Testing When Parameters are Subject to Linear Inequality Constraints*

Yanqin Fan and Xuetao Shi†

This draft: Dec 2019

Abstract

This paper introduces the concept of an implicit nuisance parameter for testing the null hypothesis of linear equality constraints against the two-sided alternative hypothesis when the parameters are subject to equality and inequality constraints in the maintained hypothesis. We propose an approach to identify the implicit nuisance parameter and provide a comprehensive study of asymptotically uniformly valid Wald, QLR, and score tests in an extremum estimation set-up. Among the two Wald tests, one QLR test, and three score tests developed in this paper, three tests fully exploit the information in the parameter space and the asymptotic distributions of their test statistics are discontinuous in the implicit nuisance parameter. The other three tests employ part of the information in the maintained hypothesis through projection and the asymptotic distributions of their test statistics are not discontinuous in any model parameter but depend on polytope projections. We present an algorithm based on Fourier-Motzkin Elimination to compute such projections. Numerical results from a Monte Carlo study of the finite sample performance of our tests and an empirical illustration are presented.

*The authors are grateful to Gaston Illanes, Hyungsik Roger Moon, Kyungchul Song, Xun Tang, and Jing Tao for insightful discussions. We also thank participants of the 2019 Seatte-Vancouver Econometrics Conference at University of British Columbia for helpful discussions. This work was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington.

†Department of Economics, University of Washington, Box 353330, Seattle, WA 98195, USA

Keywords: Implicit nuisance parameter; Wald tests; Quasi-Likelihood Ratio test; Score tests; Extremum estimation; Bonferroni-type correction; Polytope projection.

JEL Codes: C12, C18

1 Introduction

Motivation and Hypotheses When the parameter of interest is in the *interior* of its parameter space, methods for estimation and inference have been well developed. Under regularity conditions, an extremum estimator is asymptotically normally distributed. Wald, Quasi-Likelihood Ratio (QLR), and score tests for the null hypothesis of equality constraints against the two-sided alternative hypothesis are asymptotically equivalent, see e.g., Engle (1984). When the parameter is on the *boundary* of the parameter space, Andrews (1997, 1999) develops a general asymptotic theory for an extremum estimator showing that its asymptotic distribution is non-normal.

This paper provides an extensive study of Wald-type, QLR, and score-type tests for linear equality constraints against the two-sided alternative hypothesis when it may be *unknown a priori* whether some parameters are on the boundary or in the interior of the parameter space.[1] Let θ denote the parameter of interest and θ^* be the pseudo-true value. We consider the case that $\theta^* \in \Theta \subset \mathbb{R}^l$, where the parameter space Θ is defined as

$$\Theta \equiv \left\{\theta \in \mathbb{R}^l : \mathscr{R}_e\theta = \mathrm{r}_e \text{ and } \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}, \tag{1}$$

in which $\mathscr{R}_e$ and $\mathscr{R}_w$ are known matrices of dimensions $l_e \times l$ and $l_w \times l$, and r_e and r_w are known vectors of dimensions l_e and l_w respectively. The matrix $\mathscr{R}_e$ denotes equality constraints on θ^*; and the matrix $\mathscr{R}_w$ denotes weak inequality constraints on θ^* that are unknown to bind or not. Although the presence of either type of constraint is not required for the theory, the discussion in the paper caters to the case where weak inequality constraints exist. Equality and inequality constraints of the form seen in Θ are often implied by the natures of the parameters, such as weights (Fox et al. (2011) and Fox et al. (2016)), or by economic theories imposing constraints like non-negativity or monotonicity. It has long been recognized that incorporating

[1] To simplify exposition, we refer to Wald-type and score-type tests simply as Wald and score tests in the rest of the paper.

equality/inequality constraints in parameter estimation can yield an efficiency gain, e.g., Liew (1976), Judge et al. (1984), Chernozhukov and Hong (2004), and Moon and Schorfheide (2009). As noted in Andrews (2001): "in cases where the restrictions on the parameter space arise from prior information, tests that utilize this information have a considerable power advantage over tests that do not."

Under the maintained hypothesis that $\theta^* \in \Theta$, the null and alternative hypotheses we consider in this paper are expressed as

$$H_0 : \theta^* \in \Theta_0 \text{ and } H_1 : \theta^* \in \Theta_1, \tag{2}$$

where $\Theta_0 \equiv \{\theta \in \Theta : R\theta = r\}$ or equivalently,[2]

$$\Theta_0 = \left\{\theta \in \mathbb{R}^l : R\theta = r,\, \mathscr{R}_e\theta = \mathrm{r}_e, \text{ and } \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}, \tag{3}$$

in which R is a known matrix of dimension $J \times l$ and is of full row rank, r is a known vector of dimension J, and

$$\Theta_1 \equiv \Theta \backslash \Theta_0 = \{\theta \in \Theta : R\theta \neq r\}.$$

Main Contributions There are two critical steps in developing asymptotically uniformly valid tests for H_0 under Θ. The first one is to determine binding, non-binding, and undetermined inequalities in $\mathscr{R}_w\theta^* \geq \mathrm{r}_w$ under H_0; and the second one is to identify the *implicit nuisance parameter*, an important concept introduced in this paper. Our first contribution is to propose a generic algorithm for implementing both steps. Because our algorithm does not depend on any specific model, estimator, or test statistic, it is applicable to any parametric or semiparametric model in which the parameter space and the null hypothesis are specified as (1) and (2).

In the first step of our algorithm, we identify the *implicit equalities* and *strictly redundant inequalities* in the null parameter space Θ_0 by applying algorithms STREINQ and IMPLEQ in Telgen (1983), leaving the rest as *undetermined inequalities*. Then we use Gauss-Jordan elimination to identify an implicit nuisance parameter defined as a subvector of the true value of the linear function in the undetermined inequalities that corresponds to a row basis of the coefficient matrix of the linear function. Consider

[2]For the testing problem to be non-trivial, the set Θ_0 is assumed to be non-empty. Methods such as the Fourier-Motzkin Elimination discussed in Section 6.2 of this paper can be used to determine whether Θ_0 is empty or not.

the special case that the parameter vector is non-negative, i.e., $\Theta = \left\{\theta \in \mathbb{R}^l : \theta \geq \mathbf{0}\right\}$, and the null hypothesis is on a subvector of θ^*. Under the null hypothesis, inequalities in the system: $\theta^* \geq \mathbf{0}$ corresponding to the subvector under testing are known to bind (implicit equalities) or not to bind (strictly redundant inequalities) leaving the remaining inequalities undetermined and the remaining subvector as the implicit nuisance parameter.

The second contribution of the paper is to provide a comprehensive study of Wald, QLR, and score tests for H_0 against H_1 in the general class of extremum estimation problems considered in Andrews (1997, 1999) for non-trending data.[3] Although our QLR statistic takes the same form as the classical one, the classical Wald and score statistics may be extended in different ways to account for the inequality constraints in the maintained hypothesis. In the paper, we develop *two* Wald tests, one QLR test, and *three* score tests. These six tests are categorized into two groups according to how information in Θ is used. The tests in Group I, which includes one of the two Wald tests, the QLR test, and one of the three score tests, fully exploit information of the parameter space. The rest of the tests (in Group II) employ only part of the information in Θ through projection. We provide a detailed treatment of three tests in the main part of the paper: the Wald and QLR tests in Group I and one score test in Group II. Our detailed treatment of these tests highlights the different challenges in developing these tests posed by inequality constraints in the maintained hypothesis and illustrates how we address these challenges.

We first show that the null asymptotic distributions of the Wald and QLR statistics in Group I are discontinuous in the implicit nuisance parameter (when it exists),[4] but the null asymptotic distribution of the score statistic in Group II does not depend on the implicit nuisance parameter (even when it exists). In contrast to the standard case of an interior parameter value, the null asymptotic distributions of the three statistics differ except in special cases. We then develop asymptotically uniformly valid Wald and QLR tests via a two-step procedure. This approach is based on a con-

[3] A sequel to this paper will explore the applicability of this approach to dynamic models with deterministic and/or stochastic trends such as the Dickey-Fuller Regression in Andrews (1999) or the GARCH $(1, q^*)$ example in Andrews (1997).

[4] A probability measure P_ϖ is said to be continuous in the model parameter ϖ if $d\left(P_{\varpi_1}, P_{\varpi_2}\right) \to 0$ when $|\varpi_1 - \varpi_2| \to 0$, where $d(\cdot,\cdot)$ is some metric on the probability measures, such as the Kolmogorov, bounded Lipschitz, or total variation metric and $|\cdot|$ is the absolute value. To simplify exposition, we refer to parameters at which the asymptotic distribution of an estimator or the null asymptotic distribution of a test statistic is discontinuous as nuisance parameters.

fidence set for the implicit nuisance parameter and a Bonferroni-type correction. The same method applies to the score test in Group I as well. Since the null asymptotic distributions of the test statistics in Group II do not depend on the implicit nuisance parameter, the two-step approach is not needed. However, implementing such tests requires computing the *projection of a polytope*, an important problem that has been studied extensively in diverse fields, such as constraint logic programming (Huynh et al. (1992)), marginal problems (Fritz and Chaves (2012)), and robotics research (Ponce et al. (1997)). Because this problem has not been explored in the econometrics literature, we present one algorithm based on Fourier-Motzkin Elimination, and show that it can be used in our testing framework. Additionally, the consistency and local power of the tests are investigated. We show that the score tests in Group II may even be inconsistent.

The third contribution of the paper is to establish two sets of asymptotic equivalence results among the six test statistics: the Wald and score statistic in Group I; and three statistics in Group II. Based on the equivalence results, we construct three additional tests.

Lastly, we conduct a simulation study using a linear regression model to investigate and compare the finite sample performance of all six tests developed in this paper with the "classical" Wald, QLR, and score tests.[5] Results demonstrate that the full-information tests in Group I dominate the other tests. Although the three "classical" tests perform well for normally distributed errors, they perform worse than all six tests proposed in this paper when errors are skewed. As an empirical illustration, we apply our tests to the models in Autor and Handel (2013) on wage differentials related to job tasks and human capital and compare our results with the "classical" ones. For testing the significance of regression coefficients, our tests provide more consistent results across different model specifications compared to the "classical" ones.

Related Literatures Maintaining the conventional assumption that the true parameter lies in the interior of the parameter space, existing works have extended Wald, QLR, and score tests from testing equality constraints against two-sided alternatives to testing equality constraints against one-sided alternatives. For example,

[5]In practice, researchers oftentimes ignore constraints in Θ when feasible. To simplify exposition, in the rest of the paper, we refer to tests based on estimators without accounting for constraints in Θ as the "classical" tests to distinguish them from the tests developed in this paper.

Gourieroux et al. (1982) and Silvapulle and Sen (2005) study the null hypothesis of $\mathscr{R}_w\theta^* = r$ against $\mathscr{R}_w\theta^* \geq r$ in linear regression models and general parametric models respectively. This turns out to be a special case of (2) with $R = \mathscr{R}_w$, $r = \mathrm{r}_w$, and $\Theta = \left\{\theta \in \mathbb{R}^l : \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}$. Silvapulle and Sen (2005) further consider testing the null hypothesis of $R\theta^* = r$ against $\mathscr{R}_w\theta \geq \mathrm{r}_w$, where $R = (R_1', \mathscr{R}_w')'$ and $r = (r_1', \mathrm{r}_w')'$. This is another special case of (2) with $\Theta = \left\{\theta \in \mathbb{R}^l : \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}$. In both cases, the weak inequalities in the parameter space are known to bind under H_0. As a result, there is no implicit nuisance parameter. The asymptotic distributions of the test statistics under the null hypothesis are continuous in model parameters. The standard plug-in approach can be used to obtain critical values.

This paper relates closely to Andrews (2001), which studies the testing problem of the point null hypothesis on the subvector of the parameter against the two-sided alternative hypothesis under the maintained hypothesis. Uncertainty about the position of some parameters on the boundary or in the interior of the parameter space is permitted, and the presence of an unidentified nuisance parameter under the null hypothesis is also allowed. By assuming that the normalized "information" matrix is block diagonal between those parameters that are known to lie on the boundary or in the interior of the parameter space and those that are unknown, and by assuming that the approximating cone of the parameter space is a product set, the asymptotic distributions of the Wald and QLR test statistics in Andrews (2001) are continuous in the implicit nuisance parameter.

For the subvector inference problem, Ketz (2018) introduces a Conditional Likelihood Ratio statistic based on sufficient statistic in normal distribution. Under the assumption that some inequality constraints are empirically irrelevant, and that the parameter space is a product set of the space for the parameter under testing and the one for the nuisance parameter, Ketz (2018) shows that the asymptotic distribution of the Conditional Likelihood Ratio statistic is nuisance parameter free.

In this paper, we focus on the null hypothesis of linear equality constraints and the parameter space defined by linear equalities and inequalities. Without imposing any assumption on the matrices R, $\mathscr{R}_e$, and $\mathscr{R}_w$, the asymptotic distributions of some test statistics, such as the Wald and QLR statistics in Andrews (2001), are in general discontinuous in the implicit nuisance parameter. Identifying the implicit nuisance parameter is therefore essential for conducting the uniform inference.

Once the implicit nuisance parameter is found, one can adapt any existing method for uniform inference, especially for *subvector inference*, to our framework. Subvector inference has been studied extensively in the current literature. Methods for constructing asymptotically uniformly valid subvector inference in different contexts have been proposed including Bounds tests, the least favorable approach, and tests based on confidence sets for nuisance parameters, see Section 4.3.2 in Silvapulle and Sen (2005) for a brief discussion of all three approaches.[6] Among these proposals, the two-stage approach based on confidence sets for nuisance parameters and a Bonferroni-type correction has proven to perform well.

There are several works that adopt this approach. Berger and Boos (1994) and Silvapulle (1996) study some specific parametric testing problems. In a single-equation instrumental variables regression with possibly "weak" instrumental variables, Staiger and Stock (1997) construct a confidence region for the parameters based on such method. Romano and Wolf (2000) construct a confidence interval for a univariate mean that has finite sample validity. For moment equality models with overidentifying inequality moment conditions, Moon and Schorfheide (2009) propose asymptotically uniformly valid tests and confidence sets for the parameters of interest. Chernozhukov et al. (2013) construct confidence intervals for marginal effects in non-linear panel data models. For testing a finite number of moment inequalities, Romano et al. (2014) construct asymptotically uniformly valid confidence sets for parameters characterized by the moment inequalities. Finally, McCloskey (2017) considers general non-standard testing problems in which the asymptotic distribution of a test statistic is discontinuous in a nuisance parameter under the null hypothesis. We refer interested readers to Romano et al. (2014) and McCloskey (2017) for other related works using similar two-step approaches.

The test statistics in Group II originate from the score statistics in Silvapulle and Silvapulle (1995), Andrews (2001), and Silvapulle and Sen (2005), which focus on the point null hypothesis on the subvector of the parameter. In this special case, the matrix R in the null hypothesis is an identity matrix. Therefore the asymptotic distributions of the test statistics do not involve computing the projection of a polytope. This paper extends their tests to any parameter space and null hypothesis of the forms (1) and (2).

[6] Wolak (1987, 1989, 1991) develops tests for the null hypothesis of inequality constraints based on the least favorable approach. Silvapulle and Sen (2005) provide a comprehensive and systematic treatment of constrained inference via the least favorable approach.

Organization of the Rest of This Paper The rest of this paper is organized as follows. In Section 2 we introduce the set-up, test statistics, assumptions, and asymptotic distribution of the extremum estimator. In Section 3, we introduce the concept of an implicit nuisance parameter and our algorithm for identifying it. In Section 4, we first provide a detailed construction and technical treatment of an asymptotically uniformly valid Wald test for the subvector hypothesis. We then extend our result to the null hypothesis of linear equality constraints of the general form. Sections 5 and 6 develop QLR and score tests respectively. Section 7 studies the local power of all three tests. In Section 8, we present the remaining three tests and develop the equivalence results among tests studied in the paper. Section 9 reports results from a simulation study and applies the methodology developed in the paper to an empirical research question. The last section offers some concluding remarks and extensions. Technical proofs are collected in Appendix S.1. Appendix S.2 provides primitive conditions for the assumptions discussed in the paper for the linear regression model.

Notation All limits are taken as $n \to \infty$. For two vectors $v, u \in \mathbb{R}^l$, $v \geq u$ means that $v_j \geq u_j$ for $j = 1, \ldots, l$; and $\|v\|$ denotes the Euclidean norm of v. The sets $\mathbb{R}^l_{>0}$ and $\mathbb{R}^l_{\geq 0}$ denote $\left\{v \in \mathbb{R}^l : v > \mathbf{0}\right\}$ and $\left\{v \in \mathbb{R}^l : v \geq \mathbf{0}\right\}$ respectively. For A being any subset of a Euclidean space or some metric space, we use $\overline{A}$ to denote its closure. For any two subsets A and B of a Euclidean space, the Hausdorff distance is defined as

$$d_H(A, B) \equiv \max\left(\sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|a - b\|\right).$$

2 The Model and Test Statistics

In this section, we introduce our model taken from Andrews (1999) and the three test statistics that will be studied in detail in the paper. Section 8 presents the remaining three tests.

Let $l_n(\theta)$ denote the estimator objective function that depends on the data when the sample size is n for $n = 1, 2, \ldots$. The parameter space Θ is defined in (1), and takes the form of a convex polytope. The matrix $\mathscr{R}_w$ is allowed to be *row rank deficient* to incorporate constraints like $0 \leq \theta \leq 1$.

2.1 An Extremum Estimator and Asymptotic Distribution

An extremum estimator denoted as $\widehat{\theta}$ satisfies: $\widehat{\theta} \in \Theta$ and

$$l_n\left(\widehat{\theta}\right) = \sup_{\theta\in\Theta} l_n(\theta) + o_p(1). \tag{4}$$

Let $\theta^* \in \Theta$ denote the pseudo-true value of the parameter θ. The estimator objective function $l_n(\theta)$ has a quadratic expansion in θ around θ^*:

$$\begin{aligned} l_n(\theta) = {} & l_n(\theta^*) + Dl_n(\theta^*)(\theta-\theta^*) \\ & + \frac{1}{2}(\theta-\theta^*)' D^2 l_n(\theta^*)(\theta-\theta^*) + R_n(\theta), \end{aligned} \tag{5}$$

where $R_n(\theta)$, $Dl_n(\theta^*)$, and $D^2 l_n(\theta^*)$ satisfy the following assumptions:

Assumption 2.1. *For all $0 < \kappa < \infty$, $\sup_{\theta\in\Theta:\|b_n(\theta-\theta^*)\|<\kappa} |R_n(\theta)| = o_p(1)$ for some scalar constants $\{b_n : n \geq 1\}$ satisfying $b_n \to \infty$;*

Assumption 2.2. *$\left(b_n^{-1} Dl_n(\theta^*), \mathscr{T}_n\right) \xrightarrow{d} (G, \mathscr{T})$ for some random variables $G \in \mathbb{R}^l$ and $\mathscr{T} \in \mathbb{R}^{l\times l}$, where $\mathscr{T}_n \equiv -b_n^{-2} D^2 l_n(\theta^*)$ and $\mathscr{T}$ is symmetric and non-singular with probability one.*

We further impose an assumption on the convergence rate of $\widehat{\theta}$.

Assumption 2.3. *$b_n\left(\widehat{\theta} - \theta^*\right) = O_p(1)$.*

The above assumptions do not rule out the case where $l_n(\cdot)$ is non-differentiable at θ^*. When θ^* is on the boundary of Θ and the estimator objective function is not defined outside the parameter space, $Dl_n(\theta^*)$ could contain left or right partial derivatives. Andrews (1997, 1999) offers detailed discussions on the assumptions and provides sufficient conditions for them to hold. Note that instead of a general normalizing matrix denoted as B_n in Andrews (1997, 1999), we adopt the special form that $B_n = b_n I_{l\times l}$ as in Assumption 5* in Andrews (1999). This simplifies asymptotic distribution of the extremum estimator under drifting sequences essential to the construction of asymptotically uniformly valid tests. As stated in Andrews (1999), such form of the normalizing matrix is applicable to most cases with non-trending data for which $b_n = \sqrt{n}$, although it is not applicable in dynamic models with deterministic and/or stochastic trends such as the Dickey-Fuller Regression in Andrews (1999) or the GARCH (1,q*) example in Andrews (1997).

Let $Z_n \equiv \mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)$. The quadratic expansion can be alternatively expressed as

$$l_n(\theta) = l_n(\theta^*) + \frac{1}{2} Z_n' \mathscr{T}_n Z_n - \frac{1}{2} q_n(b_n(\theta - \theta^*)) + R_n(\theta), \text{ where}$$

$$q_n(\lambda) \equiv (\lambda - Z_n)' \mathscr{T}_n (\lambda - Z_n) \text{ for } \lambda \in \mathbb{R}^l.$$

Under Assumptions 2.1-2.3, the lemma below follows from Theorem 3 (a) in Andrews (1999).

Lemma 2.1. *Suppose Assumptions 2.1-2.3 hold. Then*

$$b_n\left(\widehat{\theta} - \theta^*\right) \xrightarrow{d} \arg\min_{\lambda} \left[q(\lambda) + \phi_\theta(\lambda)\right],$$

where $q(\lambda) = (\lambda - Z)' \mathscr{T} (\lambda - Z)$, $Z = \mathscr{T}^{-1} G$, *and*

$$\phi_\theta(\lambda) = \begin{cases} 0, & \text{if } \mathscr{R}_e \lambda = \mathbf{0} \text{ and } \mathscr{R}_{w,b} \lambda \geq \mathbf{0} \\ \infty, & \text{otherwise} \end{cases},$$

for $\mathscr{R}_{w,b}$ *being the submatrix of* $\mathscr{R}_w$ *composed of rows corresponding to the binding inequalities in* $\mathscr{R}_w \theta^* \geq \mathrm{r}_w$.

For the parameter space Θ defined by linear equalities and inequalities in (1), it is straightforward to show that the expression in Lemma 2.1 is the same as that in Theorem 3 (a) in Andrews (1999). The asymptotic distribution of $\widehat{\theta}$ depends on the binding inequalities in $\mathscr{R}_w \theta^* \geq \mathrm{r}_w$, and thus is discontinuous in $\mathscr{R}_w \theta^*$ at r_w.[7]

The model includes several well known examples in the literature. We present a few below adopted from Andrews (1997).

Example 2.1. [Linear Regression] The model is expressed as

$$Y_i = X_i' \theta^* + \varepsilon_i, \text{ for } i = 1, ..., n,$$

where $(X_i, Y_i)_{i=1}^n$ is the random sample and $E(\varepsilon_i | X_i) = 0$. The objective function $l_n(\cdot)$ is expressed as

$$l_n(\theta) = -\frac{1}{2} \sum_{i=1}^n (Y_i - X_i' \theta)^2.$$

[7]Different descriptions of Θ may result in different matrices $\mathscr{R}_e$ and $\mathscr{R}_{w,b}$. However, the set where $\phi_\theta(\cdot)$ equals zero is independent of the description, see Lemma S.1.6 in Appendix S.1.

We have $b_n = \sqrt{n}$, $Dl_n(\theta) = \sum_{i=1}^n (Y_i X_i - X_i X_i'\theta)$, and $D^2 l_n(\theta) = -\sum_{i=1}^n X_i X_i'$.

Example 2.2. [Generalized Method of Moments] Let $g(Z, \theta)$ be a vector of known functions of the random variable Z, which is allowed to be non-differentiable. The moment equations $E[g(Z, \theta^*)] = \mathbf{0}$ hold at θ^*. With the random sample $(Z_i)_{i=1}^n$, the sample moment functions are computed as $\frac{1}{n}\sum_{i=1}^n g(Z_i, \theta)$. For some positive definite weighting matrix Σ_n, $l_n(\cdot)$ is defined as

$$l_n(\theta) = -n\left(\frac{1}{n}\sum_{i=1}^n g(Z_i, \theta)\right)' \Sigma_n \left(\frac{1}{n}\sum_{i=1}^n g(Z_i, \theta)\right).$$

More discussion on the model can be found in Pakes and Pollard (1989) and Andrews (1997).

2.2 The Test Statistics and Asymptotic Size

The Wald and QLR test statistics take the standard forms:

$$W_n \equiv b_n^2 \left(R\widehat{\theta} - r\right)' (R\Sigma_{W,n}R')^{-1} \left(R\widehat{\theta} - r\right)$$

for some positive definite weighting matrix $\Sigma_{W,n}$ and

$$QLR_n \equiv -2\left(l_n\left(\widehat{\theta}_0\right) - l_n\left(\widehat{\theta}\right)\right),$$

where $\widehat{\theta}_0$ is a restricted (by H_0) estimator such that: $\widehat{\theta}_0 \in \Theta_0$ and

$$l_n\left(\widehat{\theta}_0\right) = \sup_{\theta \in \Theta_0} l_n(\theta) + o_p(1).$$

To account for parameters on the boundary of the parameter space, we adopt the following extension of the score test statistic introduced in Andrews (2001). It is defined as a quadratic form in the directed score. For any $\theta \in \Theta$, we call $Dl_n(\theta)$ the score function such that

$$Dl_n(\theta) = Dl_n(\theta^*) + D^2 l_n(\theta^*)(\theta - \theta^*) + R_n^D(\theta), \tag{6}$$

where $Dl_n(\theta^*)$ and $D^2 l_n(\theta^*)$ are defined in (5), and $R_n^D(\theta)$ is the remainder term satisfying Assumption 6.3 (i) in Section 6. We do not require $l_n(\theta)$ to have pointwise partial derivative with respect to θ; when it does, $Dl_n(\theta)$ equals the vector of pointwise partial derivative of $l_n(\theta)$ with respect to θ. Define the directed score ds_n as

$$\widehat{q}_R(ds_n) = \inf_{\lambda_R \in b_n(R\Theta - r)} \widehat{q}_R(\lambda_R) + o_p(1), \tag{7}$$

where

$$\widehat{q}_R(\cdot) \equiv \left(\cdot - R\widehat{\mathscr{T}}_n^{-1} b_n^{-1} Dl_n\left(\widehat{\theta}_0\right)\right)' \left(R\widehat{\mathscr{T}}_n^{-1} R'\right)^{-1} \left(\cdot - R\widehat{\mathscr{T}}_n^{-1} b_n^{-1} Dl_n\left(\widehat{\theta}_0\right)\right),$$

in which the matrix $\widehat{\mathscr{T}_n}$ is assumed to approximate $\mathscr{T}_n$. The score test statistic is defined as

$$S_n \equiv ds_n' \Sigma_{S,n}^{-1} ds_n, \tag{8}$$

where the weighting matrix $\Sigma_{S,n}$ is positive definite.

Let T_n denote one of the above test statistics and the ones introduced in Section 8. We now introduce the concept of asymptotic size of a test based upon T_n. Suppose the model of interest is fully characterized by the finite dimensional parameter $\theta^* \in \Theta$ and the infinite dimensional parameter $\psi^* \in \Psi$ characterizing the distribution of the data. The space Ψ can be restricted to be some compact metric space with a metric that induces weak convergence, see Andrews et al. (2011). Let $\omega \equiv (\theta^*, \psi^*) \in \mathcal{W}$; denote $\boldsymbol{P}_\omega$ as the probability model indexed by ω and $\Pr_\omega$ as the probability computed with respect to $\boldsymbol{P}_\omega$. Let $\mathcal{W}_0$ be the collection of elements $\omega \in \mathcal{W}$ consistent with the null hypothesis and CV_n be a (possibly) sample dependent critical value. The asymptotic size of the resulting test is defined by

$$AsySz(T_n, CV_n) \equiv \limsup_{n \to \infty} \sup_{\omega \in \mathcal{W}_0} \Pr_\omega(T_n > CV_n).$$

In the following sections, we construct critical values for each of the test statistics W_n, QLR_n, and S_n that control the asymptotic size of the resulting tests and study their local power.

3 Undetermined Inequality and Implicit Nuisance Parameter

The critical step in developing asymptotically uniformly valid tests for H_0 is to identify binding, non-binding, and undetermined inequalities in $\mathscr{R}_w \theta^* \geq \mathrm{r}_w$ under H_0.

Since the null parameter space Θ_0 contains all the available information on θ^*, identifying different types of inequalities in $\mathscr{R}_w\theta^* \geq \mathrm{r}_w$ is equivalent to identifying *implicit equalities, strictly redundant inequalities*, and *undetermined inequalities* in the null parameter space Θ_0, where

$$\Theta_0 = \left\{\theta \in \mathbb{R}^l : R\theta = r,\, \mathscr{R}_e\theta = \mathrm{r}_e, \text{ and } \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}.$$

This section proposes an approach for accomplishing this task and introduces the concept of an *implicit nuisance parameter*. Specifically, for any $\theta \in \Theta_0$, our approach identifies inequalities in $\mathscr{R}_w\theta \geq \mathrm{r}_w$ of three types: those that are known to bind (implicit equalities); those that are known not to bind (strictly redundant inequalities); and those that are undetermined.

3.1 An Algorithm for Identifying Implicit Equalities, Strictly Redundant Inequalities, and Undetermined Inequalities

We first incorporate information in equalities: $\mathscr{R}_e\theta = \mathrm{r}_e$ and $R\theta = r$ in $\mathscr{R}_w\theta \geq \mathrm{r}_w$ to obtain a new system of linear inequalities. Let Γ and γ be such that the full set of basic solutions to the system of linear equations:

$$\begin{pmatrix} \mathscr{R}_e \\ R \end{pmatrix} \theta = \begin{pmatrix} \mathrm{r}_e \\ r \end{pmatrix} \tag{9}$$

is expressed by $\theta = \Gamma\theta_f + \gamma$, where θ_f is a vector of l_f free parameters, Γ is a $l \times l_f$ matrix, and γ is a l-vector. Then under H_0, the system of linear inequalities in θ: $\mathscr{R}_w\theta \geq \mathrm{r}_w$ becomes the system of linear inequalities in θ_f:

$$\mathscr{R}_w\Gamma\theta_f \geq \mathrm{r}_w - \mathscr{R}_w\gamma. \tag{10}$$

Let $\eta \equiv \mathscr{R}_w\Gamma\theta_f^*$. After incorporating the information in H_0 and $\mathscr{R}_e\theta = \mathrm{r}_e$, some inequalities in (10) will be known to bind and some will be known not to bind. To distinguish among these three types of inequalities, we decompose η into η^b, η^{nb}, and η^u composed of rows of η, such that the inequalities given by η^b in (10) are known to bind, the inequalities given by η^{nb} are known not to bind, and finally the inequalities given by η^u are undetermined.

For systems of weak linear inequalities, Telgen (1983) introduces implicit equalities and strictly redundant inequalities and develops efficient algorithms STREINQ and IMPLEQ for finding them. For the system of weak inequalities (10), define its feasible set

$$W \equiv \left\{\theta_f \in \mathbb{R}^{l_f} : \mathscr{R}_w \Gamma \theta_f \geq \mathrm{r}_w - \mathscr{R}_w \gamma\right\}.$$

Let $\mathfrak{J} \equiv \{1, ..., l_w\}$ and the subscript (j) denote the jth row of a matrix or a vector. For any $j \in \mathfrak{J}$, let

$$W_j \quad \equiv \quad \left\{\theta_f \in \mathbb{R}^{l_f} : (\mathscr{R}_w \Gamma)_{(m)} \theta_f \geq (\mathrm{r}_w - \mathscr{R}_w \gamma)_{(m)}, \forall m \neq j, m \in \mathfrak{J}\right\}$$

Definition 3.1. *In the system of inequalities (10), for a given $j \in \mathfrak{J}$, the inequality: $(\mathscr{R}_w \Gamma)_{(j)} \theta_f \geq (\mathrm{r}_w - \mathscr{R}_w \gamma)_{(j)}$ is an implicit equality if $(\mathscr{R}_w \Gamma)_{(j)} \theta_f = (\mathrm{r}_w - \mathscr{R}_w \gamma)_{(j)}$ for all $\theta_f \in W$ and is strictly redundant if $(\mathscr{R}_w \Gamma)_{(j)} \theta_f > (\mathrm{r}_w - \mathscr{R}_w \gamma)_{(j)}$ for all $\theta_f \in W_j$.*

The following steps identify the collections of implicit equalities and strictly redundant inequalities among (10). Denote $u_j(\theta_f) \equiv (\mathscr{R}_w \Gamma)_{(j)} \theta_f - (\mathrm{r}_w - \mathscr{R}_w \gamma)_{(j)}$.

Step a. Identify the implicit equalities in (10) as

$$Sub_b \equiv \{j \in \mathfrak{J} : \max\{u_j(\theta_f) : \theta_f \in W_j\} = 0\};$$

Step b. Identify the strictly redundant inequalities in (10) as

$$Sub_{nb} \equiv \{j \in \mathfrak{J} \backslash Sub_b : \min\{u_j(\theta_f) : \theta_f \in W_j\} > 0\}.$$

Let $\mathscr{R}_w^b \in \mathbb{R}^{l_b \times l}$ ($\mathscr{R}_w^{nb} \in \mathbb{R}^{l_{nb} \times l}$) denote the submatrix of $\mathscr{R}_w$ consisting of rows with indices in Sub_b (Sub_{nb}). Denote $\mathscr{R}_w^u \in \mathbb{R}^{l_u \times l}$ as the submatrix of $\mathscr{R}_w$ consisting of rows that are not in Sub_b or Sub_{nb}; and let r_w^u be the corresponding subvector of r_w. Then $\eta^b = \mathscr{R}_w^b \Gamma \theta_f^*$, $\eta^{nb} = \mathscr{R}_w^{nb} \Gamma \theta_f^*$, $\eta^u = \mathscr{R}_w^u \Gamma \theta_f^*$, and the undetermined inequalities are $\eta^u \geq \mathrm{r}_w^u - \mathscr{R}_w^u \gamma$.

3.2 Implicit Nuisance Parameter

We now introduce the concept of an implicit nuisance parameter.

Definition 3.2. *An implicit nuisance parameter, denoted as η^k, is defined as a subvector of η^u corresponding to a row basis of $\mathscr{R}_w^u \Gamma$.*

We call η^k an implicit nuisance parameter, because it is in general a linear combination instead of a subvector of the original parameter θ^*. By definition, an implicit nuisance parameter is $\eta^k = \mathscr{R}_\Gamma^u \theta_f^*$, where $\mathscr{R}_\Gamma^u$ is a submatrix of $\mathscr{R}_w^u \Gamma$ with rows forming a row basis of $\mathscr{R}_w^u \Gamma$. When $\mathscr{R}_w^u \Gamma$ is of *full row rank*, $\mathscr{R}_\Gamma^u = \mathscr{R}_w^u \Gamma$ and the implicit nuisance parameter is $\eta^k = \eta^u$. When $\mathscr{R}_w^u \Gamma$ is *not of full row rank* with rank denoted as l_k, we compute $\mathscr{R}_\Gamma^u$ and $\Gamma^u \in \mathbb{R}^{l_u \times l_k}$ such that $\mathscr{R}_w^u \Gamma = \Gamma^u \mathscr{R}_\Gamma^u$ by Gauss-Jordan elimination on the transpose of $\mathscr{R}_w^u \Gamma$. In terms of the implicit nuisance parameter, the undetermined inequalities in (10) become:

$$\Gamma^u \eta^k \geq \mathrm{r}_w^u - \mathscr{R}_w^u \gamma. \tag{11}$$

Remark 3.1. Given Θ_0 and $\theta \in \Theta_0$, the undetermined inequalities among $\mathscr{R}_w \theta \geq \mathrm{r}_w$ are unique. On the other hand, the implicit nuisance parameter may not be unique. Although the dimensions of θ_f and η^k are uniquely determined by Θ_0, free parameters in (9) and row bases of $\mathscr{R}_w^u \Gamma$ are not unique.

We emphasize that for testing H_0 against H_1, the algorithm in this section needs to be implemented only once regardless of the model, estimator and test statistic. Once implicit equalities, strictly redundant inequalities, undetermined inequalities, and the implicit nuisance parameter in Θ_0 are identified using our algorithm, one can adapt any existing two-stage approach for subvector inference to construct uniform tests for H_0 in any model and via any test statistic.

In the rest of this paper, we demonstrate this by constructing Wald, QLR, and score tests for H_0 in the model in Section 2. Since this paper focuses on studying the effect of inequality constraints in Θ on inference, especially on the discontinuity of the null asymptotic distribution of the chosen test statistic caused by the inequality constraints, we impose the following assumption throughout the rest of this paper.

Assumption 3.1. *The distribution of $(G, \mathscr{T})$ is not discontinuous in any unknown parameters.*

Typically, G is a Gaussian distribution with zero mean, and $\mathscr{T}$ is a deterministic, symmetric and non-singular matrix.

3.3 Examples

Our approach for identifying an implicit nuisance parameter is applicable to any $\mathscr{R}_e$, $\mathscr{R}_w$, and R. We illustrate this by the following examples.

Example 3.1. Let $H_0 : \mathscr{R}_w\theta^* = \mathrm{r}_w$ and $H_1 : \mathscr{R}_w\theta^* \geq \mathrm{r}_w$. This is the hypothesis studied in Gourieroux et al. (1982) for linear regression model. Such hypotheses can be alternatively expressed as $H_0 : \mathscr{R}_w\theta^* = \mathrm{r}_w$ against $H_1 : \mathscr{R}_w\theta^* \neq \mathrm{r}_w$ for $\Theta = \left\{\theta \in \mathbb{R}^l : \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}$. It is apparent that all the inequalities in $\mathscr{R}_w\theta^* \geq \mathrm{r}_w$ are binding under the null. There is no $\mathscr{R}_w^{nb}$ or $\mathscr{R}_w^{u}$, and thus no implicit nuisance parameter.

Example 3.2. Let $H_0 : (R_1', \mathscr{R}_w')'\,\theta^* = (r_1', \mathrm{r}_w')'$ and $H_1 : (R_1', \mathscr{R}_w')'\,\theta^* \neq (r_1', \mathrm{r}_w')'$ for $\Theta = \left\{\theta \in \mathbb{R}^l : \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}$. Similar to Example 3.1, all the inequalities among $\mathscr{R}_w\theta^* \geq \mathrm{r}_w$ are binding under the null. Kodde and Palm (1986) and Silvapulle and Sen (2005) both consider such case in general parametric models.

Example 3.3. Consider testing $H_0 : \mathscr{R}_{w,1}\theta^* = r$ against $H_1 : \mathscr{R}_{w,1}\theta^* \neq r$ for $\Theta = \left\{\theta \in \mathbb{R}^l : \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}$, where $\mathscr{R}_w = \left(\mathscr{R}_{w,1}', \mathscr{R}_{w,2}'\right)'$. The set Θ_0 is expressed as

$$\Theta_0 = \left\{\theta \in \mathbb{R}^l : \mathscr{R}_{w,1}\theta = r, \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}.$$

For the first J inequalities in $\mathscr{R}_w\theta \geq \mathrm{r}_w$, we can easily determine whether they are binding or not binding by comparing values of elements in r with that of $\mathrm{r}_{w,1}$, where $\mathrm{r}_w = \left(\mathrm{r}_{w,1}', \mathrm{r}_{w,2}'\right)'$. Since the null hypothesis contains information about θ^*, the algorithm in Section 3 can be applied to identify inequalities in $\mathscr{R}_{w,2}\theta^* \geq \mathrm{r}_{w,2}$ of three types and an implicit nuisance parameter. The subvector hypothesis extensively studied in the current literature is a special case of H_0 when

$$\mathscr{R}_w = \begin{pmatrix} \mathscr{R}_{w,1} \\ \mathscr{R}_{w,2} \end{pmatrix} = \begin{pmatrix} \mathscr{R}_{w,11} & \mathbf{0} \\ \mathbf{0} & \mathscr{R}_{w,22} \end{pmatrix}$$

and $\mathscr{R}_{w,22}$ has full row rank. In this case, the implicit nuisance parameter $\eta^k = \mathscr{R}_{w,22}\theta_2$, where $\theta = (\theta_1', \theta_2')'$ is decomposed conformably.

Example 3.4. Let $l = 8$, $\theta = (\theta_1, \ldots, \theta_8)'$,

$$\begin{aligned}\Theta = \{&\theta : \theta_1 \geq 0, -\theta_1 \geq -1, \theta_1 + \theta_2 \geq -1, \theta_2 + \theta_3 \geq 0, \theta_3 + \theta_4 \geq -1, \\ &\theta_5 + \theta_7 - \theta_8 \geq 0, 2\theta_5 + \theta_6 + \theta_7 \geq 0, \theta_5 - \theta_6 + 2\theta_7 - 3\theta_8 \geq 0\},\end{aligned}$$

and $H_0 : \theta_2^* = \theta_3^* = \theta_4^* = 0$. This is a non-subvector hypothesis, with

$$\mathscr{R}_w = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 2 & -3 \end{pmatrix} \text{ and } \mathrm{r} = \begin{pmatrix} 0 \\ -1 \\ -1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

One gets $\theta_f = (\theta_1, \theta_5, \theta_6, \theta_7, \theta_8)'$ and $\gamma = \mathbf{0}$. Since

$$\begin{aligned} (\mathscr{R}_w\Gamma)_{(3)}\,\theta_f - (\mathrm{r}_w - \mathscr{R}_w\gamma)_{(3)} &= \theta_1 + 1 > 0 \text{ for } \theta_f \in S_3, \\ (\mathscr{R}_w\Gamma)_{(4)}\,\theta_f - (\mathrm{r}_w - \mathscr{R}_w\gamma)_{(4)} &= 0 + 0 = 0 \text{ for } \theta_f \in S_4, \text{ and} \\ (\mathscr{R}_w\Gamma)_{(5)}\,\theta_f - (\mathrm{r}_w - \mathscr{R}_w\gamma)_{(5)} &= 0 + 1 > 0 \text{ for } \theta_f \in S_5, \end{aligned}$$

we obtain that $Sub_b = \{4\}$, $Sub_{nb} = \{3, 5\}$, and

$$\mathscr{R}_w^u\Gamma = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 2 & 1 & 1 & 0 \\ 0 & 1 & -1 & 2 & -3 \end{pmatrix}.$$

Since the rows of $R^u\Gamma$ are linearly dependent, we proceed to the next step to find the nuisance parameter η^k.

The Gauss-Jordan elimination on the transpose of $\mathscr{R}_w^u\Gamma$ provides us that

$$(\mathscr{R}_w^u\Gamma)' = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & 0 & -3 \end{pmatrix} \xrightarrow{\text{Gauss-Jordan elimination}} \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We obtain that the first, third, and fourth rows of $\mathscr{R}_w^u\Gamma$ constitute a row basis of

$\mathscr{R}_w^u \Gamma$. Further, we get that

$$\Gamma^u = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 3 & -1 \end{pmatrix} \text{ and } \eta^k = \begin{pmatrix} \theta_1^* \\ \theta_5^* + \theta_7^* - \theta_8^* \\ 2\theta_5^* + \theta_6^* + \theta_7^* \end{pmatrix}.$$

The number of nuisance parameters is smaller than the dimension of $\mathscr{R}_w^u \Gamma \theta_f$.

Example 3.5. We introduce the following parameter space Θ that will be used in the simulation study in Section 9.1 and in the empirical application in Section 9.2. Let $\Theta = \{\theta \in \mathbb{R}^4 : \mathscr{R}_w \theta \geq \mathbf{0}\}$ with

$$\mathscr{R}_w = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix},$$

and the null hypothesis be $H_0 : R\theta^* = r$ with

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \text{ and } r = \begin{pmatrix} 0 \\ 0.1 \end{pmatrix}.$$

Solving $R\theta = r$, we get $\theta = \Gamma\theta_f + \gamma$, where

$$\Gamma = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \theta_f = \begin{pmatrix} \theta_2 \\ \theta_3 \end{pmatrix}, \text{ and } \gamma = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0.1 \end{pmatrix}.$$

Then $\mathscr{R}_w \theta \geq \mathbf{0}$ becomes $\mathscr{R}_w \Gamma \theta_f \geq \mathbf{0} - \mathscr{R}_w \gamma$ for

$$\mathscr{R}_w \Gamma = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ -1 & 1 \\ 1 & -1 \end{pmatrix} \text{ and } \mathbf{0} - \mathscr{R}_w \gamma = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -0.1 \end{pmatrix}.$$

By definition, we have $Sub_b = \{1\}$ and $Sub_{nb} = \emptyset$.

Applying the Gauss-Jordan elimination on the transpose of $\mathscr{R}_w^u \Gamma$:

$$\left(\mathscr{R}_w^u \Gamma\right)' = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \overset{\text{Gauss-Jordan elimination}}{\longrightarrow} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix},$$

we obtain $\mathscr{R}_\Gamma^u$ as the first and second row of $\mathscr{R}_w^u \Gamma$,

$$\Gamma^u = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, \text{ and } \eta^k = \mathscr{R}_\Gamma^u \theta_f^* = \begin{pmatrix} \theta_2^* \\ -\theta_2^* + \theta_3^* \end{pmatrix}.$$

4 The Wald Test

In this section, we construct an asymptotically uniformly valid Wald test using the statistic W_n introduced in Section 2.2 and provide a detailed procedure for implementing it. The following assumption is on the weighting matrix.

Assumption 4.1. *$\Sigma_{W,n} \xrightarrow{p} \Sigma_W$ for which $R\Sigma_W R'$ is positive definite with probability one.*

The asymptotic distribution of W_n under H_0 is given in the Lemma below.

Lemma 4.1. *Under H_0 and Assumptions 2.1-2.3 and 4.1, it holds that*

$$W_n \xrightarrow{d} W \equiv (R\Psi)' \left(R\Sigma_W R'\right)^{-1} (R\Psi),$$

where $\Psi \equiv \arg\min_\lambda \left[q(\lambda) + \phi(\lambda)\right]$, in which $q(\lambda)$ is defined in Lemma 2.1 and

$$\phi(\lambda) = \begin{cases} 0, & \text{if } \mathscr{R}_e \lambda = \mathbf{0},\ \mathscr{R}_w^b \lambda \geq \mathbf{0} \text{ and } \mathscr{R}_{w,b}^u \lambda \geq \mathbf{0} \\ \infty, & \text{otherwise} \end{cases},$$

with $\mathscr{R}_{w,b}^u$ being the submatrix of $\mathscr{R}_w^u$ corresponding to the binding inequalities in (11).

Lemma 4.1 implies that the asymptotic distribution of W_n under H_0 depends on the implicit nuisance parameter η^k through the undetermined inequalities in Θ_0, i.e., (11). Moreover, it is discontinuous in the implicit nuisance parameter when it exists.

Example 3.1 (continued)**.** In this example, there is no implicit nuisance parameter and $\mathscr{R}_w^b = \mathscr{R}_w$. Applying Lemma 4.1, we obtain that

$$W_n \xrightarrow{d} W = \left(\mathscr{R}_w \Psi\right)' \left(\mathscr{R}_w \Sigma_W \mathscr{R}_w'\right)^{-1} \left(\mathscr{R}_w \Psi\right),$$

where $\Psi = \arg\min_{\mathscr{R}_w \lambda \geq \mathbf{0}} (\lambda - \mathscr{T}^{-1}G)' \mathscr{T} (\lambda - \mathscr{T}^{-1}G)$. If further $G \sim \mathcal{N}(\mathbf{0}, \mathscr{T})$ and $\Sigma_W = \mathscr{T}^{-1}$, then the distribution of W is of the same form as ξ_W in Gourieroux et al. (1982), which is proved by Gourieroux et al. (1982) to follow a weighted chi-squared distribution. For more discussions and derivations on the weighted chi-squared distribution, please refer to Bartholomew (1961), Kudo (1963), Nüesch (1966), and Perlman (1969).

In the rest of this paper, we focus on the case where the implicit nuisance parameter η^k exists. For clarity and to be self-contained, we first provide a detailed treatment in Section 4.1 of the subvector hypothesis of the form:

$$H_{0S} : \theta_1^* = r \text{ against } H_{1S} : \theta_1^* \neq r,$$

where $\theta_1^* \in \mathbb{R}^J$ is a subvector of θ^* such that $\theta^* = (\theta_1^{*\prime}, \theta_2^{*\prime})'$, under the maintained hypothesis that $\Theta = \left\{\theta \in \mathbb{R}^l : \theta \geq \mathbf{0}\right\}$. Then we extend it to the general H_0 in Section 4.2.

4.1 Subvector Hypothesis

For the subvector hypothesis H_{0S}, Example 3.3 shows that the implicit nuisance parameter is $\eta^k = \theta_2^*$. The Wald statistic is calculated by

$$W_n = b_n^2 \left(\widehat{\theta}_1 - r\right)' (R\Sigma_{W,n}R')^{-1} \left(\widehat{\theta}_1 - r\right), \text{ with } R = \begin{pmatrix} I_{J\times J} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Without loss of generality, assume that $r = (\mathbf{0}', \mathrm{r}_{nb}')'$, where $\mathrm{r}_{nb} \in \mathbb{R}_{>0}^{J-J_b}$. Applying Lemma 4.1 to this case with $\mathscr{R}_w = I_{l\times l}$ and $\mathrm{r}_w = \mathbf{0}$, we obtain $W_n \xrightarrow{d} W$ with

$$\phi(\lambda) = \begin{cases} 0, & \text{if } \lambda_j \geq 0 \text{ for } j = 1, \dots, J_b \text{ and } \lambda_j \geq 0 \text{ for } j \in \mathcal{J} \\ \infty, & \text{otherwise} \end{cases},$$

where elements in $\mathcal{J}$ are the indices corresponding to zero elements in θ_2^*. The asymptotic distribution of W_n is discontinuous in $\eta^k = \theta_2^*$ at $\mathbf{0}$, unless $\mathscr{T}$ is block diagonal between θ_1 and θ_2, see Andrews (2001).

4.1.1 The Null Asymptotic Distribution Under Drifting Sequences

We decompose the model parameter $\omega \in \mathcal{W}_0$ into three groups: $\left(\eta^k, \pi_W, \xi\right)$ based on their effects on the asymptotic distribution of W_n. $\pi_W \in \Pi_W$ contains parameters in G, $\mathscr{T}$, and Σ_W; and $\xi \in \Xi$ consists of all other parameters and is infinite dimensional. From the previous discussion, the null asymptotic distribution of W_n is discontinuous in η^k; π_W affects the limiting distribution of W_n but not its continuity; ξ doesn't affect the limiting distribution of W_n given η^k and π_W.

Following Andrews et al. (2011), Andrews and Cheng (2012), Andrews and Cheng (2014), and Cheng (2015), we establish the asymptotic distribution of W_n under drifting parameter sequences $\omega_n \in \mathcal{W}_0 \to \omega \in \overline{\mathcal{W}}_0$.[8] For brevity, throughout the paper the terminology "$\omega_n \in \mathcal{W}_0$" refers to "drifting parameter sequence $\omega_n \in \mathcal{W}_0$ with limit $\omega \in \overline{\mathcal{W}}_0$". Under the null hypothesis, $\theta_n = \left(r', \theta'_{2,n}\right)'$ has the limit $\theta_\omega = \left(r', \theta'_{2,\omega}\right)'$. Let $\overline{\mathbb{R}}_{\geq 0} \equiv \mathbb{R}_{\geq 0} \cup \{+\infty\}$. In particular, we consider the parameter sequence $\left\{\left(\eta^k_n, \pi_{W,n}, \xi_n\right) \in \mathbb{R}^{l-J}_{\geq 0} \times \Pi_W \times \Xi : n \geq 1\right\}$ and the localization parameter $(c, \pi_{W,\omega})$ as the limit of $b_n \eta^k_n$ and $\pi_{W,n}$:

$$b_n \eta^k_n = b_n \theta_{2,n} \to c \in \overline{\mathbb{R}}^{l-J}_{\geq 0} \text{ and } \pi_{W,n} \to \pi_{W,\omega} \in \overline{\Pi}_W.$$

Notice that this is a definition rather than an assumption, because elements in c are not required to be finite. As shown in the lemma below, the asymptotic distribution of W_n under the null hypothesis and the drifting parameter sequence $\left(\eta^k_n, \pi_{W,n}, \xi_n\right)$ depends on c and $\pi_{W,\omega}$; whereas ξ_n (or the limiting value ξ_ω of ξ_n) does not affect the limiting distribution under any parameter sequence η^k_n and $\pi_{W,n}$.

The estimator objective function $l_n(\theta)$ has a quadratic expansion in θ around θ_n:

$$\begin{aligned} l_n(\theta) = {} & l_n(\theta_n) + Dl_n(\theta_n)(\theta - \theta_n) \\ & + \frac{1}{2}(\theta - \theta_n)' D^2 l_n(\theta_n)(\theta - \theta_n) + R_n(\theta), \end{aligned}$$

where $R_n(\theta)$, $Dl_n(\theta_n)$, and $D^2 l_n(\theta_n)$ satisfy the following assumptions.

Assumption 4.2. *For any $\boldsymbol{P}_\omega$ with $\omega \in \mathcal{W}_0$, $\sup_{\theta \in \Theta : \|\theta - \theta_\omega\| < \kappa_n} |R_n(\theta)| = o_p(1)$ for all $\kappa_n = o(1)$.*

[8] We focus our discussion on the sequence of ω_n in the main text, and later relate the result under the full sequence to that under the subsequence in the proof using Lemma 2.1 in Andrews et al. (2011).

Assumption 4.3. *For any $\omega_n \in \mathcal{W}_0$, $(b_n^{-1} Dl_n(\theta_n), \mathscr{T}_n) \xrightarrow{d} (G_\omega, \mathscr{T}_\omega)$ for some random variables $G_\omega \in \mathbb{R}^l$ and $\mathscr{T}_\omega \in \mathbb{R}^{l \times l}$, where $\mathscr{T}_n \equiv -b_n^{-2} D^2 l_n(\theta_n)$ and $\mathscr{T}_\omega$ is symmetric and non-singular with probability one.*

The next assumption is on the convergence rate of $\widehat{\theta}$ under the drifting parameter sequence.

Assumption 4.4. *For any $\omega_n \in \mathcal{W}_0$, $b_n\left(\widehat{\theta} - \theta_n\right) = O_p(1)$.*

Assumption 4.2 is slightly stronger than Assumption 2.1 by requiring the quadratic approximation to be accurate in a small neighborhood of θ_ω for each model $\boldsymbol{P}_\omega$. Nevertheless, it is still an assumption on the local property of the objective function, because we do not need the remainder term to be small uniformly over the parameter space of θ. Assumptions 4.3 and 4.4 are also stronger than their counterparts in Section 2.1. They require the normalizing constants b_n to be the same for all $\omega_n \in \mathcal{W}_0$. Tools like Lindeberg-Feller Central Limit Theorem can be employed to verify these assumptions for which the existence of bounded higher moments is often enough. In Appendix S.2, we discuss primitive conditions for Assumptions 4.2-4.4 to hold in Example 2.1.

The following assumption is on the weighting matrix. It is satisfied if $\Sigma_{W,\omega}$ is positive definite with probability one.

Assumption 4.5. *For any $\omega_n \in \mathcal{W}_0$, $\Sigma_{W,n} \xrightarrow{p} \Sigma_{W,\omega}$ for which $R\Sigma_{W,\omega}R'$ is positive definite with probability one.*

The asymptotic null distribution of W_n for any $\omega_n \in \mathcal{W}_0$ is given in the following lemma.

Lemma 4.2. *(i) If Assumptions 4.2-4.4 hold, then under $H_{0S}: \theta_1^* = r$ and any $\omega_n \in \mathcal{W}_0$,*

$$b_n\left(\widehat{\theta} - \theta_n\right) \xrightarrow{d} \Psi_\omega \equiv \arg\min_\lambda \left[q_\omega(\lambda) + \phi_\omega(\lambda)\right],$$

where $q_\omega(\lambda) = (\lambda - Z_\omega)' \mathscr{T}_\omega (\lambda - Z_\omega)$, $Z_\omega = \mathscr{T}_\omega^{-1} G_\omega$, and

$$\phi_\omega(\lambda) = \begin{cases} 0, & \text{if } \lambda_j \geq 0 \text{ for } j = 1, \ldots, J_b \\ & \text{and } \lambda_{k+J} + c_k \geq 0 \text{ for } k = 1, \ldots, l - J \\ \infty, & \text{otherwise} \end{cases}.$$

(ii) If further Assumption 4.5 holds, then $W_n \xrightarrow{d} W_\omega \equiv (R\Psi_\omega)' \left(R\Sigma_{W,\omega}R'\right)^{-1} (R\Psi_\omega)$.

4.1.2 The Testing Procedure

As shown in Lemma 4.2, the null asymptotic distribution of W_n under the drifting sequence of distributions depends on the value of $(c, \pi_{W,\omega})$. Let $\mathcal{C}^W_{c,\pi_{W,\omega}}(1-\tau)$ denote the $(1-\tau)$ quantile of the distribution of W_ω given c and $\pi_{W,\omega}$. It may not have a closed form expression but can be simulated. Building on existing work, especially McCloskey (2017), we adopt the two-step approach with Bonferroni-type correction to construct an asymptotically uniformly valid test for subvector hypothesis H_{0S}.

The detailed process consists of the following steps.

Step 1. (i) Find the consistent estimator $\widehat{\pi}_W$ such that for any $\omega_n \in \mathcal{W}_0$, $\widehat{\pi}_W \xrightarrow{p} \pi_{W,\omega}$; (ii) Construct the confidence set $\widetilde{I}_\tau$ for c such that for any $\omega_n \in \mathcal{W}_0$, $\lim_{n\to\infty} \Pr_{\omega_n}\left(c \in \widetilde{I}_\tau\right) \geq \tau$.

Consistent estimator for $\pi_{W,\omega}$ is easy to obtain in general, because $\pi_{W,\omega}$ consists of the parameters in G_ω, $\mathscr{T}_\omega$, and $\Sigma_{W,\omega}$, which are usually variance covariance matrix and Hessian matrix of the limit of the objective function. We provide one way of constructing the confidence set $\widetilde{I}_\tau$ for c. Define the unrestricted extremum estimator for $\theta_{2,n}$ as $\widetilde{\theta}_2$ such that

$$l_n\left(r, \widetilde{\theta}_2\right) = \sup_{\theta_2 \in \mathbb{R}^{l-J}} l_n\left(r, \theta_2\right) + o_p\left(1\right).$$

Let $\widetilde{c} = b_n \widetilde{\theta}_2$. It can be shown that $\widetilde{c} \xrightarrow{d} c + \mathscr{T}^{-1}_{2,\omega} G_{2,\omega}$, where $G_{2,\omega}$ and $\mathscr{T}_{2,\omega}$ are the subvector of G_ω and submatrix of $\mathscr{T}_\omega$ corresponding to θ_2. Denote $ES\left(\tau\right)$ as the set such that $\Pr\left(\mathscr{T}^{-1}_{2,\omega} G_{2,\omega} \in ES\left(\tau\right)\right) \geq 1-\tau$. Since the parameter space for c is $\overline{\mathbb{R}}^{l-J}_{\geq 0}$, we obtain a confidence set $\widetilde{I}_\tau$ as $\widetilde{I}^k_\tau \cap \overline{\mathbb{R}}^{l-J}_{\geq 0}$, where $\widetilde{I}^k_\tau \equiv \widetilde{c} - \widetilde{ES}\left(\tau\right)$ and $\widetilde{ES}\left(\tau\right)$ is the set obtained using estimators of the parameters in $\mathscr{T}^{-1}_{2,\omega} G_{2,\omega}$ rather than true values. To ensure the non-emptiness of $\widetilde{I}^k_\tau \cap \overline{\mathbb{R}}^{l-J}_{\geq 0}$, $ES\left(\tau\right)$ may not be equal-tailed.

Step 2. We construct the α level Bonferroni critical value as

$$CV^W_n\left(\alpha, \tau\right) \equiv \sup_{c \in \widetilde{I}_{\alpha-\tau}} \mathcal{C}^W_{c,\widehat{\pi}_W}\left(1-\tau\right), \tag{12}$$

for some $0 \leq \tau \leq \alpha$.

The following two theorems show that the Wald test for H_{0S} has the correct asymptotic size and is consistent.

Theorem 4.1. *Under Assumptions 3.1 and 4.2-4.5, if W_ω is continuous at $\mathcal{C}^W_{c,\pi_{W,\omega}}(1-\tau)$*

for all $(c, \pi_{W,\omega}) \in \overline{\mathbb{R}}_{\geq 0}^{l-J} \times \overline{\Pi}_W$*, then it holds that* $AsySz\left(W_n, CV_n^W(\alpha,\tau)\right) \leq \alpha$*.*

The continuity assumption in Theorem 4.1 may restrict the range of τ. In the case that $\Theta = \{(\theta_1,\theta_2) \in \mathbb{R}^2 : \theta_1 \geq 0 \text{ and } \theta_2 \geq 0\}$ and $H_0 : \theta_1 = 1$, it is satisfied for all $\tau \in (0,1)$. With the same parameter space but $H_0 : \theta_1 = 0$, this assumption is satisfied for $\tau < 0.5$.

Theorem 4.2. *Under* H_1 *and Assumptions 2.3 and 4.1,* $\Pr\left(W_n > CV_n^W(\alpha,\tau)\right) \to 1$.

4.2 General Linear Hypothesis

We extend the subvector test developed in Section 4.1 to H_0 for any $\mathscr{R}_e$, $\mathscr{R}_w$, and R. By extracting linearly independent components of η^u, we consider the model parameters $\left(\eta^k, \pi_W, \xi\right)$, where $\eta^k \in \mathrm{H}^k \subseteq \mathbb{R}^{l_k}$ is the implicit nuisance parameter, $\pi_W \in \Pi_W$ consists of parameters in G, $\mathscr{T}$, and Σ_W, and $\xi \in \Xi$ contains all other parameters and is infinite dimensional. Similar to the discussion in Section 4.1, the asymptotic distribution of W_n is discontinuous in η^k; π_W affects the limiting distribution of W_n but not its continuity; ξ doesn't affect the limiting distribution of W_n. Let $\left(\eta_n^k, \pi_{W,n}, \xi_n\right)$ be the drifting model parameters. Since the implicit nuisance parameter η_n^k satisfies inequalities in (11), we consider localization parameter c such that

$$c \equiv \lim_{n\to\infty} b_n \left(\Gamma^u \eta_n^k - \left(\mathrm{r}_w^u - \mathscr{R}_w^u \gamma\right)\right) \in C \subseteq \overline{\mathbb{R}}_{\geq 0}^{l_u}, \tag{13}$$

where

$$C \equiv \left\{c \in \overline{\mathbb{R}}_{\geq 0}^{l_u} : \exists \eta_n^k \in \mathrm{H}^k \text{ and } c = \lim_{n\to\infty} b_n \left(\Gamma^u \eta_n^k - \left(\mathrm{r}_w^u - \mathscr{R}_w^u \gamma\right)\right)\right\}.$$

The limits of $\pi_{W,n}$ and ξ_n are denoted as $\pi_{W,\omega}$ and ξ_ω respectively.

Lemma 4.3. *(i) If Assumptions 4.2-4.4 hold, then under* $H_0 : R\theta^* = r$ *and any parameter sequence* $\left(\eta_n^k, \pi_{W,n}, \xi_n\right) \in \mathrm{H}^k \times \Pi_W \times \Xi$*,*

$$b_n\left(\widehat{\theta} - \theta_n\right) \xrightarrow{d} \Psi_\omega \equiv \arg\min_\lambda \left[q_\omega(\lambda) + \phi_\omega(\lambda)\right],$$

where $q_\omega(\lambda) = (\lambda - Z_\omega)' \mathscr{T}_\omega (\lambda - Z_\omega)$*,* $Z_\omega = \mathscr{T}_\omega^{-1} G_\omega$*, and*

$$\phi_\omega(\lambda) = \begin{cases} 0, & \text{if } \mathscr{R}_e \lambda = \mathbf{0},\ \mathscr{R}_w^b \lambda \geq \mathbf{0} \text{ and } \mathscr{R}_w^u \lambda + c \geq \mathbf{0} \\ \infty, & \text{otherwise} \end{cases};$$

(ii) If further Assumption 4.5 holds, then $W_n \xrightarrow{d} W_\omega \equiv (R\Psi_\omega)' (R\Sigma_{W,\omega} R')^{-1} (R\Psi_\omega)$.

The null asymptotic distribution of W_n stated in Lemma 4.3 suggests the following procedure for computing the critical value of our test.

Step 1. (i) Find the consistent estimator $\widehat{\pi}_W$ such that for any $\omega_n \in \mathcal{W}_0$, $\widehat{\pi}_W \to_p \pi_{W,\omega}$; (ii) Construct the confidence set $\widetilde{I}_\tau$ for c such that for any $\omega_n \in \mathcal{W}_0$, $\lim_{n\to\infty} \Pr_{\omega_n} \left(c \in \widetilde{I}_\tau \right) \geq \tau$.

$\pi_{W,\omega}$ is composed of the parameters in G_ω, $\mathscr{T}_\omega$, and $\Sigma_{W,\omega}$. It is usually straightforward to obtain the consistent estimator of $\pi_{W,\omega}$. The confidence set for c can be constructed by the following procedure. By definition, $\eta_n^k = \mathscr{R}_\Gamma^u \theta_{f,n}$. Denote $\widetilde{\theta}_f$ as the unrestricted extremum estimator for $\theta_{f,n}$:

$$l_n \left(\Gamma \widetilde{\theta}_f + \gamma \right) = \sup_{\theta_f \in \mathbb{R}^{l_f}} l_n (\Gamma \theta_f + \gamma) + o_p(1).$$

Applying Lemma 4.3, one can show that $b_n \left(\widetilde{\theta}_f - \theta_{f,n} \right) \xrightarrow{d} \mathscr{T}_{f,\omega}^{-1} G_{f,\omega}$, where $G_{f,\omega}$ and $\mathscr{T}_{f,\omega}$ are the subvector of G_ω and submatrix of $\mathscr{T}_\omega$ corresponding to θ_f. Thus $b_n \left(\mathscr{R}_\Gamma^u \widetilde{\theta}_{f,n} - \mathscr{R}_\Gamma^u \theta_{f,n} \right) \xrightarrow{d} \mathscr{R}_\Gamma^u \mathscr{T}_{f,\omega}^{-1} G_{f,\omega}$. Denote $ES(\tau)$ as set such that

$$\Pr \left(\mathscr{R}_\Gamma^u \mathscr{T}_{f,\omega}^{-1} G_{f,\omega} \in ES(\tau) \right) \geq 1 - \tau.$$

We obtain $\widetilde{I}_\tau^k$ as $b_n \mathscr{R}_\Gamma^u \widetilde{\theta}_{f,n} - \widetilde{ES}(\tau)$, where $\widetilde{ES}(\tau)$ is obtained by using estimators of parameters in $\mathscr{T}_{f,\omega}^{-1} G_{f,\omega}$ rather than true values. The confidence set $\widetilde{I}_\tau$ for c is calculated as

$$\widetilde{I}_\tau = \left\{ \mathrm{c} \in \mathbb{R}_{\geq 0}^{l_u} : \mathrm{c} = \Gamma^u \iota - b_n \left(\mathrm{r}_w^u - \mathscr{R}_w^u \gamma \right), \ \iota \in \widetilde{I}_\tau^k \right\}. \tag{14}$$

To ensure that $\widetilde{I}_\tau$ is non-empty, $ES(\tau)$ may not be equal-tailed.

Step 2. Compute the α level Bonferroni critical value as

$$CV_n^W(\alpha, \tau) \equiv \sup_{c \in \widetilde{I}_{\alpha-\tau}} \mathcal{C}_{c,\widehat{\pi}_W}^W (1 - \tau)$$

for some $0 \leq \tau \leq \alpha$, where $\mathcal{C}_{c,\pi_{W,\omega}}^W (1-\tau)$ is the $(1-\tau)$ quantile of W_ω in Lemma 4.3 given $(c, \pi_{W,\omega})$. For any given $(c, \pi_{W,\omega})$, the distribution W_ω may not have a closed form expression but can be simulated.

The following two theorems establish the asymptotic validity and consistency of the Wald test.

Theorem 4.3. *Assume that W_ω is continuous at $\mathcal{C}^W_{c,\pi_{W,\omega}}(1-\tau)$ for all $(c,\pi_\omega) \in C \times \overline{\Pi}_W$. Under Assumptions 3.1 and 4.2-4.5, it holds that $AsySz\left(W_n, CV^W_n(\alpha,\tau)\right) \leq \alpha$.*

Theorem 4.4. *Under H_1 and Assumptions 2.3 and 4.1,* $\Pr\left(W_n > CV^W_n(\alpha,\tau)\right) \to 1$.

5 The Quasi Likelihood Ratio Test

Recall that Θ_0 denotes the parameter space under the null hypothesis $H_0 : R\theta^* = r$ and is given by (3). The algorithm in Section 3.1 generates $\mathscr{R}^b_w$, $\mathscr{R}^{nb}_w$, and $\mathscr{R}^u_w$, which are submatrices of $\mathscr{R}_w$ corresponding to implicit equalities, strictly redundant inequalities, and undetermined inequalities. Partition r_w conformably into subvectors r^b_w, r^{nb}_w, and r^u_w. Since inequalities defined by $\mathscr{R}^{nb}_w$ are strictly redundant, the parameter space under H_0 can be rewritten as

$$\Theta_0 = \left\{\theta \in R^l : R\theta = r,\, \mathscr{R}_e\theta = \mathrm{r}_e,\, \mathscr{R}^b_w\theta = \mathrm{r}^b_w, \text{ and } \mathscr{R}^u_w\theta \geq \mathrm{r}^u_w\right\}.$$

We impose the following assumption on the convergence rate of the restricted estimator $\widehat{\theta}_0$ defined in Section 2.2. Primitive conditions for this assumption can be found in Andrews (1997).

Assumption 5.1. $b_n\left(\widehat{\theta}_0 - \theta^*_0\right) = O_p(1)$ *for some* $\theta^*_0 \in \Theta_0$.

We call θ^*_0 the pseudo-true value of θ in Θ_0. Under H_0, it holds that $\theta^*_0 = \theta^*$; while $\theta^*_0 \neq \theta^*$ under H_1.

The lemma below provides the asymptotic distribution of QLR_n under H_0. Definitions of $q(\cdot)$ and $\phi(\cdot)$ are the same as in Lemma 4.1.

Lemma 5.1. *Under H_0 and Assumptions 2.1-2.3 and 5.1, it holds that*

$$QLR_n \xrightarrow{d} QLR \equiv \min_\lambda \left[q(\lambda) + \phi_0(\lambda)\right] - \min_\lambda \left[q(\lambda) + \phi(\lambda)\right],$$

where

$$\phi_0(\lambda) = \begin{cases} 0, & \text{if } \left(R', \mathscr{R}'_e, \mathscr{R}^{b\prime}_w\right)'\lambda = \mathbf{0} \text{ and } \mathscr{R}^u_{w,b}\lambda \geq \mathbf{0} \\ \infty, & \text{otherwise} \end{cases}.$$

Note that $\phi(\cdot)$ and $\phi_0(\cdot)$ differ in two parts. First, $\phi_0(\lambda)$ contains equalities $R\lambda = 0$, because Θ_0 is defined under the null hypothesis. Second, inequalities $\mathscr{R}^b_w\lambda \geq \mathbf{0}$ in $\phi(\lambda)$ become equalities $\mathscr{R}^b_w\lambda = \mathbf{0}$ in $\phi_0(\lambda)$. The null hypothesis allows us to determine some binding inequalities, which are represented by $\mathscr{R}_w^b$. On the other hand, inequalities $\mathscr{R}_w^b\theta \geq \mathrm{r}_w^b$ serve as equality constraints $\mathscr{R}_w^b\theta = \mathrm{r}_w^b$ when computing $\widehat{\theta}_0$ and $l_n\left(\widehat{\theta}_0\right)$.

Example 3.1 (continued)**.** There is no $\mathscr{R}_e$ or $\mathscr{R}_{w,b}^u$, and $R = \mathscr{R}_w^b$. Lemma 5.1 shows that

$$QLR_n \xrightarrow{d} QLR = \min_{\mathscr{R}_w\lambda = \mathbf{0}} \left(\lambda - \mathscr{T}^{-1}G\right)' \mathscr{T}\left(\lambda - \mathscr{T}^{-1}G\right) - \min_{\mathscr{R}_w\lambda \geq \mathbf{0}} \left(\lambda - \mathscr{T}^{-1}G\right)' \mathscr{T}\left(\lambda - \mathscr{T}^{-1}G\right).$$

When $\Sigma_W = \mathscr{T}^{-1}$, Lemma S.1.7 shows that QLR follows the same distribution as W using the duality of the optimization problem (see Ekeland (1974)). However, such result doesn't hold for general Θ and H_0. If $G \sim \mathcal{N}(\mathbf{0}, \mathscr{T})$, then QLR follows a weighted chi-squared distribution.

Comparing Lemmas 4.1 and 5.1, the asymptotic distributions of W_n and QLR_n share the similarity that they both depend on the binding inequalities in $\mathscr{R}_w^u\theta^* \geq \mathrm{r}_w^u$ and are discontinuous in the implicit nuisance parameter η^k. Because the idea for conducting uniform inference for test based upon QLR_n is analogous to that based upon W_n, certain details are omitted in the following discussion.

With Assumption 5.2 on the convergence rate of $\widehat{\theta}_0$ under $\omega_n \in \mathcal{W}_0$, the following lemma states the asymptotic distribution of QLR_n under drifting model parameters $\left(\eta_n^k, \pi_{Q,n}, \xi_n\right)$, where $\pi_Q \in \Pi_Q$ contains parameters in G and $\mathscr{T}$. The vector c is defined in (13). The asymptotic distributions of W_n and QLR_n under drifting model parameters depend on the same localization parameter vector c.

Assumption 5.2. *For any $\omega_n \in \mathcal{W}_0$, $b_n\left(\widehat{\theta}_0 - \theta_n\right) = O_p(1)$.*

Lemma 5.2. *If Assumptions 4.2-4.4 and 5.2 hold, then under $H_0 : R\theta^* = r$ and any parameter sequence $\left(\eta_n^k, \pi_{Q,n}, \xi_n\right) \in \mathrm{H}^k \times \Pi_Q \times \Xi$,*

$$QLR_n \xrightarrow{d} QLR_\omega \equiv \min_\lambda \left[q_\omega(\lambda) + \phi_{0,\omega}(\lambda)\right] - \min_\lambda \left[q_\omega(\lambda) + \phi_\omega(\lambda)\right],$$

where $q_\omega(\cdot)$ and $\phi_\omega(\cdot)$ are defined in Lemma 4.3 (i) and

$$\phi_{0,\omega}(\lambda) = \begin{cases} 0, & \text{if } \left(R', \mathscr{R}_e', \mathscr{R}_w^{b\prime}\right)'\lambda = \mathbf{0} \text{ and } \mathscr{R}_w^u\lambda + c \geq \mathbf{0} \\ \infty, & \text{otherwise} \end{cases}.$$

Let $\mathcal{C}^{Q}_{c,\pi_{Q,\omega}}(1-\tau)$ denote the $(1-\tau)$ quantile of QLR_ω given c and $\pi_{Q,\omega}$ for $0 \le \tau \le \alpha$. The α level Bonferroni critical value $CV^{Q}_{n}(\alpha,\tau)$ is defined as

$$CV^{Q}_{n}(\alpha,\tau) \equiv \sup_{c\in \widetilde{I}_{\alpha-\tau}} \mathcal{C}^{Q}_{c,\widehat{\pi}_Q}(1-\tau),$$

where $\widetilde{I}_{\alpha-\tau}$ and $\widehat{\pi}_Q$ are obtained by similar procedures presented by **Step 1** in Section 4.2. The following theorems show that $CV^{Q}_{n}(\alpha,\tau)$ controls the asymptotic size of QLR test and the test is consistent.

Theorem 5.1. *Under Assumptions 3.1, 4.2-4.4, and 5.2, if QLR_ω is continuous at $\mathcal{C}^{Q}_{c,\pi_{Q,\omega}}(1-\tau)$ for all $(c,\pi_{Q,\omega}) \in C \times \overline{\Pi}_Q$, then $AsySz\left(QLR_n, CV^{Q}_{n}(\alpha,\tau)\right) \le \alpha$ holds.*

Theorem 5.2. *Under H_1 and Assumptions 2.1-2.3 and 5.1, if $l_n(\cdot)$ is continuous at θ^*_0 and $b_n^{-2}(l_n(\theta^*) - l_n(\theta^*_0)) \xrightarrow{p} \varsigma > 0$, then $\Pr\left(QLR_n > CV^{Q}_{n}(\alpha,\tau)\right) \to 1$ holds.*

The condition $b_n^{-2}(l_n(\theta^*) - l_n(\theta^*_0)) \xrightarrow{p} \varsigma > 0$ in Theorem 5.2 is generally satisfied as the identification assumption.

Example 2.2 (continued)**.** Let $\Sigma_n \xrightarrow{p} \Sigma$ for which Σ is positive definite with probability one. We then have

$$\begin{aligned} n^{-1}(l_n(\theta^*) - l_n(\theta^*_0)) &= \left(\frac{1}{n}\sum_{i=1}^{n} g(Z_i,\theta^*_0)\right)' \Sigma_n \left(\frac{1}{n}\sum_{i=1}^{n} g(Z_i,\theta^*_0)\right) \\ &\quad - \left(\frac{1}{n}\sum_{i=1}^{n} g(Z_i,\theta^*)\right)' \Sigma_n \left(\frac{1}{n}\sum_{i=1}^{n} g(Z_i,\theta^*)\right) \\ &\xrightarrow{p} E[g(Z,\theta^*_0)]' \Sigma E[g(Z,\theta^*_0)]. \end{aligned}$$

Assumption $b_n^{-2}(l_n(\theta^*) - l_n(\theta^*_0)) \xrightarrow{p} \varsigma > 0$ in Theorem 5.2 is satisfied as long as $E[g(Z,\theta^*_0)] \neq \mathbf{0}$ for $\theta^*_0 \neq \theta^*$, which is assumed for the identification of θ^*.

6 The Score Test

Let the following two assumptions hold for the score function, directed score, and score test statistic defined in (6), (7), and (8). Discussion on the assumptions can be found in Andrews (2001).

Assumption 6.1. *(i) Assume that for all* $0 < \kappa < \infty$, $\sup_{\theta \in \Theta_0 : \left\| b_n \left(\theta - \theta_0^* \right) \right\| < \kappa} \left| b_n^{-1} R_n^D \left(\theta \right) \right| = o_p \left(1 \right)$*; (ii)* $\widehat{\mathscr{T}_n} = -b_n^{-2} D^2 l_n \left(\theta_0^* \right) + o_p \left(1 \right)$.

Assumption 6.2. $\Sigma_{S,n} \xrightarrow{p} \Sigma_S$ *for which* Σ_S *is positive definite with probability one.*

By definition, $R \in \mathbb{R}^{J \times l}$ and $J \leq l$. The polytope $R\Theta - r$ can be represented by

$$R\Theta - r = \left\{ \lambda \in \mathbb{R}^J : \exists \theta \in \Theta, \lambda = R\theta - r \right\}.$$

Let the following be a halfspace description of $R\Theta - r$:

$$R\Theta - r = \left\{ \lambda \in \mathbb{R}^J : \mathscr{R}_{R,e} \lambda = \mathrm{r}_{R,e} \text{ and } \mathscr{R}_{R,w} \lambda \geq \mathrm{r}_{R,w} \right\}. \tag{15}$$

Such description always exists, because the affine map of a polytope is a polytope, and every polytope can be represented by a halfspace description (Henk et al. (2004)). For any null hypothesis consistent with the parameter space, there exists some $\theta \in \Theta$ such that $R\theta = r$. Therefore, it holds that $\mathbf{0} \in R\Theta - r$. Consequently, we have $\mathrm{r}_{R,e} = \mathbf{0}$ and $\mathrm{r}_{R,w} \leq \mathbf{0}$. The limit of $b_n \left(R\Theta - r \right)$ in the sense of Hausdorff distance is given by

$$\Lambda_R \equiv \left\{ \lambda \in \mathbb{R}^J : \mathscr{R}_{R,e} \lambda = \mathbf{0} \text{ and } \mathscr{R}_{R,w,b} \lambda \geq \mathbf{0} \right\}, \tag{16}$$

where $\mathscr{R}_{R,w,b}$ is the submatrix of $\mathscr{R}_{R,w}$ composed of rows corresponding to the zero elements in $\mathrm{r}_{R,w}$.

As we show in Section 6.1 below, the null asymptotic distribution of S_n depends on Λ_R. When $J = l$, R is a square and invertible matrix. It is straightforward to find Λ_R. To see this, we note that the halfspace description of the polytope $R\Theta - r$ is characterized by

$$\mathscr{R}_{R,e} = \mathscr{R}_e R^{-1}, \, \mathrm{r}_{R,e} = \mathrm{r}_e - \mathscr{R}_e R^{-1} r = \mathbf{0} \text{ and}$$
$$\mathscr{R}_{R,w} = \mathscr{R}_w R^{-1}, \, \mathrm{r}_{R,w} = \mathrm{r}_w - \mathscr{R}_w R^{-1} r \leq \mathbf{0}.$$

The set Λ_R such that $d_H \left(b_n \left(R\Theta - r \right), \Lambda_R \right) \to 0$ is given by (16) with $\mathscr{R}_{R,w,b}$ being the submatrix of $\mathscr{R}_w R^{-1}$ composed of rows corresponding to the zero elements in $\mathrm{r}_w - \mathscr{R}_w R^{-1} r$.

When $J < l$, $R\Theta - r$ is an affine projection of the polytope Θ onto a lower dimensional space. Its limit Λ_R is in general not straightforward to compute. In Section 6.2, we provide an algorithm for obtaining the set Λ_R.

6.1 Asymptotic Theory

With description (16), the asymptotic distributions of ds and S_n are given in the following lemma.

Lemma 6.1. *(i) Suppose the null hypothesis and Assumptions 2.2, 5.1, and 6.1 hold. Then*

$$ds_n \xrightarrow{d} ds \equiv \arg\min_{\lambda} \left[q_R(\lambda) + \phi_R(\lambda)\right],$$

where $q_R(\lambda) = (\lambda - RZ)' \left(R\mathscr{T}^{-1}R'\right)^{-1} (\lambda - RZ)$, $Z = \mathscr{T}^{-1}G$, *and*

$$\phi_R(\lambda) = \begin{cases} 0, & \text{if } \mathscr{R}_{R,e}\lambda = \mathbf{0} \text{ and } \mathscr{R}_{R,w,b}\lambda \geq \mathbf{0} \\ \infty, & \text{otherwise} \end{cases};$$

(ii) If further Assumption 6.2 holds, then $S_n \xrightarrow{d} S \equiv ds'\Sigma_S^{-1}ds$.

The most significant difference between S in Lemma 6.1 and W in Lemma 4.1 or QLR in Lemma 5.1 is that the distribution of S is not discontinuous in the implicit nuisance parameter η^k, because $\mathscr{R}_{R,w,b}$ is known under H_0. That is, whether θ^* is on the boundary of Θ is unknown under the null hypothesis, which leads to discontinuity of the distributions of W and QLR in η^k; but whether $R\theta^*$ is on the boundary of $R\Theta$ is known, because $R\theta^* = r$ under the null hypothesis. Therefore, under H_0 the limit of $b_n(\Theta - \theta^*)$ is undetermined in general, but the limit of $b_n(R\Theta - R\theta^*) = b_n(R\Theta - r)$ is determined. Since ds_n is the projection of $R\widehat{\mathscr{T}}_n^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right)$ onto $b_n(R\Theta - r)$, its asymptotic distribution depends on the known limit of $b_n(R\Theta - r)$. Thus, parameters in G, $\mathscr{T}$, and Σ_S are the only unknown components in the distributions of ds and S. Since the distributions are continuous in those parameters, inference procedure based on the conventional plug-in approach controls the asymptotic size.

Example 3.1 (continued)**.** In this special case, the set $R\Theta - r$ can be easily obtained as

$$R\Theta - r = \left\{\lambda \in \mathbb{R}^J : \lambda \geq \mathbf{0}\right\}.$$

Lemma 6.1 implies that $S_n \xrightarrow{d} S = ds'\Sigma_S^{-1}ds$, where

$$ds = \arg\min_{\lambda \geq \mathbf{0}} \left(\lambda - \mathscr{R}_w\mathscr{T}^{-1}G\right)' \left(\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1} \left(\lambda - \mathscr{R}_w\mathscr{T}^{-1}G\right).$$

For $\Sigma_S = \mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'$, Lemma S.1.7 shows that $S = QLR$.

To study the asymptotic size of the score test, we impose the following assumption extending Assumptions 6.1 and 6.2.

Assumption 6.3. *Assume that for any* $\omega_n \in \mathcal{W}_0$, *(i)* $\sup_{\theta\in\Theta:\|\theta-\theta_\omega\|<\kappa_n} \left|b_n^{-1} R_n^D(\theta)\right| = o_p(1)$ *for all* $\kappa_n = o(1)$*; (ii)* $\widehat{\mathscr{T}_n} = \mathscr{T}_n + o_p(1)$*; and (iii)* $\Sigma_{S,n} \xrightarrow{p} \Sigma_{S,\omega}$ *for which* $\Sigma_{S,\omega}$ *is positive definite with probability one.*

Let $\mathcal{C}^S_{\pi_S}(1-\alpha)$ denote the $(1-\alpha)$ quantile of S, where $\pi_S \in \Pi_S$ contains parameters in G, $\mathscr{T}$, and Σ_S. The critical value for the α level score test is computed as $CV_n^S(\alpha) \equiv \mathcal{C}^S_{\widehat{\pi}_S}(1-\alpha)$, where $\widehat{\pi}_S$ is some consistent estimator of π_S for any $\omega_n \in \mathcal{W}_0$. For the test based upon S_n with $CV_n^S(\alpha)$, the following theorem shows that the asymptotic size is equal to α.

Theorem 6.1. *Under Assumptions 3.1, 4.3, 5.2, and 6.3, if* S *is continuous at* $\mathcal{C}^S_{\pi_S}(1-\alpha)$ *for all* $\pi_S \in \overline{\Pi}_S$*, then it holds that* $AsySz\left(S_n, CV_n^S(\alpha)\right) = \alpha$.

The consistency of the score test relies on the shape of $l_n(\cdot)$. In the following theorem, we provide sufficient conditions for the score test to be consistent.

Theorem 6.2. *Under* H_1 *and Assumptions 5.1, 6.1, and 6.2, if* $\mathscr{T}_n^{-1} b_n^{-2} Dl_n(\theta_0^*) = \upsilon(\theta^* - \theta_0^*) + o_p(1)$*, where* $0 < \upsilon \leq 1$ *and* $R\mathscr{T}^{-1}R'$ *is positive definite, then it holds that* $\Pr\left(S_n > CV_n^S(\alpha)\right) \to 1$. [9]

Since the first order derivative of $l_n(\cdot)$ at θ^* approaches zero, $\mathscr{T}_n^{-1} b_n^{-2} Dl_n(\theta^*)$ is $o_p(1)$. The condition in Theorem 6.2 requires that the difference between $\mathscr{T}_n^{-1} b_n^{-2} Dl_n(\theta_0^*)$ and $\mathscr{T}_n^{-1} b_n^{-2} Dl_n(\theta^*)$ be proportional to that between θ^* and θ_0^* up to a small order term. When l_n takes a quadratic form in θ, the condition is satisfied. However, as shown in the second example below, if $l_n(\cdot)$ takes a different form, the score test may not be consistent for certain deviations from the null hypothesis. Therefore, even though the test based upon S_n does not require choosing any tuning parameter, its consistency relies on the hypothesis and model, and is difficult to check if $l_n(\cdot)$ takes a complicated form.

Example 2.1 (continued)**.** Following the previous discussion, for $b_n = \sqrt{n}$, we have:

$$\mathscr{T}_n = -b_n^{-2} D^2 l_n(\theta_0^*) = \frac{1}{n}\sum X_i X_i' \text{ and}$$
$$b_n^{-2} Dl_n(\theta_0^*) = \frac{1}{n}\sum X_i X_i'(\theta^* - \theta_0^*) + \frac{1}{n}\sum \varepsilon_i X_i.$$

[9] If ds_n is defined as $\widehat{q}_R(ds_n) = \inf_{\lambda_R\in\Lambda_R} \widehat{q}_R(\lambda_R) + o_p(1)$, then υ can be allowed to take any positive value.

Since $\frac{1}{n}\sum \varepsilon_i X_i \xrightarrow{p} 0$, it holds that

$$\mathscr{T}_n^{-1} b_n^{-2} Dl_n\left(\theta_0^*\right) = \theta^* - \theta_0^* + \left(\frac{1}{n}\sum X_i X_i'\right)^{-1}\left(\frac{1}{n}\sum \varepsilon_i X_i\right) = \theta^* - \theta_0^* + o_p\left(1\right).$$

The assumption in the theorem is verified.

Example 6.1. For the Logit model, we have that

$$l_n\left(\theta\right) = \sum_{i=1}^{n}\left[Y_i \ln F\left(X_i'\theta\right) + \left(1 - Y_i\right)\ln\left(1 - F\left(X_i'\theta\right)\right)\right],$$

$$Dl_n\left(\theta\right) = \sum_{i=1}^{n}\left[Y_i F\left(-X_i'\theta\right) - \left(1 - Y_i\right) F\left(X_i'\theta\right)\right] X_i, \text{ and } D^2 l_n\left(\theta\right) = -\sum_{i=1}^{n} f\left(X_i'\theta\right) X_i X_i',$$

where $F\left(t\right) = \frac{1}{1+e^{-t}}$ and $f\left(t\right) = \frac{e^{-t}}{\left(1+e^{-t}\right)^2}$. Assume $\Theta = \left\{\left(\theta_1, \theta_2\right) \in \mathbb{R}^2 : \theta_1 + \theta_2 \geq 0\right\}$, $\theta^* = (1, 0)$ and $X = \left(X_1, X_2\right)'$, where X_1 has equal probability of being 1 and -1; X_2 has equal probability of being 0 and 1; and they are independent. It can be shown that S_n does not diverge to infinity when $R = \left(\frac{1+6e+e^2}{8}, e\right)$ and $r = e$. The asymptotic power of the score test is not one for testing $H_0 : R\theta^* = r$.

6.2 Implementation—Projection of Polytope

We describe one algorithm for the projection of a polytope developed in the constraint logic programming, marginal problem, and robotic research based on the Fourier-Motzkin algorithm. Other approaches such as the double description method and equality set projection are also applicable. See Fukuda and Prodon (1995) and Jones et al. (2004) for more details. Specifically we are interested in obtaining Λ_R for the asymptotic distribution of S_n. While the set $R\Theta - r$ is unique, there are infinite many different halfspace descriptions. Thanks to Lemma S.1.6, the result in Lemma 6.1 does not depend on the description. Thus, any algorithm that returns a halfspace description of $R\Theta - r$ would serve the purpose, even if the description contains many strictly redundant inequalities. Moreover, if an inequality $\mathscr{R}_{R,w(j)}\lambda \geq \mathrm{r}_{R,w(j)}$ among $\mathscr{R}_{R,w}\lambda \geq \mathrm{r}_{R,w}$ is strictly redundant, then $\mathrm{r}_{R,w(j)} < 0$, because $\mathbf{0} \in R\Theta - r$. Thus, $\mathscr{R}_{R,w(j)}$ is automatically eliminated from the submatrix $\mathscr{R}_{R,w,b}$ when we consider Λ_R.

We adopt the Fourier-Motzkin algorithm to obtain the halfspace description of $R\Theta - r$. The Fourier-Motzkin algorithm consists of two main steps. Let the combined polytope be

$$P \equiv \{(\theta, \lambda) : \mathscr{R}_e\theta = \mathrm{r}_e,\, \mathscr{R}_w\theta \geq \mathrm{r}_w, \text{ and } \lambda = R\theta - r\} .$$

First, use the equality constraints $\mathscr{R}_e\theta = \mathrm{r}_e$ and $R\theta = \lambda + r$ to eliminate as many coordinates in θ as possible: obtain the solution of the following system of linear equations

$$\begin{pmatrix} \mathscr{R}_e \\ R \end{pmatrix} \theta = \begin{pmatrix} \mathrm{r}_e \\ \lambda + r \end{pmatrix}$$

as $\theta = \Gamma\theta_f + \mathit{\Gamma}\lambda + \gamma$ by treating λ as given. The definitions of Γ, θ_f, and γ are the same as the ones for (9) and $\mathit{\Gamma}$ is some $l \times J$ matrix. This yields a reduced polytope:

$$P_f \equiv \{(\theta_f, \lambda) : \mathscr{R}_w (\Gamma\theta_f + \mathit{\Gamma}\lambda) \geq \mathrm{r}_w - \mathscr{R}_w\gamma\} .$$

Second, apply Fourier-Motzkin Elimination (FME) (Fourier (1824), Dines (1919) and Motzkin (1936)) on P_f. The procedure of FME is standard and can be implemented directly with $Matlab^{TM}$'s MPT2 or MPT3. We skip the details and refer interested readers to Dantzig and Eaves (1973), Imbert (1993), and Bastrakov and Zolotykh (2015) for more discussion on FME. Since FME usually generates many strictly redundant inequalities during the elimination, methods like Chernikov rule (Chernikov (1965)) are introduced to reduce the number of inequalities. As discussed earlier, such extra step is optional in our setting, because strictly redundant inequalities are automatically eliminated when considering Λ_R. We therefore obtain the halfspace descriptions of both $R\Theta - r$ and its limit Λ_R.

We provide several examples to illustrate the Fourier-Motzkin algorithm.

Example 6.2. Let $\Theta \equiv \left\{\theta = (\theta_1, \theta_2)' \in \mathbb{R}^2 : \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}$, where

$$\mathscr{R}_w = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -2 \end{pmatrix} \text{ and } \mathrm{r}_w = \begin{pmatrix} 0 \\ 0 \\ -6 \end{pmatrix},$$

$R = (1, 1)$, and $r = 0$. We have $R\Theta - r = \left\{\lambda = \theta_1 + \theta_2 : (\theta_1, \theta_2)' \in \Theta\right\}$. Since the set $R\Theta - r$ is one dimensional, its left and right boundary can be easily obtained by linear programming: $R\Theta - r = [0, 6]$. The Fourier-Motzkin algorithm works as the following. The solution to $\theta_1 + \theta_2 = \lambda$ is

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \theta_2 + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \lambda;$$

and the reduced polytope P_f is expressed as

$$-\theta_2 + \lambda \geq 0,\ \theta_2 \geq 0, \text{ and } -\theta_2 - \lambda \geq -6.$$

By applying FME to the above inequalities, we obtain that

$$\lambda \geq 0 \text{ and } -\lambda + 6 \geq 0.$$

There is no strictly redundant inequality and $R\Theta - r = \{\lambda : \lambda \geq 0 \text{ and } -\lambda \geq -6\}$. The limit of $b_n\,(R\Theta - r)$ is simply $\Lambda_R = \{\lambda : \lambda \geq 0\}$.

Example 6.3. Let $\Theta \equiv \left\{\theta = (\theta_1, \theta_2)' \in \mathbb{R}^2 : \mathscr{R}_w \theta \geq \mathrm{r}_w\right\}$, where

$$\mathscr{R}_w = \begin{pmatrix} 2 & 1 \\ 0 & 1 \\ -1 & -2 \end{pmatrix} \text{ and } \mathrm{r}_w = \begin{pmatrix} 0 \\ 0 \\ -6 \end{pmatrix},$$

and $R\Theta - r \equiv \left\{\lambda = \theta_1 + \theta_2 : (\theta_1, \theta_2)' \in \Theta\right\}$. With the same procedure in the previous example, we obtain $R\Theta - r = [0, 6]$ and $\Lambda_R = [0, \infty)$. This shows that for the same R and r, different parameter spaces can result in the same projection.

Example 3.5 (continued)**.** The first step of Fourier-Motzkin algorithm provides $\theta = \Gamma\theta_f + \mathit{\Gamma}\lambda + \gamma$, with Γ, θ_f, and γ being calculated in Example 3.5 and

$$\mathit{\Gamma}' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The reduced polytope is

$$P_f = \{(\theta_2, \theta_3, \lambda_1, \lambda_2) : \mathscr{R}_w \Gamma \theta_f + \mathscr{R}_w \mathit{\Gamma} \lambda \geq \mathbf{0} - \mathscr{R}_w \gamma\},$$

where values of $\mathscr{R}_w \Gamma$ and $\mathbf{0} - \mathscr{R}_w \gamma$ can be found in Example 3.5 and

$$(\mathscr{R}_w \mathit{\Gamma})' = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

FME is then applied to the following linear inequalities:

$$-\lambda_1 \geq 0,\, \theta_2 \geq 0,\, -\theta_2 + \theta_3 \geq 0 \text{ and } \theta_2 - \theta_3 + \lambda_2 \geq -0.1.$$

By eliminating θ_2 first and then θ_3, we obtain first

$$-\lambda_1 \geq 0,\, \theta_3 \geq 0 \text{ and } \theta_3 \geq \theta_3 - \lambda_2 - 0.1$$

and then

$$\lambda_1 \geq 0 \text{ and } \lambda_2 \geq -0.1.$$

Thus, $R\Theta - r = \{(\lambda_1, \lambda_2) : \lambda_1 \geq 0 \text{ and } \lambda_2 \geq -0.1\}$ and $\Lambda_R = \{(\lambda_1, \lambda_2) : \lambda_1 \geq 0\}$.

7 Local Power

In this section, we investigate the asymptotic distributions of the test statistics under sequences of local alternatives of the form

$$H_{1,n} : R\theta_n = r + b_n^{-1}\delta\left(1 + o\left(1\right)\right),$$

where $\delta \in \mathbb{R}^J$. Following Section 2.2, let $\omega_n \equiv (\theta_n, \psi_n)$ be the drifting parameter sequence consistent with $H_{1,n}$ with limit $\omega \equiv (\theta_\omega, \psi_\omega)$. Let $c_w \equiv \lim_{n\to\infty} b_n \left(\mathscr{R}_w \theta_n - \mathrm{r}_w\right) \in \overline{\mathbb{R}}^{l_w}_{\geq 0}$, and denote c and $c_{w,b}$ as the subvectors of c_w corresponding to the submatrices $\mathscr{R}^u_w$ and $\mathscr{R}^b_w$ of $\mathscr{R}_w$. Notice that the above definition of c is consistent with that in (13).

Assumptions in Lemmas 4.3, 5.2, and 6.1 are modified in Assumption 7.1 below for the drifting parameter sequence ω_n consistent with $H_{1,n}$. Similar type of assumptions have been introduced in Sections 4, 5, and 6 when the parameter sequence is consistent with H_0.

Assumption 7.1. *For the sequence ω_n consistent with $H_{1,n}$, assume the followings: (i) $\sup_{\theta\in\Theta:\|\theta-\theta_\omega\|<\kappa_n} \left|R_n\left(\theta\right)\right| = o_p\left(1\right)$ for all $\kappa_n = o\left(1\right)$; (ii) $\left(b_n^{-1}Dl_n\left(\theta_n\right), \mathscr{T}_n\right) \xrightarrow{d} \left(G_\omega, \mathscr{T}_\omega\right)$ for some random variables $G_\omega \in \mathbb{R}^l$ and $\mathscr{T}_\omega \in \mathbb{R}^{l\times l}$, where $\mathscr{T}_n \equiv -b_n^{-2}D^2 l_n\left(\theta_n\right)$ and $\mathscr{T}_\omega$ is symmetric and non-singular with probability one; (iii) $b_n\left(\widehat{\theta} - \theta_n\right) = O_p\left(1\right)$; (iv) $\Sigma_{W,n} \xrightarrow{p} \Sigma_{W,\omega}$ for which $R\Sigma_{W,\omega}R'$ is positive definite with probability one; (v) $b_n\left(\widehat{\theta}_0 - \theta_n\right) = O_p\left(1\right)$; (vi) $\sup_{\theta\in\Theta:\|\theta-\theta_\omega\|<\kappa_n} \left|b_n^{-1}R_n^D\left(\theta\right)\right| = o_p\left(1\right)$ for all $\kappa_n = o\left(1\right)$ and $\widehat{\mathscr{T}_n} = \mathscr{T}_n + o_p\left(1\right)$; and (vii) $\Sigma_{S,n} \xrightarrow{p} \Sigma_{S,\omega}$ for which $\Sigma_{S,\omega}$ is positive definite with* *probability one.*

The following lemma provides the asymptotic distributions of W_n, QLR_n, and S_n under the drifting parameter sequence consistent with the local alternative hypothesis.

Lemma 7.1. *For the parameter sequence ω_n consistent with $H_{1,n}$, if Assumption 7.1 holds, then*

(i) $W_n \xrightarrow{d} W_{1,\omega} \equiv (R\Psi_{1,\omega} + \delta)' (R\Sigma_{W,\omega} R')^{-1} (R\Psi_{1,\omega} + \delta)$, *where*

$$\Psi_{1,\omega} \equiv \arg\min_{\lambda} [q_\omega(\lambda) + \phi_{1,\omega}(\lambda)]$$

in which $q_\omega(\lambda) = (\lambda - Z_\omega)' \mathscr{T}_\omega (\lambda - Z_\omega)$, $Z_\omega = \mathscr{T}_\omega^{-1} G_\omega$, *and*

$$\phi_{1,\omega}(\lambda) = \begin{cases} 0, & \textit{if } \mathscr{R}_e \lambda = \mathbf{0} \textit{ and } \mathscr{R}_w \lambda + c_w \geq \mathbf{0} \\ \infty, & \textit{otherwise} \end{cases};$$

(ii) $QLR_n \xrightarrow{d} QLR_{1,\omega}$, *where*

$$QLR_{1,\omega} \equiv \min_{\lambda} [q_\omega(\lambda) + \phi_{0,1,\omega}(\lambda)] - \min_{\lambda} [q_\omega(\lambda) + \phi_{1,\omega}(\lambda)],$$

in which

$$\phi_{0,1,\omega}(\lambda) = \begin{cases} 0, & \textit{if } \left(R', \mathscr{R}_e', \mathscr{R}_w^{b\prime}\right)' \lambda + \left(\delta', \mathbf{0}, c_{w,b}'\right)' = \mathbf{0} \textit{ and } \mathscr{R}_w^u \lambda + c \geq \mathbf{0} \\ \infty, & \textit{otherwise} \end{cases}; \textit{ and}$$

(iii) $S_n \xrightarrow{d} S_{1,\omega} \equiv ds_{1,\omega}' \Sigma_{S,\omega}^{-1} ds_{1,\omega}$, *where*

$$ds_{1,\omega} \equiv \arg\min_{\lambda} [q_{R,\omega}(\lambda) + \phi_{R,\omega}(\lambda)],$$

in which $q_{R,\omega}(\lambda) = (\lambda - RZ_\omega - \delta)' (R\mathscr{T}_\omega^{-1} R')^{-1} (\lambda - RZ_\omega - \delta)$, *and*

$$\phi_{R,\omega}(\lambda) = \begin{cases} 0, & \textit{if } \mathscr{R}_{R,e} \lambda = \mathbf{0} \textit{ and } \mathscr{R}_{R,w,b} \lambda \geq \mathbf{0} \\ \infty, & \textit{otherwise} \end{cases}.$$

The proof for Lemma 7.1 is similar to that for Lemmas 4.3, 5.2, and 6.1. For ω_n consistent $H_{1,n}$, the asymptotic distributions of W_n and QLR_n depend on both δ and c_w. The test statistic S_n does not depend on the nuisance parameter sequence $\mathscr{R}_w \theta_n$. Thus, its asymptotic distribution relates only to δ. The local asymptotic powers of tests based upon W_n, QLR_n, and S_n with critical values $CV_n^W(\alpha,\tau)$, $CV_n^Q(\alpha,\tau)$, and $CV_n^S(\alpha)$ are given in the following corollary.

Corollary 7.1. *Let Assumption 7.1 hold and $\omega_n \in \mathcal{W}$ be the parameter sequence consistent with $H_{1,n}$.*

(i) If W_ω is continuous at $\mathcal{C}^W_{c,\pi_{W,\omega}}(1-\tau)$, then

$$\Pr_{\omega_n}\left(W_n > CV_n^W(\alpha,\tau)\right) \longrightarrow \Pr\left(W_{1,\omega} > CV^W(\alpha,\tau)\right),$$

where $CV^W(\alpha,\tau) \equiv \sup_{c\in I_{\alpha-\tau}} \mathcal{C}^W_{c,\pi_{W,\omega}}(1-\tau)$, in which

$$I_{\alpha-\tau} \equiv \left\{\mathrm{c} \in \mathbb{R}^{l_u}_{\geq 0} : \mathrm{c} = c + \Gamma^u\left(\mathscr{R}^u_\Gamma \mathscr{T}^{-1}_{f,\omega} G_{f,\omega} + \iota\right), \iota \in ES(\alpha-\tau)\right\}.$$

The random vector $G_{f,\omega}$ is the subvector of G_ω corresponding to θ_f;

(ii) If QLR_ω is continuous at $\mathcal{C}^Q_{c,\pi_{Q,\omega}}(1-\tau)$, then

$$\Pr_{\omega_n}\left(QLR_n > CV_n^Q(\alpha,\tau)\right) \longrightarrow \Pr\left(QLR_{1,\omega} > CV^Q(\alpha,\tau)\right),$$

where $CV^Q(\alpha,\tau) \equiv \sup_{c\in I_{\alpha-\tau}} \mathcal{C}^Q_{c,\pi_{Q,\omega}}(1-\tau)$; and

(iii) If S is continuous at $\mathcal{C}^S_{\pi_{S,\omega}}(1-\alpha)$, then

$$\Pr_{\omega_n}\left(S_n > CV_n^S(\alpha)\right) \longrightarrow \Pr\left(S_{1,\omega} > CV^S(\alpha)\right),$$

where $CV^S(\alpha) \equiv \mathcal{C}^S_{\pi_{S,\omega}}(1-\alpha)$.

The limiting probabilities provide the local asymptotic powers. As can be seen from the corollary, the local asymptotic powers of tests based upon W_n and QLR_n with critical values $CV_n^W(\alpha,\tau)$ and $CV_n^Q(\alpha,\tau)$ depend on δ and c_w; while the test based upon S_n and $CV_n^S(\alpha)$ has the local asymptotic power only related to δ. This is the consequence of both the test statistics and critical values. The asymptotic distributions of W_n and QLR_n and their corresponding critical values under $\omega_n \in \mathcal{W}$ all depend on δ and c_w. On the other hand, Lemma 7.1 shows that the asymptotic distribution of S_n under ω_n only depends on δ and the critical value $CV_n^S(\alpha)$ is determined solely by the estimator of model parameters π_S.

Example 3.1 (continued)**.** For the special case in Gourieroux et al. (1982), the asymptotic distributions of W_n, QLR_n, and S_n under H_0 are the same when Σ_W and Σ_S take specific forms. Moreover, since their asymptotic distributions under

$H_{1,n}$ are also the same by Lemma S.1.8, tests based upon W_n, QLR_n, and S_n are asymptotically equivalent. See Silvapulle and Sen (2005) for some related discussion.

8 More Tests and Equivalence

In this section, we introduce the remaining three tests and compare them with the previous tests. Denote $\widetilde{\theta}$ as the *unconstrained* estimator such that $\widetilde{\theta} \in \mathbb{R}^l$ and

$$l_n\left(\widetilde{\theta}\right) = \sup_{\theta \in \mathbb{R}^l} l_n(\theta) + o_p(1).$$

We impose the following convergence rate assumption on $\widetilde{\theta}$.

Assumption 8.1. $b_n\left(\widetilde{\theta} - \theta^*\right) = O_p(1)$.

Remark 8.1. When $l_n(\cdot)$ is not defined outside Θ, $\widetilde{\theta}$ may not be available and tests based on $\widetilde{\theta}$ cannot be applied. However, $\widehat{\theta}$ and $\widehat{\theta}_0$ are defined within the parameter space and always exist.

8.1 Another Wald Test

Based on $\widetilde{\theta}$, an alternative Wald test statistic can be defined as

$$W_n^1 \equiv b_n^2\left(R\widetilde{\theta} - r\right)'\left(R\Sigma_{W,n}R'\right)^{-1}\left(R\widetilde{\theta} - r\right) - \inf_{\lambda \in R\Theta} b_n^2\left(R\widetilde{\theta} - \lambda\right)'\left(R\Sigma_{W,n}R'\right)^{-1}\left(R\widetilde{\theta} - \lambda\right).$$

The first term in the expression for W_n^1 is the Wald statistic without accounting for the constraints in Θ. The second term adjusts for the fact that $\theta^* \in \Theta$ by projecting $R\widetilde{\theta}$ onto the set $R\Theta$. Consider e.g., the case where $\widetilde{\theta}$ lies outside Θ. Let $\dot{\theta} \in \Theta$ be the projection of $R\widetilde{\theta}$ onto $R\Theta$. If $R\dot{\theta} = r$, then the null hypothesis should not be rejected, because comparing all other possible values of $R\theta$ for $\theta \in \Theta$, $R\dot{\theta}$ being r is the closest to the unconstrained $R\widetilde{\theta}$. This is the evidence supporting the null hypothesis. In the case where $R\Theta$ is the whole space, the second term in W_n^1 is zero and W_n^1 becomes the "classical" Wald test statistic.

Similar to S_n, the asymptotic distribution of the second term depends on whether the value of $R\theta^*$ is in the interior of $R\Theta$ or its boundary which is known under H_0. Therefore, the asymptotic distribution of W_n^1 is not discontinuous in the implicit nuisance parameter η^k under the null.

8.2 Two More Score Tests

Define the first alternative score test statistic as

$$S_n^1 \equiv \left[\widehat{\mathscr{T}}_n^{-1} b_n^{-1}\left(Dl_n\left(\widehat{\theta}_0\right) - Dl_n\left(\widehat{\theta}\right)\right)\right]' \Sigma_{S^1,n}^{-1} \left[\widehat{\mathscr{T}}_n^{-1} b_n^{-1}\left(Dl_n\left(\widehat{\theta}_0\right) - Dl_n\left(\widehat{\theta}\right)\right)\right],$$

with $\Sigma_{S^1,n}$ being positive definite such that $\Sigma_{S^1,n} \xrightarrow{p} \Sigma_{S^1}$. S_n^1 is the same as the global score test defined in Silvapulle and Sen (2005). When no constraint is imposed on the parameter space Θ, S_n^1 equals to the well known Rao's score statistic, because $Dl_n\left(\widehat{\theta}\right)$ is zero. Since Θ is defined by equality and inequality constraints, $Dl_n\left(\widehat{\theta}\right)$ is not guaranteed to be zero. The test statistic compares values of $Dl_n(\cdot)$ evaluated at $\widehat{\theta}$ and $\widehat{\theta}_0$, which are close if the null is true. Notice that S_n^1 requires the computation of both $\widehat{\theta}$ and $\widehat{\theta}_0$.

It will be shown in the next section that the asymptotic behavior of S_n^1 is similar to that of W_n, which is discontinuous in the implicit nuisance parameter. Uniform inference is therefore needed and can proceed in a similar way in Section 4.2. For brevity, we omit details here.

The second alternative score test statistic follows from Silvapulle and Silvapulle (1995) and is defined as

$$\begin{aligned} S_n^2 \equiv & \left(R\widehat{\mathscr{T}}_n^{-1} b_n^{-1} Dl_n\left(\widehat{\theta}_0\right)\right)' \left(R\widehat{\mathscr{T}}_n^{-1} R'\right)^{-1} \left(R\widehat{\mathscr{T}}_n^{-1} b_n^{-1} Dl_n\left(\widehat{\theta}_0\right)\right) \\ & - \inf_{\lambda \in b_n(R\Theta - r)} \left(R\widehat{\mathscr{T}}_n^{-1} b_n^{-1} Dl_n\left(\widehat{\theta}_0\right) - \lambda\right)' \left(R\widehat{\mathscr{T}}_n^{-1} R'\right)^{-1} \left(R\widehat{\mathscr{T}}_n^{-1} b_n^{-1} Dl_n\left(\widehat{\theta}_0\right) - \lambda\right). \end{aligned}$$

For more detailed discussion on S_n^2, please refer to Silvapulle and Silvapulle (1995) and Section 4.6 in Silvapulle and Sen (2005) on local score tests. The asymptotic distribution of S_n^2 is the same as S_n up to the weighting matrix and is not discontinuous in the implicit nuisance parameter η^k under the null.

8.3 Comparison of Tests

The following theorem provides the equivalence in distributions among the test statistics introduced previously under the fixed model parameters.

Theorem 8.1. *Let the null hypothesis and Assumptions 2.1 and 2.2 hold.*

(i) Under Assumptions 2.3, 4.1, 5.1, 6.1, and 6.2, if $\Sigma_{S^1}^{-1} = R'\left(R\Sigma_W R'\right)^{-1} R$, *then* $W_n = S_n^1 + o_p(1)$; *(ii) Under Assumptions 5.1, 6.1, and 6.2, if* $\Sigma_S = R\mathscr{T}^{-1}R'$, *then*

$S_n^2 = S_n + o_p(1)$*; (iii) Under Assumptions 4.1, 5.1, 6.1 , 6.2, and 8.1, if* $\Sigma_W = \mathscr{T}^{-1}$*, then* $W_n^1 = S_n^2 + o_p(1)$.

From the theorem, it can be seen that there are three types of asymptotic null distributions, with QLR_n having a unique type, W_n and S_n^1 sharing the same type, and W_n^1, S_n and S_n^2 sharing one type. The asymptotic distribution of S_n^1 can be easily obtained from Lemma 4.1 and the proof of Theorem 8.1 (i). It is the same as that of W_n up to the weighting matrix. Due to the discontinuity of the asymptotic distribution in model parameters, uniform inference is needed to control the asymptotic size for the test based upon S_n^1. The same procedure in Section 4.2 applies. The three test statistics S_n, W_n^1, and S_n^2 follow the same asymptotic distribution up to the weighting matrix by Theorem 8.1 (ii) and (iii). Thus, algorithm in Section 6.2 is needed. Since their asymptotic distributions are not discontinuous in the implicit nuisance parameter η^k, the standard plug-in approach for the critical value controls the asymptotic size.

For brevity, we skip the details on how to obtain the critical values for the tests based upon W_n^1, S_n^1, and S_n^2. Under the alternative hypothesis, W_n^1 diverges to infinity if the weighting matrix is positive definite. When $l_n(\cdot)$ has no local maximum and Σ_{S^1} is positive definite, S_n^1 also diverges to infinity under H_1. However, for the test based upon S_n^2 to be consistent, similar conditions in Theorem 6.2 need to be imposed, which may not hold for some model and hypothesis.

The results of Theorem 8.1 hold for any $\mathscr{R}_e$, $\mathscr{R}_w$, and R. When the maintained hypothesis and the null hypothesis take special forms such as the ones in Gourieroux et al. (1982) and Wolak (1987), more equivalence results can be derived. Andrews (2001) also compares W_n, QLR_n, and S_n when the maintained hypothesis and the null hypothesis take special forms.

9 Numerical Results

In this section, we present two sets of numerical results. First, we conduct a small simulation study designed to examine and compare the finite sample performance of the tests developed in this paper and the three "classical" ones. Second, using the data from Princeton Data Improvement Initiative (PDII) survey, we apply our tests to some linear regression models introduced in Autor and Handel (2013) to illustrate the applicability of our method.

9.1 Monte-Carlo Simulation

The model used in this section is designed to mimic the linear regression models in Section 9.2.

9.1.1 The Data Generating Process

We consider a linear regression model of the following form:

$$Y = \beta_1 E_1 + \beta_2 E_2 + \beta_3 E_3 + \beta_4 E_4 + \mu_0 + \mu_1 X_1 + \mu_2 X_2 + \varepsilon,$$

where $\Theta = \left\{\theta = (\beta', \mu')' : \mathscr{R}_w \beta \geq \mathbf{0}\right\}$ with $\beta \equiv (\beta_1, \ldots, \beta_4)'$, $\mu \equiv (\mu_0, \mu_1, \mu_2)'$, and

$$\mathscr{R}_w = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

The discrete random vector (E_1, E_2, E_3, E_4) follows the distribution below:

$$\begin{aligned}
\Pr\left(E_1 = 1, E_2 = 0,\ E_3 = 0 \text{ and } E_4 = 0\right) = 0.09, \\
\Pr\left(E_1 = 0, E_2 = 0,\ E_3 = 0 \text{ and } E_4 = 0\right) = 0.33, \\
\Pr\left(E_1 = 0, E_2 = 1,\ E_3 = 0 \text{ and } E_4 = 0\right) = 0.26, \\
\Pr\left(E_1 = 0, E_2 = 0,\ E_3 = 1 \text{ and } E_4 = 0\right) = 0.21, \\
\Pr\left(E_1 = 0, E_2 = 0,\ E_3 = 0 \text{ and } E_4 = 1\right) = 0.11.
\end{aligned}$$

The two continuous random variables (X_1, X_2) are independent of (E_1, E_2, E_3, E_4), and follow the joint normal distribution with zero mean, unit variance, and correlation coefficient 0.2. The error term ε is independent of the observable covariates.

We consider two DGPs corresponding to two distributions for the error ε. For DGP A, ε follows the Gaussian distribution with variance $1/2$; and for DGP B, $\varepsilon \sim \mathit{Gamma}\,(2, 2) - 1$. The variance of ε is the same in both DGPs. The distribution of ε is symmetric under DGP A and has the skewness of $\sqrt{2}$ under DGP B.

We aim to test the same null hypothesis as in Example 3.5 with $H_0 : R\beta^* = r$,

Table I: Different sets of parameters

Case 1	Case 2	Case 3	Case 4	Case 5
$(\beta_1, 0.2, 0.25, 0.3)$	$(\beta_1, 0, 0.05, 0.1)$	$(\beta_1, 0, 0.1, 0.1)$	$(\beta_1, 0.01, 0.05, 0.11)$	$(\beta_1, 0.01, 0.02, 0.11)$

where

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \text{ and } r = \begin{pmatrix} 0 \\ 0.1 \end{pmatrix}.$$

There are four inequalities in the maintained hypothesis, among which the first inequality is binding under the null hypothesis. To compare the power performance of the tests under different cases, we consider five different sets of parameters corresponding to different numbers of binding inequalities, see Table I.

In Case 1, no inequality is binding except the first one and there is no close-to-binding inequality under H_0. Case 2 has the second inequality binding, and Case 3 has the second and fourth inequalities binding. There is no binding inequality in Cases 4 and 5 except the first inequality under the null hypothesis. However, the second inequality is close-to-binding in Case 4, and the second and third inequalities are close-to-binding in Case 5.

Under H_0, $\beta_1 = 0$. We consider three other values of β_1: -0.05, -0.1, and -0.15 to examine the power performance of the tests.

9.1.2 The Tests and Their Implementation

The estimator objective function $l_n(\cdot)$ takes the same form as in Example 2.1. We implement nine tests in the following three groups:

Group I. *Tests based upon W_n, QLR_n, and S_n^1.* Because the null asymptotic distributions of all three statistics exhibit discontinuity, the critical values are computed via Bonferroni correction. Example 3.5 in Section 3.2 provides steps for obtaining the implicit nuisance parameter. The confidence set for c is computed in the same way for the three tests by first constructing the confidence set for η_n^k and then applying (14). Following Romano et al. (2014) and McCloskey (2017), we set the tuning parameter $\tau = \alpha - \alpha/10$;

Group II. *Tests based upon W_n^1, S_n, and S_n^2.* The null asymptotic distributions of these three statistics do not exhibit discontinuity. However, to compute their critical values, we need to compute projections of appropriate polytopes which is done via the algorithm presented in Section 6.2. The detail can be found in the continuation

Table II: Rejection probability under H_0

		Group I			Group II			Group III		
		W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0
DGP A	Case 1	0.0412	0.0417	0.0425	0.0438	0.0498	0.0498	0.0436	0.0434	0.0431
	Case 2	0.0424	0.0419	0.0429	0.0443	0.0513	0.0513	0.0442	0.0441	0.0436
	Case 3	0.0443	0.0436	0.0441	0.0474	0.0509	0.0510	0.0483	0.0477	0.0467
	Case 4	0.0428	0.0430	0.0423	0.0467	0.0521	0.0521	0.0479	0.0472	0.0465
	Case 5	0.0430	0.0427	0.0437	0.0482	0.0502	0.0502	0.0486	0.0483	0.0480
		Group I			Group II			Group III		
		W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0
DGP B	Case 1	0.0408	0.0409	0.0421	0.0399	0.0404	0.0404	0.0396	0.0393	0.0390
	Case 2	0.0411	0.0412	0.0417	0.0402	0.0406	0.0406	0.0388	0.0379	0.0372
	Case 3	0.0423	0.0412	0.0425	0.0413	0.0410	0.0410	0.0376	0.0370	0.0361
	Case 4	0.0398	0.0407	0.0408	0.0387	0.0405	0.0405	0.0381	0.0364	0.0352
	Case 5	0.0405	0.0402	0.0406	0.0399	0.0403	0.0403	0.0392	0.0376	0.0370

of Example 3.5 in Section 6.2.

Group III. *The "classical" Wald, QLR, and score test statistics based upon* W_n^0*,* QLR_n^0*, and* S_n^0. They are computed in the standard way. The same critical value from the chi-squared distribution is used for the three tests.

The weighting matrices in W_n^0, W_n, and W_n^1 are set as the estimators of the variance covariance matrix of the asymptotic distribution of $\widetilde{\theta}$, the ordinary least square estimator of θ, except that the ones for W_n^0 and W_n^1 are calculated using $\widetilde{\theta}$, and the one for W_n uses $\widehat{\theta}$. The matrices $\Sigma_{S,n}$ and $\Sigma_{S^1,n}$ in S_n and S_n^1 are set to satisfy the equalities in Theorem 8.1 (ii) and (i).

9.1.3 Results

The results in this section are based on the sample size $n = 300$ and 5000 Monte Carlo replications. The nominal size is $\alpha = 5\%$.

Table II reports the finite sample size performance. Several observations emerge. First, for DGP A, tests based upon W_n^0, W_n^1, QLR_n^0, and S_n^0 have similar sizes, S_n and S_n^2 tests have the best performance, and all have sizes closer to the nominal size than the three tests in Group I. Second, for DGP B, tests in Group I outperform all the other tests. Tests in Group II have similar size performance and all are slightly under sized than tests in Group I. The "classical" tests in Group III are severely under sized for DGP B. Third, comparing results across the two DGPs, we see that the size performance of the tests in both Groups II and III are very sensitive to different

Table III: Finite sample size-corrected power, DGP A

	$\beta_1 = -0.05$								
	Group I			Group II			Group III		
	W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0
Case 1	0.1665	0.1640	0.1658	0.1799	0.1783	0.1784	0.1777	0.1774	0.1769
Case 2	0.1694	0.1657	0.1671	0.1806	0.1775	0.1777	0.1756	0.1751	0.1743
Case 3	0.1852	0.1814	0.1804	0.1791	0.1762	0.1763	0.1747	0.1741	0.1723
Case 4	0.1690	0.1648	0.1666	0.1800	0.1776	0.1777	0.1732	0.1728	0.1709
Case 5	0.1698	0.1681	0.1672	0.1815	0.1788	0.1788	0.1769	0.1747	0.1716
	$\beta_1 = -0.1$								
	Group I			Group II			Group III		
	W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0
Case 1	0.6167	0.6155	0.6146	0.6280	0.6273	0.6272	0.6230	0.6224	0.6169
Case 2	0.6195	0.6202	0.6189	0.6312	0.6300	0.6299	0.6256	0.6243	0.6231
Case 3	0.6382	0.6361	0.6346	0.6296	0.6291	0.6291	0.6273	0.6248	0.6229
Case 4	0.6196	0.6189	0.6183	0.6249	0.6241	0.6240	0.6219	0.6218	0.6210
Case 5	0.6222	0.6216	0.6212	0.6271	0.6259	0.6260	0.6238	0.6236	0.6217
	$\beta_1 = -0.15$								
	Group I			Group II			Group III		
	W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0
Case 1	0.9280	0.9290	0.9258	0.9408	0.9412	0.9412	0.9363	0.9360	0.9336
Case 2	0.9336	0.9342	0.9322	0.9441	0.9444	0.9444	0.9397	0.9372	0.9372
Case 3	0.9494	0.9472	0.9478	0.9422	0.9425	0.9424	0.9372	0.9354	0.9350
Case 4	0.9276	0.9275	0.9251	0.9378	0.9383	0.9383	0.9336	0.9335	0.9230
Case 5	0.9343	0.9338	0.9333	0.9424	0.9427	0.9427	0.9396	0.9380	0.9369

distributions of the error term ε. In contrast, the size performance of the uniform tests in Group I are robust to different distributions of the error term ε which is a desirable property.

Tables III and IV present the finite sample size-corrected powers of different tests. First, for both DGPs and different values of β_1, there is no significant difference among tests in the same group. Second, for both DGPs, the power for all tests increases as the value of β_1 deviates more from the null value. Third, when the error follows the Gaussian distribution, all tests perform comparably with the "classical" tests performing slightly better except in Case 3, where three of the four inequalities are binding (under the null hypothesis). In DGP A, when only a few inequalities are binding or close-to-binding, the advantage of employing the prior information through the Bonferroni-type critical value is not apparent. On the other hand, when many inequalities are binding, employing prior information in the maintained hypothesis

Table IV: Finite sample size-corrected power, DGP B

	$\beta_1 = -0.05$								
	Group I			Group II			Group III		
	W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0
Case 1	0.1476	0.1532	0.1453	0.1532	0.1519	0.1518	0.1467	0.1429	0.1439
Case 2	0.1647	0.1664	0.1632	0.1514	0.1511	0.1511	0.1433	0.1420	0.1415
Case 3	0.1715	0.1742	0.1698	0.1535	0.1528	0.1526	0.1429	0.1411	0.1408
Case 4	0.1633	0.1657	0.1604	0.1502	0.1476	0.1477	0.1411	0.1402	0.1366
Case 5	0.1641	0.1672	0.1612	0.1511	0.1487	0.1487	0.1447	0.1419	0.1402
	$\beta_1 = -0.1$								
	Group I			Group II			Group III		
	W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0
Case 1	0.5780	0.5824	0.5762	0.5834	0.5781	0.5780	0.5745	0.5732	0.5716
Case 2	0.5929	0.5956	0.5889	0.5811	0.5834	0.5834	0.5702	0.5677	0.5649
Case 3	0.5987	0.6014	0.5943	0.5877	0.5841	0.5841	0.5719	0.5702	0.5655
Case 4	0.5963	0.5976	0.5936	0.5801	0.5827	0.5826	0.5752	0.5745	0.5716
Case 5	0.5928	0.5982	0.5901	0.5813	0.5777	0.5777	0.5738	0.5695	0.5674
	$\beta_1 = -0.15$								
	Group I			Group II			Group III		
	W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0
Case 1	0.8630	0.8609	0.8593	0.8705	0.8681	0.8681	0.8627	0.8592	0.8575
Case 2	0.8836	0.8772	0.8788	0.8736	0.8759	0.8759	0.8604	0.8586	0.8571
Case 3	0.8971	0.8968	0.8954	0.8807	0.8767	0.8767	0.8695	0.8629	0.8607
Case 4	0.8928	0.8923	0.8902	0.8593	0.8618	0.8618	0.8541	0.8534	0.8502
Case 5	0.8710	0.8751	0.8721	0.8586	0.8553	0.8553	0.8524	0.8471	0.8453

does increase the finite sample power. This observation is consistent for all values of β_1 in DGP A. Fourth, for DGP B, the tests in Group I have the highest power followed by tests in Group II in all cases except Case 1, when only one inequality is binding and there is no close-to-binding inequalities under the null hypothesis. This is strong evidence that for skewed error distributions, it is important to take into account prior information in the maintained hypothesis through the Bonferroni-type critical value.

In summary, the simulation results reveal the superior performance of the tests in Group I based on W_n, QLR_n, and S_n^1: their size performance is robust to the error distribution, their power is comparable to tests in Groups II and III for DGP A and higher for DGP B. The tests in Group II based on W_n^1, S_n, and S_n^2 perform better than the "classical" tests in Group III.

9.2 An Empirical Illustration

Using original and representative survey data, Autor and Handel (2013) conduct a comprehensive study on the interaction among human capital, job tasks, and wages. Their paper includes both a conceptual framework on the causal links between human capital endowments, occupational assignment, job tasks, and wages and empirical estimations and tests. Focusing on regressions on wage differentials related to job tasks and human capital, we apply our tests and compare the results with the "classical" ones. Specifically, we consider the following three regressions in Autor and Handel (2013) on the log hourly wages: for $i = 1, ..., n$,

$$\begin{aligned}
&\text{Model 1:}\ \log Wage_i = \beta' E_i + \mu' X_i + \varepsilon_i, \\
&\text{Model 2:}\ \log Wage_i = \beta' E_i + \zeta' T_i + \mu' X_i + \varepsilon_i, \text{ and} \\
&\text{Model 3:}\ \log Wage_i = \beta' E_i + \zeta' T_i + \mu' X_i + \vartheta' Z_i + \varepsilon_i,
\end{aligned}$$

where $\beta \equiv (\beta_1, \ldots, \beta_4)'$, $E_i \equiv (E_{1,i}, \ldots, E_{4,i})'$, $\zeta \equiv (\zeta_1, \zeta_2, \zeta_3)'$, and $T_i \equiv (T_{1,i}, T_{2,i}, T_{3,i})'$. For $j = 1, \ldots, 4$, $E_{j,i}$ is the dummy variable indicating the education level, with $E_{1,i}$ being "Less than high school", $E_{2,i}$ being "Some college", $E_{3,i}$ being "College", and $E_{4,i}$ being "Postcollege". T_i vector denotes the demand of different tasks, with $T_{1,i}$ being "Abstract", $T_{2,i}$ being "Routine", and $T_{3,i}$ being "Manual". X_i is a vector of demographic variables, and Z_i includes 240 occupation dummy variables. The reference group for the regressions is high school graduates. The data source is a module of Princeton Data Improvement Initiative survey (PDII) that collects data on different types of tasks that workers regularly perform during their work. The sample size is $n = 1333$. We follow the procedure in Autor and Handel (2013) to combine items from the PDII to elicit information on the demand of three tasks: Abstract, Routine, and Manual. For instance, the Abstract job demand is calculated by combing four items in PDII into a standardized scale using the first component of a principal components analysis. The four items are: the length of longest document typically read as part of the job, frequency of mathematics tasks involving high school or higher mathematics, frequency of problem-solving tasks requiring at least 30 minutes to find a good solution, and proportion of workday managing or supervising other workers.

By interpreting the coefficients on education level as the compensating differentials for income forgone while attending school (Mincer (1974)), we expect $\beta_1 \leq 0 \leq \beta_2$ and a non-descending ordering of β_2 to β_4 under the rationality assumption. By incorporating the information from the economic theory, the parameter space is expressed as $\Theta = \left\{(\beta', \mu')' : \mathscr{R}_w \beta \geq \mathbf{0}\right\}$ for Model 1, $\Theta = \left\{(\beta', \zeta', \mu')' : \mathscr{R}_w \beta \geq \mathbf{0}\right\}$ for Model 2, and $\Theta = \left\{(\beta', \zeta', \mu', \vartheta')' : \mathscr{R}_w \beta \geq \mathbf{0}\right\}$ for Model 3, with

$$\mathscr{R}_w = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

Appendix S.2 provides primitive conditions for these linear regression models. We first conduct tests on the point null hypothesis to investigate the significance of β_1 and β_2:

$$H_0 : \beta_j = 0 \text{ against } H_1 : \beta_j \neq 0 \text{ for } j = 1, 2.$$

Under the maintained hypothesis, it holds that $\beta_1 \leq 0$ and $\beta_2 \geq 0$. Since the reference group is high school graduates, $\beta_1 = 0$ corresponds to the penalty of having education levels less than high school being zero, and $\beta_2 = 0$ means that the monetary return of attending some college is zero. For $H_0 : \beta_1 = 0$, the first inequality in $\mathscr{R}_w \beta \geq \mathbf{0}$ binds. The remaining three inequalities are undetermined and the implicit nuisance parameter is simply $(\beta_2, \beta_3, \beta_4)$. Similarly, for $H_0 : \beta_2 = 0$, the second inequality is binding and the implicit nuisance parameter is $(\beta_1, \beta_3, \beta_4)$. We compare the tests developed in the paper with the "classical" ones and the standard t test.

The results are summarized in Table V. We see that the "classical" tests in Group III provide the same conclusion as the standard t test. For the same hypothesis, the "classical" tests suggest different conclusions for different models. All four tests reject $H_0 : \beta_1 = 0$ at 5% in Model 1, and fail to reject it at 10% in Model 2 and 3. The standard OLS estimates β_1 to be 0.1 in Model 1. Based on the model specification, β_1 being positive indicates that on average people who do not finish high school receive higher wages than people with high school degrees. This clearly violates the economic theory. The "classical" tests and t test reject $\beta_1 = 0$ in Model 1, which indicates that β_1 should be positive considering that the estimator of β_1 is positive. On the other hand, our tests take the economic theory into account and fail to reject $H_0 : \beta_1 = 0$ at 10% level in all models. When testing $\beta_2 = 0$, the "classical" tests in Group III and t tests provide different results for different models, while our tests suggest that β_2 is significant no matter which model is used. In fact, all six tests developed in the

Table V: Significance results

		Group I			Group II			Group III			
		W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0	t test
	Model 1	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\ast\ast$	$\ast\ast$	$\ast\ast$	$\ast\ast$
$\beta_1 = 0$	Model 2	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
	Model 3	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
		Group I			Group II			Group III			
		W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0	t test
	Model 1	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$
$\beta_2 = 0$	Model 2	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast$	$\ast\ast$	$\ast\ast$	$\ast\ast$
	Model 3	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\star$	$\star$	$\star$	$\star$
		Group I			Group II			Group III			
		W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0	t test
$\zeta_2 = 0$	Model 2	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$
	Model 3	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
		Group I			Group II			Group III			
		W_n	QLR_n	S_n^1	W_n^1	S_n	S_n^2	W_n^0	QLR_n^0	S_n^0	
	Model 1	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	$\ast\ast\ast$	
Joint Hypo.	Model 2	$\ast$	$\ast\ast$	$\ast$	$\ast$	$\ast$	$\ast$	$\ast\ast$	$\ast\ast$	$\ast\ast$	
	Model 3	$\ast$	$\ast$	$\ast$	$\ast$	$\ast$	$\ast$	$\ast\ast$	$\ast\ast$	$\ast\ast$	

Notes: $\ast\ast\ast$ denotes to reject at 1% level; $\ast\ast$ denotes to reject at 5% level; $\ast$ denotes to reject at 10% level; and $\star$ denotes fail to reject at 10% level.

paper give the same testing results for the two hypotheses in three models and the results are consistent between the models. We also test the null hypothesis of $\beta_j = 0$ for $j = 3, 4$. The tests in all three groups give the same result, which is also the same as the one in Autor and Handel (2013).

Under the maintained hypothesis on β, researchers can test the significance of individual parameter ζ_2:

$$H_0 : \zeta_2 = 0 \text{ against } H_1 : \zeta_2 \neq 0.$$

Because the null hypothesis is imposed on the parameter different from β, the implicit nuisance parameter is β and $\Lambda_R = \mathbb{R}$. The results is summarized in Table V. The "classical" tests in Group III and t test reject the null hypothesis at 1% level in Model 2, but fail to reject it at 10% level in Model 3. At the same time, the tests in Group II provide the same results as the ones in Group III. Because the null hypothesis and maintained hypothesis are imposed on different parameters, information in Θ is lost due to projection. The Wald and score tests in Group II become or approximate their "classical" counterparts. On the other hand, the tests in Group I fully exploit the information in the maintained hypothesis and suggest consistent results for both Model 2 and Model 3. The tests in all three groups provide the same conclusions for $H_0 : \zeta_1 = 0$ and $H_0 : \zeta_3 = 0$ and the results are consistent for both Model 2 and 3.

Researchers may also be interested in testing joint hypotheses, such as $H_0 : \beta_1 = 0$ and $\beta_2 + 0.1 = \beta_4$. Under this null hypothesis, having a high school degree is not useful and the benefit of having a postcollege degree is a ten percentage increase in wage comparing to having some college education. Finding the implicit nuisance parameter is more involved in this case, where the detail can be found in Example 3.5 in Section 3.2. The continuation of Example 3.5 in Section 6.2 discusses how to obtain the polytope projection for the asymptotic distributions of the test statistics in Group II. The results of different tests are collected in the bottom part of Table V. While all the tests suggest the same for Model 1, their conclusions differ for Model 2 and 3. Except for the test based upon QLR_n in Model 2, the tests in Groups I and II tend to not reject the null hypothesis compared with the "classical" tests, which reject the null hypothesis at 10% level for Model 2 and 3.

10 Concluding Remarks

In this paper, we have proposed the concept of an implicit nuisance parameter for testing the null hypothesis of linear equality constraints against the two-sided alternative hypothesis when the true parameter is subject to equality and inequality constraints in the maintained hypothesis. Moreover we have proposed an approach for identifying the implicit nuisance parameter. This opens the door for uniform inference in a wide range of models/sample information. Using our approach, we have developed asymptotically uniformly valid Wald, QLR, and score tests in the extremum set-up for non-trending data and studied their power and equivalence results.

In a companion paper, we are working on asymptotically uniformly valid Wald, QLR, and score tests for trending data models in Andrews (2001) under the maintained hypothesis that parameters in these models are subject to equality and inequality constraints. Our approach for identifying implicit nuisance parameters can also be used for testing the null hypotheses of linear *inequality* constraints of the form H_0: $R\theta^* \geq r$ against H_1: $R\theta^* < r$ like Wolak (1987, 1989, 1991) and of linear equality constraints against one-sided alternatives such as H_0: $R\theta^* = r$ against H_1: $R\theta^* > r$. It is also worthwhile to investigate the possibility of extending our approach to allow for non-linear equality and inequality constraints.

S.1 Technical Proofs

We introduce several notations and definitions that will be used to prove Lemma 4.2. Decompose θ into three parts: $\theta \equiv (\theta_1', \theta_2')' \equiv \left(\theta_{1,b}', \theta_{1,nb}', \theta_2'\right)'$, where $\theta_{1,b} \in \mathbb{R}^{J_b}$, $\theta_{1,nb} \in \mathbb{R}^{J-J_b}$ and $\theta_2 \in \mathbb{R}^{l-J}$. Under H_0, we have $\theta_1^* = r \equiv (\mathbf{0}', r_{nb}')'$. For $\Theta = \left\{\theta \in \mathbb{R}^l : \theta \geq \mathbf{0}\right\}$ and $\theta_n = \left(r', \theta_{2,n}'\right)' \in \mathbb{R}^l_{\geq 0}$,

$$b_n\left(\Theta - \theta_n\right) = \left\{\theta \in \mathbb{R}^l : \theta_{1,b} \geq \mathbf{0},\, \theta_{1,nb} + b_n r_{nb} \geq \mathbf{0} \text{ and } \theta_2 + b_n\theta_{2,n} \geq \mathbf{0}\right\}.$$

With $\lim_{n\to\infty} b_n\theta_{2,n} = c \in \overline{\mathbb{R}}^{l-J}_{\geq 0}$, let $c_F \in \mathbb{R}^{l_F}_{\geq 0}$ be the subvector of c such that c_F contains all the finite elements of c and $\Upsilon \in \mathbb{R}^{l_F\times(l-J)}$ be the matrix such that $\Upsilon c = c_F$. Define

$$\begin{aligned} \Lambda_n &\equiv \left\{\theta \in \mathbb{R}^l : \theta_{1,b} \geq \mathbf{0} \text{ and } \Upsilon\theta_2 + \Upsilon b_n\theta_{2,n} \geq \mathbf{0}\right\}, \text{ and} \\ \Lambda &\equiv \left\{\theta \in \mathbb{R}^l : \theta_{1,b} \geq \mathbf{0} \text{ and } \Upsilon\theta_2 + \Upsilon c \geq \mathbf{0}\right\}. \end{aligned}$$

Since the inequality $a + (+\infty) \geq 0$ holds for any $a \in \mathbb{R}$, Λ can be alternatively rewritten as

$$\Lambda = \left\{\theta \in \mathbb{R}^l : \theta_{1,b} \geq \mathbf{0} \text{ and } \theta_2 + c \geq \mathbf{0}\right\}.$$

For any set $\varGamma \subset \mathbb{R}^l$ and $z \in \mathbb{R}^l$, define

$$\begin{aligned} \operatorname{dist}\left(z, \varGamma\right) &\equiv \inf_{\lambda\in\varGamma} \left\|z - \lambda\right\|, \text{ and} \\ \operatorname{dist}_n\left(z, \varGamma\right) &\equiv \inf_{\lambda\in\varGamma} \left(\left(z-\lambda\right)' \mathscr{T}_n \left(z-\lambda\right)\right)^{1/2}. \end{aligned}$$

Proof of Lemma 4.1: From Lemma 2.1, we have $b_n\left(\widehat{\theta} - \theta^*\right) \xrightarrow{d} \Psi$. By the definition of W_n and Assumption 4.1, the result of the lemma follows. ■

Lemma S.1.1. *If $Z_n = O_p(1)$ and Assumption 4.3 holds, then*

$$\inf_{\lambda\in b_n(\Theta-\theta_n)} q_n(\lambda) = \inf_{\lambda\in\Lambda_n} q_n(\lambda) + o_p(1).$$

Proof: By definition, $b_n\left(\Theta - \theta_n\right)$ is contained in Λ_n. Since $q_n(\cdot)$ takes a quadratic form and Λ_n is closed, there exists $\lambda_n \in \Lambda_n$ such that

$$\arg\inf_{\lambda\in\Lambda_n} q_n(\lambda) = q_n(\lambda_n).$$

For any give δ, we have

$$\Pr\left(\left|\inf_{\lambda\in b_n(\Theta-\theta_n)} q_n(\lambda) - \inf_{\lambda\in\Lambda_n} q_n(\lambda)\right| > \delta\right) \leq \Pr\left(\inf_{\lambda\in b_n(\Theta-\theta_n)} q_n(\lambda) \neq \inf_{\lambda\in\Lambda_n} q_n(\lambda)\right)$$
$$\leq \Pr\left(\lambda_n \in \Lambda_n \backslash b_n(\Theta-\theta_n)\right).$$

Thus, it suffices to show that $\Pr\left(\lambda_n \in \Lambda_n \backslash b_n(\Theta-\theta_n)\right) \to 0$. Let $c_I \in \mathbb{R}^{l_I}_{\geq 0}$ be the subvector of c such that c_I contains all the infinite elements of c and $\Upsilon_I \in \mathbb{R}^{l_I\times(l-J)}$ be the matrix such that $\Upsilon_I c = c_I$. By definition,

$$\Lambda_n \backslash b_n(\Theta-\theta_n) = \left\{\theta \in \mathbb{R}^l : \theta_{1,b} \geq \mathbf{0},\ \Upsilon\theta_2 + \Upsilon b_n\theta_{2,n} \geq \mathbf{0},\ \theta_{1,nb} < -b_n r_{nb} \text{ and } \Upsilon_I\theta_2 < -\Upsilon_I b_n \theta_{2,n}\right\},$$

where each element in $-b_n r_{nb}$ and $-\Upsilon_I b_n\theta_{2,n}$ goes to negative infinity when $n\to\infty$. Since $Z_n = O_p(1)$, for any $\varepsilon > 0$, there exists some $M < \infty$ and N, such that

$$\Pr\left(\mathrm{dist}^2_n(Z_n, \{\mathbf{0}\}) > M\right) < \varepsilon, \text{ for } \forall n > N.$$

There exists N_Θ, such that for $n > N_\Theta$, we have $\mathrm{dist}^2_n(\lambda, \{\mathbf{0}\}) > 2M$ for any $\lambda \in \Lambda_n \backslash b_n(\Theta-\theta_n)$. Therefore, for $n > \max(N_\Theta, N)$,

$$\Pr\left(\mathrm{dist}^2_n(Z_n, \{\mathbf{0}\}) > \mathrm{dist}^2_n(Z_n, \Lambda_n\backslash b_n(\Theta-\theta_n))\right) \leq \Pr\left(\mathrm{dist}^2_n(Z_n,\{\mathbf{0}\}) > M\right) < \varepsilon.$$

Since $\{\mathbf{0}\} \in b_n(\Theta-\theta_n)$, we have

$$\Pr\left(\lambda_n \in \Lambda_n\backslash b_n(\Theta-\theta_n)\right) \leq \Pr\left(\mathrm{dist}^2_n(Z_n,\{\mathbf{0}\}) > \mathrm{dist}^2_n(Z_n, \Lambda_n\backslash b_n(\Theta-\theta_n))\right) < \varepsilon$$

for sufficiently large n. We conclude the lemma because ε is arbitrary. ■

Lemma S.1.2. *For any $\omega_n \in \mathcal{W}_0$, if Assumptions 4.2-4.4 hold, then*

$$q_n\left(b_n\left(\widehat{\theta}-\theta_n\right)\right) = \inf_{\lambda\in b_n(\Theta-\theta_n)} q_n(\lambda) + o_p(1).$$

Proof: By the definition of the drifting sequence, $\theta_n \to \theta_\omega$ as $n\to\infty$. Assumption 4.4 implies that $\widehat{\theta}-\theta_\omega = \left(\widehat{\theta}-\theta_n\right) + (\theta_n - \theta_\omega) = o_p(1)$. Together with Assumption 4.2, we obtain that

$$l_n\left(\widehat{\theta}\right) = l_n(\theta_n) + \frac{1}{2}Z_n'\mathscr{I}_n Z_n - \frac{1}{2}q_n\left(b_n\left(\widehat{\theta}-\theta_n\right)\right) + o_p(1). \qquad \text{(S.17)}$$

Let $\widehat{\theta}_q \in \Theta$ be the approximate minimizer of $q_n\left(b_n\left(\theta-\theta_n\right)\right)$, such that

$$
\begin{aligned}
q_n\left(b_n\left(\widehat{\theta}_q-\theta_n\right)\right) &= \inf_{\theta\in\Theta} q_n\left(b_n\left(\theta-\theta_n\right)\right)+o_p\left(1\right) \\
&= \inf_{\lambda\in b_n\left(\Theta-\theta_n\right)} q_n\left(\lambda\right)+o_p\left(1\right).
\end{aligned} \tag{S.18}
$$

Since $\theta_n \in \Theta$, it holds that

$$
\begin{aligned}
\left\|\mathscr{T}_n^{1/2} b_n\left(\widehat{\theta}_q-\theta_n\right)-\mathscr{T}_n^{1/2} Z_n\right\| &= q_n\left(b_n\left(\widehat{\theta}_q-\theta_n\right)\right) \leq q_n\left(\mathbf{0}\right)+o_p\left(1\right) \\
&= Z_n' \mathscr{T}_n Z_n+o_p\left(1\right)=O_p\left(1\right).
\end{aligned}
$$

By the triangle inequality and Assumption 4.3, we have

$$
\left\|\mathscr{T}_n^{1/2} b_n\left(\widehat{\theta}_q-\theta_n\right)\right\| \leq \left\|\mathscr{T}_n^{1/2} b_n\left(\widehat{\theta}_q-\theta_n\right)-\mathscr{T}_n^{1/2} Z_n\right\|+\left\|\mathscr{T}_n^{1/2} Z_n\right\|=O_p\left(1\right).
$$

Together with $\mathscr{T}$ being non-singular with probability one, we obtain $b_n\left(\widehat{\theta}_q-\theta_n\right)=O_p\left(1\right)$. The same argument applies and we have

$$
l_n\left(\widehat{\theta}_q\right)=l_n\left(\theta_n\right)+\frac{1}{2} Z_n' \mathscr{T}_n Z_n-\frac{1}{2} q_n\left(b_n\left(\widehat{\theta}_q-\theta_n\right)\right)+o_p\left(1\right). \tag{S.19}
$$

Combing Equation (S.17) and (S.19), and the definitions of $\widehat{\theta}$ and $\widehat{\theta}_p$ in (4) and (S.18), it holds that

$$
\begin{aligned}
o_p\left(1\right) &\leq l_n\left(\widehat{\theta}\right)-l_n\left(\widehat{\theta}_q\right) \\
&= \frac{1}{2} q_n\left(b_n\left(\widehat{\theta}_q-\theta_n\right)\right)-\frac{1}{2} q_n\left(b_n\left(\widehat{\theta}-\theta_n\right)\right)+o_p\left(1\right) \leq o_p\left(1\right).
\end{aligned}
$$

The lemma holds by applying the definition of $\widehat{\theta}_p$. ■

Lemma S.1.3. *For any $\omega_n \in \mathcal{W}_0$, if Assumption 4.3 hold, then*

$$
\inf_{\lambda\in b_n\left(\Theta-\theta_n\right)} q_n\left(\lambda\right)=\inf_{\lambda\in\Lambda} q_n\left(\lambda\right)+o_p\left(1\right).
$$

Proof: By definition of the quadratic function, we have

$$
\inf_{\lambda\in b_n\left(\Theta-\theta_n\right)} q_n\left(\lambda\right)=\operatorname{dist}_n^2\left(Z_n, b_n\left(\Theta-\theta_n\right)\right)
$$

and

$$
\inf_{\lambda\in\Lambda} q_n\left(\lambda\right)=\operatorname{dist}_n^2\left(Z_n, \Lambda\right).
$$

Since $Z_n = O_p(1)$ by Assumption 2.2, Lemma S.1.1 provides that

$$\text{dist}_n^2 \left(Z_n, b_n \left(\Theta - \theta_n\right)\right) = \text{dist}_n^2 \left(Z_n, \Lambda_n\right) + o_p(1).$$

There exists some $\lambda_n \in \Lambda_n$ such that

$$\text{dist}_n \left(Z_n, \Lambda_n\right) = \text{dist}_n \left(Z_n, \{\lambda_n\}\right) + o_p(1).$$

Since Λ_n is a translation of Λ: $\Lambda_n = \Lambda + (\Upsilon c - \Upsilon b_n \theta_{2,n})$, the Hausdorff distance between Λ_n and Λ, denoted as $d_H(\Lambda_n, \Lambda)$, satisfies the inequality $d_H(\Lambda_n, \Lambda) \leq \|\Upsilon b_n \theta_{2,n} - \Upsilon c\|$. By the definition of c, $\|\Upsilon b_n \theta_{2,n} - \Upsilon c\| \to 0$ as $n \to \infty$. Therefore, we have $\text{dist}(\lambda_n, \Lambda) = o(1)$. Further with Assumption 4.3 on $\mathscr{T}_n$, we have $\text{dist}_n(\lambda_n, \Lambda) = o_p(1)$. Define λ_Λ analogously with Λ_n replaced by Λ, it holds that $\text{dist}_n(\lambda_\Lambda, \Lambda_n) = o_p(1)$, following the same argument.

By the triangle inequality,

$$\begin{aligned}
&\text{dist}_n \left(Z_n, \Lambda\right) - \text{dist}_n \left(Z_n, \Lambda_n\right) \\
\leq &\text{dist}_n \left(Z_n, \{\lambda_n\}\right) + \text{dist}_n \left(\lambda_n, \Lambda\right) - \text{dist}_n \left(Z_n, \Lambda_n\right) \\
= &\text{dist}_n \left(\lambda_n, \Lambda\right) + o_p(1) = o_p(1).
\end{aligned}$$

Similarly, we have

$$\text{dist}_n \left(Z_n, \Lambda_n\right) - \text{dist}_n \left(Z_n, \Lambda\right) \leq o_p(1).$$

Therefore, $\text{dist}_n(Z_n, \Lambda_n) - \text{dist}_n(Z_n, \Lambda) = o_p(1)$, and the lemma follows. ■

Lemma S.1.4. *Let $\widehat{\lambda}$ be the minimizer of $q_n(\lambda)$ over Λ:*

$$q_n\left(\widehat{\lambda}\right) = \min_{\lambda \in \Lambda} q_n(\lambda).$$

If Assumption 4.2-4.4 hold, then $b_n\left(\widehat{\theta} - \theta_n\right) = \widehat{\lambda} + o_p(1)$.

Proof: Let $\lambda^* \in \Lambda$ be such that $\left\|b_n\left(\widehat{\theta} - \theta_n\right) - \lambda^*\right\| = \overrightarrow{d}_H\left(b_n\left(\widehat{\theta} - \theta_n\right), \Lambda\right)$. Since Λ is a convex cone, λ^* is unique by Perlman (1969). Moreover, the closeness of Λ and the quadratic form of $q_n(\cdot)$ provide that $\widehat{\lambda}$ is well defined. By the argument in the proof of Lemma S.1.3, $d_H(\Lambda_n, \Lambda) = o(1)$. It holds that

$$\left\|b_n\left(\widehat{\theta} - \theta_n\right) - \lambda^*\right\| = \text{dist}\left(b_n\left(\widehat{\theta} - \theta_n\right), \Lambda\right) = o(1),$$

because $b_n\left(\widehat{\theta}-\theta_n\right)\in b_n\left(\Theta-\theta_n\right)\subseteq\Lambda_n$. It remains to show that $\left\|\lambda^*-\widehat{\lambda}\right\|=o_p\left(1\right)$. The rest follows from the analogous argument in the proof for Theorem 2 in Andrews (1997). All the required conditions are provided in Assumptions 4.2-4.4, Lemma S.1.2 and S.1.3. ■

Proof of Lemma 4.2: Since Λ is convex and closed, we can write $\widehat{\lambda}$ defined in Lemma S.1.4 as

$$\widehat{\lambda}=\min_{\lambda\in\Lambda}q_n\left(\lambda\right)\equiv h\left(b_n^{-1}Dl_n\left(\theta_n\right),\mathscr{T}_n\right).$$

By Assumption S.1.4, $\left(b_n^{-1}Dl_n\left(\theta_n\right),\mathscr{T}_n\right)\xrightarrow{d}\left(G_\omega,\mathscr{T}_\omega\right)$ for any drifting parameter sequences $\omega_n\in\mathcal{W}_0$ with limit $\omega\in\mathcal{W}_0$. The continuity of $h\left(\cdot,\cdot\right)$ with respect to the two arguments implies that

$$\begin{aligned}\widehat{\lambda}=h\left(b_n^{-1}Dl_n\left(\theta_n\right),\mathscr{T}_n\right)&\xrightarrow{d}h\left(G_\omega,\mathscr{T}_\omega\right)\\&=\arg\min_{\lambda\in\Lambda}\left(\lambda-\mathscr{T}_\omega^{-1}G_\omega\right)'\mathscr{T}_\omega\left(\lambda-\mathscr{T}_\omega^{-1}G_\omega\right).\end{aligned}$$

Applying Lemma S.1.4, we obtain that

$$\begin{aligned}b_n\left(\widehat{\theta}-\theta_n\right)&\xrightarrow{d}\arg\min_{\lambda\in\Lambda}\left(\lambda-\mathscr{T}_\omega^{-1}G_\omega\right)'\mathscr{T}_\omega\left(\lambda-\mathscr{T}_\omega^{-1}G_\omega\right)\\&=\arg\min_{\lambda\in\Lambda}q_\omega\left(\lambda\right)=\arg\min_{\lambda}\left[q_\omega\left(\lambda\right)+\phi_\omega\left(\lambda\right)\right].\end{aligned}$$

If further Assumption 4.5 holds, then $W_n\xrightarrow{d}\left(R\Psi_{W,\omega}\right)'\left(R\Sigma_\omega R'\right)^{-1}\left(R\Psi_\omega\right)$ by the continuous mapping theorem. ■

Proof of Theorem 4.1: We prove the theorem by verifying assumptions in McCloskey (2017). Notice that the distribution $\left(R\Psi_\omega\right)'\left(R\Sigma_{W,\omega}R'\right)^{-1}\left(R\Psi_\omega\right)$ is finite with probability 1 for all $c\in\overline{\mathbb{R}}_{\geq0}^{l-J}$ and $\pi_{W,\omega}\in\overline{\Pi}_W$. Assumption PS in McCloskey (2017) is trivially satisfied. By the expression in Lemma 4.2, $\mathcal{C}^W_{c,\pi_{W,\omega}}\left(1-\alpha\right)$ is continuous in c and $\pi_{W,\omega}$. Together with the assumption in Theorem 4.1, Assumption Cont in McCloskey (2017) is satisfied. Let Assumption DS in McCloskey (2017) hold for $\widetilde{c}\xrightarrow{d}c+\mathscr{T}_{2,\omega}^{-1}G_{2,\omega}$. For any $\omega\in\mathcal{W}_0$, the confidence set $\widetilde{ES}\left(\tau\right)$ satisfies that $\lim_{n\to\infty}\Pr_\omega\left(\mathscr{T}_{2,\omega}^{-1}G_{2,\omega}\in\widetilde{ES}\left(\tau\right)\right)\geq1-\tau$. This and the fact that $c\in\overline{\mathbb{R}}_{\geq0}^{l-J}$ imply that $\lim_{n\to\infty}\Pr_\omega\left(c\in\widetilde{I}_\tau\right)\geq1-\tau$, which fulfills Assumption CS in McCloskey (2017). It suffices to prove that Assumption DS in McCloskey (2017) is satisfied.

Lemma 4.2 provides that the asymptotic distribution of the test statistic W_n is $\left(R\Psi_\omega\right)'\left(R\Sigma_{W,\omega}R'\right)^{-1}\left(R\Psi_\omega\right)$ under the full parameter sequence $\left(\eta_n^u,\pi_{W,n},\xi_n\right)$. Since

$\widetilde{\theta}_2$ is an unconstrained estimator by definition, its asymptotic distribution can be obtained by applying Lemma 4.2 with

$$q_\omega(\lambda) = \left(\lambda - \mathscr{I}_{2,\omega}^{-1} G_{2,\omega}\right)' \mathscr{I}_{2,\omega} \left(\lambda - \mathscr{I}_{2,\omega}^{-1} G_{2,\omega}\right) \text{ and}$$
$$\phi_\omega(\lambda) = 0 \text{ for } \lambda \in \mathbb{R}^{l-J}.$$

Therefore, $b_n \left(\widetilde{\theta}_2 - \theta_{2,n}\right) = \widetilde{c} - b_n \theta_{2,n} \xrightarrow{d} \mathscr{I}_{2,\omega}^{-1} G_{2,\omega}$ and $\widetilde{c} \xrightarrow{d} c + \mathscr{I}_{2,\omega}^{-1} G_{2,\omega}$ for any parameter sequences $\omega_n \in \mathcal{W}_0$. We follow Lemma 2.1 in Andrews et al. (2011) to establish the equivalence of results under full sequences and subsequences provided that Assumption B2 in Andrews et al. (2011) holds. Therefore, the goal is to show that for any subsequence there exists a full sequence that has the same limit (possibly infinity) and has its subsequence equal to the original one. Denote the subsequence as $\left\{\eta_{p_n}^u, \pi_{W,p_n} : n \geq 1\right\}$ such that $\left(\sqrt{p_n}\eta_{p_n}^u, \pi_{W,p_n}\right) \to (c, \pi_{W,\omega})$. We aim to construct a full sequence $\left\{\eta_n^{u\star}, \pi_{W,n}^\star : n \geq 1\right\}$ satisfying that $\left(\sqrt{n}\eta_n^{u\star}, \pi_{W,n}^\star\right) \to (c, \pi_{W,\omega})$ and $\left(\eta_n^{u\star}, \pi_{W,n}^\star\right) = \left(\eta_{p_n}^u, \pi_{W,p_n}\right)$, $\forall n \geq 1$. To clarify the notation, let the full sequence be indexed by l: $\left\{\eta_l^{u\star}, \pi_{W,l}^\star : l \geq 1\right\}$. For $\forall l = p_n$, define $\left(\eta_l^{u\star}, \pi_{W,l}^\star\right) = \left(\eta_{p_n}^u, \pi_{W,p_n}\right)$; and for $\forall l \in (p_n, p_{n+1})$, define

$$\theta_{j+J,l}^\star = \begin{cases} \frac{\sqrt{p_n}\theta_{j+J,p_n}}{\sqrt{l}}, & \text{if } \sqrt{p_n}\theta_{j+J,p_n} \to c_j \in \mathbb{R}_{\geq 0} \\ \theta_{j+J,p_n}, & \text{if } \sqrt{p_n}\theta_{j+J,p_n} \to +\infty \end{cases}$$

for $j = 1, \ldots, l - J$ and $\pi_{W,l}^\star = \pi_{W,p_n}$. It is trivial that the constructed full sequence satisfies the second requirement that $\left(\eta_{p_n}^{u\star}, \pi_{W,p_n}^\star\right) = \left(\eta_{p_n}^u, \pi_{W,p_n}\right)$ for $\forall n \geq 1$. To see that the first requirement is also satisfied, please refer to page 225-226 in Cheng (2015) for a detailed derivation. ■

Proof of Theorem 4.2: It holds that

$$\begin{aligned} \widehat{\theta}_1 - r &= \widehat{\theta}_1 - \theta_1^* + \theta_1^* - r \\ &= \widehat{\theta}_1 - \theta_1^* + \theta_1^* - r, \end{aligned}$$

with $\widehat{\theta}_1 - \theta_1^* = o_p(1)$ by Assumption 2.3 and $\theta_1^* - r \neq \mathbf{0}$ by H_1. Since $R\Sigma_W R'$ is positive definite with probability one by Assumption 4.1, $\left(R\Sigma_W R'\right)^{-1}$ is also positive definite with probability one. Therefore, it holds that

$$b_n^{-2} W_n \xrightarrow{p} \left(\theta_1^* - r\right)' \left(R\Sigma_W R'\right)^{-1} \left(\theta_1^* - r\right) > 0$$

and W_n diverges to infinity when n goes to infinity with probability one. It remains to show that $CV_n^W(\alpha,\tau) = O_p(1)$. For any $c \geq \mathbf{0}$ in Lemma 4.2 and any $\pi_{W,\omega} \in \overline{\Pi}_W$, we have

$$\begin{aligned}\left\|\mathscr{T}_\omega^{1/2}\Psi_\omega - \mathscr{T}_\omega^{1/2}Z_\omega\right\| &= q_\omega(\Psi_\omega) \leq q_n(\mathbf{0}) \\ &= Z_\omega'\mathscr{T}_\omega Z_\omega = O_p(1).\end{aligned}$$

The triangular inequality implies that

$$\left\|\mathscr{T}_\omega^{1/2}\Psi_\omega\right\| \leq \left\|\mathscr{T}_\omega^{1/2}\Psi_\omega - \mathscr{T}_\omega^{1/2}Z_\omega\right\| + \left\|\mathscr{T}_\omega^{1/2}Z_\omega\right\| = O_p(1).$$

Since $\mathscr{T}_\omega$ is symmetric and non-singular with probability one, it's eigenvalue is not zero. Therefore, it holds that $\|\Psi_\omega\| = O_p(1)$ and $(R\Psi_\omega)'(R\Sigma_{W,\omega}R')^{-1}(R\Psi_\omega) = O_p(1)$. Assume that the sup in Definition (12) is achieved at $\widehat{c} \in \overline{\widetilde{I}}_{\alpha-\tau}$, where $\overline{\widetilde{I}}_{\alpha-\tau}$ is the closure of $\widetilde{I}_{\alpha-\tau}$, and $CV_n^W(\alpha,\tau) = \mathcal{C}^W_{\widehat{c},\widehat{\pi}_W}(1-\tau)$. Since $\widehat{c} \geq \mathbf{0}$ for any n, we conclude that for $\tau > 0$, $CV_n^W(\alpha,\tau) = O_p(1)$. ■

Lemma S.1.5. *For $\mathscr{R}_w^{nb}$ defined in Section 3.1, there exists some $\epsilon > 0$, such that $\mathscr{R}_w^{nb}\theta > \mathrm{r}_w^{nb} + \epsilon$ for all $\theta \in \Theta_0$.*

Proof: Assume the contrary. Then there must exist a sequence $\left(\mathscr{R}_w^{nb}\theta\right)_m \in \mathscr{R}_w^{nb}\Theta_0$ such that $\left(\mathscr{R}_w^{nb}\theta\right)_m \leq \mathrm{r}_w^{nb} + \epsilon_m$ for $\epsilon_m \equiv \frac{1}{m}$. By the definition of $\mathscr{R}_w^{nb}$, it holds that $\left(\mathscr{R}_w^{nb}\theta\right)_m > \mathrm{r}_w^{nb}$. Therefore, the converging subsequence of $\left(\mathscr{R}_w^{nb}\theta\right)_m$, denoted as $\left(\mathscr{R}_w^{nb}\theta\right)_{k_m}$, satisfies that $\mathrm{r}_w^{nb} < \left(\mathscr{R}_w^{nb}\theta\right)_{k_m} \leq \mathrm{r}_w^{nb} + \epsilon_{k_m}$ with the $\lim_{m\to\infty}\left(\mathscr{R}_w^{nb}\theta\right)_{k_m} = \mathrm{r}_w^{nb}$, because $\lim_{m\to\infty}\epsilon_m = 0$. Since the set Θ_0 is closed by definition, so is $\mathscr{R}_w^{nb}\Theta_0$. Therefore, there exists some $\mathscr{R}_w^{nb}\theta^\star \in \mathscr{R}_w^{nb}\Theta_0$ such that $\mathscr{R}_w^{nb}\theta^\star = \mathrm{r}_w^{nb}$. However, this contradicts with the definition of $\mathscr{R}_w^{nb}$. Therefore, the ϵ in the lemma always exists.

Proof of Lemma 4.3: For c defined in (13), let $c_F \in \mathbb{R}^{l_F}_{\geq 0}$ be the subvector of c such that c_F contains all the finite elements of c and $\Upsilon \in \mathbb{R}^{l_F \times l_u}$ be the matrix such that $\Upsilon c = c_F$. Define Λ_n and Λ as

$$\begin{aligned}\Lambda_n &\equiv \left\{\theta \in \mathbb{R}^l : \mathscr{R}_e\theta = \mathbf{0},\, \mathscr{R}_w^b\theta \geq \mathbf{0} \text{ and } \Upsilon\mathscr{R}_w^u\theta + \Upsilon\left(\mathscr{R}_w^u\theta_n - \mathrm{r}_w^u\right) \geq \mathbf{0}\right\} \text{ and} \\ \Lambda &\equiv \left\{\theta \in \mathbb{R}^l : \mathscr{R}_e\theta = \mathbf{0},\, \mathscr{R}_w^b\theta \geq \mathbf{0} \text{ and } \Upsilon\mathscr{R}_w^u\theta + \Upsilon c \geq \mathbf{0}\right\} \\ &= \left\{\theta \in \mathbb{R}^l : \mathscr{R}_e\theta = \mathbf{0},\, \mathscr{R}_w^b\theta \geq \mathbf{0} \text{ and } \mathscr{R}_w^u\theta + c \geq \mathbf{0}\right\}.\end{aligned}$$

It suffices to show that Λ_n, Λ and $b_n(\Theta - \theta_n)$ satisfy all the lemmas in Section S.1.

Decompose $\mathscr{R}_w\theta \geq b_n(\mathrm{r}_w - \mathscr{R}_w\theta_n)$ based the procedure in Section 3.1; define r_w^{nb} and r_w^b as the corresponding subvector of r_w. We obtain that

$$\begin{aligned}
b_n(\Theta - \theta_n) &= \left\{\theta \in \mathbb{R}^l : \mathscr{R}_e\theta = b_n(\mathrm{r}_e - \mathscr{R}_e\theta_n), \mathscr{R}_w\theta \geq b_n(\mathrm{r}_w - \mathscr{R}_w\theta_n)\right\} \\
&= \left\{\theta \in \mathbb{R}^l : \mathscr{R}_e\theta = b_n(\mathrm{r}_e - \mathscr{R}_e\theta_n), \mathscr{R}_w^{nb}\theta \geq b_n\left(\mathrm{r}_w^{nb} - \mathscr{R}_w^{nb}\theta_n\right)\right. \\
&\qquad \left.\mathscr{R}_w^b\theta \geq b_n\left(\mathrm{r}_w^b - \mathscr{R}_w^b\theta_n\right), \mathscr{R}_w^u\theta \geq b_n(\mathrm{r}_w^u - \mathscr{R}_w^u\theta_n)\right\}.
\end{aligned}$$

By incorporating the information in Θ_0, $b_n(\Theta - \theta_n)$ can be further simplified. Since $\theta_n \in \Theta$, we have $\mathscr{R}_e\theta_n = \mathrm{r}_e$; by the definition of the implicit equality, it holds that $\mathscr{R}_w^b\theta_n = \mathrm{r}_w^b$ for any $\theta_n \in \Theta_0$. By Lemma S.1.5, there exists some $\epsilon > 0$, such that $\mathscr{R}_w^{nb}\theta_n > \mathrm{r}_w^{nb} + \epsilon$ for all $\theta_n \in \Theta_0$. Thus $b_n\left(\mathrm{r}_w^{nb} - \mathscr{R}_w^{nb}\theta_n\right) < -b_n\epsilon$, where $-b_n\epsilon$ goes to negative infinity when $n \to \infty$. Since $\theta_n = \Gamma\theta_{f,n} + \gamma$, we have

$$\begin{aligned}
\lim_{n\to\infty} b_n(\mathscr{R}_w^u\theta_n - \mathrm{r}_w^u) &= \lim_{n\to\infty} b_n(\mathscr{R}_w^u(\Gamma\theta_{f,n} + \gamma) - \mathrm{r}_w^u) \\
&= \lim_{n\to\infty} b_n(\mathscr{R}_w^u\Gamma\theta_{f,n} - (\mathrm{r}_w^u - \mathscr{R}_w^u\gamma)) \\
&= \lim_{n\to\infty} b_n\left(\Gamma^u\eta_n^k - (\mathrm{r}_w^u - \mathscr{R}_w^u\gamma)\right) \\
&= c.
\end{aligned}$$

Thus, we've identified the equality constraints, binding inequality constraints, non-binding inequality constraints and undetermined inequality constraints with the associated limits. The same derivation in Section S.1 applies to $b_n(\Theta - \theta_n)$, Λ_n and Λ. The claimed result follows. ■

Proof of Theorem 4.3: The proof is essentially the same as the one of Theorem 4.1. ■

Proof of Theorem 4.4: The proof is similar to the one of Theorem 4.2. Since $R\widehat{\theta} - r \xrightarrow{p} R\theta^* - r \neq \mathbf{0}$ and $R\Sigma_W R'$ is positive definite with probability one by Assumption 4.1, we have

$$b_n^{-2}W_n \xrightarrow{p} (R\theta^* - r)'(R\Sigma_W R')^{-1}(R\theta^* - r) > 0$$

and W_n diverges to infinity with probability one. Since $CV_n^W(\alpha, \tau) = O_p(1)$ by the same argument in the proof of 4.2, the theorem holds. ■

Proof of Lemma 5.1: The proof follows from Theorem 2 in Andrews (1999) and Theorem 1 and Theorem 3 in Andrews (2001). By Assumptions 2.1, 2.3 and 5.1 and

Theorem 2 in Andrews (1999), we have

$$\begin{aligned}-2\left(l_n\left(\widehat{\theta}_0\right)-l_n\left(\widehat{\theta}\right)\right)&=q_n\left(b_n\left(\widehat{\theta}_0-\theta^*\right)\right)-q_n\left(b_n\left(\widehat{\theta}-\theta^*\right)\right)+o_p\left(1\right)\\&=\inf_{\theta\in\Theta_0}q_n\left(b_n\left(\theta-\theta^*\right)\right)-\inf_{\theta\in\Theta}q_n\left(b_n\left(\theta-\theta^*\right)\right)+o_p\left(1\right).\end{aligned}$$

Assumption 2.2, Theorem 1 and Theorem 3 in Andrews (2001) provide that

$$-2\left(l_n\left(\widehat{\theta}_0\right)-l_n\left(\widehat{\theta}\right)\right)\xrightarrow{d}\min_{\lambda\in\Lambda_0}q\left(\lambda\right)-\min_{\lambda\in\Lambda}q\left(\lambda\right),$$

where Λ_0 and Λ correspond to the set where $\phi_0\left(\lambda\right)$ and $\phi\left(\lambda\right)$ are finite. The lemma then follows. ■

Proof of Lemma 5.2: By Assumptions 4.2, 4.4 and 5.2 and Theorem 2 in Andrews (1999), it holds that

$$\begin{aligned}-2\left(l_n\left(\widehat{\theta}_0\right)-l_n\left(\widehat{\theta}\right)\right)&=q_n\left(b_n\left(\widehat{\theta}_0-\theta_n\right)\right)-q_n\left(b_n\left(\widehat{\theta}-\theta_n\right)\right)+o_p\left(1\right)\\&=\inf_{\theta\in\Theta_0}q_n\left(b_n\left(\theta-\theta_n\right)\right)-\inf_{\theta\in\Theta}q_n\left(b_n\left(\theta-\theta_n\right)\right)+o_p\left(1\right),\end{aligned}$$

under any $\omega_n\in\mathcal{W}_0$. Applying the same argument in proof of Lemma 4.3 to $b_n\left(\Theta-\theta_n\right)$ and similar argument to $b_n\left(\Theta_0-\theta_n\right)$, we obtain that

$$\begin{aligned}\Lambda&=\left\{\theta\in\mathbb{R}^l:\mathscr{R}_e\theta=\mathbf{0},\,\mathscr{R}_w^b\theta\geq\mathbf{0}\text{ and }\mathscr{R}_w^u\theta+c\geq\mathbf{0}\right\},\text{and}\\\Lambda_0&=\left\{\theta\in\mathbb{R}^l:\left(R',\mathscr{R}_e',\mathscr{R}_w^{b\prime}\right)'\theta=\mathbf{0}\text{ and }\mathscr{R}_w^u\theta+c\geq\mathbf{0}\right\}.\end{aligned}$$

The rest of proof follows from Theorem 1 and Theorem 3 in Andrews (2001) and Assumption 4.3. We obtain that

$$-2\left(l_n\left(\widehat{\theta}_0\right)-l_n\left(\widehat{\theta}\right)\right)\xrightarrow{d}\min_{\lambda\in\Lambda_0}q\left(\lambda\right)-\min_{\lambda\in\Lambda}q\left(\lambda\right),$$

and the lemma follows. ■

Proof of Theorem 5.1: The proof is similar to the proofs of Theorems 4.1 and 4.3, with the asymptotic distribution of QLR_n under $\omega_n\in\mathcal{W}_0$ being provided in Lemma 4.3. ■

Proof of Theorem 5.2: Assumptions 2.3 and 5.1 imply that $\widehat{\theta}\xrightarrow{p}\theta^*$ and $\widehat{\theta}_0\xrightarrow{p}\theta_0^*$ under H_1. By Assumptions 2.1 and 2.2, we obtain that $l_n\left(\cdot\right)$ is continuous at θ^*. Together with the assumption on the continuity of $l_n\left(\cdot\right)$ at θ_0^*, continuous mapping theorem provides that

$$b_n^{-2}\left(l_n\left(\widehat{\theta}_0\right)-l_n\left(\widehat{\theta}\right)\right)\xrightarrow{p}\varsigma>0.$$

Therefore, $-2\left(l_n\left(\widehat{\theta}_0\right)-l_n\left(\widehat{\theta}\right)\right)$ diverges to positive infinity as $n\to\infty$. We now prove that $CV_n^Q=O_p(1)$ to conclude the theorem. For any $c\geq\mathbf{0}$ in Lemma 5.2 and any $\pi_{Q,\omega}\in\overline{\Pi}_Q$, we have

$$\begin{aligned}\left|\min_\lambda\left[q_\omega(\lambda)+\phi_{0,\omega}(\lambda)\right]-\min_\lambda\left[q_\omega(\lambda)+\phi_\omega(\lambda)\right]\right| &= \min_\lambda\left[q_\omega(\lambda)+\phi_{0,\omega}(\lambda)\right]-\min_\lambda\left[q_\omega(\lambda)+\phi_\omega(\lambda)\right]\\ &\leq \min_\lambda\left[q_\omega(\lambda)+\phi_{0,\omega}(\lambda)\right]\\ &\leq q_\omega(\mathbf{0})=O_p(1),\end{aligned}$$

where the first inequality is due to the quadratic form of $q_\omega(\lambda)$, and the second inequality holds because $\phi_{0,\omega}(\mathbf{0})=0$. Assume that the sup in the definition of $CV_n^Q(\alpha,\tau)$ is achieved at $\widehat{c}\in\widetilde{I}_{\alpha-\tau}$. When $\widetilde{I}_{\alpha-\tau}$ is open, $\widehat{c}$ belongs to the closure of $\widetilde{I}_{\alpha-\tau}$. Since $\widehat{c}\geq\mathbf{0}$ for any n, we conclude that for $\tau>0$, $CV_n^Q(\alpha,\tau)=\mathcal{C}^Q_{\widehat{c},\widehat{\pi}_Q}(1-\tau)=O_p(1)$. ■

Proof of Lemma 6.1: Applying Equation (6), we obtain that

$$b_n^{-1}Dl_n\left(\widehat{\theta}_0\right)=b_n^{-1}Dl_n(\theta^*)+b_n^{-1}D^2l_n(\theta^*)\left(\widehat{\theta}_0-\theta^*\right)+b_n^{-1}R_n^D\left(\widehat{\theta}_0\right).$$

Assumptions 2.2 and 5.1 imply that $b_n^{-1}Dl_n(\theta^*)=O_p(1)$ and

$$b_n^{-1}D^2l_n(\theta^*)\left(\widehat{\theta}_0-\theta^*\right)=b_n^{-2}D^2l_n(\theta^*)\,b_n\left(\widehat{\theta}_0-\theta^*\right)=O_p(1).$$

Since $b_n\left(\widehat{\theta}_0-\theta^*\right)=O_p(1)$ by Assumption 5.1, for any $\varepsilon_1>0$, there exists some $\kappa<\infty$ such that $\Pr\left(\left\|b_n\left(\widehat{\theta}_0-\theta^*\right)\right\|\geq\kappa\right)<\varepsilon_1$ when n is sufficiently large. By Assumption 6.1 (i), for any $\delta>0$ and $\varepsilon_2>0$, $\Pr\left(\sup_{\theta\in\Theta:\|b_n(\theta-\theta^*)\|<\kappa}\left|b_n^{-1}R_n^D(\theta)\right|>\delta\right)<\varepsilon_2$

for n sufficiently large. Thus, for any $\delta > 0$, it holds that

$$\begin{aligned}
&\Pr\left(\left|b_n^{-1}R_n^D\left(\widehat{\theta}_0\right)\right| > \delta\right) \\
\leq &\Pr\left(\left|b_n^{-1}R_n^D\left(\widehat{\theta}_0\right)\right| > \delta \text{ and } \left\|b_n\left(\widehat{\theta}_0-\theta^*\right)\right\| < \kappa\right) + \Pr\left(\left\|b_n\left(\widehat{\theta}_0-\theta^*\right)\right\| \geq \kappa\right) \\
\leq &\Pr\left(\sup_{\theta\in\Theta:\|b_n(\theta-\theta^*)\|<\kappa}\left|b_n^{-1}R_n^D\left(\theta\right)\right| > \delta\right) + \Pr\left(\left\|b_n\left(\widehat{\theta}_0-\theta^*\right)\right\| \geq \kappa\right) \\
\leq &\varepsilon_2 + \varepsilon_1,
\end{aligned}$$

where ε_1 and ε_2 are both arbitrary. We obtain that $b_n^{-1}R_n^D\left(\widehat{\theta}_0\right) = o_p(1)$. Therefore, $b_n^{-1}Dl_n\left(\widehat{\theta}_0\right) = O_p(1)$ and $\widehat{\mathscr{T}_n}^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right) = \mathscr{T}_n^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right) + o_p(1)$ by Assumption 6.1 (b) and $\mathscr{T}_n \xrightarrow{d} \mathscr{T}$ for $\mathscr{T}$ being non-singular with probability one by Assumption 2.2. The definition of $\widehat{\theta}_0$ implies that $R\widehat{\theta}_0 = r$. Therefore, it holds that

$$\begin{aligned}
&R\widehat{\mathscr{T}_n}^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right) \\
=&R\mathscr{T}_n^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right) + o_p(1) \\
=&R\left(\mathscr{T}_n^{-1}b_n^{-1}Dl_n\left(\theta^*\right) + \mathscr{T}_n^{-1}b_n^{-1}D^2l_n\left(\theta^*\right)\left(\widehat{\theta}_0-\theta^*\right) + \mathscr{T}_n^{-1}b_n^{-1}R_n^D\left(\widehat{\theta}_0\right)\right) + o_p(1) \\
=&R\mathscr{T}_n^{-1}b_n^{-1}Dl_n\left(\theta^*\right) - b_nR\left(\widehat{\theta}_0-\theta^*\right) + o_p(1) \\
=&R\mathscr{T}_n^{-1}b_n^{-1}Dl_n\left(\theta^*\right) + o_p(1),
\end{aligned} \tag{S.20}$$

where the second to last equality comes from the definition of $\mathscr{T}_n$. Since

$$\left(R\widehat{\mathscr{T}_n}^{-1}R'\right)^{-1/2}R\widehat{\mathscr{T}_n}^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right) = O_p(1)$$

by Assumption 2.2 and 6.1, we obtain that $ds = O_p(1)$ by the following result:

$$\begin{aligned}
&\left\|\left(R\widehat{\mathscr{T}_n}^{-1}R'\right)^{-1/2}ds - \left(R\widehat{\mathscr{T}_n}^{-1}R'\right)^{-1/2}R\widehat{\mathscr{T}_n}^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right)\right\| \\
=&\left(ds - R\widehat{\mathscr{T}_n}^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right)\right)'\left(R\widehat{\mathscr{T}_n}^{-1}R'\right)^{-1}\left(ds - R\widehat{\mathscr{T}_n}^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right)\right) \\
\leq&R\widehat{\mathscr{T}_n}^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right)'\left(R\widehat{\mathscr{T}_n}^{-1}R'\right)^{-1}R\widehat{\mathscr{T}_n}^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right) + o_p(1) \\
=&O_p(1),
\end{aligned}$$

where the inequality holds by the definition of ds. Let

$$q_{n,R}(\cdot) \equiv \left(\cdot - R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)\right)' \left(R\mathscr{T}_n^{-1}R'\right)^{-1} \left(\cdot - R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)\right).$$

As $ds = O_p(1)$, Equation (S.20) and Assumption 6.1 (ii), we have

$$\widehat{q}_R(ds) = q_{n,R}(ds) + o_p(1).$$

Applying the same proof as for Theorem 1 (e) in Andrews (1997), we obtain that

$$\widehat{q}_R(ds) = \inf_{\lambda \in b_n(R\Theta - r)} q_{n,R}(\lambda) + o_p(1). \tag{S.21}$$

The set $R\Theta - r$ has the halfspace description (15), which is locally approximated by the set $\Lambda_R \equiv \left\{\lambda \in \mathbb{R}^l : \mathscr{R}_{R,e}\lambda = \mathbf{0}, \mathscr{R}_{R,w,b}\lambda \geq \mathbf{0}\right\}$. Λ_R is convex and closed by definition. The remaining of the proof follows from Lemma 1 in Andrews (1997) which shows that

$$\inf_{\lambda \in b_n(R\Theta - r)} q_{n,R}(\lambda) = \min_{\lambda \in \Lambda_R} q_{n,R}(\lambda) + o_p(1), \tag{S.22}$$

Theorem 2 in Andrews (1997) which provides $ds = \arg\min_{\lambda \in \Lambda_R} q_{n,R}(\lambda) + o_p(1)$, and the continuous mapping theorem which gives

$$\min_{\lambda \in \Lambda_R} q_{n,R}(\lambda) \xrightarrow{d} \min_{\lambda \in \Lambda_R} q_R(\lambda) \equiv \min_{\lambda \in \Lambda_R} \left(\lambda - R\mathscr{T}^{-1}G\right)' \left(R\mathscr{T}^{-1}R'\right)^{-1} \left(\lambda - R\mathscr{T}^{-1}G\right).$$

Conditions required for Lemma 1 and Theorem 2 in Andrews (1997) are given in the assumptions. The fact that $\mathscr{T}$ is non-singular with probability one provides the condition to apply the continuous mapping theorem. Hence, we obtain that

$$ds \xrightarrow{d} \arg\min_{\lambda \in \Lambda_R} (\lambda - RZ)' \left(R\mathscr{T}^{-1}R'\right)^{-1} (\lambda - RZ),$$

where $Z = \mathscr{T}^{-1}G$, which is equivalent to the result stated in the lemma.

Part (ii) follows immediately from part (i) and Assumption 6.2. ■

Proof of Theorem 6.1: First, we show that the asymptotic distribution of S_n under any $\omega_n \in \mathcal{W}_0$ is given by $S_n \xrightarrow{d} S_\omega \equiv ds_\omega' \Sigma_{S,\omega}^{-1} ds_\omega$, where

$$ds_\omega \equiv \arg \min_{\mathscr{R}_{R,e}\lambda = \mathbf{0} \text{ and } \mathscr{R}_{R,w,b}\lambda \geq \mathbf{0}} \left(\lambda - R\mathscr{T}_\omega^{-1}G_\omega\right)' \left(R\mathscr{T}_\omega^{-1}R'\right)^{-1} \left(\lambda - R\mathscr{T}_\omega^{-1}G_\omega\right).$$

By Assumption 5.2, $b_n\left(\widehat{\theta}_0-\theta_n\right)=O_p(1)$. Together with the fact that $\theta_n\to\theta_\omega$ as $n\to\infty$, we have $\widehat{\theta}_0-\theta_\omega=\widehat{\theta}_0-\theta_n+\theta_n-\theta_\omega=o_p(1)$. Assumption 6.3 (i) applies, and $b_n^{-1}R_n^D\left(\widehat{\theta}_0\right)=o_p(1)$ holds under any $\omega_n\in\mathcal{W}_0$ by the same argument in the proof of Lemma 6.1. Then, Equation S.20 becomes

$$\begin{aligned}R\widehat{\mathscr{T}}_n^{-1}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right)&=R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta_n)-b_nR\left(\widehat{\theta}_0-\theta_n\right)+o_p(1)\\&=R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta_n)-b_n(r-R\theta_n)+o_p(1)\\&=R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta_n)+o_p(1)\,.\end{aligned}$$

The rest of the proof is the same, with all the convergence results under $\omega_n\in\mathcal{W}_0$ provided by Assumption 2.2.

Notice that the asymptotic distribution of S_n under $\omega_n\in\mathcal{W}_0$ has the exact same form as the one given by Lemma 6.1. Therefore, $\mathcal{C}^S_{S,\pi}(1-\alpha)$ is also the $(1-\alpha)$ quantile of S_ω with π_S denoting the parameters in $\mathscr{T}_\omega$, G_ω and $\Sigma_{S,\omega}$. By the definition of the asymptotic size, we have that

$$AsySz\left(S_n,CV_n^S(\alpha)\right)=\lim_{n\to\infty}\Pr{}_{\omega_{p_n}}\left(S_{p_n}>CV_{p_n}^S(\alpha)\right),$$

where $\{p_n\}$ is some subsequence of $\{n\}$. Since the following proof goes through with p_n in place n and the convergence of the full sequence guarantees the convergence of each subsequence with same limit, we provide the following result for the full sequence $\{n\}$. The continuity of $\mathcal{C}^S_{\pi_S}(1-\alpha)$ in π by Assumption 3.1 and $\widehat{\pi}_S\xrightarrow{p}\pi_S$ imply that $\mathcal{C}^S_{\widehat{\pi}_S}(1-\alpha)\xrightarrow{p}\mathcal{C}^S_{\pi_S}(1-\alpha)$ for any $\omega_n\in\mathcal{W}_0$ by the continuous mapping theorem. The convergence in distribution result proved at the beginning and the asymptotic distribution being continuous at $\mathcal{C}^S_{\pi}(1-\alpha)$ provide that

$$\lim_{n\to\infty}\Pr{}_{\omega_n}\left(S_n>\mathcal{C}^S_{\widehat{\pi}_S}(1-\alpha)\right)=\alpha,$$

for any $\omega_n\in\mathcal{W}_0$. Therefore, the theorem follows. ■

Proof of Theorem 6.2: By Assumptions 5.1 and 6.1, under H_1, the same argument in the proof of Lemma 6.1 implies that

$$\begin{aligned}R\widehat{\mathscr{T}}_n^{-1}b_n^{-2}Dl_n\left(\widehat{\theta}_0\right)&=R\mathscr{T}_n^{-1}b_n^{-2}Dl_n(\theta_0^*)-R\left(\widehat{\theta}_0-\theta_0^*\right)+o_p\left(b_n^{-1}\right)\\&=R\mathscr{T}_n^{-1}b_n^{-2}Dl_n(\theta_0^*)+o_p\left(b_n^{-1}\right)\\&=\upsilon R(\theta^*-\theta_0^*)+o_p(1)\,,\end{aligned}$$

where the last equality follows from the assumption that $\mathscr{T}_n^{-1} b_n^{-2} Dl_n(\theta_0^*) = \upsilon(\theta^* - \theta_0^*) + o_p(1)$. Let $ds_s \equiv b_n \upsilon (R\theta^* - r)$. Since $\theta^* \in \Theta$, $b_n(R\theta^* - r) \in b_n(R\Theta - r)$; and $\mathbf{0} \in b_n(R\Theta - r)$ because there exists some $\theta \in \Theta$ such that $R\theta = r$. The convexity of $b_n(R\Theta - r)$ provides that $ds_s \in b_n(R\Theta - r)$. Moreover, since

$$\begin{aligned}
&\left(b_n^{-1} ds_s - R\widehat{\mathscr{T}}_n^{-1} b_n^{-2} Dl_n\left(\widehat{\theta}_0\right)\right)' \left(R\widehat{\mathscr{T}}_n^{-1} R'\right)^{-1} \left(b_n^{-1} ds_s - R\widehat{\mathscr{T}}_n^{-1} b_n^{-2} Dl_n\left(\widehat{\theta}_0\right)\right) \\
&= \left(b_n^{-1} ds_s - \upsilon R(\theta^* - \theta_0^*) + o_p(1)\right)' \left(R\widehat{\mathscr{T}}_n^{-1} R'\right)^{-1} \left(b_n^{-1} ds_s - \upsilon R(\theta^* - \theta_0^*) + o_p(1)\right) \\
&= (\mathbf{0} + o_p(1))' \left(R\widehat{\mathscr{T}}_n^{-1} R'\right)^{-1} (\mathbf{0} + o_p(1)) \xrightarrow{p} 0,
\end{aligned}$$

and $R\mathscr{T}^{-1}R'$ is positive definite, it must hold that

$$b_n^{-1} ds_n = b_n^{-1} ds_s + o_p(1) = \upsilon(R\theta^* - r) + o_p(1). \tag{S.23}$$

Assume the contrary. Then it holds that

$$b_n^{-1} ds_n - \upsilon(R\theta^* - r) \xrightarrow{p} \varrho \neq \mathbf{0},$$

which implies that

$$\begin{aligned}
&\left(b_n^{-1} ds_n - R\widehat{\mathscr{T}}_n^{-1} b_n^{-2} Dl_n\left(\widehat{\theta}_0\right)\right)' \left(R\widehat{\mathscr{T}}_n^{-1} R'\right)^{-1} \left(b_n^{-1} ds_n - R\widehat{\mathscr{T}}_n^{-1} b_n^{-2} Dl_n\left(\widehat{\theta}_0\right)\right) \\
&= \left(b_n^{-1} ds_n - \upsilon(R\theta^* - r) + o_p(1)\right)' \left(R\widehat{\mathscr{T}}_n^{-1} R'\right)^{-1} \left(b_n^{-1} ds_n - \upsilon(R\theta^* - r) + o_p(1)\right) \\
&\xrightarrow{p} \varrho' \left(R\mathscr{T}^{-1} R'\right)^{-1} \varrho > 0,
\end{aligned}$$

where the inequality follows from $R\mathscr{T}^{-1}R'$ being positive definite. Thus, for n sufficiently large, ds_n is not the minimizer of (7), which contradicts its definition. Applying Equation (S.23) to the definition of S_n, we have

$$\begin{aligned}
b_n^{-2} S_n &= b_n^{-2} ds_n' \Sigma_{S,n}^{-1} ds_n \\
&= (\upsilon(R\theta^* - r) + o_p(1))' \Sigma_{S,n}^{-1} (\upsilon(R\theta^* - r) + o_p(1)) \\
&\xrightarrow{p} \upsilon^2 (R\theta^* - r)' \Sigma_S^{-1} (R\theta^* - r) > 0,
\end{aligned}$$

because Σ_S is positive definite. Therefore, S_n diverges and the claimed result hold by the finiteness of $\mathcal{C}_{\pi_S}^S(1-\alpha)$. ■

Proof of Lemma 7.1: (i) The asymptotic distribution of $b_n\left(\widehat{\theta} - \theta_n\right)$ can be obtained using the same argument for the proof of Lemma 4.3 in Section S.1, with

$$b_n(\Theta - \theta_n) = \left\{\theta \in \mathbb{R}^l : \mathscr{R}_e\theta = b_n(\mathrm{r}_e - \mathscr{R}_e\theta_n) \text{ and } \mathscr{R}_w\theta \geq b_n(\mathrm{r}_w - \mathscr{R}_w\theta_n)\right\},$$
$$\Lambda \equiv \left\{\theta \in \mathbb{R}^l : \mathscr{R}_e\theta = \mathbf{0} \text{ and } \mathscr{R}_w\theta + c_w \geq \mathbf{0}\right\} \text{ and } c_w = \lim_{n\to\infty} -b_n(\mathrm{r}_w - \mathscr{R}_w\theta_n).$$

By definition, $R\theta_n = r + b_n^{-1}\delta(1 + o(1))$. Thus, we have $b_n(R\theta_n - r) \to \delta$ as $n \to \infty$, and

$$\begin{aligned}
W_n &= b_n^2\left(R\widehat{\theta} - r\right)'(R\Sigma_{W,n}R')^{-1}\left(R\widehat{\theta} - r\right) \\
&= b_n^2\left(R\widehat{\theta} - R\theta_n + R\theta_n - r\right)'(R\Sigma_{W,n}R')^{-1}\left(R\widehat{\theta} - R\theta_n + R\theta_n - r\right) \\
&= \left[b_nR\left(\widehat{\theta} - \theta_n\right) + b_n(R\theta_n - r)\right]'(R\Sigma_{W,n}R')^{-1}\left[b_nR\left(\widehat{\theta} - \theta_n\right) + b_n(R\theta_n - r)\right] \\
&\xrightarrow{d} (R\Psi_{1,\omega} + \delta)'(R\Sigma_{W,\omega}R')^{-1}(R\Psi_{1,\omega} + \delta)'.
\end{aligned}$$

(ii) The proof follows from the same argument for the proof of Lemma 5.2. Notice that $c = \lim_{n\to\infty} -b_n(\mathrm{r}_w^u - \mathscr{R}_w^u\theta_n)$ by definition. We have

$$b_n(\Theta_0 - \theta_n) = \left\{\theta \in \mathbb{R}^l : \left(R', \mathscr{R}_e', \mathscr{R}_w^{b\prime}\right)'\theta = b_n\left(\left(r', \mathrm{r}_e', \mathrm{r}_w^{b\prime}\right)' - \left(R', \mathscr{R}_e', \mathscr{R}_w^{b\prime}\right)'\theta_n\right), \mathscr{R}_w^u\theta \geq b_n(\mathrm{r}_w^u - \mathscr{R}_w^u\theta_n)\right\}$$
$$\Lambda_0 \equiv \left\{\theta \in \mathbb{R}^l : \left(R', \mathscr{R}_e', \mathscr{R}_w^{b\prime}\right)'\lambda + \left(\delta', \mathbf{0}, c_{w,b}'\right)' = \mathbf{0} \text{ and } \mathscr{R}_w^u\lambda + c \geq \mathbf{0}\right\}.$$

Applying the Λ and Λ_0 defined here to the proof of Lemma 5.2, we obtain the result.

(iii) The proof mainly follows the one of Lemma 6.1 with the following modification. First, by Assumption 7.1 (v), $b_n\left(\widehat{\theta}_0 - \theta_n\right) = O_p(1)$. Together with the fact that $\theta_n \to \theta_\omega$ as $n \to \infty$, we have $\widehat{\theta}_0 - \theta_\omega = \widehat{\theta}_0 - \theta_n + \theta_n - \theta_\omega = o_p(1)$. Assumption 7.1 (vi) applies, and $b_n^{-1}R_n^D\left(\widehat{\theta}_0\right) = o_p(1)$ holds under any $\omega_n \in \mathcal{W}$ by the same argument in the proof of Lemma 6.1. Second, Equation S.20 becomes

$$\begin{aligned}
R\widehat{\mathscr{T}_n^{-1}}b_n^{-1}Dl_n\left(\widehat{\theta}_0\right) &= R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta_n) - b_nR\left(\widehat{\theta}_0 - \theta_n\right) + o_p(1) \\
&= R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta_n) - b_n(r - R\theta_n) + o_p(1) \\
&= R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta_n) + \delta(1 + o(1)) + o_p(1) \\
&= R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta_n) + \delta + o_p(1)
\end{aligned}$$

by the definition of $H_{1,n}$. The rest of the proof is the same, with all the convergence results under $\omega_n \in \mathcal{W}$ provided by Assumption 7.1 (ii), (vi) and (vii). ■

Proof of Corollary 7.1: Part (i) of the corollary follows if we can show that

$CV_n^W(\alpha,\tau) \xrightarrow{d} CV^W(\alpha,\tau)$ and the convergence occurs jointly with $W_n \xrightarrow{d} W_{1,\omega}$. Since $\widetilde{ES}(\tau)$ is the set obtained using the estimators of parameters in $\mathscr{T}_{f,\omega}^{-1} G_{f,\omega}$ which is continuous in unknown parameters, it holds that $d_H\left(\widetilde{ES}(\tau), ES(\tau)\right) = o_p(1)$. By definition, we have

$$b_n \Gamma^u \mathscr{R}_\Gamma^u \widetilde{\theta}_{f,n} - (\mathrm{r}_w^u - \mathscr{R}_w^u \gamma) \xrightarrow{d} c + \Gamma^u \mathscr{R}_\Gamma^u \mathscr{T}_{f,\omega}^{-1} G_{f,\omega}, \tag{S.24}$$

where the random vector $G_{f,\omega}$ is the subvector of G_ω corresponding to θ_f. Therefore, $\sup_{c\in\widetilde{I}_{\alpha-\tau}} \mathcal{C}^W_{c,\pi_{W,\omega}}(1-\tau) \xrightarrow{d} \sup_{c\in I_{\alpha-\tau}} \mathcal{C}^W_{c,\pi_{W,\omega}}(1-\tau)$ for any $\pi_{W,\omega}$, because W_ω is continuous at $\mathcal{C}^W_{c,\pi_{W,\omega}}(1-\tau)$ for all $c \in C$. Moreover, $\mathcal{C}^W_{c,\pi_{W,\omega}}(1-\tau)$ is continuous in $\pi_{W,\omega}$ and $\widehat{\pi}_W \xrightarrow{p} \pi_{W,\omega}$. It holds that $CV_n^W(\alpha,\tau) \xrightarrow{d} CV^W(\alpha,\tau)$. Since the convergence of (S.24) is jointly with $\left(b_n^{-1} Dl_n(\theta^*), \mathscr{T}_n\right) \xrightarrow{d} (G, \mathscr{T})$, part (i) of the corollary follows. Similar arguments apply to part (ii). Because $\mathcal{C}^S_{\pi_{S,\omega}}(1-\alpha)$ is continuous in $\pi_{S,\omega}$ and $\widehat{\pi}_S$ consistently estimates $\pi_{S,\omega}$, we have $\mathcal{C}^S_{\widehat{\pi}_S}(1-\alpha) \xrightarrow{p} \mathcal{C}^S_{\pi_{S,\omega}}(1-\alpha)$. Part (iii) therefore holds. ■

Proof of Theorem 8.1: (i) Under Assumptions 2.2, 5.1 and 6.1, we've obtained Equation (S.20) in the proof of Lemma 6.1. Apply the same argument to $\widehat{\theta}$ with Assumption 5.1 replaced by 4.4, we obtain that

$$\widehat{\mathscr{T}}_n^{-1} b_n^{-1} Dl_n\left(\widehat{\theta}\right) = \mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*) - b_n\left(\widehat{\theta} - \theta^*\right) + o_p(1).$$

Thus,

$$\widehat{\mathscr{T}}_n^{-1} b_n^{-1}\left(Dl_n\left(\widehat{\theta}_0\right) - Dl_n\left(\widehat{\theta}\right)\right) = b_n\left(\widehat{\theta} - \widehat{\theta}_0\right) + o_p(1).$$

Substituting the above equality to the expression of S_n^1, we obtain

$$S_n^1 = \left[b_n\left(\widehat{\theta} - \widehat{\theta}_0\right) + o_p(1)\right]' \Sigma_{S^1,n}^{-1} \left[b_n\left(\widehat{\theta} - \widehat{\theta}_0\right) + o_p(1)\right].$$

Since

$$b_n\left(\widehat{\theta} - \widehat{\theta}_0\right) = b_n\left(\widehat{\theta} - \theta^*\right) - b_n\left(\widehat{\theta}_0 - \theta^*\right) = O_p(1)$$

by Assumptions 2.3 and 5.1 and the null hypothesis, Assumption 6.2 implies that

$$S_n^1 = b_n^2\left(\widehat{\theta} - \widehat{\theta}_0\right)' \Sigma_{S^1,n}^{-1}\left(\widehat{\theta} - \widehat{\theta}_0\right) + o_p(1).$$

Since $\Sigma_{S^1}^{-1} = R'\left(R\Sigma_W R'\right)^{-1} R$ and $R\widehat{\theta}_0 = r$, it holds that $W_n = S_n^1 + o_p(1)$ by

Assumptions 4.1 and 6.2.

(ii) Under Assumptions 2.2, 5.1 and 6.1, it has been shown in the proof of Lemma 6.1 that

$$ds = \arg\min_{\lambda \in \Lambda_R} q_{n,R}(\lambda) + o_p(1),$$

where $q_{n,R}(\cdot) \equiv (\cdot - R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*))' (R\mathscr{T}_n^{-1} R')^{-1} (\cdot - R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*))$. Let $ds_\Lambda \equiv \arg\min_{\lambda \in \Lambda_R} q_{n,R}(\lambda)$. Since $\Sigma_S = R\mathscr{T}^{-1}R'$, $\Sigma_{S,n} \xrightarrow{p} \Sigma_S$ and $\mathscr{T}_n \xrightarrow{d} \mathscr{T}$, the convergence of $\Sigma_{S,n} \xrightarrow{d} R\mathscr{T}^{-1}R'$ jointly with $b_n^{-1} Dl_n(\theta^*)$. Thus, we have

$$S_n = ds_\Lambda' \left(R\mathscr{T}^{-1}R'\right)^{-1} ds_\Lambda + o_p(1), \tag{S.25}$$

for

$$ds_\Lambda \equiv \arg\min_{\lambda \in \Lambda_R} \left(\lambda - R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right)' \left(R\mathscr{T}^{-1}R'\right)^{-1} \left(\lambda - R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right).$$

Applying Equations (S.20) and (S.22) to the definition of S_n^2, it holds that

$$\begin{aligned} S_n^2 &= \left(R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right)' \left(R\mathscr{T}^{-1}R'\right)^{-1} \left(R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right) \\ &\quad - \min_{\lambda \in \Lambda_R} \left(\lambda - R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right)' \left(R\mathscr{T}^{-1}R'\right)^{-1} \left(\lambda - R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right) + o_p(1) \\ &= \left(R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right)' \left(R\mathscr{T}^{-1}R'\right)^{-1} \left(R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right) \\ &\quad - \left(ds_\Lambda - R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right)' \left(R\mathscr{T}^{-1}R'\right)^{-1} \left(ds_\Lambda - R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right) + o_p(1). \end{aligned} \tag{S.26}$$

Since the set $\Lambda_R = \left\{\lambda \in \mathbb{R}^J : \mathscr{R}_{R,e}\lambda = \mathbf{0}, \mathscr{R}_{R,w,b} \geq \mathbf{0}\right\}$ is a closed convex cone with the vertex at the origin, Pythagorean Theorem provides that

$$\begin{aligned} &\left(R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right)' \left(R\mathscr{T}^{-1}R'\right)^{-1} \left(R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right) - ds_\Lambda' \left(R\mathscr{T}^{-1}R'\right)^{-1} ds_\Lambda \\ &= \left(ds_\Lambda - R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right)' \left(R\mathscr{T}^{-1}R'\right)^{-1} \left(ds_\Lambda - R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right). \end{aligned}$$

Apply the above equality to Equations (S.25) and (S.26), we obtain that $S_n = S_n^2 + o_p(1)$.

(iii) Under Assumptions 2.1, 2.2 and 8.1, we obtain

$$\begin{aligned} b_n\left(\widetilde{\theta} - \theta^*\right) &= \arg\min_{\lambda \in \mathbb{R}^l} \left(\lambda - \mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right)' \mathscr{T}_n \left(\lambda - \mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*)\right) + o_p(1) \\ &= \mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*) + o_p(1) \end{aligned}$$

by applying Theorem 2 in Andrews (1997). By the definition of W_n^1 and Assumption 4.1, it holds that

$$\begin{aligned} W_n^1 &= \left(R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)+o_p(1)\right)\left(R\Sigma_{W,n}R'\right)^{-1}\left(R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)+o_p(1)\right) \\ &\quad - \inf_{\lambda\in R\Theta}\left(R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)+r-b_n\lambda\right)'\left(R\Sigma_{W,n}R'\right)^{-1}\left(R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)+r-b_n\lambda\right) \\ &= \left(R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)\right)'\left(R\Sigma_{W,n}R'\right)^{-1}\left(R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)\right) \\ &\quad - \inf_{\lambda\in b_n(R\Theta-r)}\left(R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)-\lambda\right)'\left(R\Sigma_{W,n}R'\right)^{-1}\left(R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)-\lambda\right)+o_p(1). \end{aligned}$$

Applying Equation (S.20), we get that

$$\begin{aligned} S_n^2 &= \left(R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)\right)'\left(R\mathscr{T}^{-1}R'\right)^{-1}\left(R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)\right) \\ &\quad - \inf_{\lambda\in b_n(R\Theta-r)}\left(\lambda-R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)\right)'\left(R\mathscr{T}^{-1}R'\right)^{-1}\left(\lambda-R\mathscr{T}_n^{-1}b_n^{-1}Dl_n(\theta^*)\right)+o_p(1). \end{aligned}$$

If $\Sigma_W=\mathscr{T}^{-1}$, then $\Sigma_{W,n}=\mathscr{T}^{-1}+o_p(1)$ by Assumptions 4.1 and 6.2. Thus, $W_n^1 = S_n^2+o_p(1)$. ■

Lemma S.1.6. *Result in Lemmas 2.1, 4.1, 4.3, 5.1, 5.2 and 7.1 is independent of the description of Θ; result in Lemma 6.1 is independent of the description of $R\Theta-r$.*

Proof: Let $\Theta(1)$ and $\Theta(2)$ be two descriptions of Θ. By Andrews (1997), result in 2.1 is obtained by finding the cone that locally approximates $b_n(\Theta-\theta^*)$. Since $\Theta(1)$ and $\Theta(2)$ are two descriptions of the same set Θ, the cone is the same. Thus, Lemma 2.1 is independent of $\Theta(1)$ and $\Theta(2)$. For Lemma 4.3, the set Λ such that $\phi_\omega(\cdot)$ equals zero satisfies that $d_H(b_n(\Theta-\theta_n),\Lambda)\to 0$. Let Λ_1 and Λ_2 be two sets obtained from $\Theta(1)$ and $\Theta(2)$. Since $d_H(b_n(\Theta-\theta_n),\Lambda_1)\to 0$ and $d_H(b_n(\Theta-\theta_n),\Lambda_2)\to 0$, triangular inequality provides that $d_H(\Lambda_1,\Lambda_2)=0$, which holds if and only if the closures of Λ_1 and Λ_2 are the same. Because both Λ_1 and Λ_2 are equivalent to their closures, Λ_1 and Λ_2 are the same. Thus, Lemma 4.3 doesn't depend on the description of Θ. Result in other lemmas follows from the similar argument. ■

Lemma S.1.7. *Let*

$$\widehat{\lambda} \equiv \arg\min_{\mathscr{R}_w\lambda\geq\mathbf{0}}\left(\lambda-\mathscr{T}^{-1}G\right)'\mathscr{T}\left(\lambda-\mathscr{T}^{-1}G\right), \tag{S.27}$$

$$\overline{\lambda} \equiv \arg\min_{\mathscr{R}_w\lambda=\mathbf{0}}\left(\lambda-\mathscr{T}^{-1}G\right)'\mathscr{T}\left(\lambda-\mathscr{T}^{-1}G\right) \quad \textit{and} \tag{S.28}$$

$$\widehat{s} \equiv \arg\min_{s\geq\mathbf{0}}\left(s-\mathscr{R}_w\mathscr{T}^{-1}G\right)'\left(\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1}\left(s-\mathscr{R}_w\mathscr{T}^{-1}G\right) \tag{S.29}$$

Then the following two equations holds:

$$\left(\mathscr{R}_w\widehat{\lambda}\right)'\left(\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1}\left(\mathscr{R}_w\widehat{\lambda}\right) = \left(\overline{\lambda}-\mathscr{T}^{-1}G\right)'\mathscr{T}\left(\overline{\lambda}-\mathscr{T}^{-1}G\right) - \left(\widehat{\lambda}-\mathscr{T}^{-1}G\right)'\mathscr{T}\left(\widehat{\lambda}-\mathscr{T}^{-1}G\right) \tag{S.30}$$

$$= \widehat{s}'\left(\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1}\widehat{s}. \tag{S.31}$$

Proof: Let $\widetilde{\lambda} = \mathscr{T}^{-1}G$, and denote $\widehat{\gamma}$ and $\overline{\gamma}$ as the Lagrange multiplier vectors associated with $\mathscr{R}_w\lambda \geq \mathbf{0}$ and $\mathscr{R}_w\lambda = \mathbf{0}$ in (S.27) and (S.28) respectively. By solving the optimization problems, we have

$$\widehat{\lambda} = \widetilde{\lambda} + \mathscr{T}^{-1}\mathscr{R}_w\widehat{\gamma}/2 \text{ and} \tag{S.32}$$

$$\overline{\lambda} = \widetilde{\lambda} + \mathscr{T}^{-1}\mathscr{R}_w\overline{\gamma}/2. \tag{S.33}$$

Applying Ekeland (1974), the value of the objective function evaluated at the optimum in the primal optimization problem

$$\begin{aligned} &-\left(\widehat{\lambda}-\mathscr{T}^{-1}G\right)'\mathscr{T}\left(\widehat{\lambda}-\mathscr{T}^{-1}G\right) + \left(\overline{\lambda}-\mathscr{T}^{-1}G\right)'\mathscr{T}\left(\overline{\lambda}-\mathscr{T}^{-1}G\right) \\ &= \max_{\mathscr{R}_w\lambda\geq\mathbf{0}} -\left(\lambda-\mathscr{T}^{-1}G\right)'\mathscr{T}\left(\lambda-\mathscr{T}^{-1}G\right) + \left(\overline{\lambda}-\mathscr{T}^{-1}G\right)'\mathscr{T}\left(\overline{\lambda}-\mathscr{T}^{-1}G\right) \end{aligned}$$

is equal to that in the dual problem

$$\min_{\gamma\leq\mathbf{0}} \left(\gamma-\overline{\gamma}\right)'\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\left(\gamma-\overline{\gamma}\right)/4 = \left(\widehat{\gamma}-\overline{\gamma}\right)'\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\left(\widehat{\gamma}-\overline{\gamma}\right)/4.$$

Since $\mathscr{R}_w\overline{\lambda} = \mathbf{0}$, we have

$$\mathscr{R}_w\widehat{\lambda} = \mathscr{R}_w\left(\widehat{\lambda}-\overline{\lambda}\right) = \mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w\left(\widehat{\gamma}-\overline{\gamma}\right)/2$$

by Equations (S.32) and (S.33). Therefore, it holds that

$$\left(\mathscr{R}_w\widehat{\lambda}\right)'\left(\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1}\left(\mathscr{R}_w\widehat{\lambda}\right) = \left(\widehat{\gamma}-\overline{\gamma}\right)'\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\left(\widehat{\gamma}-\overline{\gamma}\right)/4,$$

and Equation (S.30) follows.

Equation (S.31) holds if we can show that $\widehat{s} = \mathscr{R}_w\widehat{\lambda}$. The Karush-Kuhn-Tucker

(KKT) conditions for $\widehat{\lambda}$ are the followings:

$$2\mathscr{T}\left(\widehat{\lambda}-\mathscr{T}^{-1}G\right)-\mathscr{R}_w'\widehat{\gamma}=\mathbf{0}$$
$$\left(\widehat{\lambda}\mathscr{R}_w\right)'\widehat{\gamma}=0$$
$$\mathscr{R}_w\widehat{\lambda}\geq\mathbf{0}\text{ and }\widehat{\gamma}\geq\mathbf{0}.$$

Multiplying both sides of the first equality by $\mathscr{R}_w\mathscr{T}^{-1}$, we obtain

$$2\mathscr{R}_w\left(\widehat{\lambda}-\mathscr{T}^{-1}G\right)-\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\widehat{\gamma}=\mathbf{0},$$

and thus

$$\widehat{\gamma}=2\left(\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1}\left(\mathscr{R}_w\widehat{\lambda}-\mathscr{R}_w\mathscr{T}^{-1}G\right).$$

Substituting $\widehat{\gamma}$ in the other three KKT conditions with the above equality, we get

$$\left(\widehat{\lambda}\mathscr{R}_w\right)'\left(\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1}\left(\mathscr{R}_w\widehat{\lambda}-\mathscr{R}_w\mathscr{T}^{-1}G\right)=0$$
$$\mathscr{R}_w\widehat{\lambda}\geq\mathbf{0}\text{ and }\left(\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1}\left(\mathscr{R}_w\widehat{\lambda}-\mathscr{R}_w\mathscr{T}^{-1}G\right)\geq\mathbf{0}.\tag{S.34}$$

Next, we show that the KKT conditions for $\widehat{s}$ takes the same form. Let $\widehat{\gamma}_s$ be the Lagrange multiplier vector for the optimization problem (S.29). The following KKT conditions hold for $\widehat{s}$:

$$2\left(\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1}\left(\widehat{s}-\mathscr{R}_w\mathscr{T}^{-1}G\right)-\widehat{\gamma}_s=\mathbf{0}$$
$$\widehat{s}'\widehat{\gamma}_s=0$$
$$\widehat{s}\geq\mathbf{0}\text{ and }\widehat{\gamma}_s\geq\mathbf{0}.$$

After eliminating $\widehat{\gamma}_s$ using the first equality, we obtain:

$$\widehat{s}'\left(\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1}\left(\widehat{s}-\mathscr{R}_w\mathscr{T}^{-1}G\right)=0$$
$$\widehat{s}\geq\mathbf{0}\text{ and }\left(\mathscr{R}_w\mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1}\left(\widehat{s}-\mathscr{R}_w\mathscr{T}^{-1}G\right)\geq\mathbf{0}.\tag{S.35}$$

The concavity and differentiability of the objective function provides that the KKT conditions for $\widehat{s}$ are sufficient. Therefore, System (S.35) uniquely determines $\widehat{s}$. The equivalence of Systems (S.35) and (S.34) provides that $\mathscr{R}_w\widehat{\lambda}$ is also uniquely determined and $\widehat{s}=\mathscr{R}_w\widehat{\lambda}$. Therefore, Equation (S.1.7) holds. ■

Lemma S.1.8. *For testing $H_0:\mathscr{R}_w\theta^*=\mathrm{r}_w$ against $H_1:\mathscr{R}_w\theta^*\neq\mathrm{r}_w$ under $\Theta=$*

$\left\{\theta \in \mathbb{R}^l : \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}$, $W_{1,\omega} = QLR_{1,\omega} = S_{1,\omega}$ *for* $\Sigma_{W,\omega} = \mathscr{T}_\omega^{-1}$ *and* $\Sigma_{S,\omega} = \mathscr{R}_w \mathscr{T}_\omega^{-1}\mathscr{R}_w'$, *where definitions of each term can be found in Assumption 7.1 and Lemma 7.1.*

Proof: For the null hypothesis and maintained hypothesis defined in the lemma, we obtain that $\delta = c_\omega = c_{w,b}$ and c is empty. Applying the above result, along with $\Sigma_{W,\omega} = \mathscr{T}_\omega^{-1}$ and $\Sigma_{S,\omega} = \mathscr{R}_w \mathscr{T}_\omega^{-1}\mathscr{R}_w'$, in the expressions of $W_{1,\omega}$, $QLR_{1,\omega}$ and $S_{1,\omega}$, we have that

$$\begin{aligned}
W_{1,\omega} &= \left(\mathscr{R}_w\Psi_{1,\omega} + \delta\right)' \left(\mathscr{R}_w \mathscr{T}_\omega^{-1}\mathscr{R}_w'\right)^{-1} \left(\mathscr{R}_w\Psi_{1,\omega} + \delta\right), \text{ where} \\
\Psi_{1,\omega} &= \arg\min_{\mathscr{R}_w\lambda+\delta\geq \mathbf{0}} \left(\lambda - \mathscr{T}_\omega^{-1}G_\omega\right)' \mathscr{T}_\omega \left(\lambda - \mathscr{T}_\omega^{-1}G_\omega\right); \\
QLR_{1,\omega} &= \min_{\mathscr{R}_w\lambda+\delta= \mathbf{0}} \left(\lambda - \mathscr{T}_\omega^{-1}G_\omega\right)' \mathscr{T}_\omega \left(\lambda - \mathscr{T}_\omega^{-1}G_\omega\right) \\
&\quad - \min_{\mathscr{R}_w\lambda+\delta\geq \mathbf{0}} \left(\lambda - \mathscr{T}_\omega^{-1}G_\omega\right)' \mathscr{T}_\omega \left(\lambda - \mathscr{T}_\omega^{-1}G_\omega\right); \text{ and} \\
S_{1,\omega} &= ds_{1,\omega}' \left(\mathscr{R}_w \mathscr{T}_\omega^{-1}\mathscr{R}_w'\right)^{-1} ds_{1,\omega}, \text{ where} \\
ds_{1,\omega} &= \arg\min_{\lambda\geq\mathbf{0}} \left(\lambda - \mathscr{R}_w \mathscr{T}_\omega^{-1}G_\omega - \delta\right)' \left(\mathscr{R}_w \mathscr{T}_\omega^{-1}\mathscr{R}_w'\right)^{-1} \left(\lambda - \mathscr{R}_w \mathscr{T}_\omega^{-1}G_\omega - \delta\right).
\end{aligned}$$

The proof for Lemma S.1.7 can be directly used to obtain the equality $W_{1,\omega} = QLR_{1,\omega}$. As for the equality $W_{1,\omega} = S_{1,\omega}$, similar argument applies. The KKT conditions for $ds_{1,\omega}$ are

$$\begin{aligned}
ds_{1,\omega}' \left(\mathscr{R}_w \mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1} \left(ds_{1,\omega} - \mathscr{R}_w \mathscr{T}_\omega^{-1}G_\omega - \delta\right) &= 0 \\
ds_{1,\omega} \geq \mathbf{0} \text{ and } \left(\mathscr{R}_w \mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1} \left(ds_{1,\omega} - \mathscr{R}_w \mathscr{T}_\omega^{-1}G_\omega - \delta\right) &\geq \mathbf{0};
\end{aligned} \tag{S.36}$$

and the ones for $\Psi_{1,\omega}$ lead to

$$\begin{aligned}
\left(\mathscr{R}_w\Psi_{1,\omega} + \delta\right)' \left(\mathscr{R}_w \mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1} \left(\mathscr{R}_w\Psi_{1,\omega} - \mathscr{R}_w \mathscr{T}_\omega^{-1}G_\omega\right) &= 0 \\
\mathscr{R}_w\Psi_{1,\omega} + \delta \geq \mathbf{0} \text{ and } \left(\mathscr{R}_w \mathscr{T}^{-1}\mathscr{R}_w'\right)^{-1} \left(\mathscr{R}_w\Psi_{1,\omega} - \mathscr{R}_w \mathscr{T}_\omega^{-1}G_\omega\right) &\geq \mathbf{0}.
\end{aligned} \tag{S.37}$$

System (S.36) for $ds_{1,\omega}$ and System (S.37) for $\mathscr{R}_w\Psi_{1,\omega} + \delta$ are equivalent. The same argument in the proof for Lemma S.1.7 provides that $ds_{1,\omega} = \mathscr{R}_w\Psi_{1,\omega} + \delta$. Therefore, $W_{1,\omega} = S_{1,\omega}$ holds and the lemma follows. ■

S.2 Verification of Assumptions for Linear Regression Model

We present primitive conditions for Assumptions 4.2-4.4, 5.2, and 6.3 to hold in the linear regression model in Example 2.1. Let the model be indexed by n. The sample $(X_{ni}, Y_{ni})_{i=1}^{n}$ is row-wise i.i.d. The estimator objective function $l_n(\theta)$ is calculated as:

$$
\begin{aligned}
l_n(\theta) &= -\frac{1}{2}\sum_{i=1}^{n}\left(Y_{ni} - X_{ni}'\theta\right)^2 \\
&= -\frac{1}{2}\sum_{i=1}^{n}\varepsilon_{ni}^2 + \left(\sum_{i=1}^{n}\varepsilon_{ni}X_{ni}'\right)(\theta-\theta_n) + \frac{1}{2}(\theta-\theta_n)'\left(-\sum_{i=1}^{n}X_{ni}X_{ni}'\right)(\theta-\theta_n) \\
&= l_n(\theta_n) + Dl_n(\theta_n)(\theta-\theta_n) + \frac{1}{2}(\theta-\theta_n)' D^2 l_n(\theta_n)(\theta-\theta_n),
\end{aligned}
$$

with $l_n(\theta_n) = -\frac{1}{2}\sum_{i=1}^{n}\varepsilon_{ni}^2$, $Dl_n(\theta_n) = \sum_{i=1}^{n}\varepsilon_{ni}X_{ni}'$ and $D^2 l_n(\theta_n) = -\sum_{i=1}^{n}X_{ni}X_{ni}'$. Assumption 2.2 is trivially satisfied by the quadratic form of $l_n(\theta)$. Let $E_\omega(\cdot)$ denote the expectation respect to $\boldsymbol{P}_\omega$. Under the condition that for any $\omega \in \overline{\mathcal{W}}_0$, $E_\omega\left(\|X_{ni}\|^{4+\nu}\right) < M$ and $E_\omega\left(\|\varepsilon_{ni}\|^{4+\nu}\right) < M$ for some $\nu > 0$ and $M < \infty$, we can obtain the weak convergence of $b_n^{-1}Dl_n(\theta_n)$ for any $\omega_n \in \overline{\mathcal{W}}_0$ by the Lyapunov central limit theorem, where $b_n = \sqrt{n}$. Under the same condition, $-b_n D^2 l_n(\theta_n)$ converges in probability to $E_\omega(X_{ni}X_{ni}')$ by the weak law of large numbers for triangular arrays. If further $E_\omega(X_{ni}X_{ni}')$ is non-singular for any $\omega \in \overline{\mathcal{W}}_0$, then Assumption 2.2 holds. Assumptions 2.3 and 5.2 can be verified by Theorem 1 in Andrews (1997), which extends to probability models indexed by ω_n, if Assumptions 1-4 in Andrews (1997) hold for any $\omega_n \in \overline{\mathcal{W}}_0$. Assumptions 1 and 4 in Andrews (1997) are satisfied by the quadratic form of $l_n(\theta)$; and Assumptions 2 and 3 are guaranteed by the above conditions on $E_\omega\left(\|X_{ni}\|^{4+\nu}\right)$ and $E_\omega\left(\|\varepsilon_{ni}\|^{4+\nu}\right)$ being bounded and the non-singularity on $E_\omega(X_{ni}X_{ni}')$ for any $\omega \in \overline{\mathcal{W}}_0$. Alternatively, one can use the epi-convergence argument in Pflug (1994, 1995), Geyer (1994, 1996), and Knight (1999) to verify Assumptions 2.3 and 5.2. Such tool is powerful in dealing with estimators defined by constrained optimizations. At last Assumption 6.3 (i) holds by $R_n^D(\cdot) = 0$; and 6.3 (ii) is satisfied by letting $\widehat{\mathscr{T}_n} = \mathscr{T}_n = 2/n\sum_{i=1}^{n}X_{ni}X_{ni}'$.

References

Andrews, D. W. (1997): "Estimation when a Parameter Is on a Boundary: Theory and Applications," Discussion Paper No. 1153, Cowles Foundation, Yale University.

——— (1999): "Estimation when a Parameter Is on a Boundary," *Econometrica*, 67, 1341–1383.

——— (2001): "Testing when a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

Andrews, D. W. and X. Cheng (2012): "Estimation and Inference with Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211.

——— (2014): "GMM Estimation and Uniform Subvector Inference with Possible Identification Failure," *Econometric Theory*, 30, 287–333.

Andrews, D. W., X. Cheng, and P. Guggenberger (2011): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," Discussion Paper No. 1813, Cowles Foundation, Yale University.

Autor, D. H. and M. J. Handel (2013): "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages," *Journal of Labor Economics*, 31, S59–S96.

Bartholomew, D. (1961): "A test of homogeneity of means under restricted alternatives," *Journal of the Royal Statistical Society: Series B (Methodological)*, 23, 239–272.

Bastrakov, S. I. and N. Y. Zolotykh (2015): "Fast Method for Verifying Chernikov Rules in Fourier-Motzkin Elimination," *Computational Mathematics and Mathematical Physics*, 55, 160–167.

Berger, R. L. and D. D. Boos (1994): "P Values Maximized over a Confidence Set for the Nuisance Parameter," *Journal of the American Statistical Association*, 89, 1012–1016.

Cheng, X. (2015): "Robust Inference in Nonlinear Models with Mixed Identification Strength," *Journal of Econometrics*, 189, 207–228.

Chernikov, S. N. (1965): "The Convolution of Finite Systems of Linear Inequalities," *USSR Computational Mathematics and Mathematical Physics*, 5, 1–24.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81, 535–580.

Chernozhukov, V. and H. Hong (2004): "Likelihood Estimation and Inference in a Class of Nonregular Econometric Models," *Econometrica*, 72, 1445–1480.

Dantzig, G. B. and B. C. Eaves (1973): "Fourier-Motzkin Elimination and Its Dual," *Journal of Combinatorial Theory, Series A*, 14, 288–297.

DINES, L. L. (1919): "Systems of Linear Inequalities," *Annals of Mathematics*, 191–199.

EKELAND, I. (1974): *La Théorie des Jeux et ses Applications à l'Économie Mathématique*, Presses universitaires de France.

ENGLE, R. F. (1984): "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," *Handbook of Econometrics*, 2, 775–826.

FOURIER, J. (1824): "Histoire de l'Académie, Partie Mathématique," *Mémoires de l'Académie des Sciences de l'Institut de France*, 7.

FOX, J. T., K. IL KIM, AND C. YANG (2016): "A Simple Nonparametric Approach to Estimating the Distribution of Random Coefficients in Structural Models," *Journal of Econometrics*, 195, 236–254.

FOX, J. T., K. I. KIM, S. P. RYAN, AND P. BAJARI (2011): "A Simple Estimator for the Distribution of Random Coefficients," *Quantitative Economics*, 2, 381–418.

FRITZ, T. AND R. CHAVES (2012): "Entropic Inequalities and Marginal Problems," *IEEE Transactions on Information Theory*, 59, 803–817.

FUKUDA, K. AND A. PRODON (1995): "Double Description Method Revisited," in *Franco-Japanese and Franco-Chinese Conference on Combinatorics and Computer Science*, Springer, 91–111.

GEYER, C. J. (1994): "On the Asymptotics of Constrained M-Estimation," *The Annals of Statistics*, 1993–2010.

——— (1996): "On the Asymptotics of Convex Stochastic Optimization," Unpublished Manuscript, Department of Statistics, University of Minnesota.

GOURIEROUX, C., A. HOLLY, AND A. MONFORT (1982): "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," *Econometrica*, 63–80.

HENK, M., J. RICHTER-GEBERT, AND G. M. ZIEGLER (2004): "Basic Properties of Convex Polytopes," *Handbook of Discrete and Computational Geometry*, 255–382.

HUYNH, T., C. LASSEZ, AND J.-L. LASSEZ (1992): "Practical Issues on the Projection of Polyhedral Sets," *Annals of Mathematics and Artificial Intelligence*, 6, 295–315.

IMBERT, J.-L. (1993): "Fourier's Elimination: Which to Choose?" in *Principles and Practice of Constraint Programming*, 117–129.

JONES, C., E. C. KERRIGAN, AND J. MACIEJOWSKI (2004): "Equality Set Projection: A New Algorithm for the Projection of Polytopes in Halfspace Representation," Engineering Department, Cambridge University.

JUDGE, G., T. YANCEY, M. BOCK, AND R. BOHRER (1984): "The Non-Optimality of the Inequality Restricted Estimator under Squared Error Loss," *Journal of Econometrics*, 25, 165–177.

KETZ, P. (2018): "Subvector Inference When the True Parameter Vector May Be near or at the Boundary," *Journal of Econometrics*, 207, 285–306.

KNIGHT, K. (1999): "Epi-convergence in Distribution and Stochastic Equi-semicontinuity," Unpublished Manuscript, Department of Statistics, University of Toronto.

KODDE, D. A. AND F. C. PALM (1986): "Wald Criteria for Jointly Testing Equality and Inequality Restrictions," *Econometrica*, 1243–1248.

KUDO, A. (1963): "A Multivariate Analogue of the One-Sided Test," *Biometrika*, 50, 403–418.

LIEW, C. K. (1976): "Inequality Constrained Least-squares Estimation," *Journal of the American Statistical Association*, 71, 746–751.

MCCLOSKEY, A. (2017): "Bonferroni-Based Size-Correction for Nonstandard Testing Problems," *Journal of Econometrics*, 200, 17–35.

MINCER, J. (1974): *Schooling, Experience, and Earnings*, New York: NBER Press.

MOON, H. R. AND F. SCHORFHEIDE (2009): "Estimation with Overidentifying Inequality Moment Conditions," *Journal of Econometrics*, 153, 136–154.

MOTZKIN, T. S. (1936): "Beitrage zur Theorie der Linearen Ungleichungen," Doctoral Thesis, University of Basel.

NÜESCH, P. E. (1966): "On the Problem of Testing Location in Multivariate Populations for Restricted Alternatives," *The Annals of Mathematical Statistics*, 37, 113–119.

PAKES, A. AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica: Journal of the Econometric Society*, 1027–1057.

PERLMAN, M. D. (1969): "One-Sided Testing Problems in Multivariate Analysis," *The Annals of Mathematical Statistics*, 40, 549–567.

PFLUG, G. C. (1994): "On an Argmax-Distribution Connected to the Poisson Process," in *Mandl P., Hušková M. (eds) Asymptotic Statistics*, Heidelberg: Physica-Verlag, 123–129.

——— (1995): "Asymptotic Stochastic Programs," *Mathematics of Operations Research*, 20, 769–789.

PONCE, J., S. SULLIVAN, A. SUDSANG, J.-D. BOISSONNAT, AND J.-P. MERLET (1997): "On Computing Four-Finger Equilibrium and Force-Closure Grasps of Polyhedral Objects," *The International Journal of Robotics Research*, 16, 11–35.

ROMANO, J. P., A. M. SHAIKH, AND M. WOLF (2014): "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 82, 1979–2002.

ROMANO, J. P. AND M. WOLF (2000): "Finite Sample Nonparametric Inference and Large Sample Efficiency," *Annals of Statistics*, 28, 756–778.

SILVAPULLE, M. J. (1996): "A Test in the Presence of Nuisance Parameters," *Journal of the American Statistical Association*, 91, 1690–1693.

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

Silvapulle, M. J. and P. Silvapulle (1995): "A Score Test Against One-Sided Alternatives," *Journal of the American Statistical Association*, 90, 342–349.

Staiger, D. O. and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Telgen, J. (1983): "Identifying Redundant Constraints and Implicit Equalities in Systems of Linear Constraints," *Management Science*, 29, 1209–1222.

Wolak, F. A. (1987): "An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model," *Journal of the American Statistical Association*, 82, 782–793.

——— (1989): "Local and Global Testing of Linear and Nonlinear Inequality Constraints in Nonlinear Econometric Models," *Econometric Theory*, 5, 1–35.

——— (1991): "The Local Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models," *Econometrica*, 981–995.