Uniform Inference when Parameters are Subject to Linear Inequality Constraints

Xuetao Shi

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Yanqin Fan, Chair

Fang Han

Jing Tao

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Uniform Inference when Parameters are Subject to Linear Inequality Constraints

Xuetao Shi

Chair of the Supervisory Committee:
Professor Yanqin Fan
Department of Economics

This dissertation studies the problem of uniform inference when model parameters are subject to linear inequality constraints. Linear inequality constraints such as non-negativity and monotonicity are often implied by nature of the parameters or imposed by economic theory. Chapter 1 seeks to develop Wald-type, QLR, and score-type tests for testing linear equality constraints against two-sided alternative hypotheses in a general class of extremum estimation problems. The null asymptotic distributions of Wald and QLR statistics are discontinuous in model parameters demanding for uniform inference, and the one of score statistic depends on a polytope projection. For each test, we provide steps on obtaining the critical value and conditions under which the asymptotic size is controlled and the test is consistent. Linear inequality constraints are particularly common in models for random intervals. Chapter 2 develops asymptotically uniformly valid tests for linear equality constraints on the parameters in the interval data model. Moreover, based on interval arithmetic, we introduce a new interval data model to extend and generalize the commonly used one, and propose a coefficient of determination.

More specifically, the first chapter develops Wald-type, QLR, and score-type tests for linear equality constraints against two-sided alternative hypotheses in a general class of extremum estimation problems, where the parameter space is characterized by a finite number of linear equality and inequality constraints. It shows that the null asymptotic distribu-

tions of the Wald and QLR statistics are discontinuous in an implicit nuisance parameter and proposes an algorithm to identify it. In contrast, the null asymptotic distribution of the score statistic is not discontinuous in any model parameter but depends on a polytope projection. The chapter presents an algorithm based on the Fourier-Motzkin elimination to compute such projection. It studies consistency and local power properties of the three tests, and finds that the score test may be inconsistent due to the use of partial information in the parameter space through projection. Numerical results from a Monte Carlo study of the finite sample performance of our tests are presented. An empirical illustration on Mincer earnings regression is conducted.

Via generalized interval arithmetic, the second chapter proposes a Generalized Interval Arithmetic Center and Range (GIA-CR) model for random intervals in which parameters in the model satisfy linear inequality constraints. It extends the commonly used Center and Range (CR) model in several directions. For the GIA-CR model, this chapter constructs a constrained estimator of the parameter vector and proposes a coefficient of determination. It develops asymptotically uniformly valid tests for linear equality constraints on the parameters in the model. At last, this chapter conducts a simulation study to examine the finite sample performance of the estimator and test for the correct specification of the CR model against the Interval Arithmetic Center and Range (IA-CR) model in which the parameters in the range regression satisfy non-negativity constraints.

TABLE OF CONTENTS

Page

List of Tables . . . iii

Chapter 1: Wald, QLR, and Score Tests When Parameters are Subject to Linear Inequality Constraints . . . 1

1.1 Introduction . . . 1
1.2 Extremum Estimation, Test Statistics, and Two Running Examples . . . 7
1.3 Undetermined Inequalities and Implicit Nuisance Parameter . . . 14
1.4 The Wald Test . . . 20
1.5 The Quasi Likelihood Ratio Test . . . 27
1.6 The Score Test . . . 30
1.7 Local Power . . . 37
1.8 Monte-Carlo Simulation . . . 39
1.9 An Empirical Illustration on Mincer Earnings Regression . . . 43
1.10 Concluding Remarks . . . 45

Chapter 2: Uniform Inference in a Generalized Interval Arithmetic Center and Range Linear Model . . . 47

2.1 Introduction . . . 47
2.2 The Model, Estimation, and Coefficient of Determination . . . 53
2.3 Asymptotic Properties of the Constrained Estimators . . . 59
2.4 Wald-Type Tests for Linear Hypothesis in the GIA-CR Model . . . 66
2.5 A Simulation Study . . . 77
2.6 Concluding Remarks . . . 82

Appendix A: Wald, QLR, and Score Tests When Parameters are Subject to Linear Inequality Constraints . . . 92

A.1 Technical Proofs . . . 92

A.2 Verification of Assumptions for Linear Regression Model 108

Appendix B: Uniform Inference in a Generalized Interval Arithmetic Center and Range Linear Model 110

B.1 Technical Proofs 110

B.2 A Review of Generalized Interval Arithmetic and Random Generalized Intervals 121

LIST OF TABLES

Table Number | Page

1.1 Different sets of parameters 41
1.2 Rejection probability under H_0 42
1.3 Finite sample size-corrected power 42
1.4 Significance results 44

2.1 MSE Comparison 79
2.2 Size Performance-Reject Percentage 82

ACKNOWLEDGMENTS

This thesis is the result of an intense yet totally fulfilling journey. Its completion would not have been possible without the help of many people, whom I would like to thank here.

I would like to express my deepest gratitude to Professor Yanqin Fan for her extensive support, encouragement, and insightful guidance. To me, she is not only my advisor and mentor, but also my dearest friend and family. She introduced me to the world of theoretical econometrics where I found exciting research projects, and furthermore worked selflessly to help me bring this dissertation to fruition. Her hard-working attitude and passion for research makes her a great role model from whom I will continue to learn in the future. Moreover, as a caring, responsible and enthusiastic teacher who is immensely knowledgable in her fields, she has always inspired me to become a great teacher like her.

I am truly thankful to my committee members Professor Jing Tao and Professor Fang Han, both of whom have shown great kindness and patience to me. I thank Professor Jon Wellner for having taught me several fascinating statistical courses and always making him available to have conversations with me. I want to truly thank Professor Fahad Khalil and Professor Quan Wen. I receive priceless inspiration from them. I would also like to thank my fellow students: Sunny Huang, Ming He, Ken Ho, and An-tsu Chen for their support as my companions, and Simon Reeve-Parker for all his help.

To my parents, I would like to extend my sincere gratefulness to their unconditional love, care, and support. They raised me up with love and vision; they taught me to be brave, to venture, and to pursue the life I want. I also thank my wife, Zhe Jiang, without whose companion and support, I would not have been able to complete my Ph.D. degree with joy. She is always there cheering me up, standing by me and supporting me spiritually throughout

writing the dissertation and my life.

Chapter 1

WALD, QLR, AND SCORE TESTS WHEN PARAMETERS ARE SUBJECT TO LINEAR INEQUALITY CONSTRAINTS [1]

1.1 Introduction

Motivation and Main Contributions This paper develops Wald-type (Wald hereafter), Quasi-Likelihood Ratio (QLR), and score-type (score hereafter) tests for linear equality constraints against the two-sided alternative hypotheses in a general class of extremum estimation problems, where the parameter space is characterized by a finite number of linear equality and inequality constraints. Equality and inequality constraints are often imposed by economic theory like non-negativity or monotonicity (Mincer (1974)); or implied by the nature of the parameters such as weights which are non-negative and add up to one (Fox et al. (2011) and Fox et al. (2016)). It has long been recognized that incorporating 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). Furthermore, 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."

Asymptotic theory for extremum estimation subject to such linear equality and inequality constraints follow from Andrews (1997, 1999). Specifically, Andrews (1997, 1999) show that the asymptotic distribution of an extremum estimator depends on how many and which inequalities bind in the parameter space. This paper complements Andrews (1997, 1999) by developing asymptotically uniformly valid Wald, QLR, and score tests in such an extremum estimation set-up.

[1] This Chapter is a joint work with Yanqin Fan from the University of Washington.

The null asymptotic distributions of test statistics using full information in the parameter space such as the Wald and QLR statistic depend critically on the *undetermined inequalities* in the *null parameter space*. The first contribution of this paper is to identify and characterize the set of undetermined inequalities in the null parameter space for testing a *general linear hypothesis* against the two-sided alternative hypothesis in the presence of *general linear inequality constraints* in the parameter space. Specifically, we introduce the important concept of an *implicit nuisance parameter* to characterize undetermined inequalities in the null parameter space. An implicit nuisance parameter is defined as a subvector of the linear function evaluated at the pseudo-true value of the parameter in the undetermined inequalities that corresponds to a row basis of the coefficient matrix (see Definition 1.3.2). We propose a generic algorithm for identifying the undetermined inequalities and the implicit nuisance parameter. The algorithm involves two steps: (i) we identify the implicit equalities and strictly redundant inequalities in the null parameter space by applying algorithms in Telgen (1983), leaving the rest as undetermined inequalities; (ii) we apply Gauss-Jordan elimination to identify an implicit nuisance parameter, which typically consists of linear combinations of the original parameters. We note that the implicit nuisance parameter only depends on the null hypothesis and the parameter space; it does not depend on any specific model, estimator, or test statistic.

The second contribution of this paper is to establish the null asymptotic distributions of three test statistics: the Wald and QLR statistics which take the same forms as the classical ones respectively, and the score statistic which extends the one introduced in Andrews (2001) for testing subvector hypotheses.[2] For the Wald and QLR statistics, we show that when there are undetermined inequalities in the null parameter space, their null asymptotic distributions depend on which inequality constraints in the set of undetermined inequalities bind. More concisely, we show that the null asymptotic distributions of the Wald and

[2]In Fan and Shi (2019), we develop another Wald test and two additional score tests, and establish some equivalence results among them.

QLR statistics are *discontinuous* in the implicit nuisance parameter.[3] In contrast, we show that the null asymptotic distribution of the score statistic does not depend on the implicit nuisance parameter (even when there are undetermined inequalities in the null parameter space).

Based on the null asymptotic distributions, we develop asymptotically uniformly valid Wald, QLR, and score tests. This constitutes the third contribution of this paper. For the Wald and QLR tests, their critical values are constructed via a two-step procedure based on a confidence set for the implicit nuisance parameter in the first step and a Bonferroni-type correction in the second step. Such two-step procedures have been widely adopted in the *subvector inference* in broad contexts for which there is an explicit nuisance parameter characterizing the undetermined inequalities in the null parameter space;[4] see the works cited in the literature review below. For general (non-subvector) hypotheses considered in this paper, the implicit nuisance parameter we introduce plays the same role as the explicit nuisance parameter in the subvector inference. This makes identifying the former essential. It is worthwhile emphasizing that for any given null hypothesis and parameter space, the implicit nuisance parameter only needs to be identified once. The two-step procedure can be used to construct asymptotically uniformly valid tests in any parametric or semiparametric model using any test statistic. Although the null asymptotic distribution of the score statistic does not depend on the implicit nuisance parameter, implementing the score test 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)). We present an algorithm based on Fourier-Motzkin elimination to compute such projection. Fourth, we investigate consistency and local power properties of all three tests. The Wald and QLR tests

---

[3] 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 $d\left(\cdot,\cdot\right)$ is some metric on the probability measures, such as the Kolmogorov, bounded Lipschitz, or total variation metric and $|\cdot|$ is the absolute value.

[4] To simplify the exposition, we refer to an parameter in which the asymptotic distribution of an estimator or the null asymptotic distribution of a test statistic is discontinuous as the nuisance parameter.

fully exploit information in the parameter space, and are shown to be consistent. On the other hand, since the score test only employs part of the information in the parameter space through projection, it may be inconsistent. Fifth, we provide a complete study of the Wald, QLR, and score tests for two running examples: a random coefficients model in Fox et al. (2011) and the Mincer earnings regression in Autor and Handel (2013). Lastly, we conduct a simulation study using a linear regression model to investigate the finite sample performance of the tests developed in this paper. Results demonstrate that Wald and QLR tests dominate the score test. An empirical illustration on Mincer earnings regression following Autor and Handel (2013) is also presented.

Related Literatures Asymptotic theory for the classical Wald, QLR, and score tests rely on the assumption that the true value of the parameter is in the *interior* of its parameter space. Under this assumption and mild regularity conditions, the Wald, QLR, and score statistics for testing the null hypothesis of equality constraints against the two-sided alternative hypotheses have the same asymptotic $\mathcal{X}^2$ distribution under the null hypothesis, and the three tests are asymptotically equivalent (see e.g., Engle (1984)). This paper shows that neither property holds in general when the parameter space is characterized by linear inequality constraints and it is unknown a priori how many and which inequality constraints bind under the null hypothesis.

Existing works have also studied Wald, QLR, and score tests for testing equality constraints against one-sided alternatives (see Gourieroux et al. (1982), Kodde and Palm (1986), and Silvapulle and Sen (2005)). The hypotheses in these works can be reformulated as special cases of the general hypotheses considered in this paper with inequality constraints in the parameter space. Under the reformulation, the weak inequalities in the parameter space are known to bind under the null hypothesis for all the cases in Gourieroux et al. (1982) and Kodde and Palm (1986). As a result, there is no implicit nuisance parameter and the null asymptotic distributions of the test statistics are continuous in model parameters. The standard plug-in approach can be used to obtain critical values.

Andrews (2001) and Ketz (2018) develop asymptotically uniformly valid tests for subvector hypotheses when it is unknown a priori whether some parameters are in the interior or on the boundary of a general parameter space.[5] Assumptions in Andrews (2001) and Ketz (2018) ensure that the null asymptotic distributions of the test statistics considered are continuous in all model parameters. This is achieved in Andrews (2001) by assuming that (i) the normalized "information" matrix is block diagonal between the parameters that are known to lie on the boundary or in the interior of the parameter space and the ones whose locations are completely unknown, and (ii) the approximating cone of the parameter space is a product set. Ketz (2018) introduces a Conditional Likelihood Ratio statistic. Under the assumption that (i) some inequality constraints are empirically irrelevant, and (ii) 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 null asymptotic distribution of the Conditional Likelihood Ratio statistic is nuisance parameter free.

Methods for constructing asymptotically uniformly valid subvector inference in the presence of discontinuity have been proposed in different contexts. They include 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-step 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 mod-

[5] In Andrews (2001), the presence of an unidentified nuisance parameter under the null hypothesis is also allowed.

[6] Wolak (1987, 1989, 1991) develop 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.

els 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.

Organization of the Rest of This Paper The rest of this paper is organized as follows. In Section 1.2, we introduce the extremum estimation set-up, hypotheses, test statistics, and two running examples. In Section 1.3, we introduce the concept of an implicit nuisance parameter and our algorithm for identifying undetermined inequalities in the null parameter space and the implicit nuisance parameter. The algorithm is illustrated on the two running examples. In Section 1.4, we first provide a detailed construction and technical treatment of asymptotically uniformly valid Wald tests for subvector hypotheses and then extend our results to testing null hypotheses of general linear equality constraints. Sections 1.5 and 1.6 develop QLR and score tests respectively. Section 1.7 studies the local power of all three tests. Section 1.8 reports results from a simulation study. An empirical illustration is presented in Section 1.9. The last section offers some concluding remarks and possible extensions. The technical proofs and primitive conditions for the assumptions discussed in the paper for the linear regression model are provided in an online appendix of the paper.

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. We use "$\equiv$" to denote "equals by definition". 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

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

1.2 Extremum Estimation, Test Statistics, and Two Running Examples

Let $l_n(\theta)$ denote the estimator objective function that depends on the data, where $n = 1, 2, ...$ denotes the sample size, and $\theta \in \Theta \subset \mathbb{R}^l$. The parameter space Θ is of the form of a convex polytope defined as

$$\Theta \equiv \left\{\theta \in \mathbb{R}^l : \mathscr{R}_e\theta = \mathrm{r}_e \text{ and } \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}, \tag{1.1}$$

where $\mathscr{R}_e$ 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 $\mathscr{R}_w$ is allowed to be *row rank deficient* to incorporate constraints like $0 \leq \theta \leq 1$.

1.2.1 An Extremum Estimator and Asymptotic Distribution

In this section, we present a brief review of the asymptotic distribution of the extremum estimator in Andrews (1999) denoted as $\widehat{\theta}$, i.e., $\widehat{\theta} \in \Theta$ and

$$l_n\left(\widehat{\theta}\right) = \sup_{\theta\in\Theta} l_n(\theta) + o_p(1). \tag{1.2}$$

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^*:

$$l_n(\theta) = l_n(\theta^*) + Dl_n(\theta^*)(\theta-\theta^*) + \frac{1}{2}(\theta-\theta^*)' D^2 l_n(\theta^*)(\theta-\theta^*) + R_n(\theta), \tag{1.3}$$

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

Assumption 1.2.1. *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 1.2.2. $\left(b_n^{-1}Dl_n\left(\theta^*\right), \mathscr{T}_n\right) \xrightarrow{d} \left(G, \mathscr{T}\right)$ *for some random variables* $G \in \mathbb{R}^l$ *and* $\mathscr{T} \in \mathbb{R}^{l\times l}$*, where* $\mathscr{T}_n \equiv -b_n^{-2}D^2 l_n\left(\theta^*\right)$ *and* $\mathscr{T}$ *is symmetric and non-singular with probability one.*

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

Assumption 1.2.3. $b_n\left(\widehat{\theta} - \theta^*\right) = O_p\left(1\right)$.

The above assumptions do not rule out the case where $l_n\left(\cdot\right)$ 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, $Dl_n\left(\theta^*\right)$ could contain left or right partial derivatives. Andrews (1997, 1999) offer detailed discussions on the assumptions and provide 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), where $I_{l\times l}$ denotes the identify matrix of dimension $l \times l$. 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). A sequel to this paper will explore the applicability of this approach to trending data.

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

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

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

Under Assumptions 1.2.1-1.2.3, the lemma below follows from Theorem 3 (a) in Andrews (1999).

Lemma 1.2.1. *Suppose Assumptions 1.2.1-1.2.3 hold. Then*

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

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

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

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 (2.3), it is straightforward to show that the expression in Lemma 1.2.1 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]

1.2.2 Hypotheses and Test Statistics

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

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

where $\Theta_0\equiv\{\theta\in\Theta: R\theta=r\}$ or equivalently,[8]

$$\Theta_0=\left\{\theta\in\mathbb{R}^l: R\theta=r,\ \mathscr{R}_e\theta=\mathrm{r}_e, \text{ and } \mathscr{R}_w\theta\geq\mathrm{r}_w\right\},\tag{1.4}$$

[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 the online appendix.

[8] 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 1.6.2 of this paper can be used to determine whether Θ_0 is empty or not.

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

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

The Wald and QLR test statistics take the standard forms:

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

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) for subvector hypotheses. 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

$$Dl_n(\theta) = Dl_n(\theta^*) + D^2 l_n(\theta^*)(\theta - \theta^*) + R_n^D(\theta), \tag{1.5}$$

where $Dl_n(\theta^*)$ and $D^2 l_n(\theta^*)$ are defined in (1.3), and $R_n^D(\theta)$ is the remainder term satisfying Assumption 6.3 (i) in Section 1.6. 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

$$\widehat{q}_R(ds_n) = \inf_{\lambda_R \in b_n(R\Theta - r)} \widehat{q}_R(\lambda_R) + o_p(1), \tag{1.6}$$

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

$$S_n \equiv ds_n' \Sigma_{S,n}^{-1} ds_n, \tag{1.7}$$

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

Let T_n denote any of the above test statistics. We now introduce the concept of the asymptotic size of a test based upon T_n following Andrews and Guggenberger (2010). Suppose the model of interest is fully characterized by the finite dimensional parameter $\theta^* \in \Theta$ and the infinite dimensional parameter $\psi^* \in \Psi$ consistent with the value θ^*. 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

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

1.2.3 Two Running Examples

We now introduce two running examples for which we will offer a complete treatment of the Wald, QLR, and score tests developed in the paper.

Random Coefficients Model

Fox et al. (2011) proposes an inequality-constrained least squares estimator in a random coefficients model. It considers the classical setting of the discrete choice models. Agents $i = 1, \ldots, n$ can choose between $j = 1, \ldots, J$ mutually exclusive alternatives and one outside good. The binary observable dependent variable $Y_{i,j}$ takes the value one if the choice j renders the highest utility and zero otherwise. Denote $g_j\left(X_i, \beta\right)$ as the conditional probability of observing $Y_{i,j} = 1$ conditional on the random coefficient $\beta \in \mathbb{R}^{l_\beta}$ and characteristics X_i. The functional form of g_j is known. Depending on the assumption, g_j can take different forms.

Following Fox et al. (2011), we assume that the random coefficient β has a finite number l of known "support points" β^m with weights θ^*_m, for $m = 1, \ldots, l$; see Section 3 in Fox et al. (2011) for the discussion of the choice of β^m. One can consider such an assumption as having l types of agents. The probability $Y_{i,j} = 1$ conditional on X_i denoted as $\Pr(Y_{i,j} = 1 \mid x_i)$ is given by $\Pr(Y_{i,j} = 1 \mid x_i) = \sum_{m=1}^{l} \theta^*_m g_j(X_i, \beta^m)$. Let $\theta^* \equiv (\theta^*_1, \ldots, \theta^*_l)' \in \Theta$. The parameter space Θ can be expressed as $\Theta = \left\{\theta \in \mathbb{R}^l : \mathscr{R}_e\theta = \mathrm{r}_e \text{ and } \mathscr{R}_w\theta \geq \mathrm{r}_w\right\}$, where $\mathscr{R}_e = (1, ..., 1)$, $\mathrm{r}_e = 1$, $\mathscr{R}_w = I_{l\times l}$, and $\mathrm{r}_w = (0, 0, ..., 0)'$.

The objective of the econometrician is to estimate θ^*, and more importantly, to conduct inference on the cumulative distribution function of the random coefficient β denoted as $F_\beta(\cdot)$. Let $I[\cdot]$ be the indicator function. It is straightforward to see that $F_\beta(\cdot) = \sum_{m=1}^{l} \theta^*_m I[\beta^m \leq \cdot]$. Fox et al. (2011) proposes to estimate θ^* by an inequality-constrained least squares estimator. Define $Z_{i,j} \equiv (Z_{i,j,1}, \ldots, Z_{i,j,l})'$ with $Z_{i,j,m} = g_j(X_i, \beta^m)$. The sample objective function can be written as $l_n(\theta) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}\left(Y_{i,j} - Z'_{i,j}\theta\right)^2$ for $\theta \in \Theta$. Under mild conditions discussed in Section S.2 of the online appendix, we have $b_n = \sqrt{n}$, $Dl_n(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{J} Z_{i,j}\left(Y_{i,j} - Z'_{i,j}\theta\right)$, and $D^2 l_n(\theta) = -\sum_{i=1}^{n}\sum_{j=1}^{J} Z_{i,j}Z'_{i,j}$.

For any given $b \in \mathbb{R}^{l_\beta}$, confidence intervals for $F_\beta(b)$ can be constructed by inverting tests for the linear hypothesis $H_0 : \sum_{m=1}^{l} \theta^*_m I[\beta^m \leq b] = r$ for $r \in [0, 1]$ against $H_1 : \sum_{m=1}^{l} \theta^*_m I[\beta^m \leq b] \neq r$, or equivalently for $H_0 : R\theta^* = r$ against $H_1 : R\theta^* \neq r$, where $R = \left(I[\beta^1 \leq b], ..., I\left[\beta^l \leq b\right]\right)$.

As noted in Fox et al. (2011), many estimated weights are zero or close to zero calling for inferential procedures that are asymptotically uniformly valid. However, Fox et al. (2011) also notes that "the reality is that this recent literature has not developed general-enough results that could allow us to estimate confidence intervals for our problem in a way that gives asymptotically correct coverage as defined by Andrews and Guggenberger (2010)."

This paper develops Wald, QLR, and score tests that can be used to construct asymptotically uniformly valid confidence intervals for the distribution of random coefficients at any given value of the random coefficient.

Mincer Earnings Regression

Using original and representative survey data, Autor and Handel (2013) conduct a comprehensive study on the interaction among human capital, job tasks, and wages. They first introduce a conceptual framework on the causal links between human capital endowments, occupational assignment, job tasks, and wages, and then conduct empirical estimation and tests. Focusing on regressions on wage differentials related to job tasks and human capital, we apply the tests developed in this paper to three regressions in Autor and Handel (2013), and compare the results with the t-test in Section 1.9.

To illustrate, consider one of the regressions in Autor and Handel (2013) on the log hourly wages:

$$\log Wage_i = E_i'\beta^* + X_i'\mu^* + \varepsilon_i, i = 1, ..., n,$$

where $X_i \in \mathbb{R}^{l_x}$ is a vector of demographic variables and $E_i \equiv (E_{1,i}, \ldots, E_{4,i})'$. For $j = 1, \ldots, 4$, $E_{j,i}$ indicates 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". The reference group for the regression is that of high school graduates. This regression is further discussed in Section 1.9.

Let $\beta^* \equiv (\beta_1^*, \ldots, \beta_4^*)'$. 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 economic theory, the parameter space is expressed as $\Theta = \{\theta : \mathscr{R}_w\theta \geq \mathbf{0}\}$, where $\theta = (\beta', \mu')'$ and $\mathscr{R}_w = (\mathscr{R}_{w,\beta}, \mathbf{0}_{4\times l_x})$ with

$$\mathscr{R}_{w,\beta} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

In addition to tests of significance of individual parameters, researchers may also be interested in testing joint hypotheses on β^*. As an example, consider testing $H_0 : \beta_1^* = 0$ and $\beta_2^* + 0.1 = \beta_4^*$. Under this null hypothesis, the monetary return of obtaining a high school degree is zero, and the benefit of having a postcollege degree is a ten percentage increase in wage comparing to having some college education. Under the matrix notation, the joint hypothesis can be written as $H_0 : R\theta^* = r$, where $R = (R_\beta, \mathbf{0}_{2\times l_x})$,

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

1.3 Undetermined Inequalities and Implicit Nuisance Parameter

Lemma 1.2.1 suggests that when there are undetermined inequalities in the null parameter space Θ_0 in (1.4), the null asymptotic distribution of a test statistic for H_0 may take different forms depending on which inequalities bind. To develop asymptotically uniformly valid tests for H_0, the first and a critical step 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* (those that are known to bind), *strictly redundant inequalities* (those that are known not to bind), and *undetermined inequalities* in the null parameter space Θ_0 in (1.4).

This section proposes an approach for accomplishing this task, and introduces the important concept of an implicit nuisance parameter to characterize undetermined inequalities in the null parameter space Θ_0 when they exist.

1.3.1 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:

$$\begin{pmatrix} \mathscr{R}_e \\ R \end{pmatrix} \theta = \begin{pmatrix} \mathrm{r}_e \\ r \end{pmatrix} \tag{1.8}$$

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:

$$\mathscr{R}_w\Gamma\theta_f \geq \mathrm{r}_w - \mathscr{R}_w\gamma. \tag{1.9}$$

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 (1.9) will be known to bind, some will be known not to bind, and some are undetermined. 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 (1.9) 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 (1.9), 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 j-th row of a matrix or a vector. For any $j \in \mathfrak{J}$, let

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

Definition 1.3.1. *In the system of inequalities (1.9), 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 (1.9). The detailed algorithms can be found in Telgen (1983). 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 (1.9) 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 (1.9) 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 $\eta^u \geq \mathrm{r}_w^u - \mathscr{R}_w^u\gamma$.

1.3.2 Implicit Nuisance Parameter

We now introduce the concept of an implicit nuisance parameter when there are undetermined inequalities in (1.9).

Definition 1.3.2. *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 (1.9) become:

$$\Gamma^u\eta^k \geq \mathrm{r}_w^u - \mathscr{R}_w^u\gamma. \tag{1.10}$$

Remark 1.3.1. Given Θ_0 and $\theta \in \Theta_0$, the undetermined inequalities among $\mathscr{R}_w\theta \geq \mathrm{r}_w$ are unique. However, 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 (1.8) 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-step 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 1.2. 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 1.3.1. *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. Furthermore, we assume throughout the rest of the paper that there are undetermined inequalities in the null parameter space Θ_0. If there is no undetermined inequality after applying Steps a and b in Section 1.3.1, the conventional plug-in approach would be applicable to constructing asymptotically valid tests; see Gourieroux et al. (1982), Kodde and Palm (1986), and Silvapulle and Sen (2005).

1.3.3 Running Examples (Continued)

Our approach for identifying an implicit nuisance parameter is applicable to any $\mathscr{R}_e$, $\mathscr{R}_w$, and R. We illustrate this by deriving implicit nuisance parameters for the running examples introduced in Sections 1.2.3 and 1.2.3 respectively.

Random Coefficients Model

To implement the Wald and QLR tests for $H_0 : R\theta^* = r$ developed in Sections 1.4 and 1.5 of the paper, where $R = \left(I\left[\beta^1 \leq b\right], ..., I\left[\beta^l \leq b\right]\right)$ and $r \in [0,1]$, the key step is to identify an implicit nuisance parameter. To simplify the exposition, we assume that β is one dimensional. The analysis can be directly extended to cases where β is multidimensional. For some $b \in \mathbb{R}$, $F_\beta(b) = \sum_{m=1}^{l_1} \theta_m^*$, where l_1 is the largest value of m such that $\beta^m \leq b$. Confidence intervals for the distribution of the random coefficient at any specific point can be constructed by testing $H_0 : R\theta^* = r$, for some $R = (1, \ldots, 1, 0, \ldots, 0)'$ with l_1 number of ones and $r \in [0, 1]$.

Decompose $\mathscr{R}_w = \left(\mathscr{R}_{w,1}', \mathscr{R}_{w,2}'\right)'$, where $\mathscr{R}_{w,1}$ consists of the first l_1 rows of $\mathscr{R}_w$ and $\mathscr{R}_{w,2}$ consists of the remaining rows. We solve the system of linear equations (1.8) to obtain that

$$\Gamma = \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ -1 & \cdots & -1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \\ 0 & \cdots & 0 & -1 & \cdots & -1 \end{pmatrix}, \theta_f = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_{l_1-1} \\ \theta_{l_1+1} \\ \vdots \\ \theta_{l-1} \end{pmatrix}, \text{ and } \gamma = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ r \\ 0 \\ \vdots \\ 0 \\ 1-r \end{pmatrix},$$

where $\Gamma \in \mathbb{R}^{l\times(l-2)}$, $\theta_f \in \mathbb{R}^{l-2}$, and $\gamma \in \mathbb{R}^l$. For $r = 0$, by applying Steps a and b in Section 1.3.1, we have that $Sub_b = \{1, 2, \ldots, l_1\}$ and $Sub_{nb} = \emptyset$. Thus, $\mathscr{R}_w^b = \mathscr{R}_{w,1}$ and $\mathscr{R}_w^u = \mathscr{R}_{w,2}$. The row basis of $\mathscr{R}_{w,2}\Gamma$ consists of its first $(l - l_1 - 1)$ rows. The implicit nuisance parameter is therefore obtained as $\eta^k = \left(\theta_{l_1+1}^*, \ldots, \theta_{l-1}^*\right)'$. For other values of r, the same procedure would provide that $\mathscr{R}_w^u = \mathscr{R}_w$ and $\eta^k = \left(\theta_1^*, \ldots, \theta_{l_1-1}^*, \theta_{l_1+1}^*, \ldots, \theta_{l-1}^*\right)'$ when $r \in (0, 1)$, and $\mathscr{R}_w^b = \mathscr{R}_{w,2}$, $\mathscr{R}_w^u = \mathscr{R}_{w,1}$, and $\eta^k = \left(\theta_1^*, \ldots, \theta_{l_1-1}^*\right)'$ when $r = 1$. The result shows that the implicit nuisance parameter depends on the value of r in the null hypothesis.

Mincer Earnings Regression

We focus the discussion on the joint hypothesis $H_0 : R\theta^* = r$ discussed in Section 1.2.3, where $R = (R_\beta, \mathbf{0}_{2\times l_x})$,

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

Because the inequality constraints and the joint hypothesis are both on β^*, we can ignore μ^* when identifying an implicit nuisance parameter. Solving $R_\beta\beta = r$, we get $\beta = \Gamma\beta_f + \gamma$, where

$$\Gamma = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}', \beta_f = \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix}, \text{ and } \gamma = \begin{pmatrix} 0 & 0 & 0 & 0.1 \end{pmatrix}'.$$

Then $\mathscr{R}_{w,\beta}\beta \geq \mathbf{0}$ becomes $\mathscr{R}_{w,\beta}\Gamma\beta_f \geq \mathbf{0} - \mathscr{R}_{w,\beta}\gamma$ for

$$\mathscr{R}_{w,\beta}\Gamma = \begin{pmatrix} 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix}' \text{ and } \mathbf{0} - \mathscr{R}_{w,\beta}\gamma = \begin{pmatrix} 0 & 0 & 0 & -0.1 \end{pmatrix}'.$$

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

Denote $\mathscr{R}^u_{w,\beta}$ as the submatrix of $\mathscr{R}_{w,\beta}$ excluding the first row. Applying the Gauss-Jordan elimination on the transpose of $\mathscr{R}^u_{w,\beta}\Gamma$:

$$\left(\mathscr{R}^u_{w,\beta}\Gamma\right)' = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix},$$

we obtain $\mathscr{R}^u_\Gamma$ as the first and second row of $\mathscr{R}^u_{w,\beta}\Gamma$,

$$\Gamma^u = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}' \text{ and } \eta^k = \mathscr{R}^u_\Gamma\beta^*_f = \begin{pmatrix} \beta^*_2 \\ -\beta^*_2 + \beta^*_3 \end{pmatrix}.$$

1.4 The Wald Test

In this section, we construct an asymptotically uniformly valid Wald test using the statistic W_n introduced in Section 1.2.2 and provide a detailed procedure for implementing it. The following assumption is imposed on the weighting matrix.

Assumption 1.4.1. $\Sigma_{W,n} \xrightarrow{p} \Sigma_W$ *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 1.4.1. *Under* H_0 *and Assumptions 1.2.1-1.2.3 and 1.4.1, 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 1.2.1 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 (2.18).*

Lemma 1.4.1 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., (2.18), and is discontinuous in η^k.

For clarity and to be self-contained, we first provide a detailed treatment in Section 1.4.1 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 1.4.2.

Remark 1.4.1. A more general class of subvector hypothesis takes the following form: $H_0 : \mathscr{R}_{w,1}\theta^* = r$ 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

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

and $\mathscr{R}_{w,22}$ has full row rank. For the first J inequalities in $\mathscr{R}_w\theta \geq \mathrm{r}_w$, we can easily determine whether they are binding or not 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 $\mathscr{R}_{w,22}$ has full row rank, the implicit nuisance parameter is simply $\mathscr{R}_{w,22}\theta^*_2$, where $\theta^* = (\theta^{*\prime}_1, \theta^{*\prime}_2)'$ is decomposed conformably.

1.4.1 Subvector Hypothesis

For the subvector hypothesis H_{0S}, Remark 1.4.1 shows that the implicit nuisance parameter is $\eta^k = \theta^*_2$. The Wald statistic is calculated by

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

Without loss of generality, assume that $r = (\mathbf{0}', \mathrm{r}'_{nb})'$, where $\mathrm{r}_{nb} \in \mathbb{R}^{J-J_b}_{>0}$. Applying Lemma 1.4.1 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

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

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).

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 does not 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 $\omega_n \in \mathcal{W}_0 \to \omega \in \overline{\mathcal{W}}_0$.[9] 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 $\left\{\left(\eta_n^k, \pi_{W,n}, \xi_n\right) \in \mathbb{R}_{\geq 0}^{l-J} \times \Pi_W \times \Xi : n \geq 1\right\}$ and the localization parameter $(c, \pi_{W,\omega})$ as the limit of $b_n \eta_n^k$ and $\pi_{W,n}$:

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

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 $\left(\eta_n^k, \pi_{W,n}, \xi_n\right)$ 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_n^k and $\pi_{W,n}$.

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

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

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

Assumption 1.4.2. *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* $\kappa_n = o(1)$.

Assumption 1.4.3. *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.*

[9] 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).

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

Assumption 1.4.4. *For any $\omega_n \in \mathcal{W}_0$, $b_n\left(\widehat{\theta}-\theta_n\right)=O_p(1)$.*

Assumption 1.4.2 is slightly stronger than Assumption 1.2.1 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 1.4.3 and 1.4.4 are also stronger than their counterparts in Section 1.2.1. 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 Section S.2 of the online appendix, we discuss primitive conditions for Assumptions 1.4.2-1.4.4 to hold for the linear regression model.

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

Assumption 1.4.5. *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 1.4.2. *(i) If Assumptions 1.4.2-1.4.4 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 1.4.5 holds, then $W_n \xrightarrow{d} W_\omega \equiv (R\Psi_\omega)'\left(R\Sigma_{W,\omega}R'\right)^{-1}(R\Psi_\omega)$.

The Testing Procedure

As shown in Lemma 2.4.1, 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

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

Let $\widetilde{c} = b_n \widetilde{\theta}_2$. It can be shown that $\widetilde{c} \xrightarrow{d} c + \mathscr{T}_{2,\omega}^{-1} G_{2,\omega}$, 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 $ES(\tau)$ as the set such that $\Pr\left(\mathscr{T}_{2,\omega}^{-1} G_{2,\omega} \in ES(\tau)\right) \geq 1-\tau$. 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 $\widetilde{I}^k_\tau \equiv \widetilde{c} - \widetilde{ES}(\tau)$ and $\widetilde{ES}(\tau)$ is the set obtained using estimators of the parameters in $\mathscr{T}_{2,\omega}^{-1} G_{2,\omega}$ rather than true values. In cases where $\widetilde{I}^k_\tau \cap \overline{\mathbb{R}}^{l-J}_{\geq 0}$ is an empty set, let $\widetilde{I}_\tau = \{\mathbf{0}\}$.

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

$$CV^W_n(\alpha, \tau) \equiv \sup_{c \in \widetilde{I}_{\alpha-\tau}} \mathcal{C}^W_{c,\widehat{\pi}_W}(1-\tau), \tag{1.11}$$

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 1.4.1. *Under Assumptions 1.3.1 and 1.4.2-1.4.5, if* W_ω *is continuous at* $\mathcal{C}^W_{c,\pi_{W,\omega}}(1-\tau)$ *for all* $(c, \pi_{W,\omega}) \in \overline{\mathbb{R}}^{l-J}_{\geq 0} \times \overline{\Pi}_W$*, then it holds that* $AsySz\left(W_n, CV^W_n(\alpha,\tau)\right) \leq \alpha$.

The continuity assumption in Theorem 1.4.1 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 1.4.2. *Under* H_1 *and Assumptions 1.2.3 and 1.4.1,* $\Pr\left(W_n > CV^W_n(\alpha,\tau)\right) \to 1$.

1.4.2 General Linear Hypothesis

We extend the subvector test developed in Section 1.4.1 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 1.4.1, 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 does not affect the limiting distribution of W_n. Let $\left(\eta^k_n, \pi_{W,n}, \xi_n\right)$ be the drifting model parameters. Since the implicit nuisance parameter η^k_n satisfies inequalities in (2.18), we consider localization parameter c such that

$$c \equiv \lim_{n\to\infty} b_n \left(\Gamma^u \eta^k_n - (\mathrm{r}^u_w - \mathscr{R}^u_w \gamma)\right) \in C \subseteq \overline{\mathbb{R}}^{l_u}_{\geq 0}, \tag{1.12}$$

where

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

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

Lemma 1.4.3. *(i) If Assumptions 1.4.2-1.4.4 hold, then under* $H_0 : R\theta^* = r$ *and any parameter sequence* $\left(\eta^k_n, \pi_{W,n}, \xi_n\right) \in \mathrm{H}^k \times \Pi_W \times \Xi$*,*

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

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 1.4.5 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 1.4.3 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 1.4.3, 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

$$\widetilde{I}_\tau = \left\{ \mathrm{c} \in \mathbb{R}_{\geq 0}^{l_u} : \mathrm{c} = \Gamma^u \iota - b_n\left(\mathrm{r}_w^u - \mathscr{R}_w^u \gamma\right),\ \iota \in \widetilde{I}_\tau^k \right\}. \tag{1.13}$$

In cases where (1.13) is empty, let $\widetilde{I}_\tau = \{\mathbf{0}\}$.

Step 2. Compute the α level Bonferroni critical value as

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

for some $0 \leq \tau \leq \alpha$, where $\mathcal{C}^W_{c,\pi_{W,\omega}}(1-\tau)$ is the $(1-\tau)$ quantile of W_ω in Lemma 1.4.3 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 1.4.3. *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 1.3.1 and 1.4.2-1.4.5, it holds that $AsySz\left(W_n, CV_n^W(\alpha,\tau)\right) \leq \alpha$.*

Theorem 1.4.4. *Under H_1 and Assumptions 1.2.3 and 1.4.1,* $\Pr\left(W_n > CV_n^W(\alpha,\tau)\right) \to 1$.

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 (1.4). The algorithm in Section 1.3.1 generates $\mathscr{R}_w^b$, $\mathscr{R}_w^{nb}$, and $\mathscr{R}_w^u$, which are submatrices of $\mathscr{R}_w$ corresponding to implicit equalities, strictly redundant inequalities, and undetermined inequalities. Partition r_w conformably into subvectors r_w^b, r_w^{nb}, and r_w^u. Since inequalities defined by $\mathscr{R}_w^{nb}$ are strictly redundant, the parameter space under H_0 can be rewritten as

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

We impose the following assumption on the convergence rate of the restricted estimator $\widehat{\theta}_0$ defined in Section 1.2.2. Primitive conditions for this assumption can be found in Andrews (1997).

Assumption 1.5.1. $b_n\left(\widehat{\theta}_0 - \theta_0^*\right) = O_p(1)$ *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 1.4.1.

Lemma 1.5.1. *Under H_0 and Assumptions 1.2.1-1.2.3 and 1.5.1, 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

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

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 $\mathscr{R}_w^b \lambda \geq \mathbf{0}$ in $\phi(\lambda)$ become equalities $\mathscr{R}_w^b \lambda = \mathbf{0}$ 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 $l_n\left(\widehat{\theta}_0\right)$.

Comparing Lemmas 1.4.1 and 1.5.1, 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 1.5.2 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 (1.12). The asymptotic distributions of W_n and QLR_n under drifting model parameters depend on the same localization parameter vector c.

Assumption 1.5.2. *For any $\omega_n \in \mathcal{W}_0$, $b_n\left(\widehat{\theta}_0 - \theta_n\right) = O_p(1)$.*

Lemma 1.5.2. *If Assumptions 1.4.2-1.4.4 and 1.5.2 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 1.4.3 (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 $0 \leq \tau \leq \alpha$. The α level Bonferroni critical value $CV_n^Q(\alpha,\tau)$ is defined as

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

where $\widetilde{I}_{\alpha-\tau}$ and $\widehat{\pi}_Q$ are obtained by similar procedures presented by Step 1 in Section 1.4.2. The following theorems show that $CV_n^Q(\alpha,\tau)$ controls the asymptotic size of QLR test, and that the test is consistent.

Theorem 1.5.1. *Under Assumptions 1.3.1, 1.4.2-1.4.4, and 1.5.2, 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 $AsySz\left(QLR_n, CV_n^Q(\alpha,\tau)\right) \leq \alpha$ holds.*

Theorem 1.5.2. *Under H_1 and Assumptions 1.2.1-1.2.3 and 1.5.1, if $l_n(\cdot)$ is continuous at θ_0^* and $b_n^{-2}\left(l_n(\theta^*) - l_n(\theta_0^*)\right) \xrightarrow{p} \varsigma > 0$, then $\Pr\left(QLR_n > CV_n^Q(\alpha,\tau)\right) \to 1$ holds.*

The condition $b_n^{-2}\left(l_n(\theta^*) - l_n(\theta_0^*)\right) \xrightarrow{p} \varsigma > 0$ in Theorem 1.5.2 is generally satisfied as the identification assumption. We illustrate it by the following example.

Example 1.5.1. [Inequality-constrained generalized method of moments] Let $g(X,\theta)$ be a vector of known functions of the random variable X, which is allowed to be non-differentiable. The moment equations $E\left[g(X,\theta)\right] = \mathbf{0}$ hold if and only if $\theta = \theta^* \in \Theta$. With the random sample $(X_i)_{i=1}^n$, the sample moment functions are computed as $\frac{1}{n}\sum_{i=1}^n g(X_i,\theta)$. For some positive definite weighting matrix Σ_n, $l_n(\cdot)$ is defined as

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

More discussion on the model can be found in Pakes and Pollard (1989) and Andrews (1997). Suppose $\Sigma_n \xrightarrow{p} \Sigma$ for which Σ is positive definite with probability one. Under some mild assumptions, we then have

$$\begin{aligned} n^{-1}\left(l_n(\theta^*) - l_n(\theta_0^*)\right) &= \left(\frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta_0^*)\right)' \Sigma_n \left(\frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta_0^*)\right) \\ &\quad - \left(\frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta^*)\right)' \Sigma_n \left(\frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta^*)\right) \\ &\xrightarrow{p} E\left[g(X,\theta_0^*)\right]' \Sigma E\left[g(X,\theta_0^*)\right]. \end{aligned}$$

Assumption $b_n^{-2}\left(l_n(\theta^*) - l_n(\theta_0^*)\right) \xrightarrow{p} \varsigma > 0$ in Theorem 1.5.2 is satisfied as long as $E\left[g(X,\theta_0^*)\right] \neq \mathbf{0}$ for $\theta_0^* \neq \theta^*$, which is assumed for the identification of θ^*.

1.6 The Score Test

Let the following two assumptions hold for the score function, directed score, and score test statistic defined in (1.5), (1.6), and (1.7). Discussion on the assumptions can be found in Andrews (2001).

Assumption 1.6.1. *(i) Assume that for all* $0 < \kappa < \infty$, $\sup_{\theta\in\Theta_0:\left\|b_n\left(\theta-\theta_0^*\right)\right\|<\kappa} \left|b_n^{-1}R_n^D(\theta)\right| = o_p(1)$*; (ii)* $\widehat{\mathscr{T}_n} = -b_n^{-2}D^2 l_n(\theta_0^*) + o_p(1)$.

Assumption 1.6.2. $\Sigma_{S,n} \xrightarrow{p} \Sigma_S$ *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 $R\Theta - r$:

$$R\Theta - r = \left\{\lambda \in \mathbb{R}^J : \mathscr{R}_{R,e}\lambda = \mathrm{r}_{R,e} \text{ and } \mathscr{R}_{R,w}\lambda \geq \mathrm{r}_{R,w}\right\}. \tag{1.14}$$

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 $b_n (R\Theta - r)$ in the sense of Hausdorff distance is given by

$$\Lambda_R \equiv \left\{\lambda \in \mathbb{R}^J : \mathscr{R}_{R,e}\lambda = \mathbf{0} \text{ and } \mathscr{R}_{R,w,b}\lambda \geq \mathbf{0}\right\}, \tag{1.15}$$

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 1.6.1 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 $d_H (b_n (R\Theta - r), \Lambda_R) \to 0$ is given by (1.15) 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 1.6.2, we provide an algorithm for obtaining the set Λ_R.

1.6.1 Asymptotic Theory

With description (1.15), the asymptotic distributions of ds and S_n are given in the following lemma.

Lemma 1.6.1. *(i) Suppose the null hypothesis and Assumptions 1.2.2, 1.5.1, and 1.6.1 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 1.6.2 holds, then $S_n \xrightarrow{d} S \equiv ds'\Sigma_S^{-1}ds$.

The most significant difference between S in Lemma 1.6.1 and W in Lemma 1.4.1 or QLR in Lemma 1.5.1 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.

To study the asymptotic size of the score test, we impose the following assumption extending Assumptions 1.6.1 and 1.6.2.

Assumption 1.6.3. *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 $\mathcal{C}_{\pi_S}^S(1-\alpha)$ 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 1.6.1. *Under Assumptions 1.3.1, 1.4.3, 1.5.2, and 1.6.3, 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 1.6.2. *Under H_1 and Assumptions 1.5.1, 1.6.1, and 1.6.2, 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* $\Pr\left(S_n > CV_n^S(\alpha)\right) \to 1$.[10]

Since the first order derivative of $l_n(\cdot)$ evaluated at θ^* approaches zero in the limit, $\mathscr{T}_n^{-1} b_n^{-2} Dl_n(\theta^*)$ is $o_p(1)$. The condition in Theorem 1.6.2 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; see the first example below. However, as shown in the second example, if $l_n(\cdot)$ takes a different form, the score test may not be consistent for certain deviations from the null hypothesis.

Example 1.6.1. [Inequality-constrained linear regression model] The model is expressed as $Y_i = X_i'\theta^* + \varepsilon_i$, for $i = 1, ..., n$, where $(X_i, Y_i)_{i=1}^n$ is the random sample and $E(\varepsilon_i \mid X_i) = 0$. The objective function $l_n(\cdot)$ is expressed as $l_n(\theta) = -\frac{1}{2}\sum_{i=1}^n (Y_i - X_i'\theta)^2$, $\theta \in \Theta$. We have $b_n = \sqrt{n}$, $Dl_n(\theta) = \sum_{i=1}^n (Y_i X_i - X_i X_i'\theta)$, and $D^2 l_n(\theta) = -\sum_{i=1}^n X_i X_i'$. Under some mild assumptions, we can further obtain that $\mathscr{T}_n = -b_n^{-2} D^2 l_n(\theta_0^*) = \frac{1}{n}\sum X_i X_i'$ and $b_n^{-2} Dl_n(\theta_0^*) = \frac{1}{n}\sum X_i X_i'(\theta^* - \theta_0^*) + \frac{1}{n}\sum \varepsilon_i X_i$. Since $\frac{1}{n}\sum \varepsilon_i X_i \xrightarrow{p} 0$, it holds that

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

[10] 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.

The assumption in the theorem is verified.

Example 1.6.2. For the Logit model, we have that

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

where $F(t) = \frac{1}{1+e^{-t}}$ and $f(t) = \frac{e^{-t}}{(1+e^{-t})^2}$. Assume $\Theta = \{(\theta_1, \theta_2) \in \mathbb{R}^2 : \theta_1 + \theta_2 \geq 0\}$, $\theta^* = (1, 0)$ and $X = (X_1, X_2)'$, 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 $H_0 : R\theta^* = r$.

1.6.2 Implementation—Projection of Polytope

We describe one algorithm for the projection of a polytope based on the Fourier-Motzkin algorithm. 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 in the online appendix, the result in Lemma 1.6.1 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 (1.8) 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 to eliminate θ_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.

1.6.3 Running Examples (Continued)

We apply the algorithm discussed in the previous subsection to calculate $R\Theta - r$ and Λ_R in the running examples.

Random Coefficients Model

Solving for the system of linear equations: $(\mathscr{R}_e', R')' \theta = (\mathrm{r}_e, \lambda + r)$, we obtain $\theta = \Gamma\theta_f + \mathit{\Gamma}\lambda + \gamma$, with Γ and γ being shown in Section 1.3.3 and $\mathit{\Gamma} = (0, \ldots, 0, \lambda, 0, \ldots, 0, -\lambda)'$. The reduced polytope P_f can be expressed as

$$P_f = \left\{ (\theta_f, \lambda) : \begin{array}{c} \theta_1 \geq 0, \ldots, \theta_{l_1 - 1} \geq 0, -\theta_1 - \cdots - \theta_{l_1 - 1} + \lambda + r \geq 0, \\ \theta_{l_1 + 1} \geq 0, \ldots, \theta_{l-1} \geq 0, -\theta_{l_1+1} - \cdots - \theta_{l-1} + 1 - \lambda - r \geq 0 \end{array} \right\}.$$

We first eliminate θ_1. There are two inequalities containing θ_1: $\theta_1 \geq 0$ and $-\theta_1 - \cdots - \theta_{l_1-1} + \lambda + r \geq 0$. By rearranging the second inequality, we eliminate θ_1 and obtain that $-\theta_2 - \cdots - \theta_{l_1-1} + \lambda + r \geq 0$. Together with $\theta_2 \geq 0$, we can eliminate θ_2. Continue the process, we can eliminate until θ_{l_1-1} and obtain $\lambda \geq -r$. The same procedure would eliminate $\theta_{l_1+1}, \ldots, \theta_{l-1}$ and leave us with another inequality: $\lambda \leq 1-r$. Combing the two inequalities on λ, we obtain that $R\Theta - r = \{\lambda : -r \leq \lambda \leq 1-r\}$. Thus, for $r = 0$, $\Lambda_R = \{\lambda : \lambda \geq 0\}$; for $r \in (0,1)$, $\Lambda_R = \mathbb{R}$; and for $r = 1$, $\Lambda_R = \{\lambda : \lambda \leq 0\}$.

Mincer Earnings Regression

The first step of Fourier-Motzkin algorithm provides $\beta = \Gamma\beta_f + \mathit{\Gamma}\lambda + \gamma$, with Γ, β_f, and γ being calculated in Section 1.3.3 and

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

The reduced polytope is

$$P_f = \left\{(\beta_2, \beta_3, \lambda_1, \lambda_2) : \mathscr{R}_{w,\beta}\Gamma\beta_f + \mathscr{R}_{w,\beta}\mathit{\Gamma}\lambda \geq \mathbf{0} - \mathscr{R}_{w,\beta}\gamma\right\},$$

where values of $\mathscr{R}_{w,\beta}\Gamma$ and $\mathbf{0} - \mathscr{R}_{w,\beta}\gamma$ can be found in Section 1.3.3 and

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

FME is then applied to the following linear inequalities: $-\lambda_1 \geq 0$, $\beta_2 \geq 0$, $-\beta_2 + \beta_3 \geq 0$, and $\beta_2 + \beta_3 + \lambda_2 \geq -0.1$. By eliminating β_2 first and then β_3, we obtain first $-\lambda_1 \geq 0$, $\beta_3 \geq 0$, and $\beta_3 \geq \beta_3 - \lambda_2 - 0.1$, and then $\lambda_1 \geq 0$ 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 $\Lambda_R = \{(\lambda_1, \lambda_2) : \lambda_1 \geq 0\}$.

1.7 Local Power

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 1.2.2, 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 $\mathscr{R}_w^u$ and $\mathscr{R}_w^b$ of $\mathscr{R}_w$ respectively. Notice that the above definition of c is consistent with that in (1.12).

Assumptions in Lemmas 1.4.3, 1.5.2, and 1.6.1 are modified in Assumption 1.7.1 below for the drifting parameter sequence ω_n consistent with $H_{1,n}$. Similar type of assumptions have been introduced in Sections 1.4, 1.5, and 1.6 when the parameter sequence is consistent with H_0.

Assumption 1.7.1. *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 $\mathscr{T}_n \equiv -b_n^{-2}D^2l_n\left(\theta_n\right)$ 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 1.7.1. *For the parameter sequence ω_n consistent with $H_{1,n}$, if Assumption 1.7.1 holds, then*

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

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

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

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

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

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

in which

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

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

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

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

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

The proof for Lemma 1.7.1 is similar to that for Lemmas 1.4.3, 1.5.2, and 1.6.1. For ω_n consistent $H_{1,n}$, the asymptotic distributions of W_n and QLR_n depend on both δ and c_w; while the asymptotic distribution of S_n 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 1.7.1. *Let Assumption 1.7.1 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

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

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

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

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

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

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 $CV^W_n(\alpha,\tau)$ and $CV^Q_n(\alpha,\tau)$ depend on δ and c_w; while the test based upon S_n and $CV^S_n(\alpha)$ 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 1.7.1 shows that the asymptotic distribution of S_n under ω_n only depends on δ and the critical value $CV^S_n(\alpha)$ is determined solely by the estimator of model parameters π_S.

1.8 Monte-Carlo Simulation

In this section, we conduct a small simulation study to examine and compare the finite sample performance of the Wald, QLR, and score tests developed in this paper.

1.8.1 The Data Generating Process and the Null Hypothesis

The DGPs that we design build on the linear regression model in Autor and Handel (2013); see Section 1.2.3. It is 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 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; and the error term ε is independent of the observable covariates.

Let $\theta^* \equiv (\beta^{*\prime}, \mu^{*\prime})'$ with $\beta^* \equiv (\beta_1^*, \ldots, \beta_4^*)'$ and $\mu^* \equiv (\mu_0^*, \mu_1^*, \mu_2^*)'$. Following the discussion and notation in Section 1.2.3, the parameter space is defined as $\Theta = \left\{\theta = (\beta', \mu')' : \mathscr{R}_w \theta \geq \mathbf{0}\right\}$, and the joint null hypothesis is expressed as $H_0 : R\theta^* = r$. We construct 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.

There are four inequalities in the maintained hypothesis, among which the first inequality is binding under the null hypothesis. To see the effects of the number of binding and undetermined inequalities in the null parameter space on the performance of the tests, we consider five different sets of parameter values corresponding to different numbers of binding inequalities; see Table 1.1. In Case 1, no inequality is binding except the first one, and there

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

Table 1.1: Different sets of parameters

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 $\beta_1^* = -0.05$, -0.1, and -0.15 to examine the power performance of the tests. To implement the Wald and QLR tests, we need to identify an implicit nuisance parameter which is done in Section 1.3.3; and for conducting the score test, we need to compute the polytope projection which is done in Section 1.6.3. The confidence set for c is computed in the same way for Wald and QLR tests by first constructing the confidence ellipsoid of Wald for η_n^k and then applying (1.13). Following Romano et al. (2014) and McCloskey (2017), we set the tuning parameter $\tau = \alpha - \alpha/10$. The weighting matrix in W_n is set as the estimator of the variance covariance matrix of the asymptotic distribution of the ordinary least square estimator of θ calculated using $\widehat{\theta}$. The matrix $\Sigma_{S,n}$ in S_n is calculated as $R\mathscr{I}_n^{-1}R'$.

1.8.2 Results on Size and Power

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 1.2 reports the finite sample size performance of all three tests. First, for DGP A, the score test S_n has the best performance having sizes closer to the nominal size than Wald and QLR tests; for DGP B, the three tests perform similarly and all are slightly under sized. Second, comparing results across the two DGPs, we see that the size performance of the score test is more sensitive to different distributions of the error term ε than the size

		W_n	QLR_n	S_n			W_n	QLR_n	S_n
	Case 1	0.0412	0.0417	0.0498		Case 1	0.0408	0.0409	0.0404
	Case 2	0.0424	0.0419	0.0513		Case 2	0.0411	0.0412	0.0406
DGP A	Case 3	0.0443	0.0436	0.0509	DGP B	Case 3	0.0423	0.0412	0.0410
	Case 4	0.0428	0.0430	0.0521		Case 4	0.0398	0.0407	0.0405
	Case 5	0.0430	0.0427	0.0502		Case 5	0.0405	0.0402	0.0403

Table 1.2: Rejection probability under H_0

		$\beta_1 = -0.05$			$\beta_1 = -0.1$			$\beta_1 = -0.15$		
		W_n	QLR_n	S_n	W_n	QLR_n	S_n	W_n	QLR_n	S_n
	Case 1	0.1665	0.1640	0.1783	0.6167	0.6155	0.6273	0.9280	0.9290	0.9412
	Case 2	0.1694	0.1657	0.1775	0.6195	0.6202	0.6300	0.9336	0.9342	0.9444
DGP A	Case 3	0.1852	0.1814	0.1762	0.6382	0.6361	0.6291	0.9494	0.9472	0.9425
	Case 4	0.1690	0.1648	0.1776	0.6196	0.6189	0.6241	0.9276	0.9275	0.9383
	Case 5	0.1698	0.1681	0.1788	0.6222	0.6216	0.6259	0.9343	0.9338	0.9427
	Case 1	0.1476	0.1532	0.1519	0.5780	0.5824	0.5781	0.8630	0.8609	0.8681
	Case 2	0.1647	0.1664	0.1511	0.5929	0.5956	0.5834	0.8836	0.8772	0.8759
DGP B	Case 3	0.1715	0.1742	0.1528	0.5987	0.6014	0.5841	0.8971	0.8968	0.8767
	Case 4	0.1633	0.1657	0.1476	0.5963	0.5976	0.5827	0.8928	0.8923	0.8618
	Case 5	0.1641	0.1672	0.1487	0.5928	0.5982	0.5777	0.8710	0.8751	0.8553

Table 1.3: Finite sample size-corrected power

performance of Wald and QLR tests; Third, within each DGP, Wald and QLR tests perform the best for Case 3 which has the most binding inequalities under the null hypothesis.

Table 1.3 presents the finite sample size-corrected powers of Wald, QLR, and score tests. First, for both DGPs and different values of β_1, there is no significant difference between Wald and QLR tests. Second, for both DGPs, the power of 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 score test having slightly higher power overall. Fourth, for DGP B, the Wald and QLR tests have higher power than the score test except Case 1 for which only one inequality is binding and there is no close-to-binding inequalities under the null hypothesis. This suggests that for skewed error distributions, it is important to take into account prior information in the maintained hypothesis through the Bonferroni-type correction. Lastly, within each DGP, Wald and QLR tests perform the best for Case 3 which has the most binding inequalities under the null hypothesis.

1.9 An Empirical Illustration on Mincer Earnings Regression

In this section, we apply our tests to Mincer earnings regression introduced in Section 1.2.3 and compare the results with the t-test. In addition to the regression in Section 1.2.3 denoted as Model 1 below, we consider two more regressions in Autor and Handel (2013) on the log hourly wages: for $i = 1, ..., n$,

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

where E_i , β^*, X_i, and μ^* are introduced in Section 1.2.3, $\zeta^* \equiv (\zeta_1^*, \zeta_2^*, \zeta_3^*)'$, $T_i \equiv (T_{1,i}, T_{2,i}, T_{3,i})'$ which denotes different tasks performed in a job, with $T_{1,i}$ being "Abstract", $T_{2,i}$ being "Routine", and $T_{3,i}$ being "Manual", and Z_i includes 240 occupation dummy variables.

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.

Recall that the parameter space for Model 1 is $\Theta = \left\{\theta = (\beta', \mu')' : \mathscr{R}_{w,\beta}\beta \geq \mathbf{0}\right\}$ for $\mathscr{R}_{w,\beta}$ defined in 1.2.3. Similarly, $\Theta = \left\{\theta = (\beta', \zeta', \mu')' : \mathscr{R}_{w,\beta}\beta \geq \mathbf{0}\right\}$ for Model 2 and $\Theta =$

		W_n	QLR_n	S_n	t-test			W_n	QLR_n	S_n	t-test
	Model 1	$\star$	$\star$	$\star$	$**$		Model 1	$***$	$***$	$***$	$***$
$\beta_1^* = 0$	Model 2	$\star$	$\star$	$\star$	$\star$	$\beta_2^* = 0$	Model 2	$***$	$***$	$***$	$**$
	Model 3	$\star$	$\star$	$\star$	$\star$		Model 3	$***$	$***$	$***$	$\star$
		W_n	QLR_n	S_n	t-test			W_n	QLR_n	S_n	
$\zeta_2^* = 0$	Model 2	$***$	$***$	$***$	$***$		Model 1	$***$	$***$	$***$	
	Model 3	$***$	$***$	$\star$	$\star$	Joint Hypo.	Model 2	$*$	$**$	$*$	
							Model 3	$*$	$*$	$*$	

Table 1.4: Significance results

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

$\left\{\theta = (\beta', \zeta', \mu', \vartheta')' : \mathscr{R}_{w,\beta}\beta \geq \mathbf{0}\right\}$ for Model 3.

Section S.2 of the online appendix provides primitive conditions for these linear regression models such that the proposed tests are valid. 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 $\mathscr{R}_{w,\beta}\beta^* \geq \mathbf{0}$ 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 standard t-test.

The results are summarized in Table 1.4. For the same hypothesis, the t-test suggests different conclusions for different models. It rejects $H_0 : \beta_1^* = 0$ at 5% in Model 1, but fails to reject the null hypothesis at 10% in Models 2 and 3. The standard OLS estimate of β_1^* is 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 t-test rejects $\beta_1^* = 0$ in Model 1, which indicates that β_1^* should be positive considering that the estimat 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 t-test also provides different results for different models, while our tests suggest that β_2^* is significant no matter which model is used. In fact, all three tests developed in the paper give the same testing results for each hypothesis, and the results are consistent among models. We also test the null hypothesis of $\beta_j^* = 0$ for $j = 3, 4$. The tests developed in the paper and t-test give the same results, which are the same as the ones in Autor and Handel (2013).

Under the maintained hypothesis on β, researchers can test the significance of individual parameter ζ_2^*, i.e., $H_0 : \zeta_2^* = 0$ 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 are presented in Table 1.4. The t-test rejects the null hypothesis at 1% level in Model 2, but fails to reject it at 10% level in Model 3. At the same time, the score test provides the same results as the ones based on the t-test. Because the null hypothesis and maintained hypothesis are imposed on different parameters, information in Θ is lost after the projection, and the score test acts as if there is no maintained hypothesis. On the other hand, the Wald and QLR tests fully exploit the information in the maintained hypothesis, and suggest consistent results for both Model 2 and Model 3. All four tests provide the same conclusions for $H_0 : \zeta_1^* = 0$ and $H_0 : \zeta_3^* = 0$, and the results are consistent for both Models 2 and 3.

We have also tested $H_0 : \beta_1^* = 0$ and $\beta_2^* + 0.1 = \beta_4^*$ introduced in Section 1.2.3. Table 1.4 collects the results of different tests. All three tests suggest the same for Models 1 and 3, although the results differ between models. For Model 2, QLR test rejects the null hypothesis at 5% level, while the Wald and score tests only reject it at 10% level.

1.10 Concluding Remarks

In this paper, we have developed asymptotically uniformly valid Wald, QLR, and score tests for the null hypothesis of linear equality constraints against the two-sided alternative hypotheses in an extremum estimation set-up for non-trending data where the parameter space is characterized by a finite number of linear equality and inequality constraints. In contrast to the classical textbook theory on the Wald, QLR, and score tests, when there are undetermined inequalities in the null parameter space, the null asymptotic distributions of the Wald, QLR, and score statistics are very different posing different challenges in developing asymptotically uniformly valid tests based on them. To develop Wald and QLR tests, we introduce the concept of an implicit nuisance parameter, and propose an algorithm for identifying it. Since both the concept and the algorithm depend on the null hypothesis and the parameter space only, once an implicit nuisance parameter is identified, asymptotically uniformly valid tests can be constructed from any test statistic in a wide range of models. To implement the score test, we propose an algorithm based on the Fourier-Motzkin elimination to compute polytope projections. We use a random coefficients model and Mincer earnings regression to illustrate our general 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 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 in 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 worth investigating the possibility of extending our approach to allow for non-linear equality and inequality constraints.

Chapter 2

UNIFORM INFERENCE IN A GENERALIZED INTERVAL ARITHMETIC CENTER AND RANGE LINEAR MODEL[1]

2.1 Introduction

Researchers often may have access to interval data only. For example, under the Health and Retirement Study (HRS) questionnaire protocol, a respondent is asked to report her wealth. If she does not comply, then the respondent is asked to report if her wealth falls within a sequence of brackets. The HRS thus yields a wealth interval for each respondent, see Manski and Tamer (2002) and references therein. Another example is the bid-ask price interval. In markets with microstructural frictions, we observe two sets of prices for an asset at the same time: the bid price at which an investor could sell and the ask price at which an investor could buy.

Interval data may represent uncertainty in which case they are incomplete observations on a random variable which may not always be observable; or variability in which case they are observations on a random interval. To illustrate, consider the interval data given by the bid and ask prices: they represent incomplete observations on the true price, i.e., lower and upper bounds on the true price but also represent precise observations on the random interval defined by the bid and ask prices. When the model of interest is for random variables, at least one of which is not always observable, interval data represent incomplete observations on the latent random variable and parameters in the model for random variables are typically only partially identified, see e.g., Manski and Tamer (2002), Beresteanu and Molinari (2008), and Beresteanu et al. (2011), to name only a few. On the contrary, when interval data represent variability and the model of interest is for random intervals, interval

[1]This Chapter is a joint work with Yanqin Fan from the University of Washington.

data represent precise observations on random intervals such as the bid-ask price interval and parameters in the corresponding interval model are typically point identified under standard regularity conditions.

This paper focuses on the case where interval data represent variability and the model of interest is for random intervals. Broadly speaking, there are two approaches to modeling interval data, i.e., the interval arithmetic approach and the bivariate regression approach. The interval arithmetic approach adopts interval arithmetic to model directly relations between random intervals. The most general model based on this approach in the current literature is model M_G proposed in Blanco-Fernández et al. (2015) which makes use of the canonical decomposition of an interval in terms of its center and range to model the dependent interval directly via covariate intervals with an interval error term. To ensure that the interval error is well defined, parameters in model M_G must satisfy an increasing number of random inequality constraints which are much stronger than the non-negativity constraints necessary to ensure that the predicted interval at any covariate interval is always well defined. We refer interested readers to Blanco-Fernández et al. (2015) for the constrained estimation of model M_G and a review of the interval arithmetic approach to modeling interval data.

The bivariate regression approach models jointly either the left and right end points of the dependent interval or the center and range of the dependent interval. An important example based on this approach is the Center and Range (CR) model in which one regression relates the center of the dependent interval to the centers of the covariate intervals and the other relates the range of the dependent interval to the ranges of the covariate intervals. To ensure that the predicted range of the dependent interval at any covariate interval is always non-negative, parameters in the range regression are restricted to be non-negative. Neto and de Carvalho (2010) propose to estimate the range regression of the CR model by a constrained OLS estimator restricting the coefficients in the range regression to be non-negative.[2] Applications of the CR model and methods based on modeling the left and

[2]Golan and Ullah (2017) propose an information theoretic approach to estimating linear interval models.

right end points of the dependent interval include Han et al. (2008), Han et al. (2012), and González-Rivera and Lin (2013).

The CR model and its constrained estimation have proven to be useful for forecasting random intervals. However, formal statistical inference procedures and goodness-of-fit measures for the CR model are lacking in the current literature. All the existing measures of goodness-of-fit for the CR model are based on ad hoc combinations of measures for multiple regressions, see Neto and de Carvalho (2010) and references therein. In addition, the CR model only allows the center (range) of the dependent interval to depend on the centers (ranges) of the covariate intervals and could be restrictive in some applications. In contrast to the CR model, model M_G allows the center/range of the dependent interval to depend on both centers and ranges of the covariate intervals. However the additional random constraints on the parameters in model M_G substantially complicate the asymptotic distribution of the estimators rendering inference extremely difficult if not impossible.

This paper makes three main contributions. First, we make use of generalized interval arithmetic to construct a *Generalized Interval Arithmetic Center and Range* (GIA-CR) model for random intervals which allows for general linear inequality constraints on its parameters, broadening the scope of applications of linear models for random intervals. When the constraints are non-negative constraints to ensure valid forecasts, we refer the new model simply as the *Interval Arithmetic Center and Range* (IA-CR) model. The IA-CR model extends and overcomes the drawbacks of both model M_G and the CR model. Non-negative constraints in the CR model or the IA-CR model are motivated from forecasting purposes. The GIA-CR model allows for general linear inequality constraints arising from economic theory or prior knowledge. It has long been recognized that incorporating inequality constraints in parameter estimation may yield efficiency gain, see e.g., Liew (1976), Judge et al. (1984), and more recent works of Chernozhukov and Hong (2004) and Moon and Schorfheide (2009). Moreover, 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." To accommodate general inequality constraints in the GIA-CR model, we make use of generalized interval arithmetic.[3]

The second main contribution of this paper is to propose a constrained estimator and a coefficient of determination for the GIA-CR model, and to establish asymptotic distributions of both the constrained estimator of the GIA-CR model we propose and the constrained estimator of model M_G in Blanco-Fernández et al. (2015). Although the asymptotic distribution of the constrained estimator for the GIA-CR model can be derived from results in Andrews (1999),[4] the approach in Andrews (1999) is not applicable to the constrained estimator of model M_G because of the increasing number of random inequality constraints. In this paper, we exploit the powerful tools developed in Knight (2001, 2006) for linear programming estimators and M-estimators of boundaries, i.e., epi-convergence in distribution and point process convergence for extreme values, to derive the asymptotic distribution of the constrained estimator of model M_G proposed in Blanco-Fernández et al. (2015). As a by-product, we also obtain the asymptotic distribution of the constrained estimator for the GIA-CR model via the same set of tools. We note that Chernozhukov and Hong (2004) have employed the same techniques in likelihood-based estimation and inference for two-sided and one-sided regression models and derived asymptotic properties of likelihood-based estimators as well as Bayes and Wald inference.

The third main contribution of this paper is to construct asymptotically uniformly valid tests for a class of linear constraints in the GIA-CR model, where the null hypothesis specifies the value of a subvector of $R\theta^*$ with θ^* denoting the true parameter vector satisfying linear inequality constraints of the form: $R\theta^* \geq r$ for known R and r under the maintained hypothesis. An important and motivating example for this inference set-up is that of testing the correct specification of the CR model against the IA-CR model. Due to the presence of undetermined inequalities in $R\theta^* \geq r$ under the null hypothesis of this type, the null

[3] In Han et al. (2008) and Han et al. (2012), generalized interval arithmetic is used to construct linear time series models for generalized random intervals, i.e., the observations are generalized intervals and no constraints are imposed on model parameters. Instead, this paper focuses on the case that the observations are intervals and generalized interval arithmetic is used to handle general linear inequality constraints.

[4] Other related works can be found in Silvapulle and Sen (2005) and references therein.

asymptotic distribution of the constrained estimator for the GIA-CR model is discontinuous in some model parameters posing technical challenges in constructing asymptotically uniformly valid tests. Testing the correct specification of the CR model against the IA-CR model falls within the framework of *subvector inference* with nuisance parameters.[5] Existing approaches such as Bounds tests, the least favorable approach, and tests based on confidence sets for nuisance parameters may be applied, 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. We present a detailed application of this approach to testing the CR model against the IA-CR model.[7]

For the GIA-CR model, although the null hypothesis specifies the value of a subvector of $R\theta^*$, the remaining components in $R\theta^*$ could be linearly dependent rendering direct application of the two-stage approach with Bonferroni-type correction for standard subvector inference problematic. To address this issue, we propose to use Gauss-Jordan elimination to identify nuisance parameters defined as an appropriate subvector of the remaining components in $R\theta^*$. Given the nuisance parameters and the inequalities they satisfy, we apply the two-stage method based on confidence sets for nuisance parameters with Bonferroni-type

---

[5] To simplify exposition, in this paper, we will use nuisance parameters to refer to parameters causing discontinuity of the asymptotic distribution of the test statistic under the null hypothesis.

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

[7] Work in different contexts that adopted this approach include: Berger and Boos (1994) and Silvapulle (1996) who study some specific parametric testing problem; Staiger and Stock (1997) who construct a confidence region for the parameters of a linear regression with possibly "weak" instrumental variables; Romano and Wolf (2000) who construct a confidence interval for a univariate mean that has finite sample validity; Moon and Schorfheide (2009) who propose asymptotically uniformly valid tests and confidence sets for the parameters of interest in moment equality models with overidentifying inequality moment conditions; Chernozhukov et al. (2013) who construct confidence intervals for marginal effects in nonlinear panel data models; Romano et al. (2014) who construct asymptotically uniformly valid confidence sets for parameters characterized by a finite number of moment inequalities and McCloskey (2017) who consider general nonstandard 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.

correction to constructing asymptotically uniformly valid tests for linear hypotheses in the GIA-CR model.[8]

To gauge the finite sample performance of the proposed estimator and test, we conduct a simulation study, the results from which confirm the superior performance of the proposed estimator and the asymptotically uniformly valid test for the correct specification of the CR model against the IA-CR model in finite samples.

The rest of this paper is organized as follows. Section 2 introduces the GIA-CR model and two constrained estimators: one for the GIA-CR model and one for model M_G. It also constructs a goodness-of-fit measure for the GIA-CR model. Section 3 establishes asymptotic theory for constrained estimators for the GIA-CR model and model M_G. In Section 4, we first provide a detailed construction and technical treatment of asymptotically uniformly valid tests for the correct specification of the CR model against the IA-CR model. The theoretical analysis in this part builds on Romano et al. (2014) and McCloskey (2017). Then we construct asymptotically uniformly valid tests in the GIA-CR model. Section 5 reports results from a simulation study. The last section offers some concluding remarks. Technical proofs are collected in Appendix A.

Notations All limits are taken as $n \longrightarrow \infty$. Let $|\boldsymbol{v}| \equiv (|\boldsymbol{v}_1|, \ldots, |\boldsymbol{v}_p|)'$ for any p-dimensional vector $\boldsymbol{v} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)'$; $\boldsymbol{v} \geq \boldsymbol{u}$ means that $\boldsymbol{v}_j \geq \boldsymbol{u}_j$ for $j = 1, \ldots, p$; and $\|\boldsymbol{v}\|$ denotes the Euclidean norm of $\boldsymbol{v}$. The notation $a \sim b$ means that $a/b \to 1$ as appropriate limits are taken. The remaining notations concern operations on generalized intervals and generalized random intervals, see Appendix B.2 for details. For a generalized interval $A = [a_1, a_2]$, where a_1, a_2 are real numbers, let $\mathrm{mid}A = (a_1 + a_2)/2$ and $\mathrm{sprA} = (a_2 - a_1)/2$. A can also be expressed as $A = [\mathrm{mid}A \pm \mathrm{spr}A]$. A generalized interval $A = [a_1, a_2]$ is a proper or simply an interval when $a_1 \leq a_2$; otherwise it is an improper interval. For two intervals A and B, $A -_H B$ is an interval denoting the Hukuhara difference between A and B. The Hukuhara

[8] For appropriate choices of linear inequality constraints, our tests extend those in Gourieroux et al. (1982) and Silvapulle and Sen (2005) who develop tests for special null hypotheses under which the inequalities in the parameter space are known to bind.

difference exists if and only if spr$A \geq$ spr$B$, and when it exists:

$$A -_H B \equiv [(\text{mid}A - \text{mid}B) \pm (\text{spr}A - \text{spr}B)] .$$

For two generalized intervals A and B, $A -_{GH} B$ is a generalized interval denoting the Generalized Hukuhara difference between A and B:

$$A -_{GH} B \equiv [(\text{mid}A - \text{mid}B) \pm (\text{spr}A - \text{spr}B)] .$$

In contrast to the Hukuhara difference, the Generalized Hukuhara difference always exists. Let $d_\lambda (A, B)$ denote the L_2-type metric between intervals A and B defined as

$$d_\lambda (A, B) \equiv \left((\text{mid}A - \text{mid}B)^2 + \lambda (\text{spr}A - \text{spr}B)^2 \right)^{\frac{1}{2}}$$

for some $\lambda \in (0, \infty)$. For a generalized random interval X,

$$E_A (X) \equiv [E (\text{mid}X) \pm E (\text{spr}X)] ,$$

whenever $E (\text{mid}X)$ and $E (\text{spr}X)$ exist. Define variance as

$$Var_F (X) \equiv E \left(d_\lambda^2 (X, E_A (X)) \right)$$

and conditional expectation as

$$E_A (X \mid \cdot) \equiv [E (\text{mid}X \mid \cdot) \pm E (\text{spr}X \mid \cdot)] .$$

2.2 The Model, Estimation, and Coefficient of Determination

Let $Y, X_1, \ldots, X_k$ denote $(k+1)$ random intervals, where Y is the dependent interval and $X_1, \ldots, X_k$ are covariate intervals some of which may be degenerate intervals. Let $d \geq 0$ denote the number of degenerate covariate intervals. Without loss of generality, let the last d covariates be degenerate. Further let $\{Y_i, X_{1i}, \ldots, X_{ki}\}_{i=1}^n$ denote a random sample on $(Y, X_1, \ldots, X_k)$. Define the $(2k - d)$ dimensional random vector as $\boldsymbol{X}_i \equiv (\text{mid}\boldsymbol{X}_i', \text{spr}\boldsymbol{X}_i')'$, where $\text{mid}\boldsymbol{X}_i \equiv (\text{mid}X_{1i}, \ldots, \text{mid}X_{ki})'$ and $\text{spr}\boldsymbol{X}_i \equiv \left(\text{spr}X_{1i}, \ldots, \text{spr}X_{(k-d)i}\right)'$. We note that

$\text{spr}\boldsymbol{X}_i$ only contains the ranges of $(k-d)$ non-degenerate covariate intervals.

2.2.1 Model M_G and the GIA-CR Model

To motivate the generalized interval arithmetic representation of our model, we first introduce model M_G proposed in Blanco-Fernández et al. (2015). It takes the following form:

$$Y_i = \left[\boldsymbol{X}_i'\boldsymbol{\alpha}^* \pm |\boldsymbol{X}_i|'\boldsymbol{\beta}^*\right] + \Delta_i, \tag{2.1}$$

where $\boldsymbol{\alpha}^* \in \mathbb{R}^{2k-d}$ and $\boldsymbol{\beta}^* \in \mathbb{R}^{2k-d}_{\geq 0}$ are the coefficient vectors and Δ_i is the *random interval error* defined as

$$\Delta_i \equiv Y_i -_H \left[\boldsymbol{X}_i'\boldsymbol{\alpha}^* \pm |\boldsymbol{X}_i|'\boldsymbol{\beta}^*\right]$$

satisfying $E_A(\Delta_i \mid \boldsymbol{X}_i) = [\gamma^* \pm \delta^*]$. $\theta^* \equiv (\boldsymbol{\alpha}^{*\prime}, \gamma^*, \boldsymbol{\beta}^{*\prime}, \delta^*)' \in \mathbb{R}^{2l}$ for $l = 2k-d+1$ is the parameter vector of interest. Notice that model M_G does not include an intercept term as in linear regression model. Conditioning on $\boldsymbol{X}_i$, the expectations of the midpoint and spread of Δ_i are γ^* and δ^*, which are both independent of $\boldsymbol{X}_i$. To ensure that Δ_i is an interval, $\boldsymbol{\beta}^*$ in model M_G must satisfy the additional random constraints that

$$\text{spr}Y_i - |\boldsymbol{X}_i|'\boldsymbol{\beta}^* \geq 0, \text{ for } i = 1, \ldots, n.$$

Blanco-Fernández et al. (2015) proposes an inequality constrained estimator of $(\boldsymbol{\beta}^*, \delta^*)$ without establishing its asymptotic distribution. The difficulty lies in the presence of increasing number of random inequality constraints. By exploiting tools used in Knight (2001, 2006) for linear programming estimators and M-estimators of boundaries, we derive the asymptotic distribution of the constrained estimator of model M_G in Section 2.3. However its complex nature makes inference based on it extremely difficult if not impossible.

To facilitate inference and broaden the scope of applications of interval models, we propose the GIA-CR model below. It generalizes model M_G by dispensing the random inequality constraints and by allowing for general linear inequality constraints rather than $\boldsymbol{\beta}^* \geq \mathbf{0}$ in model M_G. We make use of generalized interval arithmetic and the concept of Generalized

Hukuhara difference introduced in Section 2.1 and discussed in Appendix B.2.

Specifically, the GIA-CR model is composed of two parts: (i) the model in (2.1), where the *random generalized interval error* Δ_i is defined in terms of the Generalized Hukuhara difference:

$$\Delta_i \equiv Y_i -_{GH} \left[\boldsymbol{X}_i'\boldsymbol{\alpha}^* \pm |\boldsymbol{X}_i|' \boldsymbol{\beta}^*\right]$$

satisfying

$$E_A\left(\Delta_i \mid \boldsymbol{X}_i\right) = \left[\gamma^* \pm \delta^*\right]; \tag{2.2}$$

and (ii) the parameter space Θ for θ^* defined as

$$\Theta = \left\{\theta \in \mathbb{R}^{2l} : R\theta \geq r\right\}, \tag{2.3}$$

where R is a known matrix of dimension $l_R \times 2l$ and r is a known vector of dimension l_R. Note that unlike model M_G, both $\left[\boldsymbol{X}_i'\boldsymbol{\alpha}^* \pm |\boldsymbol{X}_i|' \boldsymbol{\beta}^*\right]$ and Δ_i in the GIA-CR model can be generalized intervals although our observations $Y_i, X_{1i}, \ldots, X_{ki}$ are all intervals.[9]

Remark 2.2.1. In the GIA-CR model, we can replace X_i and $|X_i|$ with any known transformations of X_i, but for model M_G and the important IA-CR model introduced in Example 2.2.1 below, it is convenient to use $|X_i|$ for the purpose of forecasting. To avoid introducing too many notations, we use X_i and $|X_i|$ in the GIA-CR model as well.

The GIA-CR model has an alternative bivariate regression representation:

$$\mathrm{mid}Y_i = \boldsymbol{X}_i'\boldsymbol{\alpha}^* + \mathrm{mid}\Delta_i = \mathrm{mid}\boldsymbol{X}_i'\alpha_m^* + \mathrm{spr}\boldsymbol{X}_i'\alpha_s^* + \mathrm{mid}\Delta_i \text{ and} \tag{2.4}$$

$$\mathrm{spr}Y_i = |\boldsymbol{X}_i|' \boldsymbol{\beta}^* + \mathrm{spr}\Delta_i = |\mathrm{mid}\boldsymbol{X}_i|' \beta_m^* + \mathrm{spr}\boldsymbol{X}_i'\beta_s^* + \mathrm{spr}\Delta_i, \tag{2.5}$$

where $\boldsymbol{\alpha}^* \equiv (\alpha_m^{*\prime}, \alpha_s^{*\prime})'$ with $\alpha_m^* \equiv \left(\alpha_{m,1}^*, \ldots, \alpha_{m,k}^*\right)'$ and $\alpha_s^* \equiv \left(\alpha_{s,1}^*, \ldots, \alpha_{s,(k-d)}^*\right)'$; $\boldsymbol{\beta}^* \equiv (\beta_m^{*\prime}, \beta_s^{*\prime})'$ with $\beta_m^* \equiv \left(\beta_{m,1}^*, \ldots, \beta_{m,k}^*\right)'$ and $\beta_s^* \equiv \left(\beta_{s,1}^*, \ldots, \beta_{s,(k-d)}^*\right)'$. The center and the range of Δ_i satisfy that $E\left(\mathrm{mid}\Delta_i \mid \boldsymbol{X}_i\right) = \gamma^*$ and $E\left(\mathrm{spr}\Delta_i \mid \boldsymbol{X}_i\right) = \delta^*$.

[9] In fact, all the results in this paper remain valid when the observations are generalized intervals and the same inequality constraints are imposed on the model.

The bivariate regression representation of the GIA-CR model reveals its relation to the CR model. Example 2.2.1 below presents the IA-CR model which is a direct extension of the CR model.

Example 2.2.1. [The IA-CR Model] Let $R = (\mathbf{0}_{l\times l}, \boldsymbol{I}_l)$ and $r = \mathbf{0}_{l\times 1}$. Then the parameter space Θ becomes

$$\Theta_F = \left\{\theta \in \mathbb{R}^{2l} : \boldsymbol{\beta} \geq \mathbf{0} \text{ and } \delta \geq 0\right\}.$$

Model (2.1) with parameter space Θ_F is referred to as the IA-CR model. To ensure that the predicted dependent interval at any covariate interval is an interval, the regressor in the second term on the right hand side of (2.1) is $|\boldsymbol{X}_i|$ and all elements in β^* and δ^* are required to be non-negative. When there is no degenerate covariate and $\alpha_s^* = \beta_m^* = 0$, (2.4) and (2.5) reduce to the CR model. More general than the CR model, the IA-CR model allows both the center and range of each covariate interval to affect the center and range of the dependent interval Y_i.

2.2.2 Constrained Estimation and the Coefficient of Determination

The constrained estimator of the parameters in model M_G in Blanco-Fernández et al. (2011) is defined by minimizing the d_λ metric between Y_i and $\left[\boldsymbol{X}_i'\boldsymbol{\alpha} \pm |\boldsymbol{X}_i|'\boldsymbol{\beta}\right] + [\gamma \pm \delta]$ via the following constrained minimization problem:

$$\begin{aligned}
\widehat{\theta} &= \arg\min_{\theta} \sum_{i=1}^{n} d_\lambda^2\left(Y_i, \left[\boldsymbol{X}_i'\boldsymbol{\alpha} \pm |\boldsymbol{X}_i|'\boldsymbol{\beta}\right] + [\gamma \pm \delta]\right) \qquad (2.6)\\
&= \arg\min_{\theta} \left[\sum_{i=1}^{n} \left(\text{mid}Y_i - \boldsymbol{X}_i'\boldsymbol{\alpha} - \gamma\right)^2 + \lambda \sum_{i=1}^{n}\left(\text{spr}Y_i - |\boldsymbol{X}_i|'\boldsymbol{\beta} - \delta\right)^2\right]
\end{aligned}$$

$$\text{s.t. spr}Y_i - |\boldsymbol{X}_i|'\boldsymbol{\beta} \geq 0, \text{ for } i = 1, \ldots, n, \text{ and } \boldsymbol{\beta} \geq \mathbf{0}.$$

Similarly, in the GIA-CR model, we construct the constrained estimator of θ^* by minimizing the d_λ metric between Y_i and $\left(\left[\boldsymbol{X}_i'\boldsymbol{\alpha} \pm |\boldsymbol{X}_i|'\boldsymbol{\beta}\right] + [\gamma \pm \delta]\right)$ in the parameter space Θ:

$$\begin{aligned}
\widetilde{\theta} &= \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} d_\lambda^2\left(Y_i, \left[\boldsymbol{X}_i'\boldsymbol{\alpha} \pm |\boldsymbol{X}_i|'\boldsymbol{\beta}\right] + [\gamma \pm \delta]\right) \qquad (2.7)\\
&= \arg\min_{\theta\in\Theta} \left[\sum_{i=1}^{n} (\text{mid}Y_i - \boldsymbol{X}_i'\boldsymbol{\alpha} - \gamma)^2 + \lambda \sum_{i=1}^{n} \left(\text{spr}Y_i - |\boldsymbol{X}_i|'\boldsymbol{\beta} - \delta\right)^2\right].
\end{aligned}$$

The estimators $\widehat{\theta}$ and $\widetilde{\theta}$ are solutions to quadratic minimization problems with linear inequality constraints and can be computed via built-in algorithms such as *quadprog* and *lsqlin* in Matlab.

With the obtained estimators, the interval residual $\widehat{\Delta}_i$ for model M_G is computed as

$$\widehat{\Delta}_i \equiv Y_i -_H \left[\boldsymbol{X}_i'\widehat{\boldsymbol{\alpha}} \pm |\boldsymbol{X}_i|'\widehat{\boldsymbol{\beta}}\right];$$

and the generalized interval residual $\widetilde{\Delta}_i$ for the GIA-CR model is defined as

$$\widetilde{\Delta}_i \equiv Y_i -_{GH} \left[\boldsymbol{X}_i'\widetilde{\boldsymbol{\alpha}} \pm |\boldsymbol{X}_i|'\widetilde{\boldsymbol{\beta}}\right]. \qquad (2.8)$$

For any i, $\widehat{\Delta}_i$ is an interval because of the constraints that $\text{spr}Y_i - |\boldsymbol{X}_i|'\boldsymbol{\beta} \geq 0$ for $i = 1, \ldots, n$; whereas $\widetilde{\Delta}_i$ is a generalized interval.

Remark 2.2.2. In model M_G, the constraints are only imposed on $\boldsymbol{\beta}$. Therefore, the minimization problem in (2.6) can be solved separately for the mid and range regressions. Moreover, the constant λ in the definition of d_λ metric is irrelevant. In fact, $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\gamma}$ are OLS estimators of the slope coefficient and intercept term in the linear regression of $\text{mid}Y_i$ on $\boldsymbol{X}_i$ and hence have closed-form expressions. In contrast, $\left(\widehat{\boldsymbol{\beta}}', \widehat{\delta}\right)$ solves the following constrained optimization problem and does not have closed-form expressions in general:

$$\begin{aligned}
\left(\widehat{\boldsymbol{\beta}}', \widehat{\delta}\right) &= \arg\min_{\boldsymbol{\beta},\delta} \sum_{i=1}^{n} \left(\text{spr}Y_i - |\boldsymbol{X}_i|'\boldsymbol{\beta} - \delta\right)^2 \\
\text{s.t.} &\quad \text{spr}Y_i - |\boldsymbol{X}_i|'\boldsymbol{\beta} \geq 0,\ \text{for } i = 1, \ldots, n,\ \text{and } \boldsymbol{\beta} \geq \boldsymbol{0}.
\end{aligned}$$

On the contrary, since the restrictions in the parameter space Θ for the GIA-CR model are imposed on θ, the minimization problem in (2.7) cannot be separately solved for the mid and range regressions. Moreover the constant λ matters and represents the relative importance of the mid and range regressions.

Example 2.2.1 (continued)**.** Similar to model M_G, in the IA-CR model, the constraints are imposed on $\boldsymbol{\beta}$ and δ. Therefore, the minimization problem can be done separately: $\widetilde{\boldsymbol{\alpha}}$ and $\widetilde{\gamma}$ are OLS estimators; $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\delta}$ solve the following constrained optimization problem:

$$\begin{aligned} \left(\widetilde{\boldsymbol{\beta}}', \widetilde{\delta}\right) &= \arg\min_{\boldsymbol{\beta},\delta} \sum_{i=1}^{n} \left(\mathrm{spr}Y_i - |\boldsymbol{X}_i|' \boldsymbol{\beta} - \delta\right)^2 \\ \text{s.t.} &\quad \boldsymbol{\beta} \geq \mathbf{0} \text{ and } \delta \geq 0. \end{aligned}$$

The non-negative constraints imposed in Θ_F ensure valid forecasts. Equation (2.2) implies that

$$E_A\left(Y_i \mid X_i\right) = \left[\boldsymbol{X}_i'\boldsymbol{\alpha}^* \pm |\boldsymbol{X}_i|' \boldsymbol{\beta}^*\right] + \left[\gamma^* \pm \delta^*\right].$$

Using the constrained estimator, the predictor of Y_0 at covariate interval $\boldsymbol{x}_0$ is given by

$$\widetilde{Y}_0 = \left[\boldsymbol{x}_0'\widetilde{\boldsymbol{\alpha}} \pm |\boldsymbol{x}_0|' \widetilde{\boldsymbol{\beta}}\right] + \left[\widetilde{\gamma} \pm \widetilde{\delta}\right].$$

As pointed out in Neto and de Carvalho (2010), the conceptual difficulty in constructing measures of goodness-of-fit for the CR model is due to the presence of two multiple linear regressions. As a result, some ad hoc combinations of the coefficients of determination for the center and range regressions are adopted, see Neto and de Carvalho (2010). Based on the d_λ metric and the fact that $E_A\left(\Delta_i | X_i\right) = \left[\gamma^* \pm \delta^*\right]$, we now extend the coefficient of determination for multiple linear regressions to the GIA-CR model for random intervals avoiding the ad hoc nature of existing goodness-of-fit measures for the CR model. Define the residual sum of squares for the GIA-CR model as

$$RSS_G \equiv \sum_{i=1}^{n} d_\lambda^2\left(\widetilde{\Delta}_i, \left[\widetilde{\gamma} \pm \widetilde{\delta}\right]\right) = \sum_{i=1}^{n} d_\lambda^2\left(Y_i, \widetilde{Y}_i\right),$$

where $\left[\widetilde{\gamma} \pm \widetilde{\delta}\right]$ is the estimated mean of Δ_i, $\widetilde{\Delta}_i$ is defined in (2.8), and

$$\widetilde{Y}_i = \left[\boldsymbol{X}_i'\widetilde{\boldsymbol{\alpha}} \pm |\boldsymbol{X}_i|'\widetilde{\boldsymbol{\beta}}\right] + \left[\widetilde{\gamma} \pm \widetilde{\delta}\right].$$

Together with the analogous definition of the total sum of squares as $TSS_G \equiv \sum_{i=1}^n d_\lambda^2\left(Y_i, \overline{Y}\right)$, where $\overline{Y} = n^{-1}\sum_{i=1}^n Y_i$, we define the coefficient of determination for the GIA-CR model as

$$\mathcal{R}_G^2 \equiv 1 - \frac{RSS_G}{TSS_G} = 1 - \frac{\sum_{i=1}^n d_\lambda^2\left(\widetilde{\Delta}_i, \left[\widetilde{\gamma} \pm \widetilde{\delta}\right]\right)}{\sum_{i=1}^n d_\lambda^2\left(Y_i, \overline{Y}\right)}. \tag{2.9}$$

It is worthwhile pointing out that (2.9) defines a family of goodness-of-fit measures indexed by λ. By choosing different values of λ, different weights are given to the center and range regressions. In the special case when Y is degenerate, i.e., a random variable, $\mathcal{R}_G^2$ reduces to the coefficient of determination for multiple regressions.

The coefficient of determination $\mathcal{R}_G^2$ inherits all the properties of that for multiple regression which are summarized in the following proposition. The value of $\mathcal{R}_G^2$ ranges from 0 to 1. Further, $\mathcal{R}_G^2 = 0$ indicates no linear relationship between the dependent interval and covariates; $\mathcal{R}_G^2 = 1$ indicates that the fitted model explains all variation of the dependent interval; and the value $\mathcal{R}_G^2$ is non-decreasing with inclusion of more covariates.

Proposition 2.2.1. (i) $0 \leq \mathcal{R}_G^2 \leq 1$; and (ii) $\mathcal{R}_G^2$ is non-deceasing in the number of covariates.

2.3 *Asymptotic Properties of the Constrained Estimators*

The estimators $\widetilde{\theta}$ and $\widehat{\theta}$ defined in (2.7) and (2.6) for the GIA-CR model and Model M_G are both inequality constrained estimators and their asymptotic properties are more difficult to establish than unconstrained estimators. This is particularly true for $\left(\widehat{\boldsymbol{\beta}}', \widehat{\delta}\right)$ due to the increasing number of random inequality constraints that it must satisfy, i.e., $\mathrm{spr}Y_i - |\boldsymbol{X}_i|'\boldsymbol{\beta} \geq 0$, for $i = 1, \ldots, n$. For a special case of model M_G referred to as model M with one covariate, $\left(\widehat{\boldsymbol{\beta}}', \widehat{\delta}\right)$ has a closed-form expression and its asymptotic properties are established in Blanco-Fernández et al. (2012). However, for general model M_G, there is no closed-form expression for $\left(\widehat{\boldsymbol{\beta}}', \widehat{\delta}\right)$ and the approach in Blanco-Fernández et al. (2012) breaks down.

Asymptotic properties of $\widetilde{\theta}$ can be established by applying the general results in Andrews (1999). To handle the increasing number of inequality constraints imposed on $\widehat{\theta}$, we exploit the powerful techniques used in Knight (2001, 2006) for linear programming estimators and M-estimators of boundaries, i.e., epi-convergence in distribution (Pflug (1994, 1995), Geyer (1994, 1996), and Knight (1999)) and point process convergence for extreme values (Kallenberg (1983), Resnick (1987), and Leadbetter et al. (1987)). Since the same techniques apply to $\widetilde{\theta}$ as well, we establish asymptotic properties of both $\widehat{\theta}$ and $\widetilde{\theta}$, including consistency and asymptotic distributions in this section.

Throughout the rest of this paper, we make the following assumption.

Assumption 2.3.1. $\{Y_i, X_{1i}, \ldots, X_{ki}\}_{i=1}^n$ *denotes a random sample on* $(Y, X_1, \ldots, X_k)$.

2.3.1 Consistency

Recall that $\theta \equiv (\boldsymbol{\alpha}', \gamma, \boldsymbol{\beta}', \delta)'$. Define $Z_{1n}(\cdot)$ as

$$Z_{1n}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\text{mid}Y_i - \boldsymbol{X}_i'\boldsymbol{\alpha} - \gamma\right)^2 + \frac{\lambda}{n}\sum_{i=1}^{n}\left(\text{spr}Y_i - |\boldsymbol{X}_i|'\boldsymbol{\beta} - \delta\right)^2 + \varphi_{1n}(\theta),$$

where

$$\varphi_{1n}(\theta) = \begin{cases} 0, & \text{if } \text{spr}Y_i - |\boldsymbol{X}_i|'\boldsymbol{\beta} \geq 0, \text{ for } i = 1, \ldots, n; \text{ and } \boldsymbol{\beta} \geq \mathbf{0} \\ \infty, & \text{otherwise} \end{cases}.$$

Using $Z_{1n}(\cdot)$, we can reformulate the constrained estimator of model M_G as an unconstrained estimator: $\widehat{\theta} = \arg\min_{\theta \in \mathbb{R}^{2l}} Z_{1n}(\theta)$. Similarly, let

$$Z_{2n}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\text{mid}Y_i - \boldsymbol{X}_i'\boldsymbol{\alpha} - \gamma\right)^2 + \frac{\lambda}{n}\sum_{i=1}^{n}\left(\text{spr}Y_i - |\boldsymbol{X}_i|'\boldsymbol{\beta} - \delta\right)^2 + \varphi_2(\theta),$$

where

$$\varphi_2(\theta) = \begin{cases} 0, & \text{if } R\theta \geq r \\ \infty, & \text{otherwise} \end{cases}.$$

We obtain the alternative formulation for the estimator of the GIA-CR model as: $\widetilde{\theta} = \arg\min_{\theta \in \mathbb{R}^{2l}} Z_{2n}(\theta)$.

Let $\dot{\boldsymbol{X}}_i \equiv (\boldsymbol{X}_i', 1)'$, $P_{xx} \equiv E\left(\dot{\boldsymbol{X}}_i \dot{\boldsymbol{X}}_i'\right)$, and $Q_{xx} \equiv E\left(\left|\dot{\boldsymbol{X}}_i\right| \left|\dot{\boldsymbol{X}}_i\right|'\right)$. We prove the consistency result using the notion of epi-convergence under the following assumption. The definition of $Var_F(\cdot)$ can be found in Section 2.1.

Assumption 2.3.2. *(i) P_{xx} and Q_{xx} are non-singular; (ii) $Var_F(\Delta_i) < \infty$.*

Assumption 2.3.2 imposes standard regularity conditions on the random intervals in the model. Assumption 2.3.2 (i) is a rank condition ensuring that the true parameter θ^* is point identified. The following theorem states the consistency of $\widehat{\theta}$ and $\widetilde{\theta}$ for the GIA-CR model and Model M_G.

Theorem 2.3.1. *Under Assumptions 2.3.1 and 2.3.2, $\widehat{\theta} \xrightarrow{p} \theta^*$ and $\widetilde{\theta} \xrightarrow{p} \theta^*$.*

2.3.2 Asymptotic Distribution

The asymptotic distribution of $\widehat{\theta}$ in model M_G remains unknown and is more involved than that of $\widetilde{\theta}$. We present it first.

Model $\mathbf{M}_G$ The behavior of the conditional distribution function of $\mathrm{spr}\Delta_i$ near zero is the critical component determining the asymptotic distribution of the estimator $\widehat{\theta}$, since the constraints: $\mathrm{spr}Y_i - |\boldsymbol{X}_i|' \boldsymbol{\beta} \geq 0$, for $i = 1, \ldots, n$, are equivalent to the constraint that

$$\min_{i=1,\ldots,n} \left[\mathrm{spr}Y_i - |\boldsymbol{X}_i|' \boldsymbol{\beta}\right] \geq 0, \tag{2.10}$$

where $\min_{i=1,\ldots,n} \left[\mathrm{spr}Y_i - |\boldsymbol{X}_i|' \boldsymbol{\beta}\right]$ essentially describes the conditional distribution of $\mathrm{spr}\Delta_i$ near the endpoint of its support.

The following assumption imposes restrictions on the conditional distribution of $\mathrm{spr}\Delta_i$ commonly adopted in the extreme value literature. Let the conditional distribution function of $\mathrm{spr}\Delta_i$ given $|\boldsymbol{X}_i| = \boldsymbol{x}$ be $F_s(\cdot \mid \boldsymbol{x})$.

Assumption 2.3.3. *Assume that*

$$F_s(z \mid \boldsymbol{x}) \sim g(\boldsymbol{x}) F_s(z) \ \ \text{as } z \searrow 0 \text{ uniformly in } \boldsymbol{x},$$

where $g(\cdot) > 0$ *is a continuous function.*

Assumption 2.3.3 requires that for any $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, the tail behaviors of $\text{spr}\Delta_i$ conditional on $|\boldsymbol{X}_i| = \boldsymbol{x}_1$ or $|\boldsymbol{X}_i| = \boldsymbol{x}_2$ are equivalent up to a constant. If $|\boldsymbol{X}_i|$ and $\text{spr}\Delta_i$ are independent, this assumption is trivially satisfied with $g(\cdot) \equiv 1$. Knight (2001), Chernozhukov (2005), and Chernozhukov and Fernández-Val (2011) also impose similar assumptions. However, they require $F_s(\cdot)$ to have Pareto-type tail distribution due to their specific settings, while we impose no such restriction. Let $1/\kappa \equiv \lim_{t\to\infty} tF_s\left(1/\sqrt{t}\right) \in [0, +\infty]$. The value of κ characterizes the distribution of $\text{spr}\Delta_i$ near its left endpoint and determines the effect of the random constraints on the asymptotic distribution of $\widehat{\theta}$: (i) If $\text{spr}\Delta_i$ has a relatively high probability of being close to zero, e.g., when $F_s(0) > 0$, then $\lim_{t\to\infty} tF_s\left(1/\sqrt{t}\right) = +\infty$ and $\kappa = 0$; (ii) If $\text{spr}\Delta_i$ is very 'unlikely' of being near zero, e.g., when $F_s(z) = 0$ for some $z > 0$, then $\lim_{t\to\infty} tF_s\left(1/\sqrt{t}\right) = 0$ and $\kappa = +\infty$; and (iii) when $F_s(z)$ behaves like $h(z) z^2$ near zero, where $h(z)$ is a slowly-varying function at zero,[10] then $\kappa \in (0, +\infty)$.

Assumption 2.3.4. $E\left(|\text{mid}\Delta_i|^4\right) < \infty$, $E\left(|\text{spr}\Delta_i|^4\right) < \infty$, *and* $E\left(\|\boldsymbol{X}_i\|^4\right) < \infty$.

Denote

$$M(\psi; \lambda) \equiv \psi' \begin{pmatrix} P_{xx} & \mathbf{0} \\ \mathbf{0} & \lambda Q_{xx} \end{pmatrix} \psi - 2\psi' \begin{pmatrix} \boldsymbol{I}_l & \mathbf{0} \\ \mathbf{0} & \lambda \boldsymbol{I}_l \end{pmatrix} W, \tag{2.11}$$

where $\psi \equiv (\boldsymbol{p}', q, \boldsymbol{u}', v)'$ with $\boldsymbol{p} \in \mathbb{R}^{2k-d}$, $q \in \mathbb{R}$, $\boldsymbol{u} \in \mathbb{R}^{2k-d}$, $v \in \mathbb{R}$, and $W \sim \mathcal{N}(0, \Lambda)$ in which $\Lambda = Var\left(\dot{\boldsymbol{X}}_i' \text{mid}\Delta_i, \left|\dot{\boldsymbol{X}}_i\right|' \text{spr}\Delta_i\right)$.

Theorem 2.3.2. *Under Assumptions 2.3.1-2.3.4, we obtain:*

$$\sqrt{n}\left(\widehat{\theta} - \theta^*\right) \xrightarrow{d} \arg\min_{\psi} \left[M(\psi; 1) + \phi_1(\psi)\right],$$

---

[10] A function $h(z) : (0, +\infty) \longmapsto (0, +\infty)$ is said to be slowly-varying at z_0 if for any $m > 0$, it holds that $\lim_{z\searrow z_0} [h(z)/h(mz)] = 1$.

where $M(\psi;1)$ *is defined in (2.11) and*

$$\phi_1(\psi) = \begin{cases} 0, & \text{if } \sqrt{\frac{\kappa}{g(\Upsilon_i)}}\Gamma_i \geq \Upsilon_i'\boldsymbol{u}, \text{ for } i = 1, 2, \ldots; \\ & \text{and } I\left(\boldsymbol{\beta}_j^* = 0\right)\boldsymbol{u}_j \geq 0 \text{ for } j = 1, \ldots, 2k-d \cdot \\ \infty, & \text{otherwise} \end{cases}$$

For each i, $\Gamma_i \equiv (\mathcal{E}_1 + \ldots + \mathcal{E}_i)^{\frac{1}{2}}$ *for unit mean i.i.d exponential random variables* $\mathcal{E}_1, \mathcal{E}_2, \ldots$*;* $\Upsilon_1, \Upsilon_2, \ldots$ *are independent and identically distributed with the same distribution as* $|\boldsymbol{X}_i|$*; the* Γ_i*'s are independent of* Υ_i*'s, and they are both independent of* W*.*

Several observations surface from Theorem 2.3.2. First, the estimators of the center and range regressions $(\widehat{\boldsymbol{\alpha}}', \widehat{\gamma})$ and $\left(\widehat{\boldsymbol{\beta}}', \widehat{\delta}\right)$ are asymptotically dependent through the variance-covariance matrix Λ, unless $\text{mid}\Delta_i$ and $\text{spr}\Delta_i$ are conditionally independent. Second, the asymptotic distribution does not depend on λ, which is consistent with the discussion in Remark 2.2.2. Lastly, the asymptotic distribution of $\sqrt{n}\,(\widehat{\boldsymbol{\alpha}}' - \boldsymbol{\alpha}^{*\prime}, \widehat{\gamma} - \gamma^*)$ is normal, but because of inequality constraints, the asymptotic distribution of $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}' - \boldsymbol{\beta}^{*\prime}, \widehat{\delta} - \delta^*\right)$ takes a complicated form given by the distribution of the minimizer of an inequality-constrained optimization problem. Although in general, there is no closed-form expression for the asymptotic distribution of $\left(\widehat{\boldsymbol{\beta}}', \widehat{\delta}\right)$, it is clear from the minimization problem that it is non-standard and discontinuous in model parameters including parameters characterizing the tail behavior of the conditional distribution of $\text{spr}\Delta_i$ through $\phi_1(\psi)$. For example, the value of κ determines whether the constraint in (2.10) binds or not leading to different asymptotic distributions of $\left(\widehat{\boldsymbol{\beta}}', \widehat{\delta}\right)$.

The GIA-CR Model Employing the same techniques, we obtain the following asymptotic distribution for the estimator $\widetilde{\theta}$. Denote R_b as the $l_b \times 2l$ submatrix composed of the l_b rows of R corresponding to binding inequalities in $R\theta^* \geq r$. The asymptotic distribution depends on the value of λ, because the constraints are imposed on θ directly.

Theorem 2.3.3. *Under Assumptions 2.3.1, 2.3.2, and 2.3.4, it holds that*

$$\sqrt{n}\left(\widetilde{\theta}-\theta^{*}\right)\xrightarrow{d}\arg\min_{\psi}\left[M\left(\psi;\lambda\right)+\phi_{2}\left(\psi\right)\right],$$

where $M(\psi;\lambda)$ is defined in (2.11) and

$$\phi_{2}\left(\psi\right)=\begin{cases}0, & \text{if } R_{b}\psi\geq\mathbf{0}\\ \infty, & \text{otherwise}\end{cases}.$$

Alternatively one may apply the approach in Andrews (1999) to obtain the asymptotic distribution of $\widetilde{\theta}$. Instead of studying the unconstrained minimization problem, Andrews (1999) focuses on the constrained minimization problem and shows that the asymptotic distribution of the estimator can be represented as the minimizer of a quadratic function over a convex cone.

Theorem 2.3.3 shows that the asymptotic distribution of $\widetilde{\theta}$ is discontinuous in $R\theta^{*}$ at r, because researchers do not known which inequalities among $R\theta^{*}\geq r$ are binding a priori. For hypothesis testing, asymptotic distributions of test statistics based upon $\widetilde{\theta}$ are in general also discontinuous. In Section 2.4, we provide asymptotically uniformly valid Wald-type test for testing the linear hypothesis in the GIA-CR Model.

Example 2.2.1 (continued)**.** The asymptotic distribution of the estimators for the IA-CR model can be obtained directly from Theorem 2.3.3. The matrix R_b corresponds to elements in $(\boldsymbol{\beta}^{*\prime},\delta^{*})$ that are zeros. Specifically, if Assumptions 2.3.1, 2.3.2, and 2.3.4 hold, then

$$\sqrt{n}\begin{pmatrix}\widetilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^{*}\\ \widetilde{\gamma}-\gamma^{*}\end{pmatrix}\xrightarrow{d}P_{xx}^{-1}W_{m},$$

where $W_m\sim\mathcal{N}\left(0,\Lambda_m\right)$ with Λ_m being the covariance matrix of $\dot{\boldsymbol{X}}_i^{\prime}\text{mid}\Delta_i$; and

$$\sqrt{n}\begin{pmatrix}\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^{*}\\ \widetilde{\delta}-\delta^{*}\end{pmatrix}\xrightarrow{d}\arg\min_{\boldsymbol{u},v}\left[\left(\boldsymbol{u}^{\prime},v\right)Q_{xx}\begin{pmatrix}\boldsymbol{u}\\ v\end{pmatrix}-2\left(\boldsymbol{u}^{\prime},v\right)W_{s}+\phi_{s}\left(\boldsymbol{u}^{\prime},v\right)\right],$$

where $W_s \sim \mathcal{N}(0, \Lambda_s)$ with Λ_s being the covariance matrix of $\left|\dot{\boldsymbol{X}}_i\right|' \mathrm{spr}\Delta_i$, and

$$\phi_s(\boldsymbol{u}', v) = \begin{cases} 0, & \text{if } I\left(\boldsymbol{\beta}_j^* = 0\right)\boldsymbol{u}_j \geq 0 \text{ for } j = 1, \ldots, 2k-d \text{ and } I(\delta^* = 0)\, v \geq 0 \\ \infty, & \text{otherwise} \end{cases}.$$

Like $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}' - \boldsymbol{\beta}^{*\prime}, \widehat{\delta} - \delta^*\right)$, the asymptotic distribution of $\sqrt{n}\left(\widetilde{\boldsymbol{\beta}}' - \boldsymbol{\beta}^{*\prime}, \widetilde{\delta} - \delta^*\right)$ is discontinuous in some model parameters.

For the special case of the IA-CR model, i.e., the CR model with only one covariate, we can solve the minimization problem in Example 2.2.1 by following Andrews (1999). The CR model with one covariate can be expressed as follows:

$$\begin{aligned} \mathrm{mid}Y_i &= \alpha_m^* \mathrm{mid}X_i + \gamma^* + \epsilon_i^c \text{ and} \\ \mathrm{spr}Y_i &= \beta_s^* \mathrm{spr}X_i + \delta^* + \epsilon_i^r, \end{aligned}$$

where $E(\epsilon_i^c \mid X_i) = 0$, $E(\epsilon_i^r \mid X_i) = 0$, $\alpha_m^* \in \mathbb{R}$, $\gamma^* \in \mathbb{R}$, $\beta_s^* \geq 0$, and $\delta^* \geq 0$. Let $K = \mathrm{diag}^{1/2}\left(Q_{xx}^{-1}\right)$, $Z = (Z_1, Z_2)' = K^{-1} Q_{xx}^{-1} W_s$, and $\rho_{ij} = \left[K^{-1} Q_{xx}^{-1} K^{-1}\right]_{ij}$ for $i, j = 1, 2$, where $\mathrm{diag}(A)$ returns a diagonal matrix whose elements equal to the diagonal elements of matrix A. Suppose Assumptions 2.3.1, 2.3.2, and 2.3.4 hold.

(i) When $\beta_s^* > 0$ and $\delta^* > 0$,

$$\sqrt{n}\begin{pmatrix} \widetilde{\beta}_s - \beta_s^* \\ \widetilde{\delta} - \delta^* \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, Q_{xx}^{-1} \Lambda_s Q_{xx}^{-1}\right);$$

(ii) When $\beta_s^* = 0$ and $\delta^* > 0$,

$$\sqrt{n}\begin{pmatrix} \widetilde{\beta}_s - \beta_s^* \\ \widetilde{\delta} - \delta^* \end{pmatrix} \xrightarrow{d} I(Z_1 \geq 0)\, Q_{xx}^{-1} W_s + I(Z_1 < 0)\, K \begin{pmatrix} 0 \\ Z_2 - \rho_{12} Z_1 \end{pmatrix};$$

(iii) When $\beta_s^* > 0$ and $\delta^* = 0$,

$$\sqrt{n}\begin{pmatrix} \widetilde{\beta}_s - \beta_s^* \\ \widetilde{\delta} - \delta^* \end{pmatrix} \xrightarrow{d} I(Z_2 \geq 0)\, Q_{xx}^{-1} W_s + I(Z_2 < 0)\, K \begin{pmatrix} Z_1 - \rho_{21} Z_2 \\ 0 \end{pmatrix};$$

(iv) When $\beta_s^* = 0$ and $\delta^* = 0$,

$$\sqrt{n}\begin{pmatrix} \widetilde{\beta}_s - \beta_s^* \\ \widetilde{\delta} - \delta^* \end{pmatrix} \xrightarrow{d} I\left(Z_1 > 0, Z_2 > 0\right) Q_{xx}^{-1} W_s + I\left(Z_1 - \rho_{21} Z_2 > 0, Z_2 \leq 0\right) K \begin{pmatrix} Z_1 - \rho_{21} Z_2 \\ 0 \end{pmatrix}$$

$$+ I\left(Z_1 \leq 0, Z_2 - \rho_{12} Z_1 > 0\right) K \begin{pmatrix} 0 \\ Z_2 - \rho_{12} Z_1 \end{pmatrix}.$$

The asymptotic distribution of $\left(\widetilde{\beta}_s, \widetilde{\delta}\right)$ takes a complicated form and is discontinuous in β_s^* and δ^*: it is normal when β_s^* and δ^* are both positive; non-normal otherwise.

2.4 Wald-Type Tests for Linear Hypothesis in the GIA-CR Model

In this section, we construct asymptotically uniformly valid tests for the linear hypothesis in the GIA-CR model of the following form:

$$H_0 : R_0\theta^* = r_0 \text{ against } H_1 : R_0\theta^* \neq r_0 \tag{2.12}$$

under the maintained hypothesis that $\theta^* \in \Theta = \{\theta : R\theta \geq r\}$, where R_0 and r_0 are known matrices of dimensions $J \times 2l$ and $J \times 1$ respectively. Furthermore, $R_0 \equiv (R_{00}', R_{01}')'$ is of full row rank and R_{00} is a submatrix of R. Without loss of generality, let $R = (R_{00}', R_{\Gamma}')'$ and accordingly $r = \left(r_{-\Gamma}', r_{\Gamma}'\right)'$.

One important example which motivates our testing framework is that of testing the correct specification of the CR model against the IA-CR model which corresponds to testing

$$H_0^{CR} : (\alpha_s^{*\prime}, \beta_m^{*\prime})' = \mathbf{0} \text{ against } H_1^{CR} : (\alpha_s^{*\prime}, \beta_m^{*\prime})' \neq \mathbf{0},$$

under the maintained hypothesis that $\theta^* \in \Theta_F = \{\theta : \boldsymbol{\beta} \geq \mathbf{0} \text{ and } \delta \geq 0\}$. Testing H_0^{CR} belongs to the standard subvector inference with $(\beta_s^{*\prime}, \delta^*)'$ being the vector of nuisance parameters. As another example, consider testing equality constraints in linear regression models. Let $R = (\mathbf{0}_{l_R \times l}, R^*)$ and $R_0 = (\mathbf{0}_{J \times l}, R_0^*)$ for some known matrices R^* and R_0^*. Then both the null hypothesis and linear inequality constraints in the maintained hypothesis are on parameters in the range regression of the GIA-CR model only. Our tests become tests for linear equality constraints in linear regression models with linear inequality constraints extending the Wald-type test in Gourieroux et al. (1982) for the special case that $R_0^* = R^*$ and Silvapulle and Sen (2005) for the case that R^* is a submatrix of R_0^*. In both cases considered in Gourieroux et al. (1982) and Silvapulle and Sen (2005), the inequalities in the parameter space are known to bind under H_0 and as a result, the asymptotic distribution of the inequality constrained estimator of θ^* or the test statistic under the null hypothesis is continuous in model parameters and is thus nuisance parameter free. When R_{00} is a proper submatrix of R, the asymptotic distribution of inequality constrained estimators or the null asymptotic distribution of the test statistic is typically discontinuous in nuisance parameters and asymptotically uniformly valid tests for H_0 are not available in the current literature.

In the following subsections, we first introduce our test statistic and its asymptotic size. Then we apply the two-stage approach with Bonferroni-type correction in McCloskey (2017) to construct asymptotically uniformly valid tests for H_0^{CR} against H_1^{CR}. Lastly we construct tests for H_0 in the GIA-CR model.

2.4.1 The Test Statistic and Asymptotic Size

The test statistic we adopt is of Wald-type:

$$T_n\left(R_0, r_0\right) = n\left(R_0\widetilde{\theta} - r_0\right)'\left(R_0\Sigma_n R_0'\right)^{-1}\left(R_0\widetilde{\theta} - r_0\right), \tag{2.13}$$

for some positive definite weighting matrix $\Sigma_n \to_p \Sigma$ with Σ being a deterministic positive definite matrix. Σ_n can be chosen as the identity matrix or a consistent estimator of

$$\begin{pmatrix} P_{xx}^{-1} & \mathbf{0} \\ \mathbf{0} & Q_{xx}^{-1} \end{pmatrix} \Lambda \begin{pmatrix} P_{xx}^{-1} & \mathbf{0} \\ \mathbf{0} & Q_{xx}^{-1} \end{pmatrix}.$$

When R_{00} is a proper submatrix of R, the null asymptotic distribution of $R_0\widetilde{\theta}$ or $T_n\left(R_0, r_0\right)$ is discontinuous in $R_\Gamma\theta^*$.

We now introduce the test for H_0 based on $T_n\left(R_0, r_0\right)$ and its asymptotic size. The GIA-

CR model can be fully characterized by the finite dimensional parameter $\theta^* \in \Theta$ and the infinite dimensional parameter $\mu^* \in \mathrm{M}$ characterizing the distribution of $\{(Y_i, \boldsymbol{X}_i) : 1 \leq i \leq n\}$. The space M can be restricted to be some compact metric space with a metric that induces weak convergence, see Andrews et al. (2011). Let $\omega \equiv (\theta^*, \mu^*) \in \mathcal{W}$. Denote $\boldsymbol{P}_\omega$ as the probability model indexed by ω, E_ω as the expectation, and Pr_ω 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 (CV) for the test based on the test statistic $T_n(R_0, r_0)$. The asymptotic size of the resulting test is defined by

$$AsySz\left(T_n\left(R_0, r_0\right), CV_n\right) \equiv \limsup_{n\to\infty} \sup_{\omega\in\mathcal{W}_0} Pr_\omega\left(T_n\left(R_0, r_0\right) > CV_n\right). \tag{2.14}$$

We aim to construct CV_n that controls the asymptotic size of the test based on $T_n(R_0, r_0)$. We make the following assumptions.

Assumption 2.4.1. *For any $\omega \in \mathcal{W}_0$, (i) $E_\omega\left(\dot{\boldsymbol{X}}_i\dot{\boldsymbol{X}}_i'\right)$ and $E_\omega\left(\left|\dot{\boldsymbol{X}}_i\right|\left|\dot{\boldsymbol{X}}_i\right|'\right)$ are non-singular; (ii) $E_\omega\left(\left|\mathrm{mid}\Delta_i\right|^{4+\nu}\right) < M$, $E_\omega\left(\left|\mathrm{spr}\Delta_i\right|^{4+\nu}\right) < M$ and $E_\omega\left(\left\|\boldsymbol{X}_i\right\|^{4+\nu}\right) < M$ for some $\nu > 0$ and $M < \infty$.*

Assumption 2.4.2. *For any probability model $\boldsymbol{P}_w$ with $\omega \in \mathcal{W}_0$, $\Sigma_n \xrightarrow{p} \Sigma_\omega$ for a positive definite matrix Σ_ω.*

We finish this section by introducing notations that will be used in the subsequent analysis. Let $P_{\omega,xx} \equiv E_\omega\left(\dot{\boldsymbol{X}}_i\dot{\boldsymbol{X}}_i'\right)$, $Q_{\omega,xx} \equiv E_\omega\left(\left|\dot{\boldsymbol{X}}_i\right|\left|\dot{\boldsymbol{X}}_i\right|'\right)$, and

$$M_\omega\left(\psi;\lambda\right) \equiv \psi'\begin{pmatrix} P_{\omega,xx} & \mathbf{0} \\ \mathbf{0} & \lambda Q_{\omega,xx}\end{pmatrix}\psi - 2\psi'\begin{pmatrix} \boldsymbol{I}_l & \mathbf{0} \\ \mathbf{0} & \lambda\boldsymbol{I}_l\end{pmatrix}W_\omega, \tag{2.15}$$

where $\psi = (\boldsymbol{p}', q, \boldsymbol{u}', v)$ with $\boldsymbol{p} \in \mathbb{R}^{2k-d}$, $q \in \mathbb{R}$, $\boldsymbol{u} \in \mathbb{R}^{2k-d}$, $v \in \mathbb{R}$, and $W_\omega \sim \mathcal{N}(0, \Lambda_\omega)$ in which $\Lambda_\omega = Var_\omega\left(\dot{\boldsymbol{X}}_i'\mathrm{mid}\Delta_i, \left|\dot{\boldsymbol{X}}_i\right|'\mathrm{spr}\Delta_i\right)$.

2.4.2 Testing Correct Specification of the CR Model Against the IA-CR Model

To simplify notation, we denote the test statistic $T_n(R_0, r_0)$ in (2.13) for testing H_0^{CR} as T_n, i.e.,

$$T_n = n\left(\widetilde{\alpha}_s', \widetilde{\beta}_m'\right)\left(R_0 \Sigma_n R_0'\right)^{-1}\left(\widetilde{\alpha}_s', \widetilde{\beta}_m'\right)', \tag{2.16}$$

where $\widetilde{\alpha}_s$ and $\widetilde{\beta}_m$ are the estimators defined in (2.7) and

$$R_0 = \begin{pmatrix} \mathbf{0}_{k\times k} & \mathbf{0}_{k\times(k-d)} & \mathbf{0}_{k\times 1} & \boldsymbol{I}_k & \mathbf{0}_{k\times(k-d)} & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{(k-d)\times k} & \boldsymbol{I}_{(k-d)} & \mathbf{0}_{(k-d)\times 1} & \mathbf{0}_{(k-d)\times k} & \mathbf{0}_{(k-d)\times(k-d)} & \mathbf{0}_{(k-d)\times 1} \end{pmatrix}.$$

The asymptotic distribution of T_n is discontinuous in $(\beta_m^{*\prime}, \beta_s^{*\prime}, \delta^*)'$, because the asymptotic distribution of $\sqrt{n}\left(\widetilde{\theta} - \theta^*\right)$ is discontinuous in $(\beta_m^{*\prime}, \beta_s^{*\prime}, \delta^*)'$ by applying Theorem 2.3.3. For any $\omega \in \mathcal{W}_0$, it holds that $(\alpha_s^{*\prime}, \beta_m^{*\prime})' = \mathbf{0}$. Therefore, $(\beta_m^{*\prime}, \beta_s^{*\prime}, \delta^*)' = (\mathbf{0}, \beta_s^{*\prime}, \delta^*)'$ under the null hypothesis.

We decompose the model parameter $\omega \in \mathcal{W}$ into three groups: $\omega \equiv (\eta, \pi, \xi)$ based on their roles in the asymptotic distribution of the test statistics, where $\eta \equiv (\beta_s^{*\prime}, \delta^*)' \in \mathbb{R}_{\geq 0}^{k-d+1}$, $\pi \equiv (\mathrm{vec}(P_{xx}), \mathrm{vec}(Q_{xx}), \mathrm{vec}(\Lambda), \mathrm{vec}(\Sigma))' \in \Pi$, and $\xi \in \Xi$ consists of all other parameters and is infinite-dimensional. From the previous discussion, the null asymptotic distribution of T_n is discontinuous in η; π affects the limiting distribution of T_n but not its continuity; ξ does not affect the limiting distribution of T_n.

Let $\overline{\mathbb{R}}_{\geq 0} \equiv \mathbb{R}_{\geq 0} \cup \{+\infty\}$. Following Andrews et al. (2011), Andrews and Cheng (2012), Andrews and Cheng (2014), and Cheng (2015), we establish the asymptotic distribution of T_n under drifting parameter sequences $\omega_n \in \mathcal{W}_0$. In particular, we consider the parameter sequence $\left\{(\eta_n, \pi_n, \xi_n) \in \mathbb{R}_{\geq 0}^{k-d+1} \times \Pi \times \Xi : n \geq 1\right\}$ and the localization parameter c and π_ω as the limit of $\sqrt{n}\eta_n$ and π_n:

$$\begin{aligned} \sqrt{n}\eta_n &\rightarrow c = (c_1, \ldots, c_{k-d}, c_{k-d+1})' \in \overline{\mathbb{R}}_{\geq 0}^{k-d+1} \text{ and} \\ \pi_n &\rightarrow \pi_\omega = (\mathrm{vec}(P_{\omega,xx}), \mathrm{vec}(Q_{\omega,xx}), \mathrm{vec}(\Lambda_\omega), \mathrm{vec}(\Sigma_\omega))' \in \Pi. \end{aligned}$$

As shown in the lemma below, the asymptotic distribution of T_n under the null hypothesis and the drifting parameter sequence (η_n, π_n, ξ_n) depends on c and π_ω; whereas ξ_n (or the limiting value ξ_ω of ξ_n) does not affect the limiting distribution under any parameter sequence η_n and π_n.

Let

$$\Psi = \Psi(c, \pi_\omega) \equiv \arg\min_{\psi} \left[M_\omega(\psi; 1) + \phi_\omega(\psi)\right],$$

where $M_\omega(\psi; \lambda)$ is defined in (2.15) and

$$\phi_\omega(\psi) = \begin{cases} 0, & \text{if } \boldsymbol{u}_j \geq 0 \text{ for } j = 1, \ldots, k,\ \boldsymbol{u}_{j+k} + c_j \geq 0 \text{ for } j = 1, \ldots, k-d \\ & \text{and } v + c_{k-d+1} \geq 0 \\ \infty, & \text{otherwise} \end{cases}.$$

Then $R_0\Psi$ characterizes the limiting distribution of $\sqrt{n}\left(\widetilde{\alpha}_s', \widetilde{\beta}_m'\right)'$ under H_0^{CR} and the drifting parameter sequence (η_n, π_n, ξ_n).

Lemma 2.4.1. *Under H_0^{CR} and the parameter sequence $(\eta_n, \pi_n, \xi_n) \in \mathbb{R}_{\geq 0}^{k-d+1} \times \Pi \times \Xi$ such that $\sqrt{n}\eta_n \to c$, $\pi_n \to \pi_\omega$ and $\xi_n \to \xi_\omega$, if Assumptions 2.3.1, 2.4.1, and 2.4.2 hold, then the asymptotic distribution of T_n is given by $(R_0\Psi)' (R_0\Sigma_\omega R_0')^{-1} (R_0\Psi)$.*

As shown in Lemma 2.4.1, the null asymptotic distribution of T_n under the drifting sequences of distributions depends on the value of (c, π_ω). Let $\mathcal{C}_{c,\pi_\omega}(1-\vartheta)$ denote the $(1-\vartheta)$ quantile of the distribution of $(R_0\Psi)' (R_0\Sigma_\omega R_0')^{-1} (R_0\Psi)$ given c and π_ω which can be simulated. Building on existing work, especially McCloskey (2017), we adopt the two-stage approach with Bonferroni-type correction to construct an asymptotically uniformly valid test for H_0^{CR}.

The detailed process consists of the following steps.

Step 1. (i) Construct the estimator $\widehat{\pi}$. Consistent estimator $\widehat{\pi}$ can be decomposed into two parts: $(\mathrm{vec}(P_{\omega,xx}), \mathrm{vec}(Q_{\omega,xx}))$ is estimable from the sample $\{\boldsymbol{X}_i\}_{i=1}^n$; Λ_ω and Σ_ω can be estimated by using the residuals computed with $\widetilde{\theta}$;

(ii) Construct confidence sets for $\sqrt{n}\eta_n$. Let $\widehat{\beta}_{s,OLS}$ and $\widehat{\delta}_{OLS}$ be the OLS estimators of

β_s^* and δ^*. Simple calculation shows that

$$\sqrt{n}\left(\widehat{\beta}'_{s,OLS}, \widehat{\delta}_{OLS}\right)' - \sqrt{n}\eta_n \xrightarrow{d} Z\left(\Lambda_{OLS}\right),$$

where $Z\left(\Lambda_{OLS}\right)$ is a multivariate normal distribution with zero mean and variance-covariance matrix Λ_{OLS}. Denote $ES_\Lambda\left(\tau\right)$ as a set such that $Pr\left(Z\left(\Lambda\right) \in ES_\Lambda\left(\tau\right)\right) = 1 - \tau$. The confidence set I_τ for $\sqrt{n}\eta_n$ is defined as

$$I_\tau \equiv \left\{\varsigma \in \mathbb{R}^{k-d+1}_{\geq 0} : \varsigma \in \sqrt{n}\left(\widehat{\beta}'_{s,OLS}, \widehat{\delta}_{OLS}\right)' - ES_{\widehat{\Lambda}_{OLS}}\left(\tau\right)\right\}.$$

The value $\widehat{\Lambda}_{OLS}$ can be computed using the standard approach in least squares estimation and is a consistent estimator for Λ_{OLS}. It should be pointed out that to ensure the non-emptiness of I_τ, $ES_\Lambda\left(\tau\right)$ may not be equal-tailed.

Step 2. We construct the ϑ level Bonferroni critical value CV_n^b for some $0 < \tau < \vartheta$ as

$$CV_n^b\left(\vartheta, \tau\right) \equiv \sup_{c \in I_{\vartheta - \tau}} \mathcal{C}_{c,\widehat{\pi}}\left(1 - \tau\right). \tag{2.17}$$

The following proposition establishes the asymptotic validity of the test.

Proposition 2.4.1. For some $\vartheta \in (0, 1)$, assume that $\left(R_0\Psi\right)'\left(R_0\Sigma_\omega R_0'\right)^{-1}\left(R_0\Psi\right)$ is continuous at $\mathcal{C}_{c,\pi_\omega}\left(1 - \vartheta\right)$ for all $\left(c, \pi_\omega\right) \in \overline{\mathbb{R}}^{k-d+1}_{\geq 0} \times \Pi$. Under Assumptions 2.3.1, 2.4.1, and 2.4.2, it holds that $AsySz\left(T_n, CV_n^b\left(\vartheta, \tau\right)\right) \leq \vartheta$.

Remark 2.4.1. Adjusted Bonferroni critical value or Minimum Bonferroni critical value discussed in McCloskey (2017) are also applicable. The former considers the joint distribution of T_n and $\left(\widehat{\beta}'_{s,OLS}, \widehat{\delta}_{OLS}\right)$, and computes the optimal pair of significant levels for c and T_n; while the latter combines Bonferroni critical value and Adjusted Bonferroni critical value. The computational burden for the two critical values is significant when the dimension of β_s is large. Interested readers could refer to McCloskey (2017) for detailed implementation of such methods.

2.4.3 Testing H_0 in the GIA-CR Model

In this subsection, we extend the test for the correct specification of the CR model against the IA-CR model developed in the previous subsection to the problem of testing H_0 in (2.12). Since $R_{00}\theta^*$ is a subvector of $R\theta^*$, the null asymptotic distribution of the test statistic defined in (2.13) is discontinuous in $R_\Gamma\theta^*$, where θ^* satisfies the inequalities: $R_\Gamma\theta^* \geq r_\Gamma$. In contrast to the standard subvector inference such as that considered in the previous subsection, components of $R_\Gamma\theta^*$ could be linearly dependent rendering direct application of the first stage of the two-stage approach with Bonferroni-type correction in the previous subsection problematic. To address this potential issue, we suggest a three-stage approach for constructing asymptotically uniformly valid tests for H_0.

In the first stage, we identify binding and non-binding inequalities in $R_{00}\theta^* \geq r_{-\Gamma}$; and employ the Gauss-Jordan elimination to identify a row basis of R_Γ based on which we define nuisance parameters and express the inequalities $R_\Gamma\theta^* \geq r_\Gamma$ in terms of the nuisance parameters. In the second stage, we construct confidence sets for the nuisance parameters. Lastly we construct the CV for our test using Bonferroni-type correction.

Identification of Nuisance Parameters

Under H_0, the inequalities in $R_{00}\theta^* \geq r_{-\Gamma}$ are known to bind or not to bind. Let R_{0b} denote the submatrix of R_{00} composed of rows corresponding to binding inequalities in $R_{00}\theta^* \geq r_{-\Gamma}$. R_{0b} can be identified directly from the null hypothesis.

Let $\eta = R_\Gamma\theta^* \in \mathbb{R}^{l_\Gamma}$. The vector of nuisance parameters is defined in the following.

Definition 2.4.1. *The vector of nuisance parameters, denoted as η^u, is defined as a subvector of η corresponding to a row basis of R_Γ.*

By definition, the nuisance parameters are $\eta^u = R_\Gamma^u\theta^*$, where R_Γ^u is a submatrix of R_Γ with rows forming a row basis of R_Γ. When R_Γ is of *full row rank*, $R_\Gamma^u = R_\Gamma$ and the nuisance parameters are $\eta^u = \eta$. When R_Γ is *not of full row rank*, we compute R_Γ^u and Γ such that

$R_\Gamma = \Gamma R_\Gamma^u$ by Gauss-Jordan elimination on the transpose of R_Γ. In terms of the nuisance parameters, the inequalities: $\eta = R_\Gamma \theta^* \geq r_\Gamma$ become:

$$\Gamma \eta^u \geq r_\Gamma. \tag{2.18}$$

Below we present two examples to illustrate this step. For notational compactness, we let $\theta = (\theta_1, ..., \theta_{l^*})$ for an integer l^*.

Example 2.4.1. Suppose $l^* = 4$ and $H_0 : \theta_1^* = \theta_2^*$. Then $R_0 = (1, -1, 0, 0)$.

(i) Let $\Theta = \{\theta : \theta_1 - \theta_2 \geq 0, \theta_2 - \theta_3 \geq 0, \theta_3 - \theta_4 \geq 0\}$. Then

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

Under H_0, the first inequality in Θ binds resulting in $R_{0b} = (1, -1, 0, 0)$ and

$$R_\Gamma = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

Since R_Γ is of full row rank, the nuisance parameters are given by

$$\eta^u = \eta = R_\Gamma \theta^* = \begin{pmatrix} \theta_2^* - \theta_3^* \\ \theta_3^* - \theta_4^* \end{pmatrix}.$$

(ii) Let $\Theta = \{\theta : \theta_1 - \theta_2 \geq 0, 1 \geq \theta_2 - \theta_3 \geq 0, \theta_3 - \theta_4 \geq 0\}$. Then

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

Under H_0, the first inequality in Θ binds resulting in $R_{0b} = (1, -1, 0, 0)$ and

$$R_\Gamma = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

In this case, R_Γ is not of full row rank. Applying Gauss-Jordan elimination to R'_Γ yields

$$R^u_\Gamma = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \text{ and } \Gamma = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

By definition, the nuisance parameters are given by

$$\eta^u = R^u_\Gamma \theta^* = \begin{pmatrix} \theta^*_2 - \theta^*_3 \\ \theta^*_3 - \theta^*_4 \end{pmatrix}$$

and the inequalities are

$$\Gamma \eta^u = \begin{pmatrix} \theta^*_2 - \theta^*_3 \\ -\left(\theta^*_2 - \theta^*_3\right) \\ \theta^*_3 - \theta^*_4 \end{pmatrix} \geq \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}.$$

Example 2.4.2. Let $l^* = 8$ and $H_0 : \theta^*_1 = \theta^*_2 = \theta^*_3 = 0$. Then

$$R_0 = \begin{pmatrix} 1 & 0 & 0 & \\ 0 & 1 & 0 & \mathbf{0}_{3\times 5} \\ 0 & 0 & 1 & \end{pmatrix} \text{ and } r_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Let $\Theta = \{\theta : R\theta \geq r\}$, where $R' = (R_0', R_\Gamma')$ and $r = (0, -1, 0, 0, -1, 0, 0, 0)'$ in which

$$R_\Gamma = \begin{pmatrix} & 1 & 0 & 0 & 0 & 0 \\ & -1 & 0 & 0 & 0 & 0 \\ \mathbf{0}_{5\times 3} & 0 & 1 & 0 & 1 & -1 \\ & 0 & 2 & 1 & 1 & 0 \\ & 0 & 1 & -1 & 2 & -3 \end{pmatrix}.$$

Under H_0, the first and third inequalities in Θ bind and the second inequality does not bind resulting in

$$R_{0b} = \begin{pmatrix} 1 & 0 & 0 & \\ & & & \mathbf{0}_{2\times 5} \\ 0 & 0 & 1 & \end{pmatrix}.$$

Since the rows of R_Γ are linearly dependent, we apply the Gauss-Jordan elimination to the transpose of R_Γ:

$$R_\Gamma' = \begin{pmatrix} & & \mathbf{0}_{3\times 5} & & \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & 0 & -3 \end{pmatrix} \xrightarrow{\text{Gauss-Jordan elimination}} \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -1 \\ & & \mathbf{0}_{5\times 5} & & \end{pmatrix},$$

and conclude that the first, third, and fourth rows of R_Γ constitute a row basis of R_Γ. Further, we get that

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 & \\ -1 & 0 & 0 & \\ 0 & 1 & 0 & \mathbf{0}_{5\times 5} \\ 0 & 0 & 1 & \\ 0 & 3 & -1 & \end{pmatrix}, \ \eta^u = \begin{pmatrix} \theta_4^* \\ \theta_5^* + \theta_7^* - \theta_8^* \\ 2\theta_5^* + \theta_6^* + \theta_7^* \end{pmatrix},$$

and the inequalities below:

$$\Gamma\eta^u = \begin{pmatrix} \theta_4^* \\ -\theta_4^* \\ \theta_5^* + \theta_7^* - \theta_8^* \\ 2\theta_5^* + \theta_6^* + \theta_7^* \\ 3\left(\theta_5^* + \theta_7^* - \theta_8^*\right) - \left(2\theta_5^* + \theta_6^* + \theta_7^*\right) \end{pmatrix} \geq \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

The Null Asymptotic Distribution of $T_n\left(R_0, r_0\right)$ Under Drifting Sequences and the Testing Procedure

We consider the drifting model parameters (η_n^u, π_n, ξ_n), where π_n and ξ_n are defined in the same way as in Section 2.4.2. Since the nuisance parameters η^u satisfy inequalities in (2.18), we consider local sequences η_n^u such that

$$c = \lim_{n\to\infty} \sqrt{n}\left(\Gamma\eta_n^u - r_\Gamma\right) \in \overline{\mathbb{R}}_{\geq 0}^{l_\Gamma}.$$

Let

$$\Psi = \Psi\left(c, \pi_\omega\right) \equiv \arg\min_{\psi} \left[M_\omega\left(\psi; \lambda\right) + \phi_\omega\left(\psi\right)\right],$$

where $M_\omega\left(\psi; \lambda\right)$ is defined in (2.15) and

$$\phi_\omega\left(\psi\right) = \begin{cases} 0, & \text{if } R_{0b}\psi \geq 0 \text{ and } R_\Gamma\psi + c \geq 0 \\ \infty, & \text{otherwise} \end{cases}.$$

Lemma 2.4.2. *Under H_0 and (η_n^u, π_n, ξ_n) defined above, if Assumptions 2.3.1, 2.4.1, and 2.4.2 hold, then the asymptotic distribution of $T_n\left(R_0, r_0\right)$ is given by $\left(R_0\Psi\right)'\left(R_0\Sigma_\omega R_0'\right)^{-1}\left(R_0\Psi\right)$.*

The null asymptotic distribution of $T_n\left(R_0, r_0\right)$ stated in Lemma 2.4.2 suggests the following procedure for computing the critical value of our test.

Step 1. (i) Consistently estimate π_ω using the constrained estimator $\widetilde{\theta}$, denoted as $\widehat{\pi}$;

(ii) Construct confidence sets for $\sqrt{n}\left(\Gamma\eta_n^u - r_\Gamma\right)$. By definition, $\eta_n^u = R_\Gamma^u\theta_n$. Denote $\widehat{\theta}_{OLS}$ as the OLS estimator of θ_n. Since $\sqrt{n}\left(\widehat{\theta}_{OLS} - \theta_n\right) \xrightarrow{d} \mathcal{N}\left(0, \Lambda\right)$ for some covariance matrix

Λ, it holds that

$$\sqrt{n}\left(R_{\Gamma}^{u}\widehat{\theta}_{OLS}-\eta_{n}^{u}\right)\xrightarrow{d}Z\left(\Lambda\right)\sim\mathcal{N}\left(0,R_{\Gamma}^{u}\Lambda R_{\Gamma}^{u\prime}\right).$$

The confidence set $I_{\tau}\left(\eta_{n}^{u}\right)$ for η_{n}^{u} is obtained as $R_{\Gamma}^{u}\widehat{\theta}_{OLS}-\frac{1}{\sqrt{n}}ES_{\widehat{\Lambda}}\left(\tau\right)$, where $ES_{\Lambda}\left(\tau\right)$ is the set such that $Pr\left(Z\left(\Lambda\right)\in ES_{\Lambda}\left(\tau\right)\right)=1-\tau$ and $\widehat{\Lambda}$ is a consistent estimator of Λ. The confidence set I_{τ} for $\sqrt{n}\left(\Gamma\eta_{n}^{u}-r_{\Gamma}\right)$ is calculated as

$$I_{\tau}=\left\{\varsigma\in\mathbb{R}_{\geq0}^{l_{\Gamma}}:\varsigma=\sqrt{n}\left(\Gamma\iota-r_{\Gamma}\right),\ \iota\in I_{\tau}\left(\eta_{n}^{u}\right)\right\}.$$

To ensure that I_{τ} is non-empty, $ES_{\Lambda}\left(\tau\right)$ may not be equal-tailed.

Step 2. Compute the ϑ level Bonferroni critical value CV_{n}^{b} for some $0<\tau<\vartheta$ as

$$CV_{n}^{b}\left(\vartheta,\tau\right)\equiv\sup_{c\in I_{\vartheta-\tau}}\mathcal{C}_{c,\widehat{\pi}}\left(1-\tau\right),$$

where $\mathcal{C}_{c,\pi_{\omega}}\left(1-\tau\right)$ is the $(1-\tau)$th quantile of $\left(R_{0}\Psi\right)'\left(R_{0}\Sigma_{\omega}R_{0}'\right)^{-1}\left(R_{0}\Psi\right)$ given (c,π_{ω}). For any given (c,π_{ω}), the distribution $\left(R_{0}\Psi\right)'\left(R_{0}\Sigma_{\omega}R_{0}'\right)^{-1}\left(R_{0}\Psi\right)$ may not have a closed form expression but can be simulated.

The following theorem shows that the test is asymptotic uniformly valid.

Theorem 2.4.1. *For some $\vartheta\in(0,1)$, assume that $\left(R_{0}\Psi\right)'\left(R_{0}\Sigma_{\omega}R_{0}'\right)^{-1}\left(R_{0}\Psi\right)$ is continuous at $\mathcal{C}_{c,\pi_{\omega}}\left(1-\vartheta\right)$ for all $(c,\pi_{\omega})\in\overline{\mathbb{R}}_{\geq0}^{l_{\Gamma}}\times\Pi$. Under Assumptions 2.3.1, 2.4.1, and 2.4.2, it holds that $AsySz\left(T_{n}\left(R_{0},r_{0}\right),CV_{n}^{b}\left(\vartheta,\tau\right)\right)\leq\vartheta$.*

2.5 A Simulation Study

In this section, we report results from a simulation study designed to (i) compare the finite sample performance of three estimators $\widehat{\theta}$, $\widetilde{\theta}$, and $\widehat{\theta}_{OLS}$ measured by the mean squared error (MSE), where $\widehat{\theta}_{OLS}$ denotes the OLS estimator; (ii) examine the size performance of our test referred to as the Uniform Test, and compare it with Wald test based on OLS referred to as the OLS test and a hybrid test for which the test statistic is the same as our test but the critical value is that of the OLS test. We note that the hybrid test is valid when all the parameter values are in the interior; otherwise it may not be valid.

2.5.1 Simulation Design

We generate data from the following IA-CR model with one covariate:

$$\begin{aligned}\text{mid}Y &= \alpha_m^*\text{mid}X + \alpha_s^*\text{spr}X + \text{mid}\Delta \text{ and}\\ \text{spr}Y &= \beta_m^* \left|\text{mid}X\right| + \beta_s^*\text{spr}X + \text{spr}\Delta,\end{aligned}$$

where the random interval X is defined by $\text{mid}X \sim \mathcal{N}(0,1)$, $\text{spr}X \sim \chi_1^2$, and $\text{mid}X$ is independent of $\text{spr}X$; the random generalized interval error Δ is independent of X. We consider four different DGPs: DGP1-DGP3 belong to model M_G, where Δ is a random interval; and DGP4 does not belong to model M_G.

It follows from Theorems 2.3.2 and 2.3.3 that the asymptotic distributions of the scaled and centered estimators $\widehat{\theta}$ and $\widetilde{\theta}$ are discontinuous with respect to the value of $\boldsymbol{\beta}^*$. The asymptotic distribution of $\widehat{\theta}$ further depends on the tail behavior of the distribution of $\text{spr}\Delta$ through the value of κ. To cover all the possible situations displayed in the theorems, we consider three different specifications of κ in DGP1-DGP3. In DGP4, the support of $spr\Delta$ includes negative values, so Δ and Y are random generalized intervals. In DGP1-DGP4 below, $\text{mid}\Delta$ and $\text{spr}\Delta$ are independent of each other.

DGP1 $\text{mid}\Delta \sim \mathcal{N}(0,1)$ and $\text{spr}\Delta \sim \Gamma(2,2)+1$. In this case, the distribution of $\text{spr}\Delta$ corresponds to the case when $\kappa = +\infty$.

DGP1 A: $(\alpha_m^*, \alpha_s^*, \beta_m^*, \beta_s^*) = (1,1,1,1)$; and DGP1 B: $(\alpha_m^*, \alpha_s^*, \beta_m^*, \beta_s^*) = (1,0,0,1)$.

DGP2 $\text{mid}\Delta \sim \mathcal{N}(0,1)$ and $\text{spr}\Delta \sim \Gamma(2,2)$. In this case, the distribution of $\text{spr}\Delta$ corresponds to the case when $\kappa \in (0,+\infty)$.

DGP2 A: $(\alpha_m^*, \alpha_s^*, \beta_m^*, \beta_s^*) = (1,1,1,1)$; and DGP2 B: $(\alpha_m^*, \alpha_s^*, \beta_m^*, \beta_s^*) = (1,0,0,1)$.

DGP3 $\text{mid}\Delta \sim \mathcal{N}(0,1)$ and $\text{spr}\Delta \sim \Gamma(1,2)$. The distribution of $\text{spr}\Delta$ corresponds to the case when $\kappa = 0$.

DGP3 A: $(\alpha_m^*, \alpha_s^*, \beta_m^*, \beta_s^*) = (1,1,1,1)$; and DGP3 B: $(\alpha_m^*, \alpha_s^*, \beta_m^*, \beta_s^*) = (1,0,0,1)$.

	n	$\left(\widehat{\boldsymbol{\beta}},\widehat{\delta}\right)$	$\left(\widetilde{\boldsymbol{\beta}},\widetilde{\delta}\right)$	$\left(\widehat{\boldsymbol{\beta}},\widehat{\delta}\right)_{OLS}$		n	$\left(\widehat{\boldsymbol{\beta}},\widehat{\delta}\right)$	$\left(\widetilde{\boldsymbol{\beta}},\widetilde{\delta}\right)$	$\left(\widehat{\boldsymbol{\beta}},\widehat{\delta}\right)_{OLS}$
	100	0.485	0.556	0.565		100	0.302	0.325	0.530
DGP1 A	500	0.087	0.091	0.091	DGP1 B	500	0.059	0.061	0.098
	1000	0.057	0.057	0.057		1000	0.036	0.036	0.053
	100	0.361	0.548	0.556		100	0.181	0.349	0.558
DGP2 A	500	0.082	0.116	0.116	DGP2 B	500	0.034	0.071	0.108
	1000	0.037	0.050	0.050		1000	0.016	0.034	0.053
	100	0.189	0.277	0.277		100	0.068	0.170	0.259
DGP3 A	500	0.031	0.050	0.051	DGP3 B	500	0.010	0.044	0.066
	1000	0.018	0.028	0.028		1000	0.005	0.019	0.030
	100	N/A	0.257	0.261		100	N/A	0.169	0.280
DGP4 A	500	N/A	0.048	0.048	DGP4 B	500	N/A	0.033	0.049
	1000	N/A	0.026	0.026		1000	N/A	0.015	0.026

Table 2.1: MSE Comparison

DGP4 mid$\Delta \sim \mathcal{N}(0,1)$ and spr$\Delta \sim \mathcal{N}(1,4)$.

DGP4 A: $(\alpha_m^*, \alpha_s^*, \beta_m^*, \beta_s^*) = (1,1,1,1)$; and DGP4 B: $(\alpha_m^*, \alpha_s^*, \beta_m^*, \beta_s^*) = (1,0,0,1)$.

Within each DGP, the two models differ in the value of β^* which is in the interior in model A and at the boundary in model B. DGP1-DGP3 differ in the value of κ. Moreover within each DGP in DGP1-DGP3, model B belongs to model M, where DGP1 B and DGP3 B are considered in Blanco-Fernández et al. (2012), but the tail behavior of DGP2 B is not covered in Blanco-Fernández et al. (2012). In DGP4, $spr\Delta$ has positive probability of being negative. For DGPs 1-4, $\widetilde{\theta}$ and $\widehat{\theta}_{OLS}$ are consistent estimators, whereas $\widehat{\theta}$ is consistent only for DGPs 1-3.

2.5.2 MSE of the Estimators

We focus on the comparison between estimators for $(\boldsymbol{\beta}^*, \delta^*)$ in terms of MSE for the aforementioned data set configurations, since the major difference between the three estimators lies in the constraints imposed in the minimization problem and the estimator vector for the parameter $(\boldsymbol{\alpha}^*, \gamma^*)$ is not affected by the constraints. The number of repetitions is 5000.

Table 2.1 compares the MSE between different models under different data configurations.

For all cases considered, as the sample size increases, the MSE of each estimator decreases and the most reduction in MSE occurs from $n = 100$ to $n = 500$. At any given sample size, for model M_G, which corresponds to DGP1-DPG3, $\left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)$ is the most efficient estimator among the three. More importantly, the fatter the tail distribution of sprΔ is, the more efficient the estimator $\left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)$ is compared to $\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\delta}\right)$ and $\left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)_{OLS}$. The distribution of sprΔ verifies $\kappa = +\infty$ in DGP1, which corresponds to a thin tailed distribution; the value $\kappa = 0$ in DGP3 implies a fat tailed distribution. The tail distribution of sprΔ in DGP2 is in the middle, such that $\kappa \in (0, +\infty)$. For DGP1, the estimator vector $\left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)$ is slightly better than $\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\delta}\right)$ when the sample size is small and the two are the same when the sample size is large. On the other hand, the MSE for $\left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)$ is smaller than $\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\delta}\right)$ in DGP2, especially when $\beta_m^* = 0$. This phenomenon is more pronounced in DGP3 when the tail distribution of sprΔ is fat. The MSE for $\left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)$ is almost half of that for $\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\delta}\right)$ in DGP3 A, and is less than a quarter in DGP3 B. The constraint that spr$Y_i - |\boldsymbol{X}_i|' \boldsymbol{\beta} \geq 0$ for $i = 1, \ldots, n$ provides little information on the parameter $\boldsymbol{\beta}$ when sprΔ has a relatively small probability of being close to zero. During the computation of the minimization problem, the constraint almost never binds in this case. On the other hand, if sprΔ is close to zero with a relatively high probability, the constraint binds for some $i \in 1, \ldots, n$. This improves the accuracy of the estimator vector $\left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)$, because the true parameter vector satisfies the constraint. Section 3 provides more detailed discussion about the effect of the tail of the distribution of sprΔ on the asymptotic property of $\left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)$. For the same reason, the MSEs for estimator vectors $\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\delta}\right)$ and $\left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)_{OLS}$ in DGP1 A-DGP4 A are close to each other, whereas $\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\delta}\right)$ is significantly more efficient than $\left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)_{OLS}$ when some of the parameters are at the boundary, as in DGP1 B-DGP4 B. When the correct constraint binds during the estimation, the accuracy of the constrained estimator improves.

2.5.3 Size of the Tests

We aim to study the size of the test presented in Section 2.4 when the null hypothesis involves range regression and the nuisance parameters are either at the discontinuity point or not.

For each DGP, we consider testing model A against model B, i.e., testing $H_0 : (\alpha_s^*, \beta_m^*) = \mathbf{0}$ against $H_1 : (\alpha_s^*, \beta_m^*) \neq \mathbf{0}$ under the maintained hypothesis that $(\boldsymbol{\beta}^{*\prime}, \delta^*) \geq \mathbf{0}$ using one of the three tests stated at the beginning of this section. They are the test in Section 2.4.1 denoted as Uniform Test, the test based upon OLS estimators denoted as OLS Test, and Hybrid Test which uses the statistic T_n but the critical value of the OLS Test. The test statistic for the OLS Test takes the same form as (2.16) with $\left(\widetilde{\alpha}_s', \widetilde{\beta}_m'\right)$ replaced by $\left(\widehat{\alpha}_{s,OLS}', \widehat{\beta}_{m,OLS}'\right)$. Because the critical value in the Uniform Test is computationally costly, the Hybrid Test is included in the study to examine finite sample properties of the test based upon T_n and incorrect but simple critical value. For the Uniform and Hybrid Test, Σ_n equals to an estimator of

$$\begin{pmatrix} P_{xx}^{-1} & \mathbf{0} \\ \mathbf{0} & Q_{xx}^{-1} \end{pmatrix} \Lambda \begin{pmatrix} P_{xx}^{-1} & \mathbf{0} \\ \mathbf{0} & Q_{xx}^{-1} \end{pmatrix}$$

using $\widetilde{\theta}$, while Σ_n for the OLS test is calculated with $\widehat{\theta}_{OLS}$. When implementing Uniform Test, the confidence set I_τ for c is based on the result that

$$\widehat{c} - c = \sqrt{n}\left(\widehat{\beta}_{s,OLS} - \beta_{s,n}, \widehat{\delta}_{OLS} - \delta_n\right)' \xrightarrow{d} Z_2,$$

where Z_2 is a bivariate Normal distribution with a consistently estimable variance. The Bonferroni $CV_n^b(\vartheta, \tau)$ for some $0 < \tau < \vartheta$ is defined in (2.17). Following Romano et al. (2014) and McCloskey (2017), the tuning parameter τ is set as $\vartheta - \vartheta/10$. The number of repetitions is 1000. We refer interested readers to Romano et al. (2014) and McCloskey (2017) for a general discussion of the choice of τ.

In Table 2.2, we report size performance of each test. For DGP1-DGP3, Uniform Test has more accurate size than OLS Test, whereas the Hybrid Test has the worst size performance. The asymptotic distribution of the test statistic for the OLS Test does not approximate the finite sample distribution well when the true distribution of the error is skewed. On the contrary, the test statistic in Uniform Test makes use of the constrained estimator, whose limiting distribution is truncated Gaussian. This alleviates the skewness issue and improves the finite sample performance of the test. Even in DGP4, where the distributions are indeed

		$n = 500$			$n = 1000$			$n = 5000$		
	Size	Uniform	OLS	Hybrid	Uniform	OLS	Hybrid	Uniform	OLS	Hybrid
	10%	8.4	3.8	3.2	8.6	4	3.8	8.8	7.0	4.8
DGP 1	5%	3.4	2.3	1.7	3.7	2.8	1.6	4.0	3.7	2.2
	1%	0.3	0.3	0.1	0.8	0.4	0.1	0.9	0.6	0.3
		$n = 500$			$n = 1000$			$n = 5000$		
	Size	Uniform	OLS	Hybrid	Uniform	OLS	Hybrid	Uniform	OLS	Hybrid
	10%	8.1	4.3	3.6	8.4	4.3	3.7	8.7	6.9	4.0
DGP 2	5%	3.4	2.8	1.2	3.7	2.9	1.5	4.3	3.6	2.1
	1%	0.5	0.2	0.2	0.7	0.3	0.2	0.8	0.6	0.3
		$n = 500$			$n = 1000$			$n = 5000$		
	Size	Uniform	OLS	Hybrid	Uniform	OLS	Hybrid	Uniform	OLS	Hybrid
	10%	7.9	5.3	3.1	8.2	5.7	3.7	8.7	7.4	4.9
DGP 3	5%	3.2	3.0	1.2	3.5	3.1	1.4	4.4	3.9	2.1
	1%	0.5	0.4	0.1	0.7	0.5	0.2	0.8	0.7	0.3
		$n = 500$			$n = 1000$			$n = 5000$		
	Size	Uniform	OLS	Hybrid	Uniform	OLS	Hybrid	Uniform	OLS	Hybrid
	10%	8.9	9.0	3.8	8.9	9.1	4.7	9.0	9.2	8.1
DGP 4	5%	4.2	4.1	1.6	4.2	4.2	2.0	4.4	4.5	2.2
	1%	0.7	0.8	0.2	0.8	0.8	0.2	0.9	0.9	0.4

Table 2.2: Size Performance-Reject Percentage

Gaussian, Uniform Test has comparable size performance to OLS test.

2.6 Concluding Remarks

We have made several contributions in this paper. First, we have proposed a flexible model, i.e., the GIA-CR model for random intervals via the generalized interval arithmetic approach and a constrained estimator of parameters in the GIA-CR model. As a special member of the Generalized model, the IA-CR model extends and overcomes the drawbacks of both model M_G and the CR model. Second, as a measure of goodness-of-fit, we have extended the coefficient of determination for multiple linear regressions to our GIA-CR model for random intervals. Third, we have developed asymptotically uniformly valid tests for linear hypotheses in the GIA-CR model including a test for the correct specification of the CR model against the IA-CR model. Fourth, we have conducted a simulation study to examine the finite sample performance of our estimator and test. As a separate contribution to the current literature on interval arithmetic approach to modeling interval data, we have established the asymptotic distribution of the constrained estimator of model M_G.

BIBLIOGRAPHY

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.

ANDREWS, D. W. AND P. GUGGENBERGER (2010): "Asymptotic Size and a Problem with Subsampling and with the m out of n Bootstrap," *Econometric Theory*, 426–468.

AUMANN, R. J. (1965): "Integrals of Set-Valued Functions," *Journal of Mathematical Analysis and Applications*, 12, 1–12.

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.

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.

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79, 1785–1821.

BERESTEANU, A. AND F. MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814.

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.

BERTOLUZZA, C., N. CORRAL BLANCO, AND A. SALAS (1995): "On a New Class of Distances Between Fuzzy Numbers," *Mathware & Soft Computing*, 2, 71–84.

BLANCO-FERNÁNDEZ, A., A. COLUBI, AND G. GONZÁLEZ-RODRÍGUEZ (2012): "Confidence Sets in a Linear Regression Model for Interval Data," *Journal of Statistical Planning and Inference*, 142, 1320–1329.

BLANCO-FERNÁNDEZ, A., N. CORRAL, AND G. GONZÁLEZ-RODRÍGUEZ (2011): "Estimation of a Flexible Simple Linear Model for Interval Data based on Set Arithmetic," *Computational Statistics & Data Analysis*, 55, 2568–2578.

BLANCO-FERNÁNDEZ, Á., M. GARCÍA-BÁRZANA, A. COLUBI, AND E. J. KONTOGHIORGHES (2015): "Multiple Set Arithmetic-Based Linear Regression Models for Interval-Valued Variables," Unpublished Working Paper, Department of Statistics and OR and Mathematics Didactics, University of Oviedo.

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. (2005): "Extremal Quantile Regression," *Annals of Statistics*, 806–839.

Chernozhukov, V. and I. Fernández-Val (2011): "Inference for Extremal Conditional Quantile Models, with an Application to Market and Birthweight Risks," *The Review of Economic Studies*, 78, 559–589.

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 (2002): "Likelihood Inference in a Class of Nonregular Econometric Models," Working Paper No. 02-05, Department of Economics, MIT.

——— (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.

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

Fan, Y. and X. Shi (2019): "Testing When Parameters are Subject to Linear Inequality Constraints," Working Paper, Department of Economics, University of Washington.

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.

Fréchet, M. (1948): "Les Éléments Aléatoires de Nature Quelconque dans un Espace Distancié," *Ann. Inst. H. Poincaré*, 10, 215–310.

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

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.

Golan, A. and A. Ullah (2017): "Interval estimation: An information theoretic approach," *Econometric Reviews*, 36, 781–795.

González-Rivera, G. and W. Lin (2013): "Constrained Regression for Interval-Valued Data," *Journal of Business & Economic Statistics*, 31, 473–490.

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.

Han, A., Y. Hong, and S. Wang (2012): "Autoregressive Conditional Models for Interval-Valued Time Series Data," Unpublished Manuscript, Department of Economics, Cornell Univerisity.

Han, A., H. Yongmiao, K. K. Lai, and W. Shouyang (2008): "Interval Time Series Analysis with an Application to the Sterling-Dollar Exchange Rate," *Journal of Systems Science and Complexity*, 21, 558–573.

Henk, M., J. Richter-Gebert, and G. M. Ziegler (2004): "Basic Properties of Convex Polytopes," *Handbook of Discrete and Computational Geometry*, 255–382.

Hukuhara, M. (1967): "Integration des Applications Mesurables dont la Valeur est un Compact Convexe," *Funkcialaj Ekvacioj*, 10, 205–223.

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.

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.

Kallenberg, O. (1983): *Random Measures*, Berlin: Akademie-Verlag.

Kaucher, E. (1980): "Interval Analysis in the Extended Interval Space IR," *Computing Supplementum*, 2, 33–49.

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.

——— (2001): "Limiting Distributions of Linear Programming Estimators," *Extremes*, 4, 87–103.

——— (2006): "Asymptotic Theory for M-Estimators of Boundaries," in *Sperlich S., Härdle W., Aydinli G. (eds) The Art of Semiparametrics*, Heidelberg: Physica-Verlag, 1–21.

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

Körner, R. (1997): "On the Variance of Fuzzy Random Variables," *Fuzzy Sets and Systems*, 92, 83–93.

Körner, R. and W. Näther (2002): "On the Variance of Random Fuzzy Variables," in *Bertoluzza C., Gil MÁ., Ralescu D.A. (eds) Statistical Modeling, Analysis and Management of Fuzzy Data*, Heidelberg: Physica, 25–42.

Leadbetter, M. R., G. Lindgren, and H. Rootzén (1987): *Extremes and Related Properties of Random Sequences and Processes*, Berlin: Springer-Verlag.

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

Manski, C. F. and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546.

Markov, S. (1996): "On Directed Interval Arithmetic and its Applications," *The Journal of Universal Computer Science*, 514–526.

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.

Neto, E. d. A. L. and F. d. A. de Carvalho (2010): "Constrained Linear Regression Models for Symbolic Interval-Valued Variables," *Computational Statistics & Data Analysis*, 54, 333–347.

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.

Resnick, S. I. (1987): *Extreme Values, Regular Variation, and Point Processes*, New York: Springer.

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

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

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.

Trutschnig, W., G. González-Rodríguez, A. Colubi, and M. Á. Gil (2009): "A New Family of Metrics for Compact, Convex (Fuzzy) Sets based on a Generalized Concept of Mid and Spread," *Information Sciences*, 179, 3964–3972.

Van der Vaart, A. W. (2000): *Asymptotic Statistics*, New York: Cambridge University Press.

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.

Appendix A

WALD, QLR, AND SCORE TESTS WHEN PARAMETERS ARE SUBJECT TO LINEAR INEQUALITY CONSTRAINTS

A.1 Technical Proofs

We introduce several notations and definitions that will be used to prove Lemma 2.4.1. 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 two sets: $\Lambda_n \equiv \left\{\theta \in \mathbb{R}^l : \theta_{1,b} \geq \mathbf{0} \text{ and } \Upsilon\theta_2 + \Upsilon b_n \theta_{2,n} \geq \mathbf{0}\right\}$ and $\Lambda \equiv \left\{\theta \in \mathbb{R}^l : \theta_{1,b} \geq \mathbf{0} \text{ and } \Upsilon\theta_2 + \Upsilon c \geq \mathbf{0}\right\}$. 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 $\mathit{\Gamma} \subset \mathbb{R}^l$ and $z \in \mathbb{R}^l$, define $\operatorname{dist}(z, \mathit{\Gamma}) \equiv \inf_{\lambda \in \mathit{\Gamma}} \|z - \lambda\|$, and $\operatorname{dist}_n(z, \mathit{\Gamma}) \equiv \inf_{\lambda \in \mathit{\Gamma}} \left((z-\lambda)' \mathscr{T}_n (z - \lambda)\right)^{1/2}$.

Proof of Lemma 1.4.1: From Lemma 1.2.1, we have $b_n \left(\widehat{\theta} - \theta^*\right) \xrightarrow{d} \Psi$. By the definition of W_n and Assumption 1.4.1, the result of the lemma follows. □

Lemma A.1.1. *If $Z_n = O_p(1)$ and Assumption 1.4.3 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, $b_n(\Theta - \theta_n)$ 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

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

Thus, it suffices to show that $\Pr(\lambda_n \in \Lambda_n \backslash b_n(\Theta-\theta_n)) \to 0$. 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,

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

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

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

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

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

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

Lemma A.1.2. *For any $\omega_n \in \mathcal{W}_0$, if Assumptions 1.4.2-1.4.4 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 1.4.4 implies that $\widehat{\theta} - \theta_\omega = \left(\widehat{\theta} - \theta_n\right) + (\theta_n - \theta_\omega) = o_p(1)$. Together with Assumption 1.4.2, we obtain that

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

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

$$\begin{aligned} q_n\left(b_n\left(\widehat{\theta}_q-\theta_n\right)\right) &= \inf_{\theta\in\Theta}q_n\left(b_n\left(\theta-\theta_n\right)\right)+o_p\left(1\right) \\ &= \inf_{\lambda\in b_n\left(\Theta-\theta_n\right)}q_n\left(\lambda\right)+o_p\left(1\right). \end{aligned} \tag{A.2}$$

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

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

By the triangle inequality and Assumption 1.4.3, we have

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

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

$$l_n\left(\widehat{\theta}_q\right)=l_n\left(\theta_n\right)+\frac{1}{2}Z_n'\mathscr{T}_nZ_n-\frac{1}{2}q_n\left(b_n\left(\widehat{\theta}_q-\theta_n\right)\right)+o_p\left(1\right). \tag{A.3}$$

Combing Equation (A.1) and (A.3), and the definitions of $\widehat{\theta}$ and $\widehat{\theta}_p$ in Equation (1.2) and (A.2), it holds that

$$o_p\left(1\right)\le l_n\left(\widehat{\theta}\right)-l_n\left(\widehat{\theta}_q\right)=\frac{1}{2}q_n\left(b_n\left(\widehat{\theta}_q-\theta_n\right)\right)-\frac{1}{2}q_n\left(b_n\left(\widehat{\theta}-\theta_n\right)\right)+o_p\left(1\right)\le o_p\left(1\right).$$

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

Lemma A.1.3. *For any $\omega_n\in\mathcal{W}_0$, if Assumption 1.4.3 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 the definition of the quadratic function, it holds that $\inf_{\lambda\in b_n\left(\Theta-\theta_n\right)}q_n\left(\lambda\right)=$

$\text{dist}_n^2(Z_n, b_n(\Theta - \theta_n))$ and $\inf_{\lambda\in\Lambda} q_n(\lambda) = \text{dist}_n^2(Z_n, \Lambda)$. Since $Z_n = O_p(1)$ by Assumption 1.2.2, Lemma A.1.1 provides that $\text{dist}_n^2(Z_n, b_n(\Theta - \theta_n)) = \text{dist}_n^2(Z_n, \Lambda_n) + o_p(1)$. There exists some $\lambda_n \in \Lambda_n$ such that $\text{dist}_n(Z_n, \Lambda_n) = \text{dist}_n(Z_n, \{\lambda_n\}) + o_p(1)$. 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 $d_H(\Lambda_n, \Lambda) \le \|\Upsilon b_n\theta_{2,n} - \Upsilon c\|$. 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 1.4.3 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, we have

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

Similarly, we have $\text{dist}_n(Z_n, \Lambda_n) - \text{dist}_n(Z_n, \Lambda) \le o_p(1)$. Therefore, $\text{dist}_n(Z_n, \Lambda_n) - \text{dist}_n(Z_n, \Lambda) = o_p(1)$, and the lemma follows. □

Lemma A.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 1.4.2-1.4.4 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 A.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 $b_n\left(\widehat{\theta} - \theta_n\right) \in b_n(\Theta - \theta_n) \subseteq \Lambda_n$. It remains to show that $\left\|\lambda^* - \widehat{\lambda}\right\| = o_p(1)$. The rest follows from the analogous argument in the proof for Theorem 2 in Andrews (1997). All the required conditions are provided in Assumptions 1.4.2-1.4.4, Lemma A.1.2 and A.1.3. □

Proof of Lemma 2.4.1: Since Λ is convex and closed, we can write $\widehat{\lambda}$ defined in Lemma A.1.4 as $\widehat{\lambda} = \min_{\lambda\in\Lambda} q_n(\lambda) \equiv h(b_n^{-1}Dl_n(\theta_n), \mathscr{T}_n)$. By Assumption A.1.4, $(b_n^{-1}Dl_n(\theta_n), \mathscr{T}_n) \xrightarrow{d} (G_\omega, \mathscr{T}_\omega)$ for any drifting parameter sequences $\omega_n \in \mathcal{W}_0$ with limit $\omega \in \mathcal{W}_0$. The continuity of $h(\cdot,\cdot)$ with respect to the two arguments implies that

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

Applying Lemma A.1.4, we obtain that

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

If further Assumption 1.4.5 holds, then $W_n \xrightarrow{d} (R\Psi_{W,\omega})' (R\Sigma_\omega R')^{-1} (R\Psi_\omega)$ by the continuous mapping theorem. □

Proof of Theorem 1.4.1: We prove the theorem by verifying assumptions in McCloskey (2017). Notice that the distribution $(R\Psi_\omega)' (R\Sigma_{W,\omega} R')^{-1} (R\Psi_\omega)$ is finite with probability 1 for all $c \in \overline{\mathbb{R}}_{\geq 0}^{l-J}$ and $\pi_{W,\omega} \in \overline{\Pi}_W$. Assumptions PS, Sel, and Inf in McCloskey (2017) is trivially satisfied. By the expression in Lemma 2.4.1, $\mathcal{C}^W_{c,\pi_{W,\omega}}(1-\alpha)$ is continuous in c and $\pi_{W,\omega}$. Together with the assumption in Theorem 1.4.1, 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 $\widetilde{ES}(\tau)$ satisfies that $\lim_{n\to\infty}\Pr_\omega\left(\mathscr{T}_{2,\omega}^{-1}G_{2,\omega} \in \widetilde{ES}(\tau)\right) \geq 1-\tau$. This and the fact that $c \in \overline{\mathbb{R}}_{\geq 0}^{l-J}$ imply that $\lim_{n\to\infty}\Pr_\omega\left(c \in \widetilde{I}_\tau\right) \geq 1-\tau$, which fulfills Assumption CS in McCloskey (2017). It suffices to prove that Assumption DS in McCloskey (2017) is satisfied.

Lemma 2.4.1 provides that the asymptotic distribution of the test statistic W_n is $(R\Psi_\omega)' (R\Sigma_{W,\omega} R')^{-1} (R\Psi_\omega)$ under the full parameter sequence $(\eta_n^u, \pi_{W,n}, \xi_n)$. Since $\widetilde{\theta}_2$ is an unconstrained estimator by definition, its asymptotic distribution can be obtained by applying Lemma 2.4.1 with $q_\omega(\lambda) = \left(\lambda - \mathscr{T}_{2,\omega}^{-1}G_{2,\omega}\right)' \mathscr{T}_{2,\omega}\left(\lambda - \mathscr{T}_{2,\omega}^{-1}G_{2,\omega}\right)$ and $\phi_\omega(\lambda) = 0$ 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. $\square$

Proof of Theorem 1.4.2: It holds that $\widehat{\theta}_1-r=\widehat{\theta}_1-\theta_1^*+\theta_1^*-r$, with $\widehat{\theta}_1-\theta_1^*=o_p(1)$ by Assumption 1.2.3 and $\theta_1^*-r\neq\mathbf{0}$ by H_1. Since $R\Sigma_W R'$ is positive definite with probability one by Assumption 1.4.1, $(R\Sigma_W R')^{-1}$ is also positive definite with probability one. Therefore, it holds that $b_n^{-2}W_n\xrightarrow{p}(\theta_1^*-r)'(R\Sigma_W R')^{-1}(\theta_1^*-r)>0$ 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 2.4.1 and any $\pi_{W,\omega}\in\overline{\Pi}_W$, we have

$$\left\|\mathscr{I}_\omega^{1/2}\Psi_\omega-\mathscr{I}_\omega^{1/2}Z_\omega\right\|=q_\omega(\Psi_\omega)\leq q_n(\mathbf{0})=Z_\omega'\mathscr{I}_\omega Z_\omega=O_p(1).$$

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 (2.17) 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 A.1.5. *For $\mathscr{R}_w^{nb}$ defined in Section 1.3.1, 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 1.4.3: For c defined in (1.12), 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 Lemmas A.1.1-A.1.4.

Decompose $\mathscr{R}_w\theta \geq b_n(\mathrm{r}_w - \mathscr{R}_w\theta_n)$ based the procedure in Section 1.3.1; define r_w^{nb} and

r_w^b as the corresponding subvector of r_w. We obtain that

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

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 A.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

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

Thus, we have identified the equality constraints, binding inequality constraints, non-binding inequality constraints and undetermined inequality constraints with the associated limits. The same derivation in Lemmas A.1.1-A.1.4 applies to $b_n(\Theta - \theta_n)$, Λ_n and Λ. The claimed result follows. □

Proof of Theorem 1.4.3: The proof is the same as the one of Theorem 1.4.1. □

Proof of Theorem 1.4.4: 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 1.4.1, 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 Theorem 1.4.2, the theorem holds. □

Proof of Lemma 1.5.1: The proof follows from Theorem 2 in Andrews (1999) and Theorem 1 and Theorem 3 in Andrews (2001). By Assumptions 1.2.1, 1.2.3 and 1.5.1, and Theorem 2 in Andrews (1999), we have

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

Assumption 1.2.2, Theorem 1 and Theorem 3 in Andrews (2001) provide that

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

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

Proof of Lemma 1.5.2: By Assumptions refAssump::Quadratic Expansion 1-Local, 1.4.4 and 1.5.2, and Theorem 2 in Andrews (1999), it holds that

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

under any $\omega_n\in\mathcal{W}_0$. Applying the same argument in proof of Lemma 1.4.3 to $b_n(\Theta-\theta_n)$ and similar argument to $b_n(\Theta_0-\theta_n)$, 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 1.4.3. We obtain that $-2\left(l_n\left(\widehat{\theta}_0\right)-l_n\left(\widehat{\theta}\right)\right)\xrightarrow{d}\min_{\lambda\in\Lambda_0} q(\lambda)-\min_{\lambda\in\Lambda} q(\lambda)$, and the lemma follows. □

Proof of Theorem 1.5.1: The proof is similar to the proofs of Theorems 1.4.1 and 1.4.3, with the asymptotic distribution of QLR_n under $\omega_n \in \mathcal{W}_0$ being provided in Lemma 1.4.3.

□

Proof of Theorem 1.5.2: Assumptions 1.2.3 and 1.5.1 imply that $\widehat{\theta} \xrightarrow{p} \theta^*$ and $\widehat{\theta}_0 \xrightarrow{p} \theta_0^*$ under H_1. By Assumptions 1.2.1 and 1.2.2, we obtain that $l_n(\cdot)$ is continuous at θ^*. Together with the assumption on the continuity of $l_n(\cdot)$ 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 1.5.2 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 1.6.1: Applying Equation (1.5), we obtain that

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

Assumptions 1.2.2 and 1.5.1 imply that $b_n^{-1} D l_n(\theta^*) = O_p(1)$ and

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

Since $b_n\left(\widehat{\theta}_0 - \theta^*\right) = O_p(1)$ by Assumption 1.5.1, 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 1.6.1 (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

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

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 1.6.1 (b) and $\mathscr{T}_n \xrightarrow{d} \mathscr{T}$ for $\mathscr{T}$ being non-singular with probability one by Assumption 1.2.2. The definition of $\widehat{\theta}_0$ implies that $R\widehat{\theta}_0 = r$. Therefore, it holds that

$$\begin{aligned}
& R\widehat{\mathscr{T}_n}^{-1} b_n^{-1} Dl_n\left(\widehat{\theta}_0\right) \\
=\ & R\mathscr{T}_n^{-1} b_n^{-1} Dl_n\left(\widehat{\theta}_0\right) + o_p(1) \\
=\ & R\left(\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*) + \mathscr{T}_n^{-1} b_n^{-1} D^2 l_n(\theta^*)\left(\widehat{\theta}_0 - \theta^*\right) + \mathscr{T}_n^{-1} b_n^{-1} R_n^D\left(\widehat{\theta}_0\right)\right) + o_p(1) \\
=\ & R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*) - b_n R\left(\widehat{\theta}_0 - \theta^*\right) + o_p(1) \\
=\ & R\mathscr{T}_n^{-1} b_n^{-1} Dl_n(\theta^*) + o_p(1),
\end{aligned} \tag{A.4}$$

where the second to last equality comes from the definition of $\mathscr{T}_n$. Since we have that $\left(R\widehat{\mathscr{T}_n}^{-1} R'\right)^{-1/2} R\widehat{\mathscr{T}_n}^{-1} b_n^{-1} Dl_n\left(\widehat{\theta}_0\right) = O_p(1)$ by Assumption 1.2.2 and 1.6.1, 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 (A.4) and Assumption 1.6.1 (ii) imply $\widehat{q}_R(ds) = q_{n,R}(ds) + o_p(1)$. Applying the same proof as for Theorem 1 (e) in Andrews (1997), we obtain that

$$\widehat{q}_R(ds) = \inf_{\lambda \in b_n(R\Theta - r)} q_{n,R}(\lambda) + o_p(1). \tag{A.5}$$

The set $R\Theta - r$ has the halfspace description (**??**), 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

$$\inf_{\lambda \in b_n(R\Theta - r)} q_{n,R}(\lambda) = \min_{\lambda \in \Lambda_R} q_{n,R}(\lambda) + o_p(1), \tag{A.6}$$

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 1.6.2. □

Proof of Theorem 1.6.1: 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 1.5.2, $b_n\left(\widehat{\theta}_0-\theta_n\right)=O_p(1)$. 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 1.6.3 (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 1.6.1. Then, Equation A.4 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 1.2.2.

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 1.6.1. Therefore, $\mathcal{C}_{S,\pi}^S(1-\alpha)$ 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 $\mathcal{C}_{\pi_S}^S(1-\alpha)$ in π by Assumption 1.3.1 and $\widehat{\pi}_S\xrightarrow{p}\pi_S$ imply that $\mathcal{C}_{\widehat{\pi}_S}^S(1-\alpha)\xrightarrow{p}\mathcal{C}_{\pi_S}^S(1-\alpha)$ 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 $\mathcal{C}_\pi^S(1-\alpha)$ provide that $\lim_{n\to\infty}\Pr_{\omega_n}\left(S_n>\mathcal{C}_{\widehat{\pi}_S}^S(1-\alpha)\right)=\alpha$, for any $\omega_n\in\mathcal{W}_0$. Therefore, the theorem follows. $\square$

Proof of Theorem 1.6.2: By Assumptions 1.5.1 and 1.6.1, under H_1, the same argument in the proof of Lemma 1.6.1 implies that

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

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

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

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

$$b_n^{-1}ds_n = b_n^{-1}ds_s + o_p\left(1\right) = \upsilon\left(R\theta^* - r\right) + o_p\left(1\right). \tag{A.7}$$

Assume the contrary. Then it holds that $b_n^{-1}ds_n - \upsilon\left(R\theta^* - r\right) \xrightarrow{p} \varrho \neq \mathbf{0}$, which implies that

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

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 (1.6), which contradicts its definition. Applying Equation

(A.7) to the definition of S_n, we have

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

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 1.7.1: (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 1.4.3, with

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

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

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

(ii) The proof follows from the same argument for the proof of Lemma 1.5.2. Notice that $c = \lim_{n\to\infty} -b_n\left(\mathrm{r}_w^u - \mathscr{R}_w^u \theta_n\right)$ by definition. We have

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

Applying the Λ and Λ_0 defined here to the proof of Lemma 1.5.2, we obtain the result.

(iii) The proof mainly follows the one of Lemma 1.6.1 with the following modification. First, by Assumption 1.7.1 (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 1.7.1 (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 1.6.1. Second, Equation (A.4) 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 1.7.1 (ii), (vi) and (vii). □

Proof of Corollary 1.7.1: 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

$$b_n\Gamma^u\mathscr{R}_\Gamma^u\widetilde{\theta}_{f,n}-(\mathrm{r}_w^u-\mathscr{R}_w^u\gamma)\xrightarrow{d}c+\Gamma^u\mathscr{R}_\Gamma^u\mathscr{T}_{f,\omega}^{-1}G_{f,\omega}, \tag{A.8}$$

where the random vector $G_{f,\omega}$ is the subvector of G_ω corresponding to θ_f. Therefore, $\sup_{c\in\widetilde{I}_{\alpha-\tau}}\mathcal{C}_{c,\pi_{W,\omega}}^W(1-\tau)\xrightarrow{d}\sup_{c\in I_{\alpha-\tau}}\mathcal{C}_{c,\pi_{W,\omega}}^W(1-\tau)$ for any $\pi_{W,\omega}$, because W_ω is continuous at $\mathcal{C}_{c,\pi_{W,\omega}}^W(1-\tau)$ for all $c\in C$. Moreover, $\mathcal{C}_{c,\pi_{W,\omega}}^W(1-\tau)$ 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 (A.8) is jointly with $(b_n^{-1}Dl_n(\theta^*),\mathscr{T}_n)\xrightarrow{d}(G,\mathscr{T})$, part (i) of the corollary follows. Similar arguments apply to part (ii). Because $\mathcal{C}_{\pi_{S,\omega}}^S(1-\alpha)$ is continuous in $\pi_{S,\omega}$ and $\widehat{\pi}_S$ consistently estimates $\pi_{S,\omega}$, we have $\mathcal{C}_{\widehat{\pi}_S}^S(1-\alpha)\xrightarrow{p}\mathcal{C}_{\pi_{S,\omega}}^S(1-\alpha)$. Part (iii) therefore holds. □

Lemma A.1.6. *Result in Lemmas 1.2.1, 1.4.1, 1.4.3, 1.5.1, 1.5.2 and 1.7.1 is independent of the description of* Θ*; result in Lemma 1.6.1 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 1.2.1 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 1.2.1 is independent of $\Theta(1)$ and $\Theta(2)$. For Lemma 1.4.3, 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 1.4.3 doesn't depend on the description of Θ. Result in other lemmas follows from the similar argument. □

A.2 Verification of Assumptions for Linear Regression Model

We present primitive conditions for Assumptions 1.4.2-1.4.4, 1.5.2, and 1.6.3 to hold in the linear regression model. Since the two running examples are both linear regression models, the conditions discussed in this section can be directly applied. Let the model be indexed by n: $Y_{ni} = X_{ni}'\theta^* + \varepsilon_{ni}$ with $E(\varepsilon_{ni} \mid X_{ni}) = 0$. 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:

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

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 1.4.2 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 1.4.3 holds. Assumptions 1.4.4 and 1.5.2 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 1.4.4 and 1.5.2. Such tool is powerful in dealing with estimators defined by constrained optimizations. At last Assumption 1.6.3 (i) holds by $R_n^D(\cdot) = 0$; and 1.6.3 (ii) is satisfied by letting $\widehat{\mathscr{T}_n} = \mathscr{T}_n = 2/n \sum_{i=1}^n X_{ni}X_{ni}'$.

Appendix B

UNIFORM INFERENCE IN A GENERALIZED INTERVAL ARITHMETIC CENTER AND RANGE LINEAR MODEL

B.1 Technical Proofs

Proof of Proposition 2.2.1: Using the alternative expression of RSS_G, $\mathcal{R}_G^2$ can be computed as

$$\mathcal{R}_G^2 = 1 - \frac{\sum_{i=1}^n d_\lambda^2\left(Y_i, \widetilde{Y}_i\right)}{\sum_{i=1}^n d_\lambda^2\left(Y_i, \overline{Y}\right)}.$$

Non-negativity of $\frac{\sum_i d_\lambda^2\left(Y_i, \widetilde{Y}_i\right)}{\sum_i d_\lambda^2\left(Y_i, \overline{Y}\right)}$ implies that $\mathcal{R}_G^2 \le 1$; and for $\{Y_i, X_{1i}, \ldots, X_{ki}\}_{i=1}^n$ and $\widetilde{\theta}$ such that $Y_i = \widetilde{Y}_i$ for all i, it holds that $d_\lambda^2\left(Y_i, \widetilde{Y}_i\right) = 0$ and thus $\mathcal{R}_G^2 = 1$. By (2.7) and the fact that

$$\begin{aligned}\sum_{i=1}^n d_\lambda^2\left(Y_i, \overline{Y}\right) &= \sum_{i=1}^n d_\lambda^2\left(Y_i, \left[\check{\gamma} \pm \check{\delta}\right]\right), \text{ where} \\ \left(\check{\gamma}, \check{\delta}\right)' &= \arg\min_{\gamma,\delta} \sum_{i=1}^n d_\lambda^2\left(Y_i, [\gamma \pm \delta]\right),\end{aligned}$$

it holds that

$$\begin{aligned}\sum_{i=1}^n d_\lambda^2\left(Y_i, \widetilde{Y}_i\right) &= \sum_{i=1}^n d_\lambda^2\left(Y_i, \left[\boldsymbol{X}_i'\widetilde{\boldsymbol{\alpha}} \pm |\boldsymbol{X}_i|'\,\widetilde{\boldsymbol{\beta}}\right] + \left[\widetilde{\gamma} \pm \widetilde{\delta}\right]\right) \\ &\le \sum_{i=1}^n d_\lambda^2\left(Y_i, \left[\check{\gamma} \pm \check{\delta}\right]\right) \\ &= \sum_{i=1}^n d_\lambda^2\left(Y_i, \overline{Y}\right).\end{aligned}$$

The inequality comes from property of the minimization operation: $\widetilde{Y}_i$'s are obtained by a constrained minimization problem over $\boldsymbol{\alpha}, \gamma, \boldsymbol{\beta}$ and δ, with $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\beta} = \mathbf{0}$ satisfying the constraint. Therefore, $\sum_{i=1}^{n} d_\lambda^2 \left(Y_i, \overline{Y}\right) \geq \sum_{i=1}^{n} d_\lambda^2 \left(Y_i, \widetilde{Y}_i\right)$ and $\mathcal{R}_G^2 \geq 0$, with the equality holds if $\widetilde{\boldsymbol{\alpha}}$ and $\widetilde{\boldsymbol{\beta}}$ are zero vectors. The first part of the proposition holds. The second part follows directly from the characteristics of the minimization problem that defines $\widetilde{\theta}$ in (2.7). □

Proof of Theorem 2.3.1: We first show that $\widehat{\theta} \to_p \theta^*$. Define

$$\begin{aligned} Z_1(\theta) &= \left(\boldsymbol{\alpha}^{*\prime} - \boldsymbol{\alpha}', \gamma^* - \gamma\right) P_{xx} \begin{pmatrix} \boldsymbol{\alpha}^* - \boldsymbol{\alpha} \\ \gamma^* - \gamma \end{pmatrix} + \sigma_{\mathrm{mid}\Delta}^2 \\ &\quad + \lambda \left(\boldsymbol{\beta}^{*\prime} - \boldsymbol{\beta}', \delta^* - \delta\right) Q_{xx} \begin{pmatrix} \boldsymbol{\beta}^* - \boldsymbol{\beta} \\ \delta^* - \delta \end{pmatrix} + \lambda \sigma_{\mathrm{spr}\Delta}^2 + \varphi_1(\theta), \end{aligned}$$

where $\sigma_{\mathrm{mid}\Delta}^2 = Var(\mathrm{mid}\Delta)$, $\sigma_{\mathrm{spr}\Delta}^2 = Var(\mathrm{spr}\Delta)$ and

$$\varphi_1(\theta) = \begin{cases} 0, & \text{if } Pr\left(\mathrm{spr}Y - |\boldsymbol{X}|' \boldsymbol{\beta} \geq 0\right) = 1; \text{ and } \boldsymbol{\beta} \geq 0 \\ \infty, & \text{otherwise} \end{cases}.$$

It holds that $\theta^* = \arg\min_\theta Z_1(\theta)$. First, notice that $Pr\left(\mathrm{spr}Y - |\boldsymbol{X}|' \boldsymbol{\beta}^* \geq 0\right) = 1$ and $\boldsymbol{\beta}^* \geq 0$ by the model specification. Second, by Assumption 2.3.2, both $\sigma_{\mathrm{mid}\Delta}^2$ and $\sigma_{\mathrm{spr}\Delta}^2$ are finite. Therefore, $Z_1(\theta)$ reaches its minimal value $\sigma_{\mathrm{mid}\Delta}^2 + \lambda \sigma_{\mathrm{spr}\Delta}^2$ at θ^*. At last, since P_{xx} and Q_{xx} are both non-singular by Assumption 2.3.2, they are positive definite and θ^* is the unique solution to the minimization problem. Therefore, $\theta^* = \arg\min_\theta Z_1(\theta)$.

Next we aim to show that $\arg\min_\theta Z_{1n}(\theta) \xrightarrow{p} \arg\min_\theta Z_1(\theta)$. $Z_{1n}(\theta)$ is a convex function by the convexity of the domain for which $\varphi_{1n}(\theta)$ is finite and the quadratic component in $Z_{1n}(\theta)$. Moreover, since $\arg\min_\theta Z_1(\theta)$ is unique, it suffices to show that $Z_{1n}(\theta)$ epi-converges to $Z_1(\theta)$ by the Convexity Lemma of Geyer (1996) and Knight (1999).

The finite dimensional convergence and finiteness of $Z_1(\theta)$ on an open set provide the epi-convergence, given that $Z_{1n}(\theta)$ is convex. For

$$s_0 \equiv \sup\left\{\max_{1 \leq j \leq 2k-d} \boldsymbol{b}_j : Pr\left(\mathrm{spr}Y - |\boldsymbol{X}|' \boldsymbol{b} \geq 0\right) = 1, \boldsymbol{b} \in \mathbb{R}_{\geq 0}^{2k-d}\right\},$$

if s_0 is strictly positive, one can always find an open set $O_{\boldsymbol{b}} \subset \mathbb{R}^{2k}$ such that $Pr\left(\text{spr}Y - |\boldsymbol{X}|' \boldsymbol{b} \geq 0\right) = 1$ for any $\boldsymbol{b} \in O_{\boldsymbol{b}}$. Then on the set $O_{\boldsymbol{\alpha}} \times O_{\gamma} \times O_{\boldsymbol{b}} \times O_{\delta}$, where $O_{\boldsymbol{\alpha}} \subset \mathbb{R}^{2k-d}$, $O_{\gamma} \subset \mathbb{R}$ and $O_{\delta} \subset \mathbb{R}$ are any open sets, $Z(\theta)$ is finite. We now show the finite dimensional convergence to complete the proof.

By the weak law of large numbers and the model specification, we have that

$$\frac{1}{n}\sum\left(\text{mid}Y_i - \boldsymbol{X}_i'\boldsymbol{\alpha} - \gamma\right)^2 \xrightarrow{p} \left(\boldsymbol{\alpha}^{*\prime} - \boldsymbol{\alpha}', \gamma^* - \gamma\right) P_{xx} \left(\boldsymbol{\alpha}^{*\prime} - \boldsymbol{\alpha}', \gamma^* - \gamma\right)' + \sigma^2_{\text{mid}\Delta} \text{ and}$$
$$\frac{1}{n}\sum\left(\text{spr}Y_i - |\boldsymbol{X}_i|' \boldsymbol{\beta} - \delta\right)^2 \xrightarrow{p} \left(\boldsymbol{\beta}^{*\prime} - \boldsymbol{\beta}, \delta^* - \delta\right) Q_{xx} \left(\boldsymbol{\beta}^{*\prime} - \boldsymbol{\beta}', \delta^* - \delta\right)' + \sigma^2_{\text{spr}\Delta},$$

for any pair $(\boldsymbol{\alpha}', \gamma, \boldsymbol{\beta}', \delta)$. Thus, according to Knight (2001), it suffices to show that for given $\theta^1, \ldots, \theta^m$,

$$Pr\left(\varphi_{1n}\left(\theta^1\right) = 0, \ldots, \varphi_{1n}\left(\theta^m\right) = 0\right) \longrightarrow Pr\left(\varphi_1\left(\theta^1\right) = 0, \ldots, \varphi_1\left(\theta^m\right) = 0\right)$$

when $n \to \infty$. The former probability equals to

$$\begin{aligned}
&Pr\left(\text{spr}Y_i \geq |\boldsymbol{X}_i|' \boldsymbol{\beta}^j, \text{ for } i = 1, \ldots, n \text{ and } j = 1, \ldots, m\right) \\
=&Pr^n\left(\text{spr}Y_i \geq \max_{1\leq j\leq m} |\boldsymbol{X}_i|' \boldsymbol{\beta}^j\right) \longrightarrow \begin{cases} 1, & \text{if } Pr\left(\text{spr}Y \geq \max_{1\leq j\leq m} |\boldsymbol{X}|' \boldsymbol{\beta}^j\right) = 1 \\ 0, & \text{if } Pr\left(\text{spr}Y \geq \max_{1\leq j\leq m} |\boldsymbol{X}|' \boldsymbol{\beta}^j\right) < 1 \end{cases} \\
=&Pr\left(\varphi_1\left(\theta^1\right) = 0, \ldots, \varphi_1\left(\theta^m\right) = 0\right).
\end{aligned}$$

Therefore, we can conclude that $Z_{1n}(\theta)$ epi-converges to $Z_1(\theta)$ and $\widehat{\theta} \xrightarrow{p} \theta^*$ when $s_0 > 0$.

On the other hand, if $s_0 = 0$, then $\boldsymbol{\beta}_j^* = 0$ for all $j = 1, \ldots, 2k - d$. The minimization problem can be separated into two parts with one part contains only $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\gamma}$ and the other contains only $\widehat{\boldsymbol{\beta}}$ and $\widehat{\delta}$. Since no constraints are imposed on $(\widehat{\boldsymbol{\alpha}}', \widehat{\gamma})$, the consistency of $(\widehat{\boldsymbol{\alpha}}', \widehat{\gamma})$ follows from standard arguments in the least squares estimation. Because $Z_{1n}(\boldsymbol{\alpha}', \gamma, \mathbf{0}', \delta)$ is finite, to prove that $\left(\widehat{\boldsymbol{\beta}}', \widehat{\delta}\right) \xrightarrow{p} (\boldsymbol{\beta}^{*\prime}, \delta^*) = (\mathbf{0}', \delta^*)$, it suffices to show that for any $\boldsymbol{b} \neq \mathbf{0}$, $Pr\left(Z_{1n}(\theta) < \infty\right) \to 0$ when $n \to \infty$. This follows from the fact that for $\boldsymbol{b} \neq \mathbf{0}$,

$$Pr\left(\text{spr}Y_i \geq |\boldsymbol{X}_i|' \boldsymbol{b}, \text{ for } i = 1, \ldots, n\right) = Pr^n\left(\text{spr}Y_i - |\boldsymbol{X}_i|' \boldsymbol{b} \geq 0\right) = 0,$$

where the last equality is implied by $s_0 = 0$. Therefore, $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \mathbf{0}$. The convergence of $\widehat{\delta}$ follows from the law of large numbers since $\delta^* = E\left(\text{spr}Y\right)$ when $\boldsymbol{\beta}^* = \mathbf{0}$.

The first part of the theorem then follows by combining the two cases of $s_0 > 0$ and $s_0 = 0$.

The proof for $\widetilde{\theta} \xrightarrow{p} \theta^*$ is essentially the same. Define

$$\begin{aligned} Z_2\left(\theta\right) &= \left(\boldsymbol{\alpha}^{*\prime} - \boldsymbol{\alpha}^{\prime}, \gamma^* - \gamma\right) P_{xx} \begin{pmatrix} \boldsymbol{\alpha}^* - \boldsymbol{\alpha} \\ \gamma^* - \gamma \end{pmatrix} + \sigma^2_{\text{mid}\Delta} \\ &\quad + \lambda \left(\boldsymbol{\beta}^{*\prime} - \boldsymbol{\beta}^{\prime}, \delta^* - \delta\right) Q_{xx} \begin{pmatrix} \boldsymbol{\beta}^* - \boldsymbol{\beta} \\ \delta^* - \delta \end{pmatrix} + \lambda \sigma^2_{\text{spr}\Delta} + \varphi_2\left(\theta\right), \end{aligned}$$

where $\sigma^2_{\text{mid}\Delta} = Var\left(\text{mid}\Delta\right)$ and $\sigma^2_{\text{spr}\Delta} = Var\left(\text{spr}\Delta\right)$. With Assumption 2.3.2, we have that $\theta^* = \arg\min_\theta Z_2\left(\theta\right)$, because $R\theta^* \geq r$ and P_{xx} and Q_{xx} are non-singular. The convexity of $Z_{2n}\left(\theta\right)$ is implied by its quadratic component and the geometry of the feasible set. Since $\varphi_2\left(\theta\right)$ is not random, the epi-convergence of $Z_{2n}\left(\theta\right)$ to $Z_2\left(\theta\right)$ follows from the finite dimensional convergence of the quadratic component. At last, since $Z_2\left(\theta\right)$ is finite on any open set that is contained in $\mathbb{R}^{2k-d} \times \mathbb{R} \times \mathbb{R}^{2k-d}_{\geq 0} \times \mathbb{R}_{\geq 0}$, we obtain the consistency of $\widetilde{\theta}$. □

Proof of Theorem 2.3.2: We will prove the theorem for the different cases: (i) $\kappa \in (0, +\infty)$; (ii) $\kappa = +\infty$; and (iii) $\kappa = 0$. Let $M_1\left(\psi; \lambda\right) = M\left(\psi; \lambda\right) + \phi_1\left(\psi\right)$.

Note that $\sqrt{n}\left(\widehat{\theta} - \theta^*\right)$ is the solution to the minimization problem:

$$\min_{\psi} M_{1n}\left(\psi; \lambda\right) \equiv \min_{\boldsymbol{p}, q, \boldsymbol{u}, v} \begin{bmatrix} \sum_{i=1}^n \left(\text{mid}\Delta_i - \gamma^* - \frac{1}{\sqrt{n}} \boldsymbol{X}_i^{\prime} \boldsymbol{p} - \frac{q}{\sqrt{n}}\right)^2 \\ + \lambda \sum_{i=1}^n \left(\text{spr}\Delta_i - \delta^* - \frac{1}{\sqrt{n}} \left|\boldsymbol{X}_i\right|^{\prime} \boldsymbol{u} - \frac{v}{\sqrt{n}}\right)^2 \\ - \sum_{i=1}^n \left(\text{mid}\Delta_i - \gamma^*\right)^2 - \lambda \sum_{i=1}^n \left(\text{spr}\Delta_i - \delta^*\right)^2 + \phi_{1n}\left(\psi\right) \end{bmatrix},$$

where

$$\phi_{1n}\left(\psi\right) = \begin{cases} 0, & \text{if } \sqrt{n}\text{spr}\Delta_i \geq \left|\boldsymbol{X}_i\right|^{\prime} \boldsymbol{u}, \text{ for } i = 1, 2, \ldots n; \\ & \text{and } \boldsymbol{u}_j + \sqrt{n}\boldsymbol{\beta}_j^* \geq 0, \text{ for } j = 1, \ldots, 2k - d \cdot \\ \infty, & \text{otherwise} \end{cases}$$

The goal is to show that $\arg\min_\psi M_{1n}(\psi;\lambda) \xrightarrow{d} \arg\min_\psi M_1(\psi;\lambda)$. Since the set of ψ for $\phi_{1n}(\psi)$ being finite is convex, the convexity of $M_{1n}(\psi;\lambda)$ is straightforward due to its quadratic component. Recall that by the Convexity Lemma of Geyer (1996) and Knight (1999), the following three conditions are sufficient for $\arg\min_\psi M_{1n}(\psi;\lambda) \xrightarrow{d} \arg\min_\psi M_1(\psi;\lambda)$ provided that $M_{1n}(\psi;\lambda)$ is convex: (i) $M_{1n}(\psi;\lambda)$ converges to $M_1(\psi;\lambda)$ in the finite-dimensional sense ($\xrightarrow{f.d.}$), (ii) $M_1(\psi;\lambda)$ is finite on an open set, and (iii) $M_1(\psi;\lambda)$ is uniquely minimized with probability 1.

We now prove that these three conditions are satisfied when $\lim_{t\to\infty} tF_s\left(1/\sqrt{t}\right) \in (0,+\infty)$.

Define the point process (random measure): $\upsilon_n(D) := \sum_{i=1}^{n} I\left\{\left(\sqrt{n}\text{spr}\Delta_i, |\boldsymbol{X}_i|\right) \in D\right\}$ for any Borel subsets D of $\mathbb{D} := [0,+\infty) \times \mathcal{X}$, where $\mathcal{X}$ is the support of $|\boldsymbol{X}_i|$. The point process $\upsilon_n(\cdot)$ tends in distribution with respect to the vague topology to a Poisson point process (random measure) $\upsilon(\cdot)$ in the metric space of point measure $\mathcal{M}_p(\mathbb{D})$. The limit Poisson process has the mean measure: $E[\upsilon(D)] = \int_D \frac{2}{\kappa} w g(\boldsymbol{x})\, d\mu(\boldsymbol{x})\, dw$ and can be represented by $\upsilon(D) := \sum_{i=1}^{\infty} I\left\{\left(\sqrt{\kappa} g^{-\frac{1}{2}}(\Upsilon_i)\, \Gamma_i, \Upsilon_i\right) \in D\right\}$ for all Borel subsets D of $\mathbb{D} := [0,+\infty) \times \mathcal{X}$, where $g(\cdot)$ is defined in Assumption 2.3.3, $\Gamma_i = (\mathcal{E}_1 + \ldots + \mathcal{E}_i)^{\frac{1}{2}}$ for unit mean i.i.d exponential random variables $\mathcal{E}_1, \mathcal{E}_2, \ldots$, and $\Upsilon_1, \Upsilon_2, \ldots$ are i.i.d with distribution $Pr(\Upsilon_i \in A) = \mu(A)$, where $\mu(\cdot)$ is the probability measure of $|\boldsymbol{X}_i|$. The Γ_i's are independent of Υ_i's. By Assumption 2.3.3 and the fact that $\lim_{t\to\infty} tF_s\left(1/\sqrt{t}\right) = 1/\kappa$, $\lim_{n\to\infty} E[\upsilon_n(D)] = E[\upsilon(D)]$ and $\lim_{n\to\infty} Pr\{\upsilon_n(D) = 0\} = e^{-E[\upsilon(D)]}$. The claimed weak convergence result follows from Kallenberg's theorem (Resnick (1987)).

Using the above convergence result of the point process, we are now ready to show the finite dimensional weak convergence of $M_{1n}(\psi;\lambda)$. The following convergence result is straightforward:

$$\left[\begin{array}{c}\sum_{i=1}^{n}\left(\text{mid}\Delta_i-\gamma^*-\frac{1}{\sqrt{n}}\boldsymbol{X}_i'\boldsymbol{p}-\frac{q}{\sqrt{n}}\right)^2-\sum_{i=1}^{n}\left(\text{mid}\Delta_i-\gamma^*\right)^2\\+\lambda\sum\left(\text{spr}\Delta_i-\delta^*-\frac{1}{\sqrt{n}}\left|\boldsymbol{X}_i\right|'\boldsymbol{u}-\frac{v}{\sqrt{n}}\right)^2-\lambda\sum_{i=1}^{n}\left(\text{spr}\Delta_i-\delta^*\right)^2\end{array}\right]$$
$$\xrightarrow{d}\psi'\begin{pmatrix}P_{xx}&\mathbf{0}\\\mathbf{0}&\lambda Q_{xx}\end{pmatrix}\psi-2\psi'\begin{pmatrix}\boldsymbol{I}_l&\mathbf{0}\\\mathbf{0}&\lambda\boldsymbol{I}_l\end{pmatrix}W,$$

with $W\sim\mathcal{N}(0,\Lambda)$ and Λ being the covariance matrix of $\left(\dot{\boldsymbol{X}}_i'\text{mid}\Delta_i,\left|\dot{\boldsymbol{X}}_i\right|'\text{spr}\Delta_i\right)$. The asymptotic independence between W and the point process follows from the standard proof of asymptotic independence of sample average and sample minimal order statistics, see e.g. Resnick (1987) and Lemma 21.19 in Van der Vaart (2000). A more detailed proof can be found in Chernozhukov and Hong (2002). Thus, it remains to show that for given $\psi^1,\ldots,\psi^m$,

$$Pr\left[\phi_{1n}\left(\psi^1\right)=0,\ldots,\phi_{1n}\left(\psi^m\right)=0\right]\longrightarrow Pr\left[\phi_1\left(\psi^1\right)=0,\ldots,\phi_1\left(\psi^m\right)=0\right],$$

as $n\to\infty$. Since no randomness is involved in the constraint $\boldsymbol{u}_j+\sqrt{n}\boldsymbol{\beta}_j^*\geq 0$ for $j=1,\ldots,2k-d$, its limit is straightforward. Thus, we only need to focus on the constraint $\sqrt{n}\text{spr}\Delta_i\geq\left|\boldsymbol{X}_i\right|'\boldsymbol{u}$ for $i=1,2,\ldots n$. Exploiting the convergence in distribution of $\upsilon_n(\cdot)$ to the Poisson random measure $\upsilon(\cdot)$, we have that

$$\begin{aligned}Pr\left[\phi_{1n}\left(\psi^1\right)=0,\ldots,\phi_{1n}\left(\psi^m\right)=0\right]&=Pr\left[\sum_{i=1}^{n}I\left(\sqrt{n}\text{spr}\Delta_i<\max_{1\leq j\leq m}\left|\boldsymbol{X}_i\right|'\boldsymbol{u}^j\right)=0\right]\\\to\exp\left(-\int_{\mathcal{X}}\max_{1\leq j\leq m}\frac{g(\boldsymbol{x})}{\kappa}\left(\boldsymbol{x}'\boldsymbol{u}^j\right)^2d\mu(\boldsymbol{x})\right)&=Pr\left[\phi_1\left(\psi^1\right)=0,\ldots,\phi_1\left(\psi^m\right)=0\right].\end{aligned}$$

Therefore, $M_{1n}(\psi;\lambda)\overset{f.d.}{\to}M_1(\psi;\lambda)$ as $n\to\infty$.

The other two conditions can be easily verified. On the set $O_p\times O_q\times(-\infty,0)^{2k-d}\times O_v$, where $O_p\subset\mathbb{R}^{2k-d}$, $O_q\subset\mathbb{R}$ and $O_v\subset\mathbb{R}$ are any open sets, $M_1(\psi;\lambda)$ is finite by its definition. And, for any realization of W, $\{\mathit{\Gamma}_i,\ i\geq 1\}$ and $\{\mathit{\Upsilon}_i,\ i\geq 1\}$, $M_1(\psi;\lambda)$ will be uniquely minimized due to the quadratic form of $M(\psi;\lambda)$ and the geometry of the constraint. If the minimal of $M(\psi;\lambda)$ is in the constraint set, the uniqueness is trivially satisfied. If the minimal of $M(\psi;\lambda)$ lies outside the constraint set, the solution to the minimization problem will be the intersect of the level set of $M(\psi;\lambda)$ with the boundary of the constraint set. The level set of the quadratic component of $M(\psi;\lambda)$ takes the shape of an ellipse in high dimension, while the constraint set is convex with boundary consisting of high dimensional planes. They can only intersect at one point. Thus, the latter two conditions in the Convexity Lemma are satisfied. Hence, when $\lim_{t\to\infty} tF_s\left(1/\sqrt{t}\right) \in (0,+\infty)$, $\arg\min_\psi M_{1n}(\psi;\lambda) \xrightarrow{d} \arg\min_\psi M_1(\psi;\lambda)$ as $n\to\infty$.

The above result also holds when $\lim_{t\to\infty} tF_s\left(1/\sqrt{t}\right) = 0$. Rewrite $M_1(\psi;\lambda)$ by substituting $\kappa = +\infty$: $M_1(\psi;\lambda) = M(\psi;\lambda) + \phi_1(\psi)$, with

$$\phi_1(\psi) = \begin{cases} 0, & \text{if } \Upsilon_i'\boldsymbol{u} \leq +\infty, \text{ for } i=1,2,\ldots; \\ & \text{and } I\left(\boldsymbol{\beta}_j^* = 0\right)\boldsymbol{u}_j \geq 0 \text{ for } j=1,\ldots,2k-d\, \cdot \\ \infty, & \text{otherwise} \end{cases}$$

Since Υ_i follows a tight probability measure, we can further simplify

$$\phi_1(\psi) = \begin{cases} 0, & \text{if } I\left(\boldsymbol{\beta}_j^* = 0\right)\boldsymbol{u}_j \geq 0 \text{ for } j=1,\ldots,2k \\ \infty, & \text{otherwise} \end{cases}.$$

The finite dimensional convergence of the quadratic component of $M_{1n}(\psi;\lambda)$ follows the same argument. We now show that for any given $\psi^1,\ldots,\psi^m$, as $n\to\infty$,

$$Pr\left[\phi_{1n}\left(\psi^1\right)=0,\ldots,\phi_{1n}\left(\psi^m\right)=0\right] \longrightarrow Pr\left[\phi_1\left(\psi^1\right)=0,\ldots,\phi_1\left(\psi^m\right)=0\right].$$

By writing the probability as an expectation of a conditional probability, we have that

$$Pr\left[\phi_{1n}\left(\psi^1\right)=0,\ldots,\phi_{1n}\left(\psi^m\right)=0\right] = Pr\left[\sum_{i=1}^n I\left(\sqrt{n}\mathrm{spr}\Delta_i < \max_{1\leq j\leq m}|\boldsymbol{X}_i|'\boldsymbol{u}^j\right)=0\right]$$

$$= Pr^n\left[\sqrt{n}\mathrm{spr}\Delta \geq \max_{1\leq j\leq m}|\boldsymbol{X}|'\boldsymbol{u}^j\right] = E^n\left[Pr\left[\mathrm{spr}\Delta \geq \frac{1}{\sqrt{n}}\max_{1\leq j\leq m}|\boldsymbol{X}|'\boldsymbol{u}^j \mid \boldsymbol{X}\right]\right]$$

$$= E^n\left[1 - g\left(|\boldsymbol{X}|\right)F_s^-\left(\frac{1}{\sqrt{n}}\max_{1\leq j\leq m}|\boldsymbol{X}|'\boldsymbol{u}^j\right)\right] = \left[1 - E\left[g\left(|\boldsymbol{X}|\right)F_s^-\left(\frac{1}{\sqrt{n}}\max_{1\leq j\leq m}|\boldsymbol{X}|'\boldsymbol{u}^j\right)\right]\right]^n,$$

where $F_s^-(a) \equiv \lim_{z\uparrow a} F_s(z)$. The indeterminate form has the limit of

$$\exp\left(-\lim_{n\to\infty} nE\left[g(|\boldsymbol{X}|)\, F_s^-\left(\frac{1}{\sqrt{n}}\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j\right)\right]\right).$$

For $|\boldsymbol{X}| \in \mathcal{X}$ such that $\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j \le 0$, we have that $F_s^-\left(\frac{1}{\sqrt{n}}\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j\right) = 0$; if $\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j > 0$, the inequality $\lim_{n\to\infty} nF_s^-\left(\frac{1}{\sqrt{n}}\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j\right) \le \lim_{n\to\infty} nF_s\left(\frac{1}{\sqrt{n}}\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j\right) = 0$ holds. By the dominated convergence theorem, we obtain that

$$Pr\left[\phi_{1n}\left(\psi^1\right) = 0, \ldots, \phi_{1n}\left(\psi^m\right) = 0\right] \to \exp(0) = 1 = Pr\left[\phi_1\left(\psi^1\right) = 0, \ldots, \phi_1\left(\psi^m\right) = 0\right].$$

Therefore, the finite dimensional convergence of $M_{1n}(\psi;\lambda)$ is verified. The finiteness of $M_1(\psi;\lambda)$ on an open set and with probability one the uniqueness of its minimizer can be verified using the same statement as in the first case when $\kappa \in (0,+\infty)$. The Convexity Lemma then provides that $\arg\min_\psi M_{1n}(\psi;\lambda) \xrightarrow{d} \arg\min_\psi M_1(\psi;\lambda)$ when $\kappa = +\infty$.

The proof for the case when $\lim_{t\to\infty} tF_s\left(1/\sqrt{t}\right) = +\infty$ is essentially the same. Since the part $I\left(\boldsymbol{\beta}_j^* = 0\right)\boldsymbol{u}_j \ge 0$ for $j = 1, \ldots, 2k-d$ does not contain any randomness, we focus the proof on the constraint $\sqrt{n}\mathrm{spr}\Delta_i \ge |\boldsymbol{X}_i|'\boldsymbol{u}$ for $i = 1, 2, \ldots n$. For any given $\psi^1, \ldots, \psi^m$, it holds that

$$Pr\left[\phi_{1n}\left(\psi^1\right) = 0, \ldots, \phi_{1n}\left(\psi^m\right) = 0\right] = \left[1 - E\left[g(|\boldsymbol{X}|)\, F_s^-\left(\frac{1}{\sqrt{n}}\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j\right)\right]\right]^n$$

with the limit

$$\exp\left(-\lim_{n\to\infty} nE\left[g(|\boldsymbol{X}|)\, F_s^-\left(\frac{1}{\sqrt{n}}\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j\right)\right]\right).$$

If $\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j \le 0$, $F_s^-\left(\frac{1}{\sqrt{n}}\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j\right) \le F_s^-(0) = 0$, because $F_s^-(0) = 0$. If $\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j > 0$, we have that $\lim_{n\to\infty} nF_s^-\left(\frac{1}{\sqrt{n}}\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j\right) \ge \lim_{n\to\infty} nF_s\left(\frac{1}{2\sqrt{n}}\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j\right) = \infty$ by condition $\lim_{t\to\infty} tF_s\left(1/\sqrt{t}\right) = +\infty$. Therefore for the given $\boldsymbol{u}$'s, when $Pr\left(\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j \le 0\right) = 1$,

$$\exp\left(-\lim_{n\to\infty} nE\left[g(|\boldsymbol{X}|)\, F_s^-\left(\frac{1}{\sqrt{n}}\max_{1\le j\le m} |\boldsymbol{X}|'\boldsymbol{u}^j\right)\right]\right) \to 1;$$

and when $Pr\left(\max_{1\le j\le m}|\boldsymbol{X}|'\boldsymbol{u}^j\le 0\right)<1$,

$$\exp\left(-\lim_{n\to\infty}nE\left[g\left(|\boldsymbol{X}|\right)F_s^-\left(\frac{1}{\sqrt{n}}\max_{1\le j\le m}|\boldsymbol{X}|'\boldsymbol{u}^j\right)\right]\right)\to 0.$$

Now consider $Pr\left[\phi_1\left(\psi^1\right)=0,\ldots,\phi_1\left(\psi^m\right)=0\right]$. The probability can be calculated as

$$\begin{aligned}Pr\left[\phi_1\left(\psi^1\right)=0,\ldots,\phi_1\left(\psi^m\right)=0\right] &= Pr\left[\max_{1\le j\le m}\Upsilon_i'\boldsymbol{u}^j\le 0,:\text{for}:i=1,2,\ldots\right]\\ &= \begin{cases}1, & \text{if } Pr\left(\max_{1\le j\le m}|\boldsymbol{X}|'\boldsymbol{u}^j\le 0\right)=1\\ 0, & \text{if } Pr\left(\max_{1\le j\le m}|\boldsymbol{X}|'\boldsymbol{u}^j\le 0\right)<1\end{cases},\end{aligned}$$

where the last equality follows from the fact that the distribution of Υ_i is the same as $|\boldsymbol{X}|$ and i goes to infinity. Therefore, we have shown that for any given $\psi^1,\ldots,\psi^m$,

$$Pr\left[\phi_{1n}\left(\psi^1\right)=0,\ldots,\phi_{1n}\left(\psi^m\right)=0\right]\longrightarrow Pr\left[\phi_1\left(\psi^1\right)=0,\ldots,\phi_1\left(\psi^m\right)=0\right],$$

as $n\to\infty$. The rest is the same as in case $\lim_{t\to\infty}tF_s\left(1/\sqrt{t}\right)=0$.

Hence, we have shown that $\arg\min_\psi M_{1n}\left(\psi;\lambda\right)$ converges in law to $\arg\min_\psi M_1\left(\psi;\lambda\right)$ for all different values of $\lim_{t\to\infty}tF_s\left(1/\sqrt{t}\right)$. Moreover, it follows directly from the definition that $M_1\left(\psi;\lambda\right)=M_1\left(\psi;1\right)$. The claimed theorem can be concluded. □

Proof of Theorem 2.3.3: The proof is similar to the proof for Theorem 2.3.2. $\sqrt{n}\left(\widetilde{\theta}-\theta^*\right)$ is the solution to the minimization problem:

$$\min_{\boldsymbol{p},q,\boldsymbol{u},v}\left[\begin{array}{c}\sum_{i=1}^n\left(\text{mid}\Delta_i-\gamma^*-\frac{1}{\sqrt{n}}\boldsymbol{X}_i'\boldsymbol{p}-\frac{q}{\sqrt{n}}\right)^2-\sum_{i=1}^n\left(\text{mid}\Delta_i-\gamma^*\right)^2\\ +\lambda\sum_{i=1}^n\left(\text{spr}\Delta_i-\delta^*-\frac{1}{\sqrt{n}}|\boldsymbol{X}_i|'\boldsymbol{u}-\frac{v}{\sqrt{n}}\right)^2-\lambda\sum_{i=1}^n\left(\text{spr}\Delta_i-\delta^*\right)^2+\phi_{2n}\left(\psi\right)\end{array}\right],$$

where

$$\phi_{2n}\left(\psi\right)=\begin{cases}0, & \text{if } R\psi\ge\sqrt{n}\left(r-R\theta^*\right)\\ \infty, & \text{otherwise}\end{cases}.$$

For the vector $\left(r-R\theta^*\right)$, some elements are zero and the rest are strictly negative. The zero elements correspond to the submatrix R_b of R by definition. This implies that $\phi_{2n}\left(\psi\right)\to$

$\phi_2(\psi)$ pointwise as $n \to \infty$. The rest of the proof is the same for Theorem 2.3.2 by showing finite-dimensional convergence, $M(\psi;\lambda) + \phi_2(\psi)$ being finite on an open set, and $M(\psi;\lambda) + \phi_2(\psi)$ being uniquely minimized with probability 1. □

Proof of Lemma 2.4.1: By Assumption 2.4.2, the claimed result follows if we show that $R_0\Psi$ represents the asymptotic distribution of $\sqrt{n}\left(\widetilde{\alpha}_s', \widetilde{\beta}_m'\right)'$ under the null hypothesis and the parameter sequence (η_n, π_n, ξ_n). Under H_0, it holds that $\beta_m^* = \mathbf{0}$. Simple manipulation of the proof for Theorem 2.3.3 would provide that $\sqrt{n}\left(\widetilde{\alpha}_s', \widetilde{\beta}_m'\right)' \xrightarrow{d} R_0\Psi$, with the Lindeberg-Lévy Central Limit Theorem replaced by Lyapunov Central Limit Theorem under Assumption 2.4.1, see Lemma A.3 in Cheng (2015). ■

Proof of Proposition 2.4.1: We prove the proposition by verifying assumptions in McCloskey (2017). Notice that the distribution $(R_0\Psi)'(R_0\Sigma_\omega R_0')^{-1}(R_0\Psi)$ is finite with probability 1 for all c and π_ω in the localization parameter space. Assumption PS in McCloskey (2017) is trivially satisfied. By the expression in Lemma 2.4.1, $\mathcal{C}_{c,\pi_\omega}(1-\vartheta)$ is continuous in c and π_ω. Together with the assumption in Proposition 2.4.1, Assumption Cont in McCloskey (2017) is satisfied. The requirement for the confidence set I_τ that $\lim_{n\to\infty} Pr\left(\sqrt{n}\eta_n \in I_\tau\right) \geq 1-\tau$ fulfills Assumption CS in McCloskey (2017). It suffices to prove that Assumption DS in McCloskey (2017) is satisfied for the first claim of the proposition.

Lemma 2.4.1 provides that the asymptotic distribution of the test statistic T_n is $(R_0\Psi)'(R_0\Sigma_\omega R_0')^{-1}(R_0\Psi)$ under the full parameter sequence (η_n, π_n, ξ_n); and the asymptotic convergence of $\sqrt{n}\left(\widehat{\beta}_{s,OLS}', \widehat{\delta}_{OLS}\right)' - \sqrt{n}\eta_n \xrightarrow{d} Z(\Lambda_{OLS})$ is straightforward. 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 $\{\eta_{p_n}, \pi_{p_n} : n \geq 1\}$ such that $\left(\sqrt{p_n}\eta_{p_n}, \pi_{p_n}\right) \to (c, \pi_\omega)$. We aim to construct a full sequence

$\{\eta_n^\star, \pi_n^\star : n \geq 1\}$ satisfying that $(\sqrt{n}\eta_n^\star, \pi_n^\star) \to (c, \pi_\omega)$ and $(\eta_n^\star, \pi_n^\star) = (\eta_{p_n}, \pi_{p_n})$, $\forall n \geq 1$. To clarify the notation, let the full sequence be indexed by m: $\{\eta_m^\star, \pi_m^\star : m \geq 1\}$. For $\forall m = p_n$, define $(\eta_m^\star, \pi_m^\star) = (\eta_{p_n}, \pi_{p_n})$; and for $\forall m \in (p_n, p_{n+1})$, define

$$\delta_m^\star = \begin{cases} \frac{\sqrt{p_n}\delta_{p_n}}{\sqrt{m}}, & \text{if } \sqrt{p_n}\delta_{p_n} \to c_{k-d+1} \in \mathbb{R}_{\geq 0} \\ \delta_{p_n}, & \text{if } \sqrt{p_n}\delta_{p_n} \to +\infty \end{cases} \quad \text{and}$$

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

for $j = 1, \ldots, k-d$ and $\pi_m^\star = \pi_{p_n}$. It is trivial that the constructed full sequence satisfies the second requirement that $\left(\eta_{p_n}^\star, \pi_{p_n}^\star\right) = (\eta_{p_n}, \pi_{p_n})$ 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 Lemma 2.4.2: With Assumption 2.4.2, the lemma follows if we can show that $\sqrt{n}\left(\widetilde{\theta} - \theta_n\right) \xrightarrow{d} \Psi$ under the model parameters (η_n^u, π_n, ξ_n). The estimator is defined as $\widetilde{\theta} = \arg\min_{\theta \in \mathbb{R}^{2l}} Z_{2n}(\theta)$. Note that $\sqrt{n}\left(\widetilde{\theta} - \theta_n\right)$ is the solution to the minimization problem:

$$\min_{\boldsymbol{p},q,\boldsymbol{u},v} \left[\begin{array}{c} \sum_{i=1}^n \left(\text{mid}\Delta_i - \gamma_n - \frac{1}{\sqrt{n}}\boldsymbol{X}_i'\boldsymbol{p} - \frac{q}{\sqrt{n}}\right)^2 - \sum_{i=1}^n (\text{mid}\Delta_i - \gamma_n)^2 \\ +\lambda \sum_{i=1}^n \left(\text{spr}\Delta_i - \delta_n - \frac{1}{\sqrt{n}}|\boldsymbol{X}_i|'\boldsymbol{u} - \frac{v}{\sqrt{n}}\right)^2 - \lambda \sum_{i=1}^n (\text{spr}\Delta_i - \delta_n)^2 + \phi_n(\psi) \end{array} \right],$$

where

$$\phi_{2n}(\psi) = \begin{cases} 0, & \text{if } R\psi + \sqrt{n}(R\theta_n - r) \geq 0 \\ \infty, & \text{otherwise} \end{cases}.$$

By Lyapunov Central Limit Theorem, we obtain that

$$\left[\begin{array}{c} \sum_{i=1}^n \left(\text{mid}\Delta_i - \gamma_n - \frac{1}{\sqrt{n}}\boldsymbol{X}_i'\boldsymbol{p} - \frac{q}{\sqrt{n}}\right)^2 - \sum_{i=1}^n (\text{mid}\Delta_i - \gamma_n)^2 \\ +\lambda \sum_{i=1}^n \left(\text{spr}\Delta_i - \delta_n - \frac{1}{\sqrt{n}}|\boldsymbol{X}_i|'\boldsymbol{u} - \frac{v}{\sqrt{n}}\right)^2 - \lambda \sum_{i=1}^n (\text{spr}\Delta_i - \delta_n)^2 \end{array} \right] \xrightarrow{d} M_\omega(\psi; \lambda).$$

Decompose R into three submatrices: R_Γ, R_{0b} and R_{0s}, where R_{0s} denotes the non binding inequalities in $R_{00}\theta \geq r_{-\Gamma}$. By the definition of c, it holds that $\Gamma\eta_n^u - r_\Gamma = R_\Gamma\theta_n - r_\Gamma \to c$

when $n \to \infty$; under the null hypothesis, R_{0b} and R_{0s} represent the binding and non binding inequalities $R_{00}\theta \geq r_{-\Gamma}$. Therefore, we have $\phi_{2n}(\psi) \to \phi_{\omega}(\psi)$ pointwise as $n \to \infty$. We obtain that $\sqrt{n}\left(\widetilde{\theta} - \theta_n\right) \xrightarrow{d} \Psi$ and the claimed lemma. □

Proof of Theorem 2.4.1: The theorem follows from the same proof for Proposition 2.4.1. □

B.2 A Review of Generalized Interval Arithmetic and Random Generalized Intervals

Interval Arithmetic Given $a_1, a_2 \in \mathbb{R}$ and $a_1 \leq a_2$, an interval A is defined by its left and right end points: $A = [a_1, a_2] = \{x \in \mathbb{R} : a_1 \leq x \leq a_2\}$, or by its center and range: $A = [\text{mid}A \pm \text{spr}A]$, where $\text{mid}A = (a_1 + a_2)/2$ and $\text{spr}A = (a_2 - a_1)/2 \geq 0$. The set of all intervals is denoted by $I(\mathbb{R})$. For all $A, B \in I(\mathbb{R})$ and $\lambda \in \mathbb{R}$, it holds that

(i) $A + B \equiv [(\text{mid}A + \text{mid}B) \pm (\text{spr}A + \text{spr}B)]$ and

(ii) $\lambda A \equiv [\lambda \text{mid}A \pm |\lambda| \text{spr}A]$.

Combining (i) and (ii), we obtain that for all $A, B \in I(\mathbb{R})$ and $\lambda \in \mathbb{R}$,

$$A + \lambda B = [(\text{mid}A + \lambda \text{mid}B) \pm (\text{spr}A + |\lambda| \text{spr}B)]. \tag{B.1}$$

It follows from (ii) that $-A \equiv (-1)A = [-\text{mid}A \pm \text{spr}A] = [-a_2, -a_1]$. Subtraction between two intervals A and B is defined as

$$A - B = A + (-B) = [(\text{mid}A - \text{mid}B) \pm (\text{spr}A + \text{spr}B)].$$

As a result, we have:

$$A - A = [0 \pm (2\text{spr}A)] \neq [0, 0] \text{ and}$$
$$A - B + B = [\text{mid}A \pm (\text{spr}A + 2\text{spr}B)] \neq A.$$

To partly remedy this situation, Hukuhara (1967) introduces an alternative difference operation on intervals referred to as Hukuhara difference and denoted as $(-_H)$. Specifically, for any $A, B \in I(\mathbb{R})$, $A -_H B = C$ if there exists $C \in I(\mathbb{R})$ such that $A = B + C$. It can be shown that Hukuhara difference $A -_H B$ exists if and only if spr$A \geq$ spr$B$ and when it exists,

$$A -_H B \equiv [(\text{mid}A - \text{mid}B) \pm (\text{spr}A - \text{spr}B)] .$$

In contrast to subtraction $(-)$, Hukuhara difference satisfies: $A -_H A = [0, 0]$ and $A -_H B + B = A$. However, Hukuhara difference between two intervals may not exist which limits the scope of applications of the interval arithmetic approach to modeling interval data.

Generalized Interval Arithmetic and an L_2-type Metric for Generalized Intervals
In the paper, we make use of generalized intervals studied in the mathematics literature, see e.g., Kaucher (1980) and Markov (1996), to fully explore advantages of the interval arithmetic approach to modeling interval data. Specifically, for $a_1, a_2 \in \mathbb{R}$, a generalized interval is an ordered couple denoted as $A = [a_1, a_2]$: it is a proper or simply an interval when $a_1 \leq a_2$; otherwise it is an improper interval. A generalized interval can also be represented as $A = [\text{mid}A \pm \text{spr}A]$: it is proper if spr$A \geq 0$; improper if spr$A < 0$. Denote $K(\mathbb{R})$ as the space of generalized intervals. For $A, B \in K(\mathbb{R})$, it turns out that the operations: addition and scalar product can be computed in the same way as in (B.1), see Kaucher (1980) and Markov (1996) for details.

With generalized intervals, we can extend Hukuhara difference to any two intervals. Let A and B be two intervals. Generalized Hukuhara difference[1] is defined as follows:

$$A -_{GH} B \equiv A + \left(-\overline{B}\right) = [(\text{mid}A - \text{mid}B) \pm (\text{spr}A - \text{spr}B)] ,$$

where $\overline{B} = [\text{mid}B \mp \text{spr}B]$ is the conjugation or dual of B. In contrast to Hukuhara difference, the Generalized Hukuhara difference between two intervals A and B always exists: when spr$A \geq$ spr$B$, $A -_{GH} B$ is an interval and $A -_{GH} B = A -_H B$; otherwise it is an

[1]It is sufficient for our purpose to define Generalized Hukuhara difference for intervals only. It turns out that the Generalized Hukuhara difference for intervals is the $-_h$ operation defined in Markov (1996) for generalized intervals when applied to intervals.

improper interval. It is easy to see that $A -_{GH} A = [0, 0]$ and $A -_{GH} B + B = A$.

Let A and B be two generalized intervals. We define an L_2-type metric d_λ between A and B as

$$d_\lambda (A, B) \equiv \left((\text{mid}A - \text{mid}B)^2 + \lambda (\text{spr}A - \text{spr}B)^2\right)^{\frac{1}{2}}$$

for some $\lambda \in (0, \infty)$ and the norm of $A \in K(\mathbb{R})$ as $\|A\|^2 = (\text{mid}A)^2 + \lambda (\text{spr}A)^2$. It is easy to verify that the d_λ metric satisfies non-negativity, identity of indiscernibles, symmetry and triangle inequality. By choosing different values of λ, one can assign different relative importance for the squared distance between the ranges with respect to the square distance between the midpoints. When A and B are both intervals, the d_λ metric generalizes the well-known Bertoluzza metric denoted as d_W in Bertoluzza et al. (1995) if the metric is required to be invariant to rigid motion (Trutschnig et al. (2009)). One common choice for λ is 1/3, as it corresponds to d_W when W is chosen as the Lebesgue measure. For more discussions on different metrics on $I(\mathbb{R})$, see Bertoluzza et al. (1995) and Trutschnig et al. (2009).

Random Generalized Intervals Let (Ω, Σ, P) be an abstract probability space.

Definition B.2.1. *(i) A random generalized interval* $X : \Omega \to K(\mathbb{R})$ *is a map from the sample space* Ω *to the space of generalized intervals such that* $\text{mid}X : \Omega \to \mathbb{R}$ *and* $\text{spr}X : \Omega \to \mathbb{R}$ *are random variables; (ii) The expected value of a random generalized interval* X*, denoted as* $E_A(X) \in K(\mathbb{R})$*, is defined as*

$$E_A(X) = E_A([\text{mid}X \pm \text{spr}X]) \equiv [E(\text{mid}X) \pm E(\text{spr}X)], \tag{B.2}$$

whenever $E(\text{mid}X)$ *and* $E(\text{spr}X)$ *exist.*

When $\text{spr}X \geq 0$ with probability one, the random generalized interval X becomes a random interval, which is a measurable map from Ω to $I(\mathbb{R})$. When X is a random interval, the expectation defined in (B.2) agrees with the well-known Aumann expectation (Aumann (1965)). Let $\mathfrak{A}$ be a sub-σ-algebra of Σ, the conditional expectation of a generalized random interval X given $\mathfrak{A}$ is defined accordingly as

$$E_A(X \mid \mathfrak{A}) = E_A([\mathrm{mid}X \pm \mathrm{spr}X] \mid \mathfrak{A}) \equiv [E(\mathrm{mid}X \mid \mathfrak{A}) \pm E(\mathrm{spr}X \mid \mathfrak{A})].$$

We follow the approach of Fréchet (1948) to define the variance of a random generalized interval as:

$$Var_F(X) \equiv \inf_{A \in K(\mathbb{R})} E\left(d_\lambda^2(X, A)\right) \text{ whenever } E\left(\|X\|^2\right) < \infty.$$

Since the expectation defined in (B.2) agrees with Fréchet expectation with respect to the metric d_λ (Körner (1997), Körner and Näther (2002)), Fréchet variance of a random generated interval X in the metric space $K(\mathbb{R})$ endowed with the d_λ metric is simplified as $Var_F(X) = E(d_\lambda^2(X, E_A(X)))$.