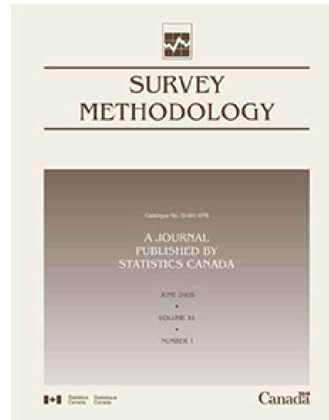
Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

One-sided testing of population domain means in surveys



by Xiaoming Xu and Mary C. Meyer

Release date: June 30, 2023

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2023
Vol. 49, No. 1, pp. 117-138
Statistics Canada, Catalogue No. 12-001-X

One-sided testing of population domain means in surveys

Xiaoming Xu and Mary C. Meyer[1]

Abstract

Recent work in survey domain estimation allows for estimation of population domain means under *a priori* assumptions expressed in terms of linear inequality constraints. For example, it might be known that the population means are non-decreasing along ordered domains. Imposing the constraints has been shown to provide estimators with smaller variance and tighter confidence intervals. In this paper we consider a formal test of the null hypothesis that all the constraints are binding, versus the alternative that at least one constraint is non-binding. The test of constant versus increasing domain means is a special case. The power of the test is substantially better than the test with the same null hypothesis and an unconstrained alternative. The new test is used with data from the National Survey of College Graduates, to show that salaries are positively related to the subject's father's educational level, across fields of study and over several years of cohorts.

Key Words: Survey domain; Order constraints; Monotone; Block monotone.

1. Introduction

Methods for estimation of population domain means under *a priori* assumptions in the form of linear inequality constraints have been recently established. Suppose interest is in estimating $\overline{\mathbf{y}}_U \in \mathbb{R}^D$, a vector of population domain means, where D is the number of domains. Wu, Meyer and Opsomer (2016) derived an isotonic survey estimator of $\overline{\mathbf{y}}_U$, where it is assumed that $\bar{y}_{U_1} \leq \ldots \leq \bar{y}_{U_D}$. They showed that the constrained estimator is equivalent to a "pooled" estimator, where weighted averages of adjacent sample domain means are used to form an isotonic vector of domain mean estimates. Advantages to the ordered mean estimates are that they "make sense" in terms of satisfying the assumptions, and the confidence intervals for the estimates are typically reduced in length. Oliva-Aviles, Meyer and Opsomer (2019) proposed an information criterion to check the validity of the monotone assumption; that is, determining whether the domain means are ordered or unordered.

Oliva-Aviles, Meyer and Opsomer (2020) proposed a framework for estimation and inference with more general shape and order constraints in survey contexts. Examples include block orderings, and orderings of domain means arranged in grids. For example, average cholesterol level may be assumed to be increasing in age category and body mass index (BMI) level, but decreasing in exercise category. In another context, suppose average salary is to be estimated by job rank, job type, and location, with average salary assumed to be increasing with rank, and block orderings imposed on job type and location. More recently, Xu, Meyer and Opsomer (2021) formulated a mixture covariance matrix for constrained estimation that was shown to improve coverage of confidence intervals while retaining the smaller lengths.

1. Xiaoming Xu, Department of Biostatistics and Bioinformatics, Duke University, Durham, NC 27710, U.S.A. E-mail: xiaoming.xu197@duke.edu; Mary C. Meyer, Department of Statistics, Colorado State University, Fort Collins, CO 80523, U.S.A. E-mail: mary.meyer@colostate.edu.

The desired linear inequality constraints may be formulated using an $M \times D$ constraint matrix $\mathbf{A}$, where the assumption is $\mathbf{A}\bar{\mathbf{y}}_U \geq \mathbf{0}$. For the isotonic domain means, $M = D - 1$, and the nonzero elements of the constraint matrix are $\{\mathbf{A}\}_{m,m} = -1$ and $\{\mathbf{A}\}_{m,m+1} = 1$. For block orderings, where domains are grouped by ordered blocks, each domain in block one, for example, is assumed to have a population mean not larger than each domain in block two, and in block two, each population domain mean does not exceed any of those in block three, etc. Here the number of constraints is $M = \sum_{b=1}^{B-1} \sum_{b'=b+1}^{B} D_b D_{b'}$, where B is the number of blocks and D_b is the number of domains in the b^{th} block, $b = 1, \ldots, B$. For example, suppose interest is in mean salaries at an institution, where the domains are four "fields", and it is assumed that fields 3 and 4 have higher salaries than fields 1 and 2. In this case $B = 2$, $D_1 = D_2 = 2$, and the constraint matrix is

$$\mathbf{A} = \begin{pmatrix} -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}.$$

For a third example, consider domains arranged in a grid; for a context suppose the population units are lakes in a state, and y_i is the level of a certain pollutant in lake i. We are interested in average levels by county and by distance from an industrial plant. If there are 60 counties and 5 categories of distance, there are 300 domains. If we know that the level of pollutant is non-increasing in the distance variable, then there are $60 \times 4 = 240$ constraints, formulated as antitonic within each county.

We propose a test where the null hypothesis is that $\mathbf{A}\bar{\mathbf{y}}_U = \mathbf{0}$, versus the alternative $\mathbf{A}\bar{\mathbf{y}}_U \geq \mathbf{0}$, and $\mathbf{A}\bar{\mathbf{y}}_U$ has at least one positive element. The simplest example is the null hypothesis of constant domain means, versus the alternative of increasing domain means. (Note that these hypotheses are different from the alternatives in Oliva-Aviles, Meyer and Opsomer (2019), who were deciding between monotone and non-monotone domain means.) For the industrial plant example above, we can test the null hypothesis that, within each county, the domain means are constant in distance. Using the constraints for a one-sided alternative results in improved power over the equivalent two-sided test.

This test has been widely studied outside of the survey context; see Bartholomew (1959); Bartholomew (1961); Chacko (1963); McDermott and Mudholkar (1993); Robertson, Wright and Dykstra (1988); Meyer (2003); Silvapulle and Sen (2005); Sen and Meyer (2017) and others. The null distribution of the likelihood-ratio test statistic for the one-sided test has been derived based on the normal-errors model. In brief, when the error terms are independently and identically distributed with known model variance, the null distribution of the likelihood ratio statistic is shown to be a mixture of chi-square distributions, while for the unknown model variance, the test statistic has the null distribution of a mixture of beta distributions. If the error terms are not independently and identically distributed, the results, based on principles of generalized least squares, still hold provided the covariance structure for the error terms is available. Similar results for the one-sided likelihood ratio test were obtained by Perlman (1969) where the completely unknown covariance matrix was considered. Meyer and Wang (2012) formally proved that the one-sided test will provide higher power than the test using the unconstrained alternative.

In this paper we extend the one-sided test to the survey context. In the Section 2, the test is derived, and in Section 3 some large sample theory is given. Simulations in Section 4 show that the test performs well compared to the test with the unconstrained alternative, with better power and a test size closer to the target. In Section 5 the methods are applied to the National Survey of College Graduates (NSCG), to test whether salaries are higher for people whose father's education level is higher, controlling for field of study, highest degree attained, and year of degree. The test is available in the `R` package `csurvey`.

2. Formulation of the test statistic

To establish the notation, let $U = \{1, 2, \ldots, N\}$ be the finite population. A sample $s \subset U$ of size n is to be drawn based on a probability sampling design p, where $p(s)$ is the probability of drawing the sample s. The first order inclusion probability $\pi_i = \Pr(i \in s) = \sum_{i \in s} p(s)$ and the second order inclusion probability $\pi_{ij} = \Pr(i, j \in s) = \sum_{i, j \in s} p(s)$, determined by the sampling design, are both assumed to be positive. The assumed positive π_i and π_{ij} ensure that the design-based estimator of the population parameter and the associated design-based variance estimator can be obtained, respectively. In terms of the domains of interest, let $\{U_d : d = 1, \ldots, D\}$ be a partition of the population U and N_d be the population size of domain d, where D is the number of domains. We denote by s_d the intersection of s and U_d, and let n_d be the sample size for s_d. Sample size n_d arises from a random sampling procedure and thus is not fixed in general.

Let y be the variable of interest and denote by y_i the value for the i^{th} unit in the population. The population domain means are $\bar{\mathbf{y}}_U = \left(\bar{y}_{U_1}, \ldots, \bar{y}_{U_D}\right)^\top$, and $\bar{y}_{U_d}$ is given by:

$$\bar{y}_{U_d} = \frac{\sum_{i \in U_d} y_i}{N_d} \quad d = 1, \ldots, D.$$

Two common design-based estimators of the population means are the Horvitz-Thompson (HT) estimator (Horvitz and Thompson, 1952) or the Hájek estimator (Hájek, 1971); because the Hájek estimator $\tilde{y}_{s_d}$ does not require information about the population domain size N_d and has other advantages in practice, we will focus on the Hájek estimator. The results for the Horvitz-Thompson estimator, however, can be derived analogously. The Hájek estimator for domain means is $\tilde{\mathbf{y}}_s = \left(\tilde{y}_{s_1}, \ldots, \tilde{y}_{s_D}\right)$, where

$$\tilde{y}_{s_d} = \frac{\sum_{i \in s_d} y_i / \pi_i}{\hat{N}_d}$$

and $\hat{N}_d = \sum_{i \in s_d} 1/\pi_i$.

We are concerned with testing:

$$H_0 : \bar{\mathbf{y}}_U \in V \quad \text{versus} \quad H_1 : \bar{\mathbf{y}}_U \in C \setminus V \tag{2.1}$$

where $V = \{\mathbf{y} : \mathbf{A}\mathbf{y} = \mathbf{0}\}$ is the null space of $\mathbf{A}$ and the alternative set is the convex cone $C = \{\mathbf{y} : \mathbf{A}\mathbf{y} \geq \mathbf{0}\}$ excluding the set V. A set C is a convex cone if for any $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in C, $\alpha_1\boldsymbol{\theta}_1 + \alpha_2\boldsymbol{\theta}_2$ is in C for any non-negative α_1 and α_2.

We start with a brief review of the properties of the unconstrained estimator $\tilde{\mathbf{y}}_s$. By the Taylor expansion, we can linearize the $\tilde{\mathbf{y}}_s$ as follows:

$$\tilde{\mathbf{y}}_s = \bar{\mathbf{y}}_U + \hat{\mathbf{y}}^{\text{center}} + O_p\left(n^{-1}\right)$$

where

$$\hat{\mathbf{y}}^{\text{center}} = \left(\frac{1}{N_1} \sum_{i \in s_1} \frac{\left(y_i - \bar{y}_{U_1}\right)}{\pi_i}, \ldots, \frac{1}{N_D} \sum_{i \in s_D} \frac{\left(y_i - \bar{y}_{U_D}\right)}{\pi_i} \right)^{\top}.$$

The properties of $\tilde{\mathbf{y}}_s - \bar{\mathbf{y}}_U$ can be approximated by $\hat{\mathbf{y}}^{\text{center}}$ and we have that $\mathrm{E}\left(\hat{\mathbf{y}}^{\text{center}}\right) = 0$ and the variance of $\hat{\mathbf{y}}^{\text{center}}$ is $\boldsymbol{\Sigma}$, where the dd'^{th} element of $\boldsymbol{\Sigma}$ is:

$$\{\boldsymbol{\Sigma}\}_{dd'} = \frac{1}{N_d N_{d'}} \sum_{i \in U_d} \sum_{j \in U_{d'}} \Delta_{ij} \frac{\left(y_i - \bar{y}_{U_d}\right)\left(y_j - \bar{y}_{U_{d'}}\right)}{\pi_i \pi_j}, \quad d, d' = 1, 2, \ldots, D$$

where $\Delta_{ij} = \operatorname{cov}\left(I_i, I_j\right) = \pi_{ij} - \pi_i \pi_j$ and I_i is the indicator variable of whether unit i is selected by sampling design. By the design normal assumption (A5) in the appendix, we have $\boldsymbol{\Sigma}^{-1/2}\hat{\mathbf{y}}^{\text{center}} \overset{\mathcal{D}}{\rightarrow} N(\mathbf{0}, \mathbf{I})$, hence:

$$\boldsymbol{\Sigma}^{-1/2}\left(\tilde{\mathbf{y}}_s - \bar{\mathbf{y}}_U\right) = \boldsymbol{\Sigma}^{-1/2}\hat{\mathbf{y}}^{\text{center}} + o_p(1) \overset{\mathcal{D}}{\rightarrow} N(\mathbf{0}, \mathbf{I}).$$

We denote by $\hat{\boldsymbol{\Sigma}}$ a consistent estimator of $\boldsymbol{\Sigma}$, in the sense that $n\left(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right) = o_p(1)$. For testing (2.1), we propose the following weighted least squares test statistic:

$$\hat{T} = \frac{\min_{\boldsymbol{\theta}_0 \in V} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^{\top} \hat{\boldsymbol{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right) - \min_{\boldsymbol{\theta}_1 \in C} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^{\top} \hat{\boldsymbol{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)}{\min_{\boldsymbol{\theta}_0 \in V} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^{\top} \hat{\boldsymbol{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)}.$$

Assuming the second order inclusion probability π_{ij} to be known, the dd'^{th} element of the design based consistent estimator $\hat{\boldsymbol{\Sigma}}$ has the following expression:

$$\{\hat{\boldsymbol{\Sigma}}\}_{dd'} = \frac{1}{\hat{N}_d \hat{N}_{d'}} \sum_{i \in s_d} \sum_{j \in s_{d'}} \frac{\Delta_{ij}}{\pi_{ij}} \frac{\left(y_i - \tilde{y}_{s_d}\right)\left(y_j - \tilde{y}_{s_{d'}}\right)}{\pi_i \pi_j}, \quad d, d' = 1, 2, \ldots, D. \tag{2.2}$$

See Särndal, Swensson and Wretman (1992) Chapter 5 on page 185 for more details. Particularly, under a fixed size design, the Sen-Yates-Grundy variance estimator, derived as an alternative form of (2.2), can also be used. In addition, under many complex survey designs, the second order inclusion probability π_{ij} might be zero or unknown so that the design based covariance estimator $\hat{\mathbf{\Sigma}}$ cannot be obtained. In such cases, the use of consistent replication-based variance estimators (such as Jackknife estimator, bootstrap estimator) can be considered, since the calculation of replication variance estimator does not involve the second order inclusion probabilities. As long as the replication-based variance estimators are good approximation for $\mathbf{\Sigma}$, the asymptotic properties of $\hat{T}$, which will be developed shortly, would hold.

We will reject H_0 if $\hat{T}$ is large. This is similar in structure to the classical test (as was presented in, for example, Silvapulle and Sen (2005) Chapter 3). If $\tilde{\mathbf{y}}_s$ were normal with $\operatorname{cov}(\tilde{\mathbf{y}}_s) = \hat{\mathbf{\Sigma}}$, then $\hat{T}$ would be distributed as a mixture of beta random variables, under the null hypothesis. In the survey context, we approximate the distribution of $\hat{T}$.

3. Asymptotic distribution of the test statistic

The assumptions needed to derive an approximate distribution of $\hat{T}$ are listed in Appendix, and are similar to those in Xu et al. (2021).

To derive the asymptotic null distribution of $\hat{T}$, we first show the following result.

Lemma 1. *The test statistic* $\hat{T}$ *can be written as:*

$$\begin{aligned}\hat{T} &= \frac{\min_{\boldsymbol{\theta}_0 \in V} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top \hat{\mathbf{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right) - \min_{\boldsymbol{\theta}_1 \in C} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^\top \hat{\mathbf{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)}{\min_{\boldsymbol{\theta}_0 \in V} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top \hat{\mathbf{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)} \\ &= \frac{\min_{\boldsymbol{\theta}_0 \in V} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top \mathbf{\Sigma}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right) - \min_{\boldsymbol{\theta}_1 \in C} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^\top \mathbf{\Sigma}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)}{\min_{\boldsymbol{\theta}_0 \in V} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top \mathbf{\Sigma}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)} + o_p(1).\end{aligned}$$

Proof. Let $\hat{\mathbf{A}} = \mathbf{A}\hat{\mathbf{\Sigma}}^{1/2}$, $\hat{\mathbf{Z}}_s = \hat{\mathbf{\Sigma}}^{-1/2}\tilde{\mathbf{y}}_s$, $\hat{\boldsymbol{\theta}}_0 = \hat{\mathbf{\Sigma}}^{-1/2}\boldsymbol{\theta}_0$, $\hat{\boldsymbol{\theta}}_1 = \hat{\mathbf{\Sigma}}^{-1/2}\boldsymbol{\theta}_1$, $\hat{\mathbf{V}} = \left\{\hat{\boldsymbol{\theta}}_0 : \hat{\mathbf{A}}\hat{\boldsymbol{\theta}}_0 = 0\right\}$ and $\hat{C} = \left\{\hat{\boldsymbol{\theta}}_1 : \hat{\mathbf{A}}\hat{\boldsymbol{\theta}}_1 \geq 0\right\}$. Then by a transformation, we have:

$$\begin{aligned}\hat{T} &= \frac{\min_{\boldsymbol{\theta}_0 \in V} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top \hat{\mathbf{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right) - \min_{\boldsymbol{\theta}_1 \in C} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^\top \hat{\mathbf{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)}{\min_{\boldsymbol{\theta}_0 \in V} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top \hat{\mathbf{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)} \\ &= \frac{\min_{\hat{\boldsymbol{\theta}}_0 \in \hat{V}} \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0\right)^\top \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0\right) - \min_{\hat{\boldsymbol{\theta}}_1 \in \hat{C}} \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_1\right)^\top \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_1\right)}{\min_{\hat{\boldsymbol{\theta}}_0 \in \hat{V}} \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0\right)^\top \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0\right)} \\ &= 1 - \frac{\min_{\hat{\boldsymbol{\theta}}_1 \in \hat{C}} \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_1\right)^\top \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_1\right)}{\min_{\hat{\boldsymbol{\theta}}_0 \in \hat{V}} \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0\right)^\top \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0\right)}.\end{aligned}$$

Let $\hat{V}^\perp$ be the linear space of vectors in $\mathbb{R}^D$ that are orthogonal to vectors in $\hat{V}$. Note that $\min_{\hat{\boldsymbol{\theta}}_0 \in \hat{V}} \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0\right)^\top \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0\right)$ is the squared length of the projection of $\hat{\mathbf{Z}}_s$ onto $\hat{V}^\perp$ and the projection of

$\hat{\mathbf{Z}}_s$ onto $\hat{V}$ has the explicit expression $\hat{\boldsymbol{\theta}}_0^* = \left(\mathbf{I} - \hat{\mathbf{A}}^\top \left(\hat{\mathbf{A}}\hat{\mathbf{A}}^\top\right)^- \hat{\mathbf{A}}\right)\hat{\mathbf{Z}}_s,$ where $\left(\hat{\mathbf{A}}\hat{\mathbf{A}}^\top\right)^-$ is the generalized inverse of $\hat{\mathbf{A}}\hat{\mathbf{A}}^\top$. Hence, by the consistency of $\hat{\boldsymbol{\Sigma}}$, we have the following:

$$\begin{aligned}
\min_{\hat{\boldsymbol{\theta}}_0 \in \hat{V}} \frac{1}{n}\left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0\right)^\top \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0\right) &= \frac{1}{n}\left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0^*\right)^\top \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0^*\right) \\
&= \frac{1}{n}\left(\hat{\mathbf{A}}^\top \left(\hat{\mathbf{A}}\hat{\mathbf{A}}^\top\right)^- \hat{\mathbf{A}}\hat{\mathbf{Z}}_s\right)^\top \hat{\mathbf{A}}^\top \left(\hat{\mathbf{A}}\hat{\mathbf{A}}^\top\right)^- \hat{\mathbf{A}}\hat{\mathbf{Z}}_s \\
&= \frac{1}{n}\hat{\mathbf{Z}}_s^\top \hat{\mathbf{A}}^\top \left(\hat{\mathbf{A}}\hat{\mathbf{A}}^\top\right)^- \hat{\mathbf{A}}\hat{\mathbf{Z}}_s \\
&= \frac{1}{n}\tilde{\mathbf{y}}_s^\top \mathbf{A}^\top \left(\mathbf{A}\hat{\boldsymbol{\Sigma}}\mathbf{A}^\top\right)^- \mathbf{A}\tilde{\mathbf{y}}_s \\
&= \tilde{\mathbf{y}}_s^\top \mathbf{A}^\top \left(\mathbf{A}n\boldsymbol{\Sigma}\mathbf{A}^\top\right)^- \mathbf{A}\tilde{\mathbf{y}}_s + o_p(1) \\
&= \min_{\boldsymbol{\theta}_0 \in V} \frac{1}{n}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top \boldsymbol{\Sigma}^{-1}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right) + o_p(1).
\end{aligned} \tag{3.1}$$

By (3.1) and the result that $\min_{\boldsymbol{\theta}_1 \in C} \frac{1}{n}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^\top \hat{\boldsymbol{\Sigma}}^{-1}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right) - \min_{\boldsymbol{\theta}_1 \in C} \frac{1}{n}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^\top \boldsymbol{\Sigma}^{-1}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right) = o_p(1)$ by Lemma 4 in the Appendix, we get

$$\begin{aligned}
\hat{T} &= 1 - \frac{\min_{\boldsymbol{\theta}_1 \in C} \frac{1}{n}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^\top \hat{\boldsymbol{\Sigma}}^{-1}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)}{\min_{\boldsymbol{\theta}_0 \in V} \frac{1}{n}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top \hat{\boldsymbol{\Sigma}}^{-1}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)} \\
&= 1 - \frac{\min_{\boldsymbol{\theta}_1 \in C} \frac{1}{n}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^\top \boldsymbol{\Sigma}^{-1}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)}{\min_{\boldsymbol{\theta}_0 \in V} \frac{1}{n}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top \boldsymbol{\Sigma}^{-1}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)} + o_p(1)
\end{aligned}$$

the proof is complete.

The denominator in above expression must be bounded away from zero in probability, which is indeed the case because it can be shown that the $\min_{\boldsymbol{\theta}_0 \in V}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top (n\boldsymbol{\Sigma})^{-1}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)$ has, asymptotically, $\chi^2(M)$ distribution under the null and design normal assumption.

Next, let $\tilde{\mathbf{Z}}_s = \boldsymbol{\Sigma}^{-1/2}\tilde{\mathbf{y}}_s, \mathbf{Z}_U = \boldsymbol{\Sigma}^{-1/2}\bar{\mathbf{y}}_U, \tilde{\boldsymbol{\theta}}_0 = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_1 = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}_1$ and define $\tilde{V} = \left\{\tilde{\boldsymbol{\theta}} : \tilde{\mathbf{A}}\tilde{\boldsymbol{\theta}} = 0\right\}, \tilde{C} = \left\{\tilde{\boldsymbol{\theta}} : \tilde{\mathbf{A}}\tilde{\boldsymbol{\theta}} \geq 0\right\},$ where $\tilde{\mathbf{A}} = \mathbf{A}\boldsymbol{\Sigma}^{1/2}$. Then, we have the following main result of the paper.

Theorem 1. *Define*

$$T = \frac{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} \left\|\mathbf{Z} - \tilde{\boldsymbol{\theta}}_0\right\|^2 \min_{\tilde{\boldsymbol{\theta}}_1 \in \tilde{C}} \left\|\mathbf{Z} - \tilde{\boldsymbol{\theta}}_1\right\|^2}{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} \left\|\mathbf{Z} - \tilde{\boldsymbol{\theta}}_0\right\|^2}$$

where $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$. *Then under the null,* $\hat{T}$ *converges in distribution to* T. *That is,*

$$\hat{T} \xrightarrow{\mathcal{D}} T.$$

Proof. According to the transformation above, we can express $\hat{T}$ as:

$$\hat{T} = \frac{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} \left\| \tilde{\mathbf{Z}}_s - \tilde{\boldsymbol{\theta}}_0 \right\|^2 - \min_{\tilde{\boldsymbol{\theta}}_1 \in \tilde{C}} \left\| \tilde{\mathbf{Z}}_s - \tilde{\boldsymbol{\theta}}_1 \right\|^2}{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} \left\| \tilde{\mathbf{Z}}_s - \tilde{\boldsymbol{\theta}}_0 \right\|^2} + o_p(1)$$

$$= \frac{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} \left\| \tilde{\mathbf{Z}}_s - \mathbf{Z}_U + \mathbf{Z}_U - \tilde{\boldsymbol{\theta}}_0 \right\|^2 - \min_{\tilde{\boldsymbol{\theta}}_1 \in \tilde{C}} \left\| \tilde{\mathbf{Z}}_s - \mathbf{Z}_U + \mathbf{Z}_U - \tilde{\boldsymbol{\theta}}_1 \right\|^2}{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} \left\| \tilde{\mathbf{Z}}_s - \mathbf{Z}_U + \mathbf{Z}_U - \tilde{\boldsymbol{\theta}}_0 \right\|^2} + o_p(1)$$

$$= \frac{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} \left\| \mathbf{Z}^{\text{center}} - \tilde{\boldsymbol{\theta}}_0 \right\|^2 - \min_{\tilde{\boldsymbol{\theta}}_1 \in \tilde{C}} \left\| \mathbf{Z}^{\text{center}} - \tilde{\boldsymbol{\theta}}_1 \right\|^2}{\min_{\tilde{\boldsymbol{\theta}}_0 \in \tilde{V}} \left\| \mathbf{Z}^{\text{center}} - \tilde{\boldsymbol{\theta}}_0 \right\|^2} + o_p(1)$$

where $\mathbf{Z}^{\text{center}} = \tilde{\mathbf{Z}}_s - \mathbf{Z}_U$, and recall that under H_0, $\mathbf{Z}_U \in \tilde{V}$, so that, in the above expression, minimizing over $\tilde{\boldsymbol{\theta}}_0$ is equivalent to minimizing over $-\mathbf{Z}_U + \tilde{\boldsymbol{\theta}}_0$, and similarly for minimizing over $\tilde{\boldsymbol{\theta}}_1$.

Then, we have $\hat{T} \xrightarrow{\mathcal{D}} T$. This follows from the Lipschitz continuity of the projection of $\mathbf{Z}$ onto a convex cone; that is, if $\hat{\boldsymbol{\theta}}$ is the projection of $\mathbf{Z}$ onto the cone C, then $\hat{\boldsymbol{\theta}}$ is a continuous function of $\mathbf{Z}$; see Proposition 1 and its proof in Meyer and Woodroofe (2000).

The random variable T defined in Theorem 1 has been shown to be distributed as a mixture of beta random variables under H_0. See Robertson et al. (1988) in Chapter 2 and Meyer (2003) for more details. Also, the mixing distribution can be found (to within a desired precision) via simulation. Specifically, if $M_0 \le M$ is the rank of the constraint matrix $\mathbf{A}$,

$$\Pr(T \le c) = \sum_{m=0}^{M_0} \Pr\left\{ \text{Be}\left(\frac{M_0 - m}{2}, \frac{m}{2} \right) \le c \right\} p_m,$$

where the mixing probabilities $p_0, \ldots, p_{M_0}$ are approximated through simulations, and $\text{Be}(\alpha, \beta)$ represents a Beta random variable with parameters α and β, respectively. By convention, $\text{Be}(0, \beta) = 0$ and $\text{Be}(\alpha, 0) = 1$.

If m is the dimension of the space spanned by the rows of $\mathbf{A}$ that represent binding constraints, then each p_m represents the probability that m constraints are binding, $m = 0, \ldots, M_0$. Each row of $\hat{\mathbf{A}}$ represents a constraint, and we say that the j^{th} constraint is binding if the j^{th} element of $\hat{\mathbf{A}}\hat{\boldsymbol{\theta}}$ is zero. The quantity $D - m$, where m is the number of binding constraints, can be thought of as the observed degrees of freedom of the fit. For more information about this mixing distribution, see Silvapulle and Sen (2005), Chapter 3. The mixing probabilities are approximated as follows:

(1) Generate $\mathbf{Z}$ from a standard multivariate normal distribution $N(\mathbf{0}, \mathbf{I})$.
(2) Project the generated $\mathbf{Z}$ onto the convex cone $\hat{C} = \{\boldsymbol{\theta} : \hat{\mathbf{A}}\boldsymbol{\theta} \ge 0\}$ to obtain the J set, where $\hat{\mathbf{A}} = \mathbf{A}\hat{\boldsymbol{\Sigma}}^{1/2}$. Specifically, let $\hat{\boldsymbol{\theta}}$ be the projection of $\mathbf{Z}$ onto the $\hat{C}$, then $J = \{ j : \hat{\mathbf{A}}_j \hat{\boldsymbol{\theta}} = 0 \}$, where $\hat{\mathbf{A}}_j$ is the j^{th} row of $\hat{\mathbf{A}}$. That is, J indexes the set of "binding constraints". The `R`

package `coneproj` (Liao and Meyer (2014)) finds $\hat{\boldsymbol{\theta}}$ given the generated $\mathbf{Z}$ and $\hat{\mathbf{A}}$, and also returns the set of binding constraints J.

(3) Repeat the previous steps R times (say $R = 1{,}000$).

(4) Estimate p_m by the proportion of times that the set J has m elements, $m = 0, 1, \ldots, M_0$. When the matrix $\mathbf{A}$ has more constraints than dimensions, then, the cone projection routine in `coneproj` can always find a minimal unique J set. (See Meyer (2013) for details.)

3.1 The properties of asymptotic power of the test

In this section, we prove consistency and monotonicity of the power function of this test. First, we show that if the alternative hypothesis is true, then the probability of rejecting the null hypothesis increases to one as N and n increase without bound.

Theorem 2. *Let α be the test size and c_α be the corresponding critical value of the test. Then, the power of the test converges to 1 under the alternative. That is:*

$$P\left(\hat{T} > c_\alpha \mid \overline{\mathbf{y}}_U \in \mathcal{C} \setminus V\right) \to 1, \quad as \quad N \to \infty.$$

Proof. Since $\hat{T} = 1 - \frac{\min_{\boldsymbol{\theta}_1 \in C} (\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1)^\top \hat{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1)}{\min_{\boldsymbol{\theta}_0 \in V} (\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0)^\top \hat{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0)}$, it suffices to show that:

$$\frac{\min_{\boldsymbol{\theta}_1 \in C} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^\top \hat{\boldsymbol{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)}{\min_{\boldsymbol{\theta}_0 \in V} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top \hat{\boldsymbol{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)} = o_p(1)$$

under the the alternative. For the numerator, we have

$$\begin{aligned}
\min_{\boldsymbol{\theta}_1 \in C} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^\top \hat{\boldsymbol{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right) &\leq \left(\tilde{\mathbf{y}}_s - \overline{\mathbf{y}}_U\right)^\top \hat{\boldsymbol{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \overline{\mathbf{y}}_U\right) \\
&= O_p\left(\frac{1}{\sqrt{n}}\right) O_p(n)\, O_p\left(\frac{1}{\sqrt{n}}\right) = O_p(1)
\end{aligned}$$

where we use the fact that $\tilde{\mathbf{y}}_s - \overline{\mathbf{y}}_U = O_p\left(n^{-1/2}\right)$ and $\hat{\boldsymbol{\Sigma}} = O_p\left(n^{-1}\right)$ element-wise. For the denominator, we have:

$$\begin{aligned}
\min_{\boldsymbol{\theta}_0 \in V} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right)^\top \hat{\boldsymbol{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_0\right) &= \min_{\hat{\boldsymbol{\theta}}_0 \in \hat{V}} \left\| \hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_0 \right\|^2 \\
&= \left[\hat{\mathbf{Z}}_s - \left(\mathbf{I} - \hat{\mathbf{A}}^\top \left(\hat{\mathbf{A}}\hat{\mathbf{A}}^\top\right)^- \hat{\mathbf{A}}\right) \hat{\mathbf{Z}}_s\right]^\top \left[\hat{\mathbf{Z}}_s - \left(\mathbf{I} - \hat{\mathbf{A}}^\top \left(\hat{\mathbf{A}}\hat{\mathbf{A}}^\top\right)^- \hat{\mathbf{A}}\right) \hat{\mathbf{Z}}_s\right] \\
&= \hat{\mathbf{Z}}_s^\top \hat{\mathbf{A}}^\top \left(\hat{\mathbf{A}}\hat{\mathbf{A}}^\top\right)^- \hat{\mathbf{A}} \hat{\mathbf{Z}}_s \\
&= \tilde{\mathbf{y}}_s^\top \mathbf{A}^\top \left(\mathbf{A}\hat{\boldsymbol{\Sigma}}\mathbf{A}^\top\right)^- \mathbf{A}\tilde{\mathbf{y}}_s.
\end{aligned}$$

Hence, we have:

$$\frac{\min_{\boldsymbol{\theta}_1 \in C}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)^{\top} \hat{\boldsymbol{\Sigma}}^{-1}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)}{\min_{\boldsymbol{\theta}_0 \in V}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_0\right)^{\top} \hat{\boldsymbol{\Sigma}}^{-1}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_0\right)} = \frac{\min_{\boldsymbol{\theta}_1 \in C}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)^{\top} \left(n\hat{\boldsymbol{\Sigma}}\right)^{-1}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)}{\min_{\boldsymbol{\theta}_0 \in V}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_0\right)^{\top} \left(n\hat{\boldsymbol{\Sigma}}\right)^{-1}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_0\right)}$$

$$= O_p\left(n^{-1}\right) \frac{1}{\tilde{\mathbf{y}}_s^{\top} \mathbf{A}^{\top}\left(\mathbf{A} n \hat{\boldsymbol{\Sigma}} \mathbf{A}^{\top}\right)^{-} \mathbf{A} \tilde{\mathbf{y}}_s}$$

$$= O_p\left(n^{-1}\right) \frac{1}{\overline{\mathbf{y}}_U^{\top} \mathbf{A}^{\top}\left(\mathbf{A} n \boldsymbol{\Sigma} \mathbf{A}^{\top}\right)^{-} \mathbf{A} \overline{\mathbf{y}}_U + o_p(1)}$$

$$= O_p\left(n^{-1}\right) O_p(1) = o_p(1)$$

because $\tilde{\mathbf{y}}_s$ and $\hat{\boldsymbol{\Sigma}}$ are consistent for $\overline{\mathbf{y}}_U$ and $\boldsymbol{\Sigma}$ respectively. Therefore, under the alternative, $\hat{T}$ goes to 1 asymptotically.

4. Simulation studies

The simulations involve one or two dimensional grids, with several constraints and population domain means. We present the results in table form from three scenarios: for each, we record the proportions of times the null is rejected in various cases, with different sample sizes, significance levels and the variances for generating the study variables. In each case, we generate a population of size N, then we draw 10,000 samples from the population according to a sampling design. For each sample, we compute the test statistic value and the estimated covariance matrix. We compare the test statistic with the critical values under different significance levels, where the critical values are obtained from the asymptotic null distribution of the test statistics. Further, we compare the power of this one-sided test with that of ANOVA F test using the unconstrained alternative. That is,

$$H_0: \mathbf{A}\overline{\mathbf{y}}_U = \mathbf{0} \quad \text{versus} \quad H_2: \mathbf{A}\overline{\mathbf{y}}_U \neq \mathbf{0}.$$

Here, we use `svyglm` function in `survey` package to fit the ANOVA model and compute the P-values of the ANOVA F test by applying the `anova` function in `survey` package.

4.1 Monotonicity in one variable

As in Xu et al. (2021) and Oliva-Aviles et al. (2020), the limiting domain means for generating the study variables are given by the functions as follows:

$$\mu_d^{(0)} \equiv 1, \quad \mu_d^{(1)} = \frac{\exp(12d/D-6)}{3.5\left(1+\exp(12d/D-6)\right)}, \quad \mu_d^{(2)} = \frac{\exp(12d/D-6)}{2.5\left(1+\exp(12d/D-6)\right)}$$

for $d = 1, 2, \ldots, D,$ where $D = 12$ is the number of domains. The study variables $y_1, \ldots, y_N$ are generated by adding independent and identically distributed $N\left(0, \sigma_i^2\right)\left(i = 1, 2\right)$ errors to the μ_d values from above three functions, respectively, with $\sigma_1 = 1$ and $\sigma_2 = 1.5.$ We compare the test size and power for the test of constant versus increasing domain means, with the standard ANOVA test of constant versus non-constant domain means. Notice that under $\boldsymbol{\mu}^{(0)} = \left(\mu_1^{(0)}, \ldots, \mu_D^{(0)}\right)^{\mathrm{T}},$ the null hypothesis is true, while under $\boldsymbol{\mu}^{(0)} = \left(\mu_1^{(1)}, \ldots, \mu_D^{(1)}\right)^{\mathrm{T}}$ and $\boldsymbol{\mu}^{(2)} = \left(\mu_1^{(2)}, \ldots, \mu_D^{(2)}\right)^{\mathrm{T}},$ the population domain means have increasing order and thus the alternative is true, with $\boldsymbol{\mu}^{(2)}$ having larger effect size.

We draw the samples from a stratified random sampling design without replacement, with $H = 4$ strata that cut across the D domains. The strata are determined using an auxiliary variable $z,$ which is correlated with study variable $y.$ The values of z are created by adding i.i.d. standard normal errors to $\left(d/D\right).$ By ranking the values of $z,$ we can create 4 blocks of N/H elements. Then, the stratum membership of the population element is determined by the corresponding ranked $z.$ Finally, the population sizes are set to be $N = 9{,}600,$ $N = 19{,}200,$ $N = 57{,}600$ and $N = 76{,}800$ with population domain size $N_d = N/D$ for $d = 1, \ldots, D.$ The total sample sizes $n = 200,$ $n = 400,$ $n = 1{,}200$ and $n =$ 1,600 are assigned to the 4 strata with sample size (25, 50, 50, 75), (50, 100, 100, 150), (150, 300, 300, 450), (200, 400, 400, 600) in each stratum, respectively.

The results in Table 4.1 show that the test size for the proposed one-sided test is closer to the target, while the two-sided test size is somewhat inflated even for the larger sample sizes. For the simulations where the alternative hypothesis is true, the one-sided test has substantially higher power.

Table 4.1
Monotonicity in one variable: the proportions of times null is rejected under various settings and power comparison between the constrained one-sided test (top half) and the unconstrained test (bottom half)

	σ	n	$\alpha = 0.1$			$\alpha = 0.05$			$\alpha = 0.01$		
			$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$
One-sided test	$\sigma = 1$	n = 200	0.0996	0.4689	0.6686	0.0533	0.3218	0.5055	0.0134	0.1194	0.2230
		n = 400	0.0840	0.6352	0.8529	0.0403	0.4780	0.7268	0.0085	0.2028	0.4054
		n = 1,200	0.1039	0.9657	0.9986	0.0537	0.9027	0.9941	0.0121	0.6444	0.9133
		n = 1,600	0.0981	0.9867	0.9999	0.0489	0.9550	0.9988	0.0110	0.7533	0.9654
	$\sigma = 1.5$	n = 200	0.0994	0.3128	0.4370	0.0528	0.2008	0.2938	0.0133	0.0625	0.1056
		n = 400	0.0839	0.4101	0.5946	0.0402	0.2740	0.4338	0.0084	0.0873	0.1770
		n = 1,200	0.1037	0.7838	0.9461	0.0532	0.6327	0.8679	0.0120	0.3142	0.5773
		n = 1,600	0.0980	0.8544	0.9751	0.0488	0.7253	0.9334	0.0109	0.3900	0.6928
ANOVA F test	$\sigma = 1$	n = 200	0.1412	0.2677	0.4017	0.0746	0.1627	0.2685	0.0147	0.0457	0.0973
		n = 400	0.1280	0.3618	0.6034	0.0658	0.2385	0.4627	0.0147	0.0835	0.2259
		n = 1,200	0.1123	0.8139	0.9854	0.0590	0.7121	0.9694	0.0117	0.4736	0.8943
		n = 1,600	0.1111	0.9253	0.9986	0.0576	0.8633	0.9964	0.0126	0.6868	0.9814
	$\sigma = 1.5$	n = 200	0.1412	0.1909	0.2502	0.0746	0.1087	0.1495	0.0147	0.0261	0.0408
		n = 400	0.1280	0.2195	0.3278	0.0658	0.1296	0.2094	0.0147	0.0313	0.0661
		n = 1,200	0.1123	0.4670	0.7538	0.0590	0.3320	0.6361	0.0117	0.1397	0.3902
		n = 1,600	0.1111	0.5947	0.8795	0.0576	0.4602	0.8014	0.0126	0.2367	0.5932

4.2 Block monotonic in one variable

In “block monotonic” ordering case, we assume the population means are ordered among blocks, but there is no ordering imposed within the blocks. Specifically, we organize the limiting domain means in four blocks of three domains as following:

$$\boldsymbol{\mu}^{(0)} = \left(0.05\ \ 0.05\ \ 0.05 \,|\, 0.05\ \ 0.05\ \ 0.05 \,|\, 0.05\ \ 0.05\ \ 0.05 \,|\, 0.05\ \ 0.05\ \ 0.05\right)$$

$$\boldsymbol{\mu}^{(1)} = \left(-0.06\ \ 0\ \ 0.06 \,|\, 0.12\ \ 0.06\ \ 0.18 \,|\, 0.18\ \ 0.24\ \ 0.30 \,|\, 0.30\ \ 0.36\ \ 0.30\right)$$

$$\boldsymbol{\mu}^{(2)} = \left(-0.08\ \ 0\ \ 0.08 \,|\, 0.16\ \ 0.08\ \ 0.24 \,|\, 0.24\ \ 0.32\ \ 0.40 \,|\, 0.40\ \ 0.48\ \ 0.40\right)$$

where the blocks are separated by the vertical lines. Hence, under the alternative, we expect the population mean for each of the domains in block b would be at least as large as those in block $b-1$, for $b = 2, 3, 4$. The effect size of $\overline{\mathbf{y}}_U^{(2)}$ generated from $\boldsymbol{\mu}^{(2)}$ would be larger than that of $\overline{\mathbf{y}}_U^{(1)}$ from $\boldsymbol{\mu}^{(1)}$. We use the same stratified simple random sampling design as in the previous example.

The results in Table 4.2 show again that one-sided test has substantially higher power for simulations where the alternative is true, and for simulations under the null hypothesis, the test size is approximately correct for the one-sided test and the two-sided ANOVA test has inflated test size.

Table 4.2
Block monotonicity in one variable: the proportions of times null is rejected under various settings and power comparison between the constrained one-sided test (top half) and the unconstrained test (bottom half)

	σ	n	α_1 = 0.1			α_2 = 0.05			α_3 = 0.01		
			$\mu^{(0)}$	$\mu^{(1)}$	$\mu^{(2)}$	$\mu^{(0)}$	$\mu^{(1)}$	$\mu^{(2)}$	$\mu^{(0)}$	$\mu^{(1)}$	$\mu^{(2)}$
One-sided test	σ = 1	n = 200	0.1013	0.5114	0.6795	0.0568	0.3590	0.5216	0.0119	0.1397	0.2391
		n = 400	0.1036	0.7368	0.8856	0.0534	0.5838	0.7878	0.0109	0.2840	0.4722
		n = 1,200	0.0964	0.9718	0.9978	0.0487	0.9224	0.9880	0.0089	0.6671	0.8801
		n = 1,600	0.0976	0.9877	0.9998	0.0492	0.9635	0.9958	0.0098	0.7668	0.9339
	σ = 1.5	n = 200	0.1014	0.3421	0.4535	0.0567	0.2191	0.3124	0.0117	0.0731	0.1144
		n = 400	0.1031	0.4992	0.6616	0.0534	0.3544	0.5028	0.0109	0.1335	0.2235
		n = 1,200	0.0965	0.8187	0.9422	0.0485	0.6794	0.8672	0.0091	0.3474	0.5661
		n = 1,600	0.0974	0.8830	0.9743	0.0497	0.7652	0.9232	0.0099	0.4367	0.6746
ANOVA F test	σ = 1	n = 200	0.1412	0.2941	0.4368	0.0746	0.1847	0.2951	0.0147	0.0551	0.1155
		n = 400	0.1280	0.4220	0.6556	0.0658	0.2912	0.5231	0.0147	0.1123	0.2712
		n = 1,200	0.1123	0.8940	0.9921	0.0590	0.8177	0.9840	0.0117	0.6099	0.9363
		n = 1,600	0.1111	0.9678	0.9995	0.0576	0.9293	0.9986	0.0126	0.8094	0.9911
	σ = 1.5	n = 200	0.1412	0.2052	0.2611	0.0746	0.1173	0.1583	0.0147	0.0281	0.0431
		n = 400	0.1280	0.2445	0.3543	0.0658	0.1457	0.2333	0.0147	0.0389	0.0787
		n = 1,200	0.1123	0.5399	0.8012	0.0590	0.4099	0.6932	0.0117	0.1926	0.4549
		n = 1,600	0.1111	0.6799	0.9091	0.0576	0.5539	0.8468	0.0126	0.3153	0.6589

4.3 Monotonicity in two variables

Here we take into consideration a grid of domains, which represent two variables. The null hypothesis is that the population domain means are constant in one of the variables, and the alternative is that the population means are increasing in that variable, while the domain means unconstrained in the other variable. In other words, we test for monotonicity in one variable while "controlling for" the effects of the other. In particular, we set the limiting domain means as follows:

$$\boldsymbol{\mu}^{(0)} = \begin{pmatrix} 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.02 & 0.02 & 0.02 & 0.02 & 0.02 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.04 & 0.04 & 0.04 & 0.04 & 0.04 \end{pmatrix},$$

while

$$\boldsymbol{\mu}^{(1)} = \begin{pmatrix} 0 & 0.04 & 0.16 & 0.24 & 0.28 \\ 0.04 & 0.08 & 0.20 & 0.32 & 0.40 \\ 0.04 & 0.12 & 0.12 & 0.20 & 0.28 \\ 0.04 & 0.04 & 0.12 & 0.24 & 0.28 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\mu}^{(2)} = \begin{pmatrix} 0 & 0.05 & 0.20 & 0.30 & 0.35 \\ 0.05 & 0.10 & 0.25 & 0.40 & 0.50 \\ 0.05 & 0.15 & 0.15 & 0.25 & 0.35 \\ 0.05 & 0.05 & 0.15 & 0.30 & 0.35 \end{pmatrix}.$$

The sampling mechanism and the way we generate the study variable y are the same as that in one dimensional case. However, because there are more domains in this case, we set the sample size to be $n = 400,$ $n = 800,$ $n = 1{,}200$ and $n = 2{,}000,$ respectively, corresponding to the population size $N = 8{,}000,$ $N = 16{,}000,$ $N = 24{,}000$ and $N = 40{,}000,$ where the sample sizes are divided among the strata as $(50, 100, 100, 150),$ $(100, 200, 200, 300),$ $(150, 300, 300, 450)$ and $(250, 500, 500, 750),$ respectively. The simulation results in Table 4.3 demonstrate similar properties as those in the previous scenarios: the tests have higher power as sample size gets larger and the effect size of the population domain means is larger.

Table 4.3
Monotonicity in two variables: the proportions of times null is rejected under various settings and power comparison between the constrained one-sided test (top half) and the unconstrained test (bottom half)

	σ	n	$\alpha_1 = 0.1$			$\alpha_2 = 0.05$			$\alpha_3 = 0.01$		
			$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\mu}^{(0)}$	$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\mu}^{(2)}$
One-sided test	$\sigma = 1$	n = 400	0.1770	0.7738	0.8755	0.1000	0.6415	0.7757	0.0255	0.3460	0.4907
		n = 800	0.1203	0.8732	0.9576	0.0590	0.7677	0.8975	0.0129	0.4706	0.6598
		n = 1,200	0.1097	0.9571	0.9921	0.0562	0.8972	0.9762	0.0102	0.6556	0.8523
		n = 2,000	0.1093	0.9929	0.9994	0.0558	0.9794	0.9975	0.0103	0.8661	0.9700
	$\sigma = 1.5$	n = 400	0.1778	0.5837	0.6840	0.1006	0.4301	0.5382	0.0255	0.1844	0.2586
		n = 800	0.1210	0.6512	0.7783	0.0594	0.4967	0.6399	0.0133	0.2257	0.3421
		n = 1,200	0.1098	0.7701	0.8908	0.0565	0.6247	0.7881	0.0100	0.3235	0.4909
		n = 2,000	0.1089	0.9019	0.9725	0.0560	0.8040	0.9292	0.0103	0.5150	0.7236
ANOVA F test	$\sigma = 1$	n = 400	0.1584	0.4337	0.5642	0.0828	0.3005	0.4255	0.0184	0.1075	0.1886
		n = 800	0.1338	0.5817	0.7748	0.0703	0.4407	0.6600	0.0154	0.2165	0.4058
		n = 1,200	0.1273	0.7028	0.8922	0.0662	0.5773	0.8149	0.0140	0.3224	0.6055
		n = 2,000	0.1289	0.9174	0.9912	0.0697	0.8577	0.9789	0.0149	0.6664	0.9198
	$\sigma = 1.5$	n = 400	0.1584	0.2899	0.3578	0.0828	0.1759	0.2285	0.0184	0.0510	0.0732
		n = 800	0.1338	0.3283	0.4443	0.0703	0.2138	0.3133	0.0154	0.0717	0.1274
		n = 1,200	0.1273	0.3803	0.5358	0.0662	0.2606	0.4009	0.0140	0.1014	0.1883
		n = 2,000	0.1289	0.5759	0.7811	0.0697	0.4434	0.6683	0.0149	0.2148	0.4215

5. Application to NSCG 2019 data

To demonstrate the utility of the proposed one-sided test procedure in real survey data, we consider the 2019 National Survey of College Graduates, which is conducted by the U.S. Census Bureau. NSCG is a repeated cross-sectional biennial complex survey that provides data on the characteristics of the nation's college graduates, with a focus on those in the science and engineering workforce. In all survey cycles, NSCG used a stratified sampling design to select its sample from the eligible sampling frame, which is the American Community Survey (ACS). Specifically, sample cases were selected from the returning sample members in 2013 NSCG (originally selected from the 2011 ACS), 2015 NSCG (originally selected from the 2013 ACS), 2017 NSCG (originally selected from the 2015 ACS) and the 2017 ACS. Within the sampling strata, probability proportional to size (PPS) or systematic random sampling techniques was used to select the NSCG sample. Due to its various complexities, NSCG implemented replication based approach to variance estimation. The variance-covariance matrix is computed by using the 2019 NSCG replicate weights, which are based on Successive Difference and Jackknife replication methods. The number of replicate weights is 320, which is a decent number to provide a stable variance estimate. Both the replicate weights and replicate adjustment factors were calculated by NSCG and are available upon request. The public use files and relevant documentation are available to the public on the NCSES website (https://www.nsf.gov/statistics/srvygrads/).

The annual salary is the study variable (denoted by SALARY in the dataset), restricted to observations with an annual salary between $30,000 and $900,000. As the annual salary variable distribution is skewed, a log transformation is implemented. Four variables are considered:

- Field (denoted by NDGMEMG in the dataset): This nominal variable defines the field of study for the highest degree. There are six levels: (1) Computer and mathematical sciences; (2) Biological, agricultural and environmental life sciences; (3) Physical and related sciences; (4) Social and related sciences; (5) Engineering; (6) Other.
- Father's education level (denoted by EDDAD in the dataset): This ordinal variable denotes the highest level of education completed by the respondents' father (or male guardian). The six levels are: (1) Less than high school completed; (2) High school diploma or equivalent; (3) Some college, vocational, or trade school (including 2-year degrees); (4) Bachelors degree (e.g., BS, BA, AB); (5) Masters degree (e.g., MS, MA, MBA); (6) Professional degree (e.g., JD, LLB, MD, DDS, etc.) and Doctorate (e.g., PhD, DSc, EdD, etc.).
- Academic year of award for the highest degree (denoted by HDACYR).
- Highest degree type (denoted by DGRDG): The four levels are: (1) Bachelor's; (2) Master's; (3) Doctorate; (4) Professional.

Suppose interest is in the question: for wage-earners whose highest degree is a bachelor's, does the father's education level influence the salary, when controlling for field of study and time since degree? To answer this, we perform separate tests for cohorts in years that the degree was attained, as in Table 5.1.

Within each cohort, there are 36 domains, with six levels each of field and father's education level. The sample sizes for the five cohorts in Table 5.1 are 2,021; 4,032; 5,259; 2,969 and 1,813, respectively. So, the domain sample sizes are generally not small within each cohort. We test the null hypothesis that the salary is constant over father's education level, within each field, against the alternative that the salary is increasing in father's education level. We compare the p-values for this test with constrained alternative to the ANOVA test with unconstrained alternative. The `svyglm` function in `survey` package is used for the unconstrained alternative, and the F test by applying the `anova` function in `survey` package gives the p-value. The results of the tests for five recent cohorts are in Table 5.1.

Table 5.1
***p*-values for the null hypothesis that salary is constant in father's education level, controlling for field of study**

year	2006-2007	2008-2010	2011-2013	2014-2015	2016-2017
one-sided test	0.01951	0.00248	0.00029	0.00622	0.00052
ANOVA F test	0.15198	0.10045	0.01357	0.22231	0.06551

For each cohort, the p-value for the one-sided test is below 0.05, indicating that salaries increase significantly with father's education level, consistently across years. In contrast, the p-value for the two-sided test is consistently larger, and does not indicate a significant trend for some of the cohorts, and for other years the test results could be considered "borderline". Using the *a priori* knowledge that if father education level affects salary, it must be a positive effect, helps increase the power to see the trend.

6. Discussion

In this paper, we developed a testing procedure for testing the linear inequality restrictions of the population domain means within the survey context. Under the design normal assumption of the survey domain means, the proposed test statistic $\hat{T}$ has the asymptotic mixture beta densities, where the mixing probabilities (or the weights) can be easily computed via simulations. The covariance estimator $\hat{\mathbf{\Sigma}}$ and the unconstrained estimator $\tilde{\mathbf{y}}_s$ are obtained from the `survey` package in `R` and the constrained least square projection obtained by using the `coneA` function in `coneproj` package. We showed that the power of the test tends to one as the sample size increases, when the alternative hypothesis is true. Simulations show that the test behaves well, with both increased power and improved test size.

The proposed test procedure can be applied to all kinds of complex sampling designs, including stratified sampling, multistage sampling and so on. In practice, though the total sample size n is large, n_d, the number of randomly selected sample in domain d, may be small, or even zero. In such a case, the degrees of freedom (DF) on the estimate of the covariance matrix is small. The degrees of freedom associated with variance estimators was suggested to be (the number of sampled Primary Sampling Units (PSU) with sampled observations in domain d) minus (the number of strata with sampled observations in domain d), see Graubard and Korn (1996) for more details. Thus, neither the design based variance estimator nor the replication based variance estimator can provide accurate covariance estimate, which may undermine the effectiveness of the proposed test. To address the issue of n_d being small or zero, one might need to apply appropriate imputation methods to create proxy responses for domain d (Haziza and Vallée (2020) considered the use of imputed data in variance estimation). The proposed procedure is expected to work properly as long as the estimated covariance matrix $\hat{\mathbf{\Sigma}}$ accounts for the complex design and the sample size for each domain is not too small. Taking the stratified design as an example, even if the sample size is zero for domain d within certain strata, the test procedure is still applicable provided a decent number of samples for domain d were selected from other strata and the covariance estimate $\hat{\mathbf{\Sigma}}$ properly took into account the specific stratified sampling design being considered. In addition, the simulations gave a partial guide for minimum sample sizes needed for the proposed test under stratified simple random sampling design. For more complex sampling design, the effective sample size, defined as the original sample size divided by their design effect, can be considered. Also, it is important to check the weights for units with very low selection probabilities, because extremely small π_i's or π_{ij}'s will result in rather unstable covariance estimate and thus make the proposed test invalid.

Another related issue is that the covariance matrix estimate $\hat{\mathbf{\Sigma}}$ may not be positive definite or even positive semidefinite in finite samples. This problem is not uncommon in survey practice, see Théberge (2022), Haslett (2019), Haslett (2016) for more information. This practical issue will have an impact on the inverses of covariance matrix estimates and thus affect the stability of the proposed test procedure. Hence, we suggest survey practitioners check if covariance estimate is positive definite before applying the proposed test in real application.

The implementation of the test in the `csurvey` package borrows from the `survey` package. For example, suppose we have a grid of domains in two variables `x1` and `x2` and study variable `y`. The survey design is specified with the `svydesign` command in the `survey` package, and the design object `ds` is used in the implementation of the test. The p-value for the test of constant versus increasing domain means along the `x1` variable, without constraining the domain means in the `x2` variable, is obtained as follows.

```
ansc=csvy(y~incr(x1)*x2, data=data_set_name, design=ds, nD=M, test=TRUE)
ansc$pval
```

The `csurvey` package also provides the cone information criteron (CIC) for the fitted model, with and without constraints. The CIC was proposed by Oliva-Aviles, Meyer and Opsomer (2019), for checking monotonicity assumptions in the estimation of order-restricted survey domain means, but is valid for any type of constraints. The command `ansc$CIC`, using the above `csurvey` object, returns the CIC for the data fitted with the constraints. The command `ansc$CIC.un` returns the CIC for the data fitted with no constraints. If the CIC is smaller for the constrained fit, this is evidence that the constraints hold. On the other hand, if the unconstrained CIC is larger, this indicates that the assumptions may be incorrect.

For more information and examples, see the `csurvey` manual.

Acknowledgements

This work was partially supported by NSF MMS 1533804.

Appendix

A. Assumptions

(A1) The number of domains D is a known fixed integer and $\liminf_{N\to\infty} \frac{N_d}{N} > 0$, $\limsup_{N\to\infty} \frac{N_d}{N} < 1$ for $d = 1, 2, \ldots, D$.

(A2) The boundedness property of the finite population fourth moment holds. That is, we have:

$$\limsup_{N\to\infty} N^{-1} \sum_{i\in U} y_i^4 < \infty.$$

(A3) The sample size n is non-random and for a sequence of finite populations U_N with corresponding sequence of samples of size n (for simplicity in notation, we omit the subscript N from n_N), we have $n/N \to \lambda$, as $N \to \infty$, where $0 < \lambda < 1$. There exists a constant vector $\boldsymbol{\mu} \in \mathbb{R}^D$ called the "limiting domain means" so that $\bar{y}_{U_d} \to \mu_d$, for $d = 1, \ldots, D$. In addition, there exists a $\pi \in (0, 1)$ such that, $\min_d \frac{n_d}{N_d} \geq \pi$, as $N \to \infty$, for $d = 1, \ldots, D$.

(A4) For all N, $\min_{i\in U} \pi_i \geq \lambda_1 > 0$ and $\min_{i,j\in U} \pi_{ij} \geq \lambda_2 > 0$, and

$$\limsup_{N\to\infty} n \max_{i,j\in U, i\neq j} \left| \Delta_{ij} \right| < \infty$$

where $\Delta_{ij} = \text{cov}(I_i, I_j) = \pi_{ij} - \pi_i \pi_j$ and I_i is the sample membership indicator for subject i.

(A5) For any vector $\mathbf{x} \in \mathbb{R}^D$ with finite fourth population moment, we have:

$$\text{var}(\hat{\mathbf{x}}_s)^{-1/2} (\hat{\mathbf{x}}_s - \bar{\mathbf{x}}_U) \xrightarrow{\mathcal{D}} N(0, \mathbf{I}_{D\times D})$$

where $\hat{\mathbf{x}}_s$ is the HT estimator of $\bar{\mathbf{x}}_U = \left(N_1^{-1} \sum_{i\in U_1} x_i, \ldots, N_D^{-1} \sum_{i\in U_D} x_i \right)^\top$, $\mathbf{I}_{D\times D}$ is the identity matrix of dimension D, the design covariance matrix $\text{var}(\hat{\mathbf{x}}_s)$ is positive definite.

The assumption (A1) states that the number of domains remains constant as the population size N changes and ensures that there is no asymptotically vanishing domains. Assumption (A2) is a condition needed for showing the variance consistency of the Horvitz-Thompson estimator and this condition generally can be satisfied for most survey data.

In (A3), the assumption of $n/N \to \lambda$ asymptotically ensures that the sample and the population size are of the same order. In addition, by assuming $\min_d \frac{n_d}{N_d} \geq \pi$, as $N \to \infty$, we guarantee that there is no vanishing sampling fraction for each domain d asymptotically, which is a mild condition in the design-based context. Further, the non-random sample size assumption can be relaxed to accommodate a random sample size by imposing particular conditions on the expected sample size $\mathrm{E}_p(n)$.

Assumption (A4) illustrates that the design is both a probability sampling design and a measurable design. The assumption on the Δ_{ij} states that the covariance between sample membership indicators is sufficiently small, which goes to zero at rate of n^{-1}. These conditions hold in many classical sampling designs, including simple random sampling with and without replacement, and many other unequal probability sampling designs.

The asymptotic normal assumption in (A5) is usually assumed explicitly and it is satisfied for many specific sampling designs, including simple random sampling with or without replacement. Also, it holds for Poisson sampling and unequal probability sampling with replacement. The design asymptotic normal assumption, taken together with the variance consistency of the Horvitz-Thompson estimator, can be used to derive the asymptotic distribution of the constrained domain mean estimator. More importantly, it is this normal assumption that makes it possible for us to take advantage of the available techniques in the one-sided test literatures and obtain the null distribution of the test statistics approximately. Otherwise, we have to resort to the bootstrap method to get the empirical distribution of the test statistics when the properties of the design estimator are completely unknown.

It is useful to note that all the results developed in this paper remains design-based. Only the design variability is accounted for by the asymptotic variance in the main results. While the design normal assumption can be viewed as "model-like" assumption, it does not imply a random structure for the population and the inference does not involve any type of model variability. The distributional properties derived in the main text follow from the design and sample size assumptions (A3)-(A5).

B. Supplemental materials for Section 3

In this section, we will show the following result

$$\min_{\boldsymbol{\theta}_1 \in C} \frac{1}{n} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^{\top} \hat{\boldsymbol{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right) - \min_{\boldsymbol{\theta}_1 \in C} \frac{1}{n} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^{\top} \boldsymbol{\Sigma}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right) = o_p(1)$$

to complete the proof for Lemma 1. Based on the result from (2.1) in Xu et al. (2021), for the term $\min_{\boldsymbol{\theta}_1 \in C} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^{\top} \hat{\boldsymbol{\Sigma}}^{-1} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right) = \min_{\hat{\boldsymbol{\theta}}_1 \in \hat{C}} \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_1\right)^{\top} \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_1\right)$, the projection of $\hat{\mathbf{Z}}_s$ onto the cone $\hat{C}$ can be expressed as:

$$\hat{\boldsymbol{\theta}}_1^* = \sum_J \left(\mathbf{I} - \hat{\mathbf{A}}_J^{\top} \left(\hat{\mathbf{A}}_J \hat{\mathbf{A}}_J^{\top}\right)^{-} \hat{\mathbf{A}}_J\right) \hat{\mathbf{Z}}_s I_J(s) \tag{B.1}$$

where the sum is over $J \subseteq \{1, \ldots, M\}$ such that the rows of $\hat{\mathbf{A}}_J$ form a linearly independent set and for each sample s, there is only one subset J for which $I_J(s) = 1$. Using the above explicit form of $\hat{\boldsymbol{\theta}}_1^*$, we prove the following results.

Lemma 2. *Let* $\boldsymbol{\mu}$ *be the limiting domain means. Let* J *be the set that is associated with* $\hat{\boldsymbol{\theta}}_1^*$ *in (B.1) and* $J_{\boldsymbol{\mu}}^0$ *be the corresponding set for the solution* $\boldsymbol{\theta}_{\boldsymbol{\mu}}^*$ *that minimizes* $(\mathbf{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}_1)^\top (\mathbf{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}_1)$ *subject to* $\boldsymbol{\theta}_1 \in C_{\boldsymbol{\mu}} = \{\boldsymbol{\theta} : \mathbf{A}_{\boldsymbol{\mu}}\boldsymbol{\theta} \geq 0\}$, *where* $\mathbf{Z}_{\boldsymbol{\mu}} = \boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{-1/2}\boldsymbol{\mu}$, $C_{\boldsymbol{\mu}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ *are limiting versions of* $\hat{\mathbf{Z}}_s$, $\hat{C}$, $\hat{\boldsymbol{\Sigma}}$ *and* $\mathbf{A}_{\boldsymbol{\mu}} = \mathbf{A}\boldsymbol{\Sigma}_{\boldsymbol{\mu}}^{1/2}$. *Define* $J_{\boldsymbol{\mu}}^1 = \{j : \mathbf{A}_j\boldsymbol{\mu} = 0\}$ *and let* $J_{\boldsymbol{\mu}} = J_{\boldsymbol{\mu}}^0 \cup J_{\boldsymbol{\mu}}^1$, *Then, we have:*

$$\Pr\left(J \nsubseteq J_{\boldsymbol{\mu}}\right) = o(1) \quad \textit{and} \quad \Pr\left(J_{\boldsymbol{\mu}}^0 \nsubseteq J\right) = o(1).$$

Proof. Firstly, consider the event $J \nsubseteq J_{\boldsymbol{\mu}}$. Define

$$\begin{aligned}
\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_1^*\right) &= \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_1^*\right)^\top \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_1^*\right) \\
&= \left[\hat{\mathbf{Z}}_s - \left(\mathbf{I} - \hat{\mathbf{A}}_J^\top \left(\hat{\mathbf{A}}_J \hat{\mathbf{A}}_J^\top\right)^- \hat{\mathbf{A}}_J\right)\hat{\mathbf{Z}}_s\right]^\top \left[\hat{\mathbf{Z}}_s - \left(\mathbf{I} - \hat{\mathbf{A}}_J^\top \left(\hat{\mathbf{A}}_J \hat{\mathbf{A}}_J^\top\right)^- \hat{\mathbf{A}}_J\right)\hat{\mathbf{Z}}_s\right] \\
&= \hat{\mathbf{Z}}_s^\top \hat{\mathbf{A}}_J^\top \left(\hat{\mathbf{A}}_J \hat{\mathbf{A}}_J^\top\right)^- \hat{\mathbf{A}}_J \hat{\mathbf{Z}}_s \\
&= \tilde{\mathbf{y}}_s^\top \mathbf{A}_J^\top \left(\mathbf{A}_J \hat{\boldsymbol{\Sigma}} \mathbf{A}_J^\top\right)^- \mathbf{A}_J \tilde{\mathbf{y}}_s
\end{aligned}$$

similarly, we define:

$$\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right) = \left(\mathbf{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right)^\top \left(\mathbf{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right) = \boldsymbol{\mu}^\top \mathbf{A}_{J_{\boldsymbol{\mu}}^0}^\top \left(\mathbf{A}_{J_{\boldsymbol{\mu}}^0} \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mathbf{A}_{J_{\boldsymbol{\mu}}^0}^\top\right)^- \mathbf{A}_{J_{\boldsymbol{\mu}}^0} \boldsymbol{\mu}.$$

Note that the projection of $\mathbf{Z}_{\boldsymbol{\mu}}$ onto the linear space spanned by rows of $\mathbf{A}_{\boldsymbol{\mu}}$ in position $J_{\boldsymbol{\mu}}^0$ is the same as the projection onto the linear space spanned by rows of $\mathbf{A}_{\boldsymbol{\mu}}$ in position $J_{\boldsymbol{\mu}}$, so we have:

$$\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right) = \boldsymbol{\mu}^\top \mathbf{A}_{J_{\boldsymbol{\mu}}^0}^\top \left(\mathbf{A}_{J_{\boldsymbol{\mu}}^0} \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mathbf{A}_{J_{\boldsymbol{\mu}}^0}^\top\right)^- \mathbf{A}_{J_{\boldsymbol{\mu}}^0} \boldsymbol{\mu} = \boldsymbol{\mu}^\top \mathbf{A}_{J_{\boldsymbol{\mu}}}^\top \left(\mathbf{A}_{J_{\boldsymbol{\mu}}} \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mathbf{A}_{J_{\boldsymbol{\mu}}}^\top\right)^- \mathbf{A}_{J_{\boldsymbol{\mu}}} \boldsymbol{\mu}.$$

Further, denote:

$$\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}\right) = \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}\right)^\top \left(\hat{\mathbf{Z}}_s - \hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}\right) = \tilde{\mathbf{y}}_s^\top \mathbf{A}_{J_{\boldsymbol{\mu}}}^\top \left(\mathbf{A}_{J_{\boldsymbol{\mu}}} \hat{\boldsymbol{\Sigma}} \mathbf{A}_{J_{\boldsymbol{\mu}}}^\top\right)^- \mathbf{A}_{J_{\boldsymbol{\mu}}} \tilde{\mathbf{y}}_s$$

$$\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu},J}\right) = \left(\mathbf{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}_{\boldsymbol{\mu},J}\right)^\top \left(\mathbf{Z}_{\boldsymbol{\mu}} - \boldsymbol{\theta}_{\boldsymbol{\mu},J}\right) = \boldsymbol{\mu}^\top \mathbf{A}_J^\top \left(\mathbf{A}_J \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mathbf{A}_J^\top\right)^- \mathbf{A}_J \boldsymbol{\mu}$$

where $\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}} = \left(\mathbf{I} - \hat{\mathbf{A}}_{J_{\boldsymbol{\mu}}}^\top \left(\hat{\mathbf{A}}_{J_{\boldsymbol{\mu}}} \hat{\mathbf{A}}_{J_{\boldsymbol{\mu}}}^\top\right)^- \hat{\mathbf{A}}_{J_{\boldsymbol{\mu}}}\right)\hat{\mathbf{Z}}_s$ and $\boldsymbol{\theta}_{\boldsymbol{\mu},J} = \left(\mathbf{I} - \mathbf{A}_{\boldsymbol{\mu},J}^\top \left(\mathbf{A}_{\boldsymbol{\mu},J} \mathbf{A}_{\boldsymbol{\mu},J}^\top\right)^- \mathbf{A}_{\boldsymbol{\mu},J}\right)\mathbf{Z}_{\boldsymbol{\mu}}$. Then, we must have

$$\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right) < \mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu},J}\right) \quad \text{and} \quad \widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_1^*\right) < \widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}\right)$$

and due to the consistency of $\tilde{\mathbf{y}}_s$ and $\hat{\boldsymbol{\Sigma}}$, respectively, we also have:

$$\frac{1}{n}\left(\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_1^*\right) - \mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu},J}\right)\right) = o_p(1) \quad \text{and} \quad \frac{1}{n}\left(\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}\right) - \mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right)\right) = o_p(1).$$

Finally, by Markov's inequality, we get:

$$\begin{aligned}\Pr\left(J \not\subseteq J_{\boldsymbol{\mu}}\right) &\le \Pr\left(\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}\right)-\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_1^*\right)+\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu},J}\right)-\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right)>\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu},J}\right)-\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right)\right)\\
&\le \frac{E\left(\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}\right)-\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_1^*\right)+\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu},J}\right)-\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right)\right)}{\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu},J}\right)-\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right)}\\
&= \frac{E\left(\frac{1}{n}\left(\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}\right)-\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right)\right)\right)-E\left(\frac{1}{n}\left(\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_1^*\right)-\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu},J}\right)\right)\right)}{\frac{1}{n}\left(\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu},J}\right)-\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right)\right)}\\
&\to 0\end{aligned}$$

since $E\left(\frac{1}{n}\left(\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_{1,J_{\boldsymbol{\mu}}}\right)-\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right)\right)\right)=o(1)$ and $E\left(\frac{1}{n}\left(\widetilde{\mathrm{SSE}}\left(\hat{\boldsymbol{\theta}}_1^*\right)-\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu},J}\right)\right)\right)=o(1)$. Using the similar argument, we can also show that:

$$\Pr\left(J_{\boldsymbol{\mu}}^0 \not\subseteq J\right)=o(1)$$

this completes the proof.

By the same argument as in Lemma 2, we also have the following result.

Lemma 3. *Let* $J_{\boldsymbol{\Sigma}}$ *(unknown) be the corresponding set of the solution* $\tilde{\boldsymbol{\theta}}_1^*$ *that minimizes* $\left(\tilde{\mathbf{Z}}_s-\boldsymbol{\theta}_1\right)^{\top}\left(\tilde{\mathbf{Z}}_s-\boldsymbol{\theta}_1\right)$ *subject to* $\boldsymbol{\theta}_1 \in \tilde{C}$. *Then, we have:*

$$\Pr\left(J_{\boldsymbol{\Sigma}} \not\subseteq J_{\boldsymbol{\mu}}\right)=o(1) \quad \textit{and} \quad \Pr\left(J_{\boldsymbol{\mu}}^0 \not\subseteq J_{\boldsymbol{\Sigma}}\right)=o(1),$$

where $J_{\boldsymbol{\mu}}$ *and* $J_{\boldsymbol{\mu}}^0$ *are defined in Lemma 2.*

Lemma 4. *We have:*

$$\min_{\boldsymbol{\theta}_1 \in C}\frac{1}{n}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)^{\top}\hat{\boldsymbol{\Sigma}}^{-1}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)-\min_{\boldsymbol{\theta}_1 \in C}\frac{1}{n}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)=o_p(1)$$

with respect to the sampling mechanism.

Proof. Let J be the observed set for a given sample s. We can write the difference as follows:

$$\begin{aligned}&\min_{\boldsymbol{\theta}_1 \in C}\frac{1}{n}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)^{\top}\hat{\boldsymbol{\Sigma}}^{-1}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)-\min_{\boldsymbol{\theta}_1 \in C}\frac{1}{n}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\tilde{\mathbf{y}}_s-\boldsymbol{\theta}_1\right)\\
&\quad= \frac{1}{n}\hat{\mathbf{Z}}_s^{\top}\hat{\mathbf{A}}_J^{\top}\left(\hat{\mathbf{A}}_J\hat{\mathbf{A}}_J^{\top}\right)^{-}\hat{\mathbf{A}}_J\hat{\mathbf{Z}}_s-\frac{1}{n}\tilde{\mathbf{Z}}_s^{\top}\tilde{\mathbf{A}}_{J_{\boldsymbol{\Sigma}}}^{\top}\left(\tilde{\mathbf{A}}_{J_{\boldsymbol{\Sigma}}}\tilde{\mathbf{A}}_{J_{\boldsymbol{\Sigma}}}^{\top}\right)^{-}\tilde{\mathbf{A}}_{J_{\boldsymbol{\Sigma}}}\tilde{\mathbf{Z}}_s\\
&\quad= \tilde{\mathbf{y}}_s^{\top}\mathbf{A}_J^{\top}\left(\mathbf{A}_J n\hat{\boldsymbol{\Sigma}}\mathbf{A}_J^{\top}\right)^{-}\mathbf{A}_J\tilde{\mathbf{y}}_s-\tilde{\mathbf{y}}_s^{\top}\mathbf{A}_{J_{\boldsymbol{\Sigma}}}^{\top}\left(\mathbf{A}_{J_{\boldsymbol{\Sigma}}}n\boldsymbol{\Sigma}\mathbf{A}_{J_{\boldsymbol{\Sigma}}}^{\top}\right)^{-}\mathbf{A}_{J_{\boldsymbol{\Sigma}}}\tilde{\mathbf{y}}_s\\
&\quad= \tilde{\mathbf{y}}_s^{\top}\mathbf{A}_J^{\top}\left(\mathbf{A}_J n\hat{\boldsymbol{\Sigma}}\mathbf{A}_J^{\top}\right)^{-}\mathbf{A}_J\tilde{\mathbf{y}}_s\left(I_{\left(J_{\boldsymbol{\mu}}^0 \subseteq J \subseteq J_{\boldsymbol{\mu}}\right)}+I_{\left(J \not\subseteq J_{\boldsymbol{\mu}} \text{ or } J_{\boldsymbol{\mu}}^0 \not\subseteq J\right)}\right)\\
&\quad\quad- \tilde{\mathbf{y}}_s^{\top}\mathbf{A}_{J_{\boldsymbol{\Sigma}}}^{\top}\left(\mathbf{A}_{J_{\boldsymbol{\Sigma}}}n\boldsymbol{\Sigma}\mathbf{A}_{J_{\boldsymbol{\Sigma}}}^{\top}\right)^{-}\mathbf{A}_{J_{\boldsymbol{\Sigma}}}\tilde{\mathbf{y}}_s\left(I_{\left(J_{\boldsymbol{\mu}}^0 \subseteq J_{\boldsymbol{\Sigma}} \subseteq J_{\boldsymbol{\mu}}\right)}+I_{\left(J_{\boldsymbol{\Sigma}} \not\subseteq J_{\boldsymbol{\mu}} \text{ or } J_{\boldsymbol{\mu}}^0 \not\subseteq J_{\boldsymbol{\Sigma}}\right)}\right)\end{aligned}$$

by Lemma 2 and Lemma 3, we have that $I_{\left(J \nsubseteq J_{\boldsymbol{\mu}} \text{ or } J_{\boldsymbol{\mu}}^0 \nsubseteq J\right)} = o_p(1)$ and $I_{\left(J_{\Sigma} \nsubseteq J_{\boldsymbol{\mu}} \text{ or } J_{\boldsymbol{\mu}}^0 \nsubseteq J_{\Sigma}\right)} = o_p(1)$. Then, we have:

$$\begin{aligned}
&\min_{\boldsymbol{\theta}_1 \in C} \frac{1}{n}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^\top \hat{\boldsymbol{\Sigma}}^{-1}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right) - \min_{\boldsymbol{\theta}_1 \in C} \frac{1}{n}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right)^\top \boldsymbol{\Sigma}^{-1}\left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}_1\right) \\
&\quad = \tilde{\mathbf{y}}_s^\top \mathbf{A}_J^\top \left(\mathbf{A}_J n \hat{\boldsymbol{\Sigma}} \mathbf{A}_J^\top\right)^- \mathbf{A}_J \tilde{\mathbf{y}}_s I_{\left(J_{\boldsymbol{\mu}}^0 \subseteq J \subseteq J_{\boldsymbol{\mu}}\right)} - \tilde{\mathbf{y}}_s^\top \mathbf{A}_{J_\Sigma}^\top \left(\mathbf{A}_{J_\Sigma} n \boldsymbol{\Sigma} \mathbf{A}_{J_\Sigma}^\top\right)^- \mathbf{A}_{J_\Sigma} \tilde{\mathbf{y}}_s I_{\left(J_{\boldsymbol{\mu}}^0 \subseteq J_\Sigma \subseteq J_{\boldsymbol{\mu}}\right)} + o_p(1) \\
&\quad = \boldsymbol{\mu}^\top \mathbf{A}_J^\top \left(\mathbf{A}_J n \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mathbf{A}_J^\top\right)^- \mathbf{A}_J \boldsymbol{\mu} I_{\left(J_{\boldsymbol{\mu}}^0 \subseteq J \subseteq J_{\boldsymbol{\mu}}\right)} - \boldsymbol{\mu}^\top \mathbf{A}_{J_\Sigma}^\top \left(\mathbf{A}_{J_\Sigma} n \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mathbf{A}_{J_\Sigma}^\top\right)^- \mathbf{A}_{J_\Sigma} \boldsymbol{\mu} I_{\left(J_{\boldsymbol{\mu}}^0 \subseteq J_\Sigma \subseteq J_{\boldsymbol{\mu}}\right)} + o_p(1) \\
&\quad = \boldsymbol{\mu}^\top \mathbf{A}_{J_{\boldsymbol{\mu}}^0}^\top \left(\mathbf{A}_{J_{\boldsymbol{\mu}}^0} n \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mathbf{A}_{J_{\boldsymbol{\mu}}^0}^\top\right)^- \mathbf{A}_{J_{\boldsymbol{\mu}}^0} \boldsymbol{\mu} - \boldsymbol{\mu}^\top \mathbf{A}_{J_{\boldsymbol{\mu}}^0}^\top \left(\mathbf{A}_{J_{\boldsymbol{\mu}}^0} n \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mathbf{A}_{J_{\boldsymbol{\mu}}^0}^\top\right)^- \mathbf{A}_{J_{\boldsymbol{\mu}}^0} \boldsymbol{\mu} + o_p(1) \\
&\quad = o_p(1)
\end{aligned}$$

where we use the fact that for any set J with $J_{\boldsymbol{\mu}}^0 \subseteq J \subseteq J_{\boldsymbol{\mu}}$, we have that

$$\begin{aligned}
\mathrm{SSE}\left(\boldsymbol{\theta}_{\boldsymbol{\mu}}^*\right) &= \boldsymbol{\mu}^\top \mathbf{A}_J^\top \left(\mathbf{A}_J \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mathbf{A}_J^\top\right)^- \mathbf{A}_J \boldsymbol{\mu} \\
&= \boldsymbol{\mu}^\top \mathbf{A}_{J_{\boldsymbol{\mu}}^0}^\top \left(\mathbf{A}_{J_{\boldsymbol{\mu}}^0} \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mathbf{A}_{J_{\boldsymbol{\mu}}^0}^\top\right)^- \mathbf{A}_{J_{\boldsymbol{\mu}}^0} \boldsymbol{\mu} \\
&= \boldsymbol{\mu}^\top \mathbf{A}_{J_{\boldsymbol{\mu}}}^\top \left(\mathbf{A}_{J_{\boldsymbol{\mu}}} \boldsymbol{\Sigma}_{\boldsymbol{\mu}} \mathbf{A}_{J_{\boldsymbol{\mu}}}^\top\right)^- \mathbf{A}_{J_{\boldsymbol{\mu}}} \boldsymbol{\mu}.
\end{aligned}$$

References

Bartholomew, D. (1959). A test of homogeneity for ordered alternatives. *Biometrika*, 46, 36-48.

Bartholomew, D. (1961). A test of homogeneity of means under order restrictions. *Journal of the Royal Statistical Society, Series B*, 23(1), 239-272.

Chacko, V.J. (1963). Testing homogeneity against ordered alternatives. *The Annals of Mathematical Statistics*, 34(3), 945-956.

Graubard, B.I., and Korn, E.L. (1996). Survey inference for subpopulations. *American Journal of Epidemiology*, 144, 102-106.

Hájek, J. (1971). Comment on a paper by D. Basu. In *Foundations of Statistical Inference*, (Eds., V.P. Godambe and D.A. Sprott), Toronto: Holt, Rinehartand Winston, 236.

Haslett, S. (2016). Positive semidefiniteness of estimated covariance matrices in linear models for sample survey data. *Special Matrices*, 4, 218-224.

Haslett, S. (2019). Best linear unbiased estimation for varying probability with and without replacement sampling. *Special Matrices*, 7, 78-91.

Haziza, D., and Vallée, A.-A. (2020). Variance estimation procedures in the presence of singly imputed survey data: A critical review. *Japanese Journal of Statistics and Data Science*, 3, 583-623.

Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

Liao, X., and Meyer, M.C. (2014). coneproj: an R package for the primal or dual cone projections with routines for constrained regression. *Journal of Statistical Software*, 61, 1-22.

McDermott, M.P., and Mudholkar, G.S. (1993). A simple approach to testing homogeneity of order-constrained means. *Journal of the American Statistical Association*, 88, 1371-1379.

Meyer, M.C. (2003). A test for linear versus convex regression function using shape-restricted regression. *Biometrika*, 90(1), 223-232.

Meyer, M.C. (2013). A simple new algorithm for quadratic programming with applications in statistics. *Communications in Statistics*, 42, 1126-1139.

Meyer, M.C., and Wang, J. (2012). Improved power of one-sided tests. *Statistics and Probability Letters*, 82, 1619-1622.

Meyer, M.C., and Woodroofe, M. (2000). On the degrees of freedom in shape-restricted regression. *Annals of Statistics*, 28(4), 1083-1104.

Oliva-Aviles, C., Meyer, M.C. and Opsomer, J.D. (2019). Checking validity of monotone domain mean estimators. *The Canadian Journal of Statistics/La revue canadienne de statistique*, 47(2), 315-331.

Oliva-Aviles, C., Meyer, M.C. and Opsomer, J.D. (2020). Estimation and inference of domain means subject to qualitative constraints. *Survey Methodology*, 46, 2, 145-180. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2020002/article/00002-eng.pdf.

Perlman, M.D. (1969). One-sided testing problems in multivariate analysis. *The Annals of Mathematical Statistics*, 40(2), 549-567.

Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order Restricted Statistical Inference*. New York: John Wiley & Sons, Inc.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Sen, B., and Meyer, M.C. (2017). Testing against a linear regression model using ideas from shape-restricted estimation. *Journal of the Royal Statistical Society, Series B*, 79, 423-448.

Silvapulle, M.J., and Sen, P. (2005). *Constrained Statistical Inference*. Hoboken, New Jersey: Wiley.

Théberge, A. (2022). A generalization of inverse probability weighting. *Survey Methodology*, 48, 1, 177-190. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2022001/article/00009-eng.pdf.

Wu, J., Meyer, M.C. and Opsomer, J.D. (2016). Survey estimation of domain means that respect natural orderings. *The Canadian Journal of Statistics/La revue canadienne de statistique*, 44(4), 431-444.

Xu, X., Meyer, M.C. and Opsomer, J.D. (2021). Improved variance estimation for inequality-constrained domain mean estimators using survey data. *Journal of Statistical Planning and Inference*, 215, 47-71.