}^{-1})^T - 2(\pi_2 V_{Y_1 X_1} V_{X_1}^{-1} + \pi_1 V_{Y_2 X_2} V_{X_2}^{-1}) V_{XY}.$$

Since $V_{X_1} = V_{X_2}$, then $V_X = (\pi_1^{-1} + \pi_2^{-1}) V_{X_1}$; hence, $V_{X_1} = V_{X_2} = (\pi_1 \pi_2) V_X$. Replacing V_{X_1} and V_{X_2} in terms of V_X , yields $\Sigma_M = V_Y - V_{YX} V_X^{-1} V_{XY}$. Hence $\hat{\Delta}_Y^M$ is first order equivalent to Koch's estimator.

5 Simulation study

5.1 Simulation models

The computations were done in the R language (R Core Team, 2020). The EL weights were calculated using the R package `emplik` (Zhou and Yang, 2018) and the UL weights were calculated using the R package `quadprog` (Turlach and Weingessel, 2019). The WEL method for one sample was implemented using the R package `nleqslv` (Hasselman, 2019). We implemented the methods in an R library, called WELCAM (Weighted empirical and Euclidean likelihood covariate adjustment methods), which is available from the authors upon request. In the simulation study, we compared the EL and the UL with Koch's method, the unadjusted method, and MANCOVA in terms of root mean squared errors (RMSE) of the estimators and size distortions of the Wald tests, the ELRT, the ULRT and their weighted versions.

We simulated data according to a randomized experiment with two groups, a bivariate response, and a bivariate covariate which may be regarded as a multivariate version of the simulation model of Lesaffre and Senn (2003). Specifically, we generated samples $Z_{1:n_k}^{(k)}$ of i.i.d. draws from a multivariate normal distribution with mean 0 and covariance matrix V_Z , where $k = 1, 2$, $p = q = 2$, $n_1, n_2 \in \{50, 100, 200\}$, and

$$V_Z = \begin{pmatrix} 1 & \rho_1 & \rho_3 & \rho_3 \\ \rho_1 & 1 & \rho_3 & \rho_3 \\ \rho_3 & \rho_3 & 1 & \rho_2 \\ \rho_3 & \rho_3 & \rho_2 & 1 \end{pmatrix},$$

with $(\rho_1, \rho_2, \rho_3) \in \{(0, 0, 0.4), (0.3, 0.3, 0.6), (0.5, 0.5, 0.7)\}$. We also generated data from a centered and scaled Gamma distribution, but since the results were qualitatively similar to the normal case, we include the results for the normal case only; the results for the gamma case are available from the authors upon request.

The parameter of interest is the treatment effect $\Delta_Y = 0 \in \mathbb{R}^2$. For each simulation model, we generated $S = 10,000$ samples to approximate the RMSE of the estimators and the sizes of the tests.

To improve the small sample performance of the tests, we use an F-distribution approximation of the null distribution of the test statistics. Specifically, let \hat{T} be a test statistic for the treatment effect with a limiting χ_p^2 distribution under H_0 , with two groups of sizes n_1 and n_2 , and a covariate vector $X \in \mathbb{R}^q$. First, compute the scaled test statistic $p^{-1}\hat{T}$; then, instead of using the critical value of a scaled χ_p^2 distribution (i.e., the distribution of $p^{-1}Z^2$, where $Z^2 \sim \chi_p^2$), which, with a slight abuse of notation, is the F-distribution $F_{p,\infty}$, we use the critical values of F_{p,n_0} , where $n_0 = \min(n_1, n_2) - p - q$. That is, reject H_0 if $p^{-1}\hat{T} \geq F_{p,n_0;1-\alpha}$, where $F_{p,n_0;1-\alpha}$ is the $(1 - \alpha)$ quantile of F_{p,n_0} . We have also used a small sample adjustment based on a Wishart approximation for the distribution of the sum of independent sample covariance matrices (Nel and van der Merwe, 1986). Since both methods have similar small sample behavior, we did not include the simulation results for the latter.

5.2 Simulation results

Table 1 shows the RMSE of the unadjusted estimator ($\bar{\Delta}_Y$), Koch's estimator ($\hat{\Delta}_Y^K$), MANCOVA estimator ($\hat{\Delta}_Y^M$), MELE ($\hat{\Delta}_Y^E$), and MULE ($\hat{\Delta}_Y^U$). We did not include the results for the pseudo MULE since it is equivalent to Koch's estimator (see section 4.1). Note that $\bar{\Delta}_Y$ has the largest RMSE for all sample sizes and correlation coefficients, while $\hat{\Delta}_Y^M$, $\hat{\Delta}_Y^K$, $\hat{\Delta}_Y^E$, and $\hat{\Delta}_Y^U$ have almost identical values (with slightly larger values for $\hat{\Delta}_Y^U$ when $n_1 = n_2 = 50$). While the RMSE of $\bar{\Delta}_Y$ are similar for the same sample size and different correlation coefficients, the RMSE of $\hat{\Delta}_Y^K$, $\hat{\Delta}_Y^M$, $\hat{\Delta}_Y^E$, and $\hat{\Delta}_Y^U$ are smaller for higher correlation coefficients. This shows, as expected, that the adjusted estimators are more efficient for stronger correlations between the response and the covariates.

Table 1: Simulation estimates of root mean squared errors of the unadjusted estimator ($\bar{\Delta}_Y$), Koch's estimator ($\hat{\Delta}_Y^K$), MANCOVA estimator ($\hat{\Delta}_Y^M$), maximum empirical likelihood estimator ($\hat{\Delta}_Y^E$), and maximum Euclidean likelihood estimator ($\hat{\Delta}_Y^U$). The number of simulated samples was $S = 10,000$.

n_1, n_2	ρ	$\bar{\Delta}_Y$	$\hat{\Delta}_Y^K$	$\hat{\Delta}_Y^M$	$\hat{\Delta}_Y^E$	$\hat{\Delta}_Y^U$
50,50	0,0,.4	0.282	0.235	0.235	0.235	0.236
50,50	.3,.3,.6	0.283	0.191	0.191	0.191	0.193
50,50	.5,.5,.7	0.282	0.168	0.168	0.168	0.17
100,100	0,0,.4	0.201	0.166	0.166	0.166	0.166
100,100	.3,.3,.6	0.202	0.135	0.135	0.135	0.135
100,100	.5,.5,.7	0.2	0.118	0.118	0.118	0.118
150,50	0,0,.4	0.233	0.195	0.195	0.195	0.196
150,50	.3,.3,.6	0.23	0.155	0.155	0.155	0.157
150,50	.5,.5,.7	0.232	0.139	0.139	0.139	0.142
150,150	0,0,.4	0.162	0.134	0.134	0.134	0.134
150,150	.3,.3,.6	0.162	0.109	0.109	0.109	0.109
150,150	.5,.5,.7	0.162	0.096	0.096	0.096	0.096
200,100	0,0,.4	0.174	0.144	0.144	0.144	0.144
200,100	.3,.3,.6	0.172	0.116	0.116	0.116	0.116
200,100	.5,.5,.7	0.173	0.103	0.103	0.103	0.104
200,200	0,0,.4	0.141	0.116	0.116	0.116	0.116
200,200	.3,.3,.6	0.142	0.095	0.095	0.095	0.095
200,200	.5,.5,.7	0.142	0.083	0.083	0.083	0.083

Table 2 shows the simulation estimates of the sizes of the Wald tests based on the unadjusted estimator (\hat{T}_Y), Koch's estimator (\hat{T}_K), and MANCOVA estimator (\hat{T}_M), the empirical likelihood ratio test ($\hat{\phi}_Z^E$), the Euclidean likelihood ratio test ($\hat{\phi}_Z^U$), the pseudo Euclidean likelihood ratio test ($\hat{\phi}_Z^P$), the weighted empirical likelihood test ($\hat{\phi}_Z^{wE}$), the weighted Euclidean likelihood ratio test ($\hat{\phi}_Z^{wU}$), and the weighted pseudo Euclidean likelihood test ($\hat{\phi}_Z^{wP}$). The critical values of the tests were calculated using the small sample adjustment based on the F-distribution. Note that that all tests maintain their nominal levels very well for all sample sizes, with most empirical sizes between 0.045 and 0.055. The results for the Wald test based on Koch's estimator are slightly different from the pseudo ULRT because we used the unbiased estimators of covariance matrices for the former. The results using the chi-square critical values were inferior for smaller and moderate sample sizes ($n_1, n_2 \leq 150$), while they behaved similarly for larger sample sizes (results available from the authors upon request).

Table 2: Empirical sizes of the Wald tests based on the unadjusted estimator (\hat{T}_Y), Koch's estimator (\hat{T}_K), and MANCOVA estimator (\hat{T}_M), the empirical likelihood ratio test ($\hat{\phi}_Z^E$), the Euclidean likelihood ratio test ($\hat{\phi}_Z^U$), the pseudo Euclidean likelihood ratio test ($\hat{\phi}_Z^P$), the weighted empirical likelihood ratio test ($\hat{\phi}_Z^{wE}$), the weighted Euclidean likelihood ratio test ($\hat{\phi}_Z^{wU}$), and the weighted pseudo Euclidean likelihood ratio test ($\hat{\phi}_Z^{wP}$). The number of simulated samples was $S = 10,000$.

n_1, n_2	ρ	\hat{T}_Y	\hat{T}_K	\hat{T}_M	$\hat{\phi}_Z^E$	$\hat{\phi}_Z^U$	$\hat{\phi}_Z^P$	$\hat{\phi}_Z^{wE}$	$\hat{\phi}_Z^{wU}$	$\hat{\phi}_Z^{wP}$
50,50	0,0,.4	0.046	0.053	0.047	0.052	0.054	0.054	0.054	0.058	0.056
50,50	.3,.3,.6	0.047	0.054	0.049	0.054	0.056	0.055	0.056	0.06	0.057
50,50	.5,.5,.7	0.046	0.052	0.047	0.051	0.055	0.054	0.054	0.057	0.055
100,100	0,0,.4	0.05	0.05	0.048	0.049	0.051	0.051	0.051	0.052	0.052
100,100	.3,.3,.6	0.049	0.049	0.046	0.049	0.05	0.05	0.049	0.051	0.051
100,100	.5,.5,.7	0.047	0.05	0.047	0.05	0.051	0.051	0.05	0.052	0.052
150,50	0,0,.4	0.05	0.058	0.052	0.057	0.06	0.06	0.059	0.066	0.061
150,50	.3,.3,.6	0.046	0.054	0.047	0.052	0.055	0.055	0.054	0.062	0.056
150,50	.5,.5,.7	0.05	0.058	0.054	0.058	0.061	0.061	0.06	0.068	0.062
150,150	0,0,.4	0.048	0.052	0.051	0.051	0.052	0.052	0.051	0.053	0.053
150,150	.3,.3,.6	0.047	0.049	0.047	0.048	0.049	0.049	0.048	0.05	0.05
150,150	.5,.5,.7	0.047	0.047	0.044	0.045	0.047	0.047	0.046	0.048	0.048
200,100	0,0,.4	0.051	0.054	0.05	0.052	0.055	0.055	0.053	0.056	0.056
200,100	.3,.3,.6	0.053	0.053	0.051	0.052	0.053	0.053	0.053	0.055	0.054
200,100	.5,.5,.7	0.05	0.054	0.05	0.052	0.055	0.055	0.054	0.056	0.056
200,200	0,0,.4	0.05	0.049	0.047	0.048	0.049	0.049	0.049	0.05	0.05
200,200	.3,.3,.6	0.051	0.049	0.048	0.049	0.05	0.05	0.049	0.05	0.05
200,200	.5,.5,.7	0.049	0.049	0.048	0.049	0.05	0.05	0.05	0.05	0.05

6 Data analysis

The data set used for this analysis is from the IMPROvE study (Amris et al., 2014), a randomized, controlled, single-center trial with two groups that evaluated the functional and psychological outcomes of a two-week, group-based, multimodal, interdisciplinary rehabilitation program, tailored for patients with chronic widespread pain (CWP). Participants were recruited from the outpatient clinic of the Department of Rheumatology, Frederiksberg Hospital, Denmark. Female patients aged 18 years and older, with CWP diagnosed according to the American College of Rheumatology 1990 definition of widespread pain Wolfe (1990), were eligible for the trial. Exclusion criteria were concurrent psychiatric disorders not related to the pain and other uncontrolled rheumatic or medical diseases that may cause CWP. Based on an a priori sample size

calculation, the trial enrolled 192 female patients with CWP randomized 1:1 to either treatment or a waiting list control group, stratified by the baseline level of functional ability as measured with the Assessment of Motor and Process Skills (AMPS) Fisher (2010). One patient withdrew before receiving the result of the randomization, and thus the study sample had 191 female patients diagnosed with CWP who were randomly assigned to either the intervention group ($n_1 = 96$) or the control group ($n_2 = 95$).

The AMPS is an individualized, observation-based evaluation of functional ability performed by occupational therapists, developed to establish the extent of an individual's ability to execute and complete familiar and relevant activities of daily living (ADL). Two ADL ability measures are provided by the AMPS, one for ADL motor ability (ADLm) and another for ADL process ability (ADLp). The ADLm ability measure is an indication of how much effort and/or fatigue the person demonstrated when performing ADL tasks, while the ADLp ability measure is an indication of how timely and well-organized the person was observed to be during task performance. The AMPS is based on a Rasch measurement model and it provides equal-interval linear measures of ADL ability expressed in logits. The ADLm and ADLp measures evaluated at the 6-months follow-up represent the bivariate response of interest, while ADLm and ADLp measures evaluated at baseline represent the two covariates. The average ADLm at baseline were not significantly different between the treated and controls (1.14 vs 1.13, $p = 0.89$), and the average ADLp at baseline were also not significantly different between the groups (1.06 vs 1.05, $p = 0.96$). However, the average ADLm at the 6-months follow-up was significantly higher in the treated group (1.36 vs 1.16, $p < 0.01$), and the average ADLp at 6-months follow-up was also significantly higher in the treated group (1.13 vs 0.93, $p < 0.01$).

We performed the data analysis using the covariate adjustment methods described in this paper. The unadjusted treatment effect estimates for ADLm and ADLp were 0.21 and 0.20, respectively, with a Wald test statistic for the overall treatment effect of $\hat{T}_Y = 23.98$ ($p < 0.01$). The treatment effect estimates for ADLm and ADLp were approximately 0.20 and 0.20 for all covariate adjustment methods. The MANCOVA Wald test statistic was $\hat{T}_M = 34.96$ ($p < 0.01$) and the Wald test based on Koch's estimator was $\hat{T}_K = 35.18$ ($p < 0.01$). The ELRT statistic was $\hat{\phi}_Z^E = 32.66$ ($p < 0.01$), the ULRT statistic was $\hat{\phi}_Z^U = 35.36$ ($p < 0.01$), the weighted ELRT statistic was $\hat{\phi}_Z^{WE} = 32.70$ ($p < 0.01$), and the weighted ULRT was $\hat{\phi}_Z^{WU} = 36.17$ ($p < 0.01$). All covariate adjustment methods provide the same conclusion that there is a significant positive effect of the multicomponent therapy for both ADL scales. Note that the unadjusted method is the least efficient (as evidenced by the smallest values of the Wald test statistic), while all covariate adjustment methods perform similarly well (as evidenced by similarly large values of the test statistics). Note also that the weighted test statistics were slightly higher than their unweighted versions (32.70 vs 32.66 for WEL vs EL and 36.17 vs 35.36

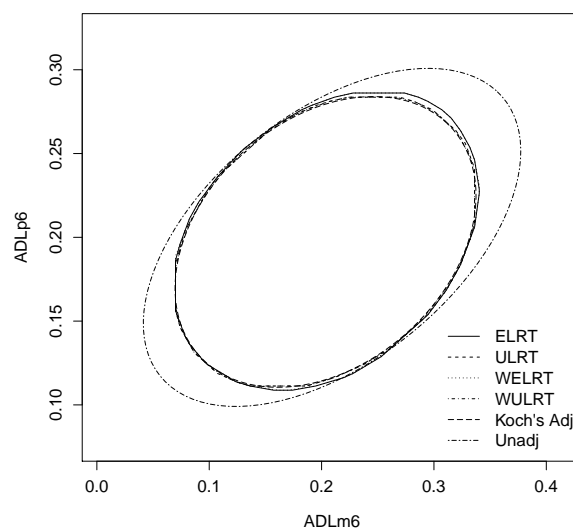


Figure 1: Test inversion 95% confidence regions for the treatment effect of ADLm and ADLp based on empirical likelihood ratio test (ELRT), Euclidean likelihood ratio test (ULRT), weighted empirical likelihood ratio test (WELRT), weighted Euclidean likelihood ratio test (WULRT), and Wald tests based on Koch's and the unadjusted estimators.

for WUL vs UL) and we could speculate that the weighted test statistics may have more power. However, a local asymptotic power calculation for the weighted methods and comparison with unweighted methods (Qin and Lawless, 1994; Peng and Schick, 2018) is a topic of future research and beyond the scope of this paper.

Figure 1 shows the 95% confidence regions for the treatment effects for the ADLm and ADLp scales obtained by inverting the ELRT, the ULRT, the weighted ELRT (WELRT), the weighted ULRT (WULRT), and the Wald tests based on $\hat{\Delta}_Y^K$ and $\bar{\Delta}_Y$. The confidence regions are all similar with the exception of inverting the Wald test using the unadjusted estimator which was the largest. This shows that, in agreement with our theoretical and simulation results, the unadjusted method is the least efficient and the empirical and Euclidean likelihood methods and their weighted version behave similarly. Lastly, although the confidence regions produced by different covariate adjustment methods are not the same, they all provide evidence of a significant positive effect of the multicomponent therapy on the ADLm and ADLp scales.

7 Conclusion

In this article, we proposed weighted empirical and Euclidean likelihood methods of inference for the mean with auxiliary information. In the particular case of randomized clinical trials, the auxiliary information consists of that the covariates have equal means among the treatment groups due to randomization. We studied the asymptotic properties of the proposed methods and showed that they are first order equivalent to their unweighted versions. We showed that allowing the weights to be negative for the UL, the resulting pseudo UL is equivalent to Koch's method. We further showed that the weighted pseudo UL can be viewed as a weighted version of Koch's method. We showed that our method can be applied for inference in sample surveys with auxiliary information and that it blends naturally the survey weights with the weights from the auxiliary information. In a simulation study, we compared the finite sample properties of these methods with alternative methods, such as the Koch's method and multivariate analysis of covariance. We illustrated an application of these methods to the practical situation of data analysis of a clinical trial data set.

Based on the results of our simulation study and its clear computational advantage, we recommend the use of Koch's method of inference for the treatment effect in randomized clinical trials. We have investigated additional simulation scenarios, e.g., when the multivariate response in one group has a normal distribution and in the other group has a scaled gamma distribution, and Koch's method has performed the best overall (results available from the authors upon request). Future research needs to evaluate the performance of these methods for generalized empirical likelihood (Baggerly, 1998), misspecification of auxiliary information (Giurcanu and Presnell, 2018, 2019), for high dimensional case (Giurcanu, 2016), and for more complex models such as causal inference in presence of confounders as indicated by a directed acyclic graph (Rubin, 1974).

Conflict of interest: Prof. George Luta is a member of the Editorial Board of Open Statistics. He was not involved in the review process of this article.

References

- Adimari, G. (1995), "Empirical likelihood confidence intervals for the difference between means," *Statistica*, 55, 87–94.
- Amris, K., Wærens, E. E., Christensen, R., Bliddal, H., and Danneskiold-Samsøe, B. (2014), "Interdisciplinary rehabilitation of patients with chronic widespread pain: Primary endpoint of the randomized, nonblinded, parallel-group IMPROvE trial," *PAIN*, 155(7), 1356–1364.
- Baggerly, K. A. (1998), "Empirical Likelihood as a Goodness-of-Fit Measure," *Biometrika*, 85, 535–547.
- Boyd, S., and Vandenberghe, L. (2004), *Convex Optimization*, Cambridge, New York: Cambridge.
- Chen, J., and Sitter, R. R. (1999), "A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys," *Statistica Sinica*, 9, 385–406.
- Deville, J. C., and Sarndal, C. E. (1992), "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, 87, 376–382.

- DiCiccio, T. J., Hall, P., and Romano, J. P. (1991), "Empirical Likelihood is Bartlett-correctable," *The Annals of Statistics*, 19, 1053–1061.
- Elze, M. C., Gregson, J., Baber, U., Williamson, E., Sartori, S., Mehran, R., Nichols, M., Stone, G. W., and Pocock, S. J. (2017), "Comparison of propensity score methods and covariate adjustment," *Journal of the American College of Cardiology*, 69(3), 345–347.
- Fienberg, S. E., and Tanur, J. M. (1987), "Experimental and sampling structures: parallels diverging and meeting," *International Statistical Review*, 55, 75–96.
- Fisher, A. G. (2010), *Assessment of Motor and Process Skills: Development, Standardization, and Administration Manual*, 7 edn, Fort Collins, Colorado: Three Star Press.
- Giurcanu, M. (2016), "Thresholding least-squares inference in high-dimensional regression models," *Electronic Journal of Statistics*, 10, 2124–2156.
- Giurcanu, M., and Presnell, B. (2018), "Bootstrap inference for misspecified moment condition models," *Annals of the Institute of Statistical Mathematics*, 70, 605–630.
- Giurcanu, M., and Presnell, B. (2019), "Oracle GMM estimation for misspecified models via thresholding," *Communications in Statistics - Theory and Methods*, 48(11), 2657–2686.
- Glenn, N. L., and Zhao, Y. (2006), "Weighted empirical likelihood estimates and their robustness properties," *Computational Statistics & Data Analysis*, 51, 5130–5141.
- Hall, P., and La Scala, B. (1990), "Methodology and Algorithms of Empirical Likelihood," *International Statistical Review*, 58, 109–127.
- Harville, D. A. (2008), *Matrix Algebra From A Statistician's Perspective*, New York: Springer.
- Hasselmann, B. (2019), *nleqslv: Solve Systems of Nonlinear Equations*. R package version 3.3.2.
- Jing, B. Y. (1995), "Two-sample empirical likelihood method," *Statistics & Probability Letters*, 24, 315–319.
- Kawaguchi, A., Koch, G. G., and Wang, X. (2011), "Stratified multivariate Mann-Whitney estimators for the comparison of two treatments with randomization based covariance adjustment," *Statistics in Biopharmaceutical Research*, 3(2), 217–231.
- Koch, G. G., Amara, I. A., Davis, G. W., and Gillings, D. B. (1982), "A review of some statistical methods for covariance analysis of categorical data," *Biometrics*, 38, 563–595.
- Koch, G. G., Tangen, C. M., Jung, J. W., and Amara, I. A. (1998), "Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them," *Statistics in Medicine*, 17, 1863–1892.
- Lesaffre, E., Bogærts, K., Li, X., and Bluhmki, E. (2002), "On the variability of covariate adjustment: experience with Koch's method for evaluating the absolute difference in proportions in randomized clinical trials," *Controlled Clinical Trials*, 23, 127–142.
- Lesaffre, E., and Senn, S. (2003), "A note on non-parametric ANCOVA for covariate adjustment in randomized clinical trials," *Statistics in Medicine*, 22, 3583–3596.
- Liu, Y., Zou, C., and Zhang, R. (2008), "Empirical likelihood for the two-sample mean problem," *Statistics & Probability Letters*, 78, 548–556.
- Nel, D. G., and van der Merwe, C. A. (1986), "A solution to the multivariate Behrens-Fisher problem," *Communications in Statistics. Theory and Methods*, 15, 3719–3735.
- Owen, A. (1988), "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, 75, 237–249.
- Owen, A. (1991), "Empirical Likelihood for Linear Models," *The Annals of Statistics*, 19, 1725–1747.
- Owen, A. (2001), *Empirical Likelihood*, Boca Raton: Chapman & Hall / CRC.
- Peng, H., and Schick, A. (2018), "Maximum empirical likelihood estimation and related topics," *Electronic Journal of Statistics*, 12, 2962–2994.
- Pocock, S. J., Assmann, S. E., Enos, L. E., and Kasten, L. E. (2002), "Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems," *Statistics in Medicine*, 21, 2917–2930.
- Puri, M. L., and Sen, P. K. (1971), *Nonparametric methods in multivariate analysis*, New York: Wiley.
- Qin, J., and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22(1), 300–325.
- Quade, D. (1967), "Rank analysis of covariance," *Journal of the American Statistical Association*, 62, 1187–1200.
- R Core Team (2020), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. R version 3.6.3.
- Ren, J. J. (2001), "Weighted Empirical Likelihood Ratio Confidence Intervals for the Mean with Censored Data," *Annals of the Institute of Statistical Mathematics*, 53, 498–516.
- Robinson, L. D., and Jewell, N. P. (1991), "Some surprising results about covariate adjustment in logistic regression models," *International Statistical Review*, 59(2), 227–240.
- Rubin, D. B. (1974), "Estimating Causal Effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66(5), 688–701.
- Saville, B. R., and Koch, G. G. (2013), "Estimating covariate-adjusted log hazard ratios in randomized clinical trials using Cox proportional hazards models and nonparametric randomization based analysis of covariance," *Journal of Biopharmaceutical Statistics*, 23(2), 477–490.
- Saville, B. R., LaVange, L. M., and Koch, G. G. (2011), "Estimating covariate-adjusted incidence density ratios for multiple time intervals in clinical trials using nonparametric randomization-based ANCOVA," *Statistics in Biopharmaceutical Research*, 3(2), 242–252.

- Snedecor, G. W., and Cochran, W. G. (1980), *Statistical Methods, 7th Edition*, Ames: Ames: Iowa State University Press.
- Tangen, C. M., and Koch, G. G. (2001), “Non-parametric analysis of covariance for confirmatory randomized clinical trials to evaluate dose-response relationships,” *Statistics in Medicine*, 20, 2585–2607.
- Tsao, M., and Wu, C. (2006), “Empirical likelihood inference for a common mean in the presence of heteroscedasticity,” *Canadian Journal of Statistics*, 34, 45–59.
- Tsiatis, A. A., Davidian, M., Zhang, M., and X., L. (2008), “Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach,” *Statistics in Medicine*, 27(23), 4658–4677.
- Turlach, B., and Weingessel, A. (2019), *quadprog: Functions to solve Quadratic Programming Problems*. R package version 1.5-7.
- Wilks, S. S. (1938), “The large-sample distribution of the likelihood ratio for testing composite hypotheses,” *The Annals of Mathematical Statistics*, pp. 60–62.
- Willan, A. R., Briggs, A. H., and Hoch, J. S. (2004), “Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data,” *Health Economics*, 13, 461–475.
- Wolfe, F. (1990), “The American College of Rheumatology 1990 Criteria for the classification of Fibromyalgia. Report of the Multicenter Criteria Committee,” *Arthritis and Rheumatism*, 33, 160–1172.
- Wu, C. (2002), “Empirical likelihood method for finite populations,” *Recent Advances in Statistical Methods*.
- Wu, C. (2005), “Algorithms and R codes for the pseudo empirical likelihood method in survey sampling,” *Survey Methodology*, 31, 239–243.
- Wu, C., and Rao, J. N. K. (2006), “Pseudo-empirical likelihood ratio confidence intervals for complex surveys,” *The Canadian Journal of Statistics*, 34(3), 359–375.
- Zhong, B., and Rao, J. N. K. (2000), “Empirical likelihood inference under stratified random sampling using auxiliary population information,” *Biometrika*, 87, 929–938.
- Zhou, M., and Yang, Y. (2018), *emplik: Empirical likelihood ratio for Censored/Truncated data*. R package version 1.0-4.3.

Appendix Proofs of theoretical results

Proof of (5). The Lagrangian is given by:

$$\mathcal{L}(q, \eta, \lambda) = -(1/2) \sum_{i=1}^n (nq_i - 1)^2 + \eta \left(\sum_{i=1}^n q_i - 1 \right) - n\lambda^T \left(\sum_{i=1}^n q_i X_i - \mu \right).$$

Setting the partial derivatives of the Lagrangian to zero, yields:

$$(a) -n^2 q_i + n + \eta - n\lambda^T X_i = 0, \quad (b) \sum_{i=1}^n q_i = 1, \quad (c) \sum_{i=1}^n q_i X_i = \mu.$$

Summing (a) over $1 \leq i \leq n$, yields $n\eta - n^2 \lambda^T \bar{X} = 0$. Solving for η in terms of λ , yields $\hat{\eta}(\lambda) = n\lambda^T \bar{X}$. Substituting $\hat{\eta}(\lambda)$ for η in (a), we obtain the solution for q_i in terms of λ :

$$\hat{q}_i(\lambda) = n^{-1} (1 - \lambda^T (X_i - \bar{X})). \quad (66)$$

Substituting $\hat{q}_i(\lambda)$ for q_i in (c), yields $\bar{X} - \hat{V}_X \lambda = \mu$. Solving for λ in terms of μ , yields $\hat{\lambda}(\mu) = \hat{V}_X^{-1} (\bar{X} - \mu)$. Substituting $\hat{\lambda}(\mu)$ for λ in (66) implies that (5) holds, as stated. \square

Proof of (10). Using the formula for a matrix inverse in block form (Harville, 2008, 5.16a, p. 98), by (8) we obtain:

$$\begin{aligned} -2\hat{\ell}_{Y|X}^{\text{P}}(\mu|\mu_X) &= n(\bar{X} - \mu_X)^T (\hat{V}_X^{-1} + \hat{V}_X^{-1} \hat{V}_{XY} \hat{\Sigma}_Y^{-1} \hat{V}_{YX} \hat{V}_X^{-1}) (\bar{X} - \mu_X) \\ &\quad - 2n(\bar{X} - \mu_X)^T \hat{V}_X^{-1} \hat{V}_{XY} \hat{\Sigma}_Y^{-1} (\bar{Y} - \mu) + n(\bar{Y} - \mu)^T \hat{\Sigma}_Y^{-1} (\bar{Y} - \mu) \\ &\quad - n(\bar{X} - \mu_X)^T \hat{V}_X^{-1} (\bar{X} - \mu_X) = n(\bar{X} - \mu_X)^T \hat{V}_X^{-1} \hat{V}_{XY} \hat{\Sigma}_Y^{-1} \hat{V}_{YX} \hat{V}_X^{-1} (\bar{X} - \mu_X) \\ &\quad - 2n(\bar{X} - \mu_X)^T \hat{V}_X^{-1} \hat{V}_{XY} \hat{\Sigma}_Y^{-1} (\hat{\mu}_Y^{\text{P}} - \mu + \hat{V}_{YX} \hat{V}_X^{-1} (\bar{X} - \mu_X)) \\ &\quad + n(\hat{\mu}_Y^{\text{P}} - \mu + \hat{V}_{YX} \hat{V}_X^{-1} (\bar{X} - \mu_X))^T \hat{\Sigma}_Y^{-1} (\hat{\mu}_Y^{\text{P}} - \mu + \hat{V}_{YX} \hat{V}_X^{-1} (\bar{X} - \mu_X)) \\ &= n(\hat{\mu}_Y^{\text{P}} - \mu)^T \hat{\Sigma}_Y^{-1} (\hat{\mu}_Y^{\text{P}} - \mu). \end{aligned}$$

\square

Proof of (12). The Lagrangian is given by

$$\mathcal{L}(p, \eta, \lambda) = \sum_{i=1}^n (n\hat{p}_i^{\text{E}}) \log(p_i / \hat{p}_i^{\text{E}}) + \eta \left(\sum_{i=1}^n p_i - 1 \right) - n\lambda^T \left(\sum_{i=1}^n p_i Z_i - \nu \right).$$

Setting the partial derivatives of the Lagrangian to zero, yields:

$$(a) \frac{n\hat{p}_i^{\text{E}}}{p_i} + \eta - n\lambda^T Z_i = 0, \quad (b) \sum_{i=1}^n p_i = 1, \quad (c) \sum_{i=1}^n p_i Z_i = \nu. \quad (67)$$

Multiplying (a) by p_i and summing over i , by (c) yields $n + \eta - n\lambda^T \nu = 0$. Solving for η in terms of λ , yields $\hat{\eta}^{\text{W}}(\lambda) = n\lambda^T \nu - n$. Substituting $\hat{\eta}^{\text{W}}(\lambda)$ for η in (a), then (12) holds, where by (c) the Lagrange multiplier $\hat{\lambda}^{\text{wE}}(\mu)$ satisfies (13). \square

Proof of Theorem 1. By Lemma C.2 of Giurcanu and Presnell (2018), any $o(1)$ neighborhood of $\mu_Z = (\mu_X, \mu_Y) \in \mathbb{R}^{p+q}$ is inside the convex hull of $Z_{1:n}$ with probability tending to one. Similarly to Qin and Lawless (1994), $\|\hat{\mu}_Z^{\text{E}} - \mu_Z\| = O_p(n^{-1/2})$. Let $\hat{\lambda}^{\text{wE}}(\mu) = \|\hat{\lambda}^{\text{wE}}(\mu)\| \hat{\theta}$, where $\hat{\theta} \in \mathbb{R}^{p+q}$ is a unit vector, and let further

$$\tilde{\nu}_Z^{\text{wE}}(\mu) = \sum_{i=1}^n \hat{p}_i^{\text{wE}}(\mu) (Z_i - \nu)(Z_i - \nu)^T \quad \text{and} \quad \tilde{V}_Z^{\text{E}}(\mu) = \sum_{i=1}^n \hat{p}_i^{\text{E}}(Z_i - \nu)(Z_i - \nu)^T.$$

Since

$$\frac{\hat{p}_i^E \hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v)}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v)} = \hat{p}_i^E \hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v) - \frac{\hat{p}_i^E (\hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v))^2}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v)}, \quad (68)$$

summing both sides of (68) over $1 \leq i \leq n$, by (12) and (13), yields:

$$0 = \hat{\lambda}^{\text{wE}}(\mu)^T (\hat{\mu}_Z^E - v) - \hat{\lambda}^{\text{wE}}(\mu)^T \tilde{V}_Z^{\text{wE}}(\mu) \hat{\lambda}^{\text{wE}}(\mu).$$

Hence, $\|\hat{\lambda}^{\text{wE}}(\mu)\| \hat{\theta}^T \tilde{V}_Z^{\text{wE}}(\mu) \hat{\theta} = \hat{\theta}^T (\hat{\mu}_Z^E - v)$. By (12),

$$\begin{aligned} \|\hat{\lambda}^{\text{wE}}(\mu)\| \hat{\theta}^T \tilde{V}_Z^{\text{wE}}(\mu) \hat{\theta} &\leq \|\hat{\lambda}^{\text{wE}}(\mu)\| \hat{\theta}^T \tilde{V}_Z^{\text{wE}}(\mu) \hat{\theta} (1 + \|\hat{\lambda}^{\text{wE}}(\mu)\| \max_{1 \leq i \leq n} \|Z_i - v\|) \\ &= \hat{\theta}^T (\hat{\mu}_Z^E - v) (1 + \|\hat{\lambda}^{\text{wE}}(\mu)\| \max_{1 \leq i \leq n} \|Z_i - v\|). \end{aligned} \quad (69)$$

Maximizing both sides of (69) with respect to μ , yields:

$$\left\{ \max_{\mu \in \mathbb{K}_n} \|\hat{\lambda}^{\text{wE}}(\mu)\| \right\} \left(\min_{\mu \in \mathbb{K}_n} \{ \hat{\theta}^T \tilde{V}_Z^{\text{wE}}(\mu) \hat{\theta} \} - \max_{\mu \in \mathbb{K}_n} \{ \max_{1 \leq i \leq n} \|Z_i - v\| \} \max_{\mu \in \mathbb{K}_n} \{ \hat{\theta}^T (\hat{\mu}_Z^E - v) \} \right) \leq \max_{\mu \in \mathbb{K}_n} \{ \hat{\theta}^T (\hat{\mu}_Z^E - v) \}. \quad (70)$$

Let α_1 and α_p denote the smallest and the largest eigenvalues of V_Z . Then,

$$\alpha_1 + o_P(1) \leq \min_{\mu \in \mathbb{K}_n} \{ \hat{\theta}^T \tilde{V}_Z^{\text{wE}}(\mu) \hat{\theta} \} \leq \alpha_p + o_P(1).$$

Since $\max_{1 \leq i \leq n} \|Z_i - \mu_Z\| = o_P(n^{1/2})$ (see, e.g., Owen, 2001, Lemma 11.22),

$$\max_{\mu \in \mathbb{K}_n} \max_{1 \leq i \leq n} \|Z_i - v\| \leq \max_{1 \leq i \leq n} \|Z_i - \mu_Z\| + \max_{\mu \in \mathbb{K}_n} \|\mu_Z - v\| = o_P(n^{1/2}).$$

Since $\max_{\mu \in \mathbb{K}_n} \|\hat{\mu}_Z^E - v\| = O_P(n^{-1/2})$, by (70),

$$\max_{\mu \in \mathbb{K}_n} \|\hat{\lambda}^{\text{wE}}(\mu)\| = O_P(n^{-1/2}).$$

Hence, we further obtain

$$\max_{\mu \in \mathbb{K}_n} \left\{ \max_{1 \leq i \leq n} |\hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v)| \right\} \leq \left\{ \max_{\mu \in \mathbb{K}_n} \|\hat{\lambda}^{\text{wE}}(\mu)\| \right\} \max_{\mu \in \mathbb{K}_n} \left\{ \max_{1 \leq i \leq n} \|Z_i - v\| \right\} = O_P(n^{-1/2}) o_P(n^{1/2}) = o_P(1).$$

Since

$$\frac{\hat{p}_i^E}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v)} = \hat{p}_i^E - \hat{p}_i^E (\hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v)) + \frac{\hat{p}_i^E (\hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v))^2}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v)},$$

multiplying both sides by $(Z_i - v)$ and summing over $1 \leq i \leq n$, by (13), yields:

$$\hat{\mu}_Z^E - v = \tilde{V}_Z^{\text{wE}}(\mu) \hat{\lambda}^{\text{wE}}(\mu) - \sum_{i=1}^n \hat{p}_i^E \frac{(Z_i - v) (\hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v))^2}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v)}.$$

Hence

$$V_Z \hat{\lambda}^{\text{wE}}(\mu) - (\hat{\mu}_Z^E - v) = (V_Z - \tilde{V}_Z^{\text{wE}}(\mu)) \hat{\lambda}^{\text{wE}}(\mu) + \sum_{i=1}^n \hat{p}_i^E \frac{(Z_i - v) (\hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v))^2}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v)}.$$

Since $\max_{\mu \in \mathbb{K}_n} \|(\tilde{V}_Z^{\text{wE}}(\mu) - V_Z) \hat{\lambda}^{\text{wE}}(\mu)\| = o_P(n^{-1/2})$ and

$$\begin{aligned} \max_{\mu \in \mathbb{K}_n} \left\| \sum_{i=1}^n \hat{p}_i^E \frac{(Z_i - v) (\hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v))^2}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v)} \right\| &\leq \left(\max_{\mu \in \mathbb{K}_n} \left\{ \max_{1 \leq i \leq n} \|Z_i - v\| \right\} \right) \\ &\times \left(\max_{\mu \in \mathbb{K}_n} \left\{ \hat{\lambda}^{\text{wE}}(\mu)^T \tilde{V}_Z^{\text{wE}}(\mu) \hat{\lambda}^{\text{wE}}(\mu) \right\} \right) = o_P(n^{1/2}) o_P(n^{-1}) = o_P(n^{-1/2}), \end{aligned}$$

then (14) holds. To establish (15), note first that by (11) and (12),

$$-2\hat{\ell}_{Y|X}^{\text{wE}}(\mu|\mu_X) = 2n \sum_{i=1}^n \hat{p}_i^E \log(1 + \hat{\lambda}^{\text{wE}}(\mu)^T (Z_i - v)).$$

One term Taylor expansion of $\log(x + 1)$ about x in a neighborhood of 0, yields:

$$\log(1 + \hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu)) = \hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu) - (1/2)(\hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu))^2 + o_P((\hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu))^2). \quad (71)$$

Multiplying both sides of (71) by \hat{p}_i^{E} and summing over $1 \leq i \leq n$, by (14), yields:

$$\begin{aligned} \sum_{i=1}^n \hat{p}_i^{\text{E}} \log(1 + \hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu)) &= (\hat{\mu}_Z^{\text{E}} - \nu)^T V_Z^{-1}(\hat{\mu}_Z^{\text{E}} - \nu) \\ &\quad - (1/2)(\hat{\mu}_Z^{\text{E}} - \nu)^T V_Z^{-1}(\hat{\mu}_Z^{\text{E}} - \nu) + o_P(n^{-1}) = (1/2)(\hat{\mu}_Y^{\text{E}} - \mu)^T \Sigma_Y^{-1}(\hat{\mu}_Y^{\text{E}} - \mu) + o_P(n^{-1}). \end{aligned}$$

Hence,

$$\max_{\mu \in \mathbb{K}_n} |-2\hat{\ell}_{Y|X}^{\text{wE}}(\mu|\mu_X) - n(\hat{\mu}_Y^{\text{E}} - \mu)^T \Sigma_Y^{-1}(\hat{\mu}_Y^{\text{E}} - \mu)| = o_P(1),$$

as stated. To establish (16), note first that by (12), we have:

$$\begin{aligned} \hat{p}_i^{\text{wE}}(\mu) &= \hat{p}_i^{\text{E}} \left(1 - \hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu) + \frac{(\hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu))^2}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu)} \right) \\ &= \hat{p}_i^{\text{E}} \left(1 - (\hat{\mu}_Z^{\text{E}} - \nu)^T V_Z^{-1}(Z_i - \nu) - (Z_i - \nu)^T (\hat{\lambda}^{\text{wE}}(\mu) - V_Z^{-1}(\hat{\mu}_Z^{\text{E}} - \nu)) + \frac{(\hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu))^2}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu)} \right). \end{aligned}$$

Hence,

$$\left| \hat{p}_i^{\text{wE}}(\mu) - \hat{p}_i^{\text{E}} \left(1 - (\hat{\mu}_Z^{\text{E}} - \nu)^T V_Z^{-1}(Z_i - \nu) \right) \right| = \hat{p}_i^{\text{E}} \left| \frac{(\hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu))^2}{1 + \hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu)} - (Z_i - \nu)^T (\hat{\lambda}^{\text{wE}}(\mu) - V_Z^{-1}(\hat{\mu}_Z^{\text{E}} - \nu)) \right|.$$

Since $\max_{1 \leq i \leq n} |\hat{p}_i^{\text{E}}| = O_P(n^{-1})$ (see Giurcanu and Presnell, 2018, Lemma C.3) and

$$\begin{aligned} \left(\max_{\mu \in \mathbb{K}_n} \left\{ \max_{1 \leq i \leq n} \|Z_i - \nu\| \right\} \right) \left(\max_{\mu \in \mathbb{K}_n} \|\hat{\lambda}^{\text{wE}}(\mu) - V_Z^{-1}(\hat{\mu}_Z^{\text{E}} - \nu)\| \right) &= o_P(1), \\ \max_{\mu \in \mathbb{K}_n} \left\{ \max_{1 \leq i \leq n} |\hat{\lambda}^{\text{wE}}(\mu)^T(Z_i - \nu)| \right\} &= o_P(1), \end{aligned}$$

then (16) holds. Lastly, by (15), (17) holds, as stated. This completes the proof. \square

Proof of (20). The Lagrangian is given by:

$$\mathcal{L}(q, \eta, \lambda) = -(1/2) \sum_{i=1}^n n \hat{q}_i (q_i / \hat{q}_i - 1)^2 + \eta \left(\sum_{i=1}^n q_i - 1 \right) - n \lambda^T \left(\sum_{i=1}^n q_i Z_i - \nu \right).$$

Setting the partial derivatives of the Lagrangian to zero, yields:

$$(a) - \frac{nq_i}{\hat{q}_i} + n + \eta - n\lambda^T Z_i = 0, \quad (b) \sum_{i=1}^n q_i = 1, \quad (c) \sum_{i=1}^n q_i Z_i = \nu. \quad (72)$$

Multiplying (a) by \hat{q}_i and summing over $1 \leq i \leq n$, yields $\eta - n\lambda^T \sum_{i=1}^n \hat{q}_i Z_i = 0$. Solving for η in terms of λ , yields $\hat{\eta}^{\text{w}}(\lambda) = n\lambda^T \hat{\mu}_Z^{\text{P}}$. Substituting $\hat{\eta}^{\text{w}}(\lambda)$ for η in (a), yields

$$\hat{q}_i^{\text{w}}(\lambda) = \hat{q}_i (1 - \lambda^T (Z_i - \hat{\mu}_Z^{\text{P}})). \quad (73)$$

Substituting $\hat{q}_i^{\text{w}}(\lambda)$ for q_i in (c), yields $\hat{\mu}_Z^{\text{P}} - \tilde{V}_Z \lambda = \nu$. Solving for λ , yields $\hat{\lambda}^{\text{w}}(\nu) = \tilde{V}_Z^{-1}(\hat{\mu}_Z^{\text{P}} - \nu)$. Since $\hat{\mu}_X = \mu_X$, substituting $\hat{\lambda}^{\text{w}}(\nu)$ for λ in (73), then (20) holds, as stated. \square

Proof of (21). By (19) and (20), we have

$$\begin{aligned} \hat{\ell}_{Y|X}^{\text{wP}}(\mu|\mu_X) &= -(1/2) \sum_{i=1}^n (n \hat{q}_i) ((\hat{\mu}_Y^{\text{P}} - \mu)^T \tilde{\Sigma}_Y^{-1} \tilde{V}_{YX} \tilde{V}_X^{-1} (X_i - \mu_X) - (\hat{\mu}_Y^{\text{P}} - \mu)^T \tilde{\Sigma}_Y^{-1} (Y_i - \hat{\mu}_Y^{\text{P}}))^2 \\ &= -(n/2) (\hat{\mu}_Y^{\text{P}} - \mu)^T \tilde{\Sigma}_Y^{-1} \tilde{V}_{YX} \tilde{V}_X^{-1} \tilde{V}_{XY} \tilde{\Sigma}_Y^{-1} (\hat{\mu}_Y^{\text{P}} - \mu) + n (\hat{\mu}_Y^{\text{P}} - \mu)^T \tilde{\Sigma}_Y^{-1} \tilde{V}_{YX} \tilde{V}_X^{-1} \tilde{V}_{XY} \tilde{\Sigma}_Y^{-1} (\hat{\mu}_Y^{\text{P}} - \mu) \\ &\quad - (n/2) (\hat{\mu}_Y^{\text{P}} - \mu)^T \tilde{\Sigma}_Y^{-1} \tilde{V}_Y \tilde{\Sigma}_Y^{-1} (\hat{\mu}_Y^{\text{P}} - \mu) = -(n/2) (\hat{\mu}_Y^{\text{P}} - \mu)^T \tilde{\Sigma}_Y^{-1} (\hat{\mu}_Y^{\text{P}} - \mu). \end{aligned}$$

\square

Proof of (25). The Lagrangian is given by:

$$\mathcal{L}(p_1, p_2, \eta_1, \eta_2, \lambda_1, \lambda_2) = \sum_{k=1}^2 \sum_{i=1}^{n_k} \log(n_k p_{k,i}) + \sum_{k=1}^2 \eta_k \left(\sum_{i=1}^{n_k} p_{k,i} - 1 \right) - \sum_{k=1}^2 n_k \lambda_k^T \left(\sum_{i=1}^{n_k} p_{k,i} X_{k,i} - \mu \right).$$

Setting the partial derivatives of the Lagrangian to zero, yields:

$$(a) p_{k,i}^{-1} + \eta_k - n_k \lambda_k^T X_{k,i} = 0, (b) \sum_{i=1}^{n_k} p_{k,i} = 1, (c) \sum_{i=1}^{n_k} p_{k,i} X_{k,i} = \mu.$$

Multiplying (a) by $p_{k,i}$ and summing over $1 \leq i \leq n_k$, by (c), yields $n_k + \eta_k - n_k \lambda_k^T \mu = 0$. Solving for η_k in terms of λ_k , yields $\hat{\eta}_k^E(\lambda_k) = n_k \lambda_k^T \mu - n_k$. Substituting $\hat{\eta}_k^E(\lambda_k)$ for η_k in (a), then (25) holds, as stated. \square

Proof of (29). The Lagrangian is given by:

$$\mathcal{L}(q_1, q_2, \eta_1, \eta_2, \lambda_1, \lambda_2) = -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k q_{k,i} - 1)^2 + \sum_{k=1}^2 \eta_k \left(\sum_{i=1}^{n_k} q_{k,i} - 1 \right) - \sum_{k=1}^2 n_k \lambda_k^T \left(\sum_{i=1}^{n_k} q_{k,i} X_{k,i} - \mu \right).$$

Setting the partial derivatives of the Lagrangian to zero, yields:

$$(a) -n_k^2 q_{k,i} + n_k + \eta_k - n_k \lambda_k^T X_{k,i} = 0, (b) \sum_{i=1}^{n_k} q_{k,i} = 1, (c) \sum_{i=1}^{n_k} q_{k,i} X_{k,i} = \mu.$$

Summing (a) over $1 \leq i \leq n_k$, yields $n_k \eta_k - n_k^2 \lambda_k^T \bar{X}_k = 0$. Solving for η_k in terms of λ_k , yields $\hat{\eta}_k(\lambda_k) = n_k \lambda_k^T \bar{X}_k$. Substituting $\hat{\eta}_k(\lambda_k)$ for η_k in (a), yields:

$$\hat{q}_{k,i}(\lambda_k) = n_k^{-1} (1 - \lambda_k^T (X_{k,i} - \bar{X}_k)). \quad (74)$$

Substituting $\hat{q}_{k,i}(\lambda_k)$ for $q_{k,i}$ in (c), yields:

$$\sum_{i=1}^{n_k} n_k^{-1} (X_{k,i} - X_{k,i} (X_{k,i} - \bar{X}_k)^T \lambda_k) = \mu.$$

Hence $\hat{V}_{X_k} \lambda_k = \bar{X}_k - \mu$, and thus, $\hat{\lambda}_k(\mu) = \hat{V}_{X_k}^{-1} (\bar{X}_k - \mu)$. Substituting $\hat{\lambda}_k(\mu)$ for λ_k in (74) shows that (29) holds, as stated. \square

Proof of (33). The Lagrangian is given by

$$\begin{aligned} \mathcal{L}(p_1, p_2, \eta_1, \eta_2, \lambda_1, \lambda_2) &= \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k \hat{p}_{k,i}^E) \log(p_{k,i} / \hat{p}_{k,i}^E) + \sum_{k=1}^2 \eta_k \left(\sum_{i=1}^{n_k} p_{k,i} - 1 \right) \\ &\quad - \sum_{k=1}^2 n_k \lambda_k^T \left(\sum_{i=1}^{n_k} p_{k,i} X_{k,i} - \mu \right). \end{aligned}$$

Setting the partial derivatives of the Lagrangian to 0, yields:

$$(a) \frac{n_k \hat{p}_{k,i}^E}{p_{k,i}} + \eta_k - n_k \lambda_k^T X_{k,i} = 0, (b) \sum_{i=1}^{n_k} p_{k,i} = 1, (c) \sum_{i=1}^{n_k} p_{k,i} X_{k,i} = \mu.$$

Multiplying (a) by $p_{k,i}$ and summing over $1 \leq i \leq n_k$, by (c) yields $n_k + \eta_k - n_k \lambda_k^T \mu = 0$. Solving for η_k , yields $\hat{\eta}_k^{wE}(\lambda_k) = n_k \lambda_k^T \mu - n_k$. Substituting $\hat{\eta}_k^{wE}(\lambda_k)$ for η_k in (a), then (33) holds, as stated. \square

Proof of Theorem 2. Consider the corresponding one-sample WEL problem (43). By (14), then (34) holds. By (15), then (35) holds. By (16) and (17), then (36) and (37) hold. This completes the proof of the theorem. \square

Proof of (40). The Lagrangian is given by:

$$\begin{aligned} \mathcal{L}(q_1, q_2, \eta_1, \eta_2, \lambda_1, \lambda_2) = & -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k \hat{q}_{k,i}) (q_{k,i} / \hat{q}_{k,i} - 1)^2 \\ & + \sum_{k=1}^2 \eta_k \left(\sum_{i=1}^{n_k} q_{k,i} - 1 \right) - \sum_{k=1}^2 n_k \lambda_k^T \left(\sum_{i=1}^{n_k} q_{k,i} X_{k,i} - \mu \right). \end{aligned}$$

Setting the partial derivatives of the Lagrangian to zero, yields:

$$(a) - \frac{n_k q_{k,i}}{\hat{q}_{k,i}} + n_k + \eta_k - n_k \lambda_k^T X_{k,i} = 0, \quad (b) \sum_{i=1}^{n_k} q_{k,i} = 1, \quad (c) \sum_{i=1}^{n_k} q_{k,i} X_{k,i} = \mu.$$

Multiplying (a) by $\hat{q}_{k,i}$ and summing over $1 \leq i \leq n_k$, yields $\eta_k - n_k \lambda_k^T \hat{\mu}_X^P = 0$. Solving for η_k in terms of λ_k , yields $\hat{\eta}_k^w(\lambda_k) = n_k \lambda_k^T \hat{\mu}_X^P$. Substituting $\hat{\eta}_k^w(\lambda_k)$ for η_k in (a), yields:

$$\hat{q}_{k,i}^w(\lambda_k) = \hat{q}_{k,i} (1 - \lambda_k^T (X_{k,i} - \hat{\mu}_X^P)). \quad (75)$$

Substituting $\hat{q}_{k,i}^w(\lambda_k)$ from (75) for $q_{k,i}$ in (c), yields $\hat{\mu}_X^P - \tilde{V}_{X_k} \lambda_k = \mu$. Solving for λ_k in terms of μ , yields $\hat{\lambda}_k^w(\mu) = \tilde{V}_{X_k}^{-1} (\hat{\mu}_X^P - \mu)$. Substituting $\hat{\lambda}_k^w(\mu)$ for λ_k in (75), then (40) holds, as stated. \square

Proof of (46). The Lagrangian corresponding to (45) is given by:

$$\mathcal{L}(p_1, p_2, \eta_1, \eta_2, \lambda) = \sum_{k=1}^2 \sum_{i=1}^{n_k} \log(n_k p_{k,i}) + \sum_{k=1}^2 \eta_k \left(\sum_{i=1}^{n_k} p_{k,i} - 1 \right) - n \lambda^T \left(\sum_{i=1}^{n_1} p_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} p_{2,i} Z_{2,i} - v \right).$$

Setting the partial derivatives of the Lagrangian to zero, yields:

$$(a) (1/p_{k,i}) + \eta_k + (-1)^k n \lambda^T Z_{k,i} = 0, \quad (b) \sum_{i=1}^{n_k} p_{k,i} = 1, \quad (c) \sum_{i=1}^{n_1} p_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} p_{2,i} Z_{2,i} = v. \quad (76)$$

By (a) of (76), it follows that (46) holds, as stated. \square

Proof of (49). The Lagrangian corresponding to (48) is given by:

$$\begin{aligned} \mathcal{L}(q_1, q_2, \eta_1, \eta_2, \lambda) = & -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k q_{k,i} - 1)^2 + \sum_{k=1}^2 \eta_k \left(\sum_{i=1}^{n_k} q_{k,i} - 1 \right) \\ & - n \lambda^T \left(\sum_{i=1}^{n_1} q_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} q_{2,i} Z_{2,i} - v \right). \end{aligned}$$

Setting the partial derivatives of the Lagrangian to zero, yields:

$$(a) -n_k^2 q_{k,i} + n_k + \eta_k + (-1)^k n \lambda^T Z_{k,i} = 0, \quad (b) \sum_{i=1}^{n_k} q_{k,i} = 1, \quad (c) \sum_{i=1}^{n_1} q_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} q_{2,i} Z_{2,i} = v,$$

where $v = (0, \Delta) \in \mathbb{R}^{p+q}$. Summing (a) of over $1 \leq i \leq n_k$, yields $n_k \eta_k + (-1)^k n n_k \lambda^T \bar{Z}_k = 0$. Solving for η_k in terms of λ , yields $\hat{\eta}_k(\lambda) = (-1)^{k+1} n \lambda^T \bar{Z}_k$. Substituting $\hat{\eta}_k(\lambda)$ for η_k in (a), we obtain the solution for $q_{k,i}$ in terms of λ :

$$\hat{q}_{k,i}(\lambda) = n_k^{-1} (1 + (-1)^k (n/n_k) \lambda^T (Z_{k,i} - \bar{Z}_k)). \quad (77)$$

Substituting $\hat{q}_{k,i}(\lambda)$ for $q_{k,i}$ in (c), yields $\hat{V}_Z \lambda = \bar{\Delta}_Z - v$. Solving for λ in terms of Δ , yields $\hat{\lambda}(\Delta) = \hat{V}_Z^{-1} (\bar{\Delta}_Z - v)$. Substituting $\hat{\lambda}(\Delta)$ for λ in (77), yields:

$$\hat{q}_{k,i}(\Delta) = n_k^{-1} (1 + (-1)^k (n/n_k) (\bar{\Delta}_Z - v)^T \hat{V}_Z^{-1} (Z_{k,i} - \bar{Z}_k)).$$

as stated at (49). \square

Proof of (50). Substituting $\hat{q}_{k,i}(\Delta)$ for $q_{k,i}$ in (48), yields:

$$-2\hat{\phi}_{Z_1, Z_2}^P(\Delta) = \sum_{k=1}^2 \sum_{i=1}^{n_k} ((-1)^k n/n_k (\bar{\Delta}_Z - \nu)^T \hat{V}_Z^{-1} (Z_{k,i} - \bar{Z}_k))^2 = n(\bar{\Delta}_Z - \nu)^T \hat{V}_Z^{-1} (\bar{\Delta}_Z - \nu).$$

Hence, (50) holds, as stated. To prove (51), solving $\nabla \hat{\phi}_{Z_1, Z_2}^P(\Delta) = 0$ for Δ , yields: $\hat{\Delta}_Y^P = \bar{\Delta}_Y - \hat{V}_{YX} \hat{V}_X^{-1} \bar{\Delta}_X$. Similarly to the proof of (10), yields:

$$-2\hat{\phi}_{Z_1, Z_2}^P(\Delta) + 2\hat{\phi}_{Z_1, Z_2}^P(\hat{\Delta}_Y^P) = n(\hat{\Delta}_Y^P - \Delta)^T \hat{\Sigma}_Y^{-1} (\hat{\Delta}_Y^P - \Delta).$$

□

Proof of (53). The Lagrangian is given by:

$$\begin{aligned} \mathcal{L}(p_1, p_2, \eta_1, \eta_2, \lambda) &= \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k \hat{p}_{k,i}^E) \log(p_{k,i} / \hat{p}_{k,i}^E) + \sum_{k=1}^2 \eta_k \left(\sum_{i=1}^{n_k} p_{k,i} - 1 \right) \\ &\quad - n\lambda^T \left(\sum_{i=1}^{n_1} p_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} p_{2,i} Z_{2,i} - \nu \right). \end{aligned}$$

Setting the partial derivatives of the Lagrangian to zero, yields:

$$(a) \frac{n_k \hat{p}_{k,i}^E}{p_{k,i}} + \eta_k + (-1)^k n\lambda^T Z_{k,i} = 0, \quad (b) \sum_{i=1}^{n_k} p_{k,i} = 1, \quad (c) \sum_{i=1}^{n_1} p_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} p_{2,i} Z_{2,i} = \nu. \quad (78)$$

Solving (a) for $p_{k,i}$ in terms of λ , then (53) holds, as stated. □

Proof of (60). The Lagrangian corresponding to (59) is given by:

$$\begin{aligned} \mathcal{L}(q_1, q_2, \eta_1, \eta_2, \lambda) &= -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} (n_k \hat{q}_{k,i}) (q_{k,i} / \hat{q}_{k,i} - 1)^2 + \sum_{k=1}^2 \eta_k \left(\sum_{i=1}^{n_k} q_{k,i} - 1 \right) \\ &\quad - n\lambda^T \left(\sum_{i=1}^{n_1} q_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} q_{2,i} Z_{2,i} - \nu \right). \end{aligned}$$

Setting the partial derivatives of the Lagrangian to zero, yields:

$$(a) -\frac{n_k q_{k,i}}{\hat{q}_{k,i}} + n_k + \eta_k + (-1)^k n\lambda^T Z_{k,i} = 0, \quad (b) \sum_{i=1}^{n_k} q_{k,i} = 1, \quad (c) \sum_{i=1}^{n_1} q_{1,i} Z_{1,i} - \sum_{i=1}^{n_2} q_{2,i} Z_{2,i} = \nu.$$

Multiplying (a) by $\hat{q}_{k,i}$ and summing over i , yields $\eta_k + (-1)^k n\lambda^T \tilde{Z}_k = \nu$, where $\tilde{Z}_k = \sum_{i=1}^{n_k} \hat{q}_{k,i} Z_{k,i}$. Solving for η_k in terms of λ , yields $\hat{\eta}_k^w(\lambda) = (-1)^{k+1} n\lambda^T \tilde{Z}_k$. Substituting $\hat{\eta}_k^w(\lambda)$ for η_k in (a), yields:

$$\hat{q}_{k,i}^w(\lambda) = \hat{q}_{k,i} (1 + (-1)^k (n/n_k) \lambda^T (Z_{k,i} - \tilde{Z}_k)). \quad (79)$$

Substituting $\hat{q}_{k,i}^w(\lambda)$ for $q_{k,i}$ in (c), yields $\hat{\Delta}_Z^P - \tilde{V}_Z \lambda = \nu$, where $\hat{\Delta}_Z^P = \tilde{Z}_1 - \tilde{Z}_2$. Solving for λ , yields $\hat{\lambda}^w(\Delta) = \tilde{V}_Z^{-1} (\hat{\Delta}_Z^P - \nu)$. Substituting $\hat{\lambda}^w(\Delta)$ for λ in (79), yields that (60) holds, as stated. □

Proof of (61). Substituting $\hat{q}_{k,i}^w(\Delta)$ from (60) for $q_{k,i}$ in (59), using the formula for matrix inverse in block form, yields:

$$\begin{aligned} \hat{\phi}_{Z_1, Z_2}^{WP}(\Delta) &= -(1/2) \sum_{k=1}^2 \sum_{i=1}^{n_k} n_k \hat{q}_{k,i} ((n/n_k) (\hat{\Delta}_Y^P - \Delta)^T \tilde{\Sigma}_Y^{-1} \tilde{V}_{YX} \tilde{V}_X^{-1} (X_{k,i} - \tilde{X}_k) \\ &\quad - (n/n_k) (\hat{\Delta}_Y^P - \Delta)^T \tilde{\Sigma}_Y^{-1} (Y_{k,i} - \tilde{Y}_k))^2 = -(n/2) (\hat{\Delta}_Y^P - \Delta)^T \tilde{\Sigma}_Y^{-1} \tilde{V}_Y \tilde{\Sigma}_Y^{-1} (\hat{\Delta}_Y^P - \Delta) \\ &\quad + n (\hat{\Delta}_Y^P - \Delta)^T \tilde{\Sigma}_Y^{-1} \tilde{V}_{YX} \tilde{V}_X^{-1} \tilde{V}_{XY} \tilde{\Sigma}_Y^{-1} (\hat{\Delta}_Y^P - \Delta) \\ &\quad - (n/2) (\hat{\Delta}_Y^P - \Delta)^T \tilde{\Sigma}_Y^{-1} \tilde{V}_{YX} \tilde{V}_X^{-1} \tilde{V}_{XY} \tilde{\Sigma}_Y^{-1} (\hat{\Delta}_Y^P - \Delta) = (n/2) (\hat{\Delta}_Y^P - \Delta)^T \tilde{\Sigma}_Y^{-1} (\hat{\Delta}_Y^P - \Delta), \end{aligned} \quad (80)$$

and thus (61) holds, as stated. □

Proof of Theorem 3. Consider the corresponding one-sample WEL problem (63). By (14), it follows that (54) holds. By (15), then (55) holds. By (16) and (17), then (56) and (57) hold, as stated. This completes the proof of the theorem. \square

Proof of (65). Let $(\hat{\mu}_1, \hat{\mu}_2, \hat{\beta}_1, \hat{\beta}_2) = \operatorname{argmin}_{\mu_k, \beta_k} \mathcal{L}(\mu_1, \mu_2, \beta_1, \beta_2)$ be the least squares estimator of $(\mu_1, \mu_2, \beta_1, \beta_2)$, where

$$\mathcal{L}(\mu_1, \mu_2, \beta_1, \beta_2) = \sum_{k=1}^2 \sum_{i=1}^{n_k} (Y_{k,i} - \mu_k - (I_p \otimes \hat{X}_{k,i}^T) \beta_k)^T (Y_{k,i} - \mu_k - (I_p \otimes \hat{X}_{k,i}^T) \beta_k).$$

Setting the partial derivatives of $\mathcal{L}(\mu_1, \mu_2, \beta_1, \beta_2)$ to zero, yields:

$$(a) \sum_{i=1}^{n_k} (Y_{k,i} - \hat{\mu}_k - (I_p \otimes \hat{X}_{k,i}^T) \hat{\beta}_k) = 0, \quad (b) \sum_{i=1}^{n_k} (I_p \otimes \hat{X}_{k,i}) (Y_{k,i} - \hat{\mu}_k - (I_p \otimes \hat{X}_{k,i}^T) \hat{\beta}_k) = 0.$$

From (a), we obtain $\hat{\mu}_k = \bar{Y}_k - (I_p \otimes \hat{X}_k^T) \hat{\beta}_k$ where $\hat{X}_k = n_k^{-1} \sum_{i=1}^{n_k} \hat{X}_{k,i} = \bar{X}_k - \bar{X}$. Substituting $\hat{\mu}_k$ for μ_k in (b), yields:

$$\sum_{i=1}^{n_k} (I_p \otimes \hat{X}_{k,i}) (\hat{Y}_{k,i} - (I_p \otimes \hat{X}_{k,i}^T) \hat{\beta}_k) = 0,$$

where $\hat{X}_{k,i} = \hat{X}_{k,i} - \hat{X}_k = X_{k,i} - \bar{X}_k$ and $\hat{Y}_{k,i} = Y_{k,i} - \bar{Y}_k$. Using the identity $\operatorname{vec}(AB) = (I_m \otimes A) \operatorname{vec}(B)$, where $A \in \mathbb{R}^{k \times l}$ and $B \in \mathbb{R}^{l \times m}$ are arbitrary matrices, we obtain:

$$\begin{aligned} \hat{\beta}_k &= \left(n_k^{-1} \sum_{i=1}^{n_k} I_p \otimes (\hat{X}_{k,i} \hat{X}_{k,i}^T) \right)^{-1} n_k^{-1} \sum_{i=1}^{n_k} (I_p \otimes \hat{X}_{k,i}) \hat{Y}_{k,i} \\ &= (I_p \otimes \hat{V}_{X_k}^{-1}) n_k^{-1} \sum_{i=1}^{n_k} (I_p \otimes \hat{X}_{k,i}) \hat{Y}_{k,i} \\ &= n_k^{-1} \sum_{i=1}^{n_k} (I_p \otimes (\hat{V}_{X_k}^{-1} \hat{X}_{k,i})) \hat{Y}_{k,i} = n_k^{-1} \sum_{i=1}^{n_k} (I_p \otimes (\hat{V}_{X_k}^{-1} \hat{X}_{k,i})) \operatorname{vec}(\hat{Y}_{k,i}^T) \\ &= n_k^{-1} \sum_{i=1}^{n_k} \operatorname{vec}(\hat{V}_{X_k}^{-1} \hat{X}_{k,i} \hat{Y}_{k,i}^T) = \operatorname{vec}(\hat{V}_{X_k}^{-1} \hat{V}_{X_k Y_k}). \end{aligned}$$

\square