cemmap

centre for microdata methods and practice

Testing multiple inequality hypotheses: A smoothed indicator approach

Le-Yu Chen
Jerzy Szroeter

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP16/12

Testing Multiple Inequality Hypotheses : A Smoothed Indicator Approach*

Le-Yu Chen†
Institute of Economics, Academia Sinica

Jerzy Szroeter‡
Department of Economics, University College London

Revision : June 2012

Abstract

This paper proposes a class of origin-smooth approximators of indicators underlying the sum-of-negative-part statistic for testing multiple inequalities. The need for simulation or bootstrap to obtain test critical values is thereby obviated. A simple procedure is enabled using fixed critical values. The test is shown to have correct asymptotic size in the uniform sense that supremum finite-sample rejection probability over null-restricted data distributions tends asymptotically to nominal significance level. This applies under weak assumptions allowing for estimator covariance singularity. The test is unbiased for a wide class of local alternatives. A new theorem establishes directions in which the test is locally most powerful. The proposed procedure is compared with predominant existing tests in structure, theory and simulation.

KEYWORDS : Test, Multiple inequalities, One-sided hypothesis, Composite null, Binding constraints, Asymptotic exactness, Covariance singularity, Indicator smoothing

JEL SUBJECT AREA : C1, C4

---

*This paper is a substantial revision of Chapter 3 of the first author's doctoral dissertation (Chen 2009) and the subsequent Cemmap working paper (Chen and Szroeter 2009). We thank Oliver Linton, Sokbae Lee, Yoon-Jae Whang, Chungmin Kuan, Hidehiko Ichimura and Joon Park for helpful comments. We are also grateful to seminar participants for various insightful discussions of this work presented in 2009 International Symposium on Econometric Theory and Applications and 2010 Royal Economic Society Annual Conference.

†Corresponding author. E-mail : lychen@econ.sinica.edu.tw

‡E-mail : j.szroeter@ucl.ac.uk

1 Introduction

This paper is concerned with the problem of testing the null hypothesis H_0 that the true value of a finite p-dimensional parameter vector μ is non-negative versus the alternative that at least one element of μ is strictly negative. A major problem for testing such hypotheses has been dependence of null rejection probability on the unknown subset of binding inequalities (zero-valued μ_j). Under H_0, the asymptotic distribution of a nontrivial test statistic is typically degenerate at interior points (all elements of μ strictly positive) of parameter space. But at boundary points (one or more elements zero), that distribution is non-degenerate and may depend on the number and position of the zero elements but not on strict positives. In consequence, determining the critical value to be used for the test at some nominal significance level α is a nontrivial issue. The classic least favorable configuration (LFC) approach seeks the parameter point in the null that maximizes the rejection probability (e.g., see Perlman (1969) and Robertson, Wright and Dykstra (1988)). This principle risks yielding tests which have comparatively low power against sequences of alternatives converging to boundary points which are not LFC. To improve test power, recent literature has proposed using data-driven selection of the true binding inequalities in place of the LFC point to compute test critical values. Whatever the critical value, it is important to demonstrate that null rejection probability does not exceed α uniformly over all H_0-compliant data generating processes for sample size large enough. Such uniformity has been emphasized in recent literature (e.g., see Mikusheva (2007), Romano and Shaikh (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010) and Linton et al. (2010)) to ensure validity of asymptotic approximation to actual finite sample test size especially when the test statistic has a limiting distribution which is discontinuous on parameter space. Regardless of whether the binding inequalities are fixed according to the LFC or determined via a stochastic selection mechanism, the functional forms of test statistics proposed in this literature are generally non-smooth and hence computation of test critical values requires simulation or bootstrap.

The contributions of the present paper are as follows. We develop a multiple inequality test whose implementation does not require computer intensive methods. The central idea is to construct a sequence of origin-smooth approximators of indicators underlying the sum-of-negative-part statistic for testing multiple inequalities. The approximation is a form of indicator smoothing in the spirit of Horowitz (1992), enabling standard asymptotic distribution results and obviating simulation and bootstrap computation of test critical values. Moreover, the test allows for estimator covariance singularity.

The test statistic of this paper has a non-degenerate asymptotic distribution of simple analytic form at boundary points of the null hypothesis but becomes degenerate at interior points. Despite this type of discontinuity, the test critical value can be fixed ex ante without compromising asymptotic validity in the uniform sense that the limit of finite sample test size (defined as supremal rejection probability over all H_0-compatible data generating processes) is equal to the nominal size. We prove that this uniformity property holds for every approximator in a wide class allowed by the paper.

The smoothing design of this paper embodies a data driven weighting scheme which automatically concentrates the test statistic onto those parameter estimates signaling binding inequalities. This feature is connected to methods of binding inequality selection used in Hansen (2005), Chernozhukov et al. (2007), Andrews and Soares (2010) and Linton et al. (2010). Indeed, the smoother can also be interpreted as an asymptotic selector and the key component of our test statistic coincides with the sum of elements of the difference between the estimated and recentered null-compatible mean used to obtained the simulated test critical values for Andrews and Soares (2010)'s generalized moment selection (GMS) based tests. The difference itself, however, is not within the class of test statistics covered by the theory of these authors but its properties emerge from the theory developed in the present paper.

The relative computational ease of the test of this paper might be expected to carry a cost in terms of power. However, as we show, the test is consistent against all fixed alternatives and is unbiased for a wide class of local alternatives. In comparison with existing tests, its relative strength varies with the particular direction of local alternative. We provide a new theorem establishing directions in which the test is locally most powerful. Monte Carlo results support the theory and reveal that finite sample performance of the present test is not dominated by the GMS based tests.

We now review relevant test methods in addition to the works cited above. The QLR test has been well developed in the inequality test literature. See, e.g. Perlman (1969), Kodde and Palm (1986), Wolak (1987, 1988, 1989, 1991), Gourieroux and Monfort (1995, chapter 27) and Silvapulle and Sen (2005, chapters 3-4). This test is also applied in the moment inequality literature (see Rosen (2008), Andrews and Guggenberger (2009) and Andrews and Soares (2010)). The asymptotic null distribution of the QLR test statistic generally has no analytical form. Since computing this test statistic requires solution of a quadratic optimization program subject to non-negativity constraints, simulation and bootstrapping for the test critical value is particularly heavy.

An extreme value (EV) form of test statistic was developed by White (2000) in the context of comparing predictive abilities among forecasting models. Such a statistic is lighter on computation but its asymptotic null distribution remains non-standard. Hansen (2005) incorporates estimation of actual binding inequalities to bootstrap null distribution of the extreme value statistic. Hansen's refinement is a special case of the GMS based critical value estimation proposed by Andrews and Soares (2010) who also consider a broad class of test functions including both the QLR and other simpler forms using negative-part functions.

The rest of the paper is organized as follows. Section 2 summarizes the method of Andrews and Soares (2010) for testing with estimated critical values which embody the GMS procedure for estimation of binding inequalities. We contrast that with the smoothing approach of this paper and highlight connecting features. Section 3 sets out functional assumptions on the class of smoothers and completes construction of the test statistic. Section 4 states basic distributional assumptions on parameter estimators and presents asymptotic null distribution of the test statistic. Section 5 establishes key results on asymptotic size of the test. Section 6 studies test consistency and local power. Section 7 presents results of some Monte Carlo simulation studies. Section 8 concludes. Appendix A derives the details of an adjustment component of the test statistic. Appendix B provides proofs of theoretical results of the paper. Appendix C gives examples of covariance matrix singularity and illustrates how they can fit into our framework.

2 Recentering, Selection and Smoothing in Inequality Tests

Let $\mu = (\mu_1, \mu_2, ..., \mu_p)'$ be a column vector of (functions of) parameters appearing in an econometric model. We are interested in testing :

$$H_0 : \mu_j \geq 0 \text{ for all } j \in \{1, 2, ..., p\} \text{ versus } H_1 : \mu_j < 0 \text{ for at least one } j. \tag{2.1}$$

We assume that there exists a vector $\widehat{\mu}$ of parameter estimators based on sample size T such that $\sqrt{T}(\widehat{\mu} - \mu)$ is asymptotically multivariate normal with mean 0 and covariance V consistently estimated by $\widehat{V}$. The vector μ and matrix V may depend on common parameters but this is generally kept implicit for notational simplicity.

2.1 Recentering and Generalized Moment Selection in Critical Value Estimation

Recent improved tests developed by Andrews and Soares (2010) of the hypothesis (2.1) are distinguished by their use of estimated critical values embodying a selection rule to statistically decide which inequalities are binding ($\mu_j = 0$). In brief, these tests proceed operationally as follows. A statistic $S(\sqrt{T}\widehat{\mu}, \widehat{V})$ is first computed for some fixed function $S(.,.)$. The asymptotic critical value of the statistic is then obtained by simulation (or resampling) as the appropriate quantile of the distribution of $S(Z + K(T)\widetilde{\mu}, \widehat{V})$ where Z is an artificially generated vector such that $Z \sim N(0, \widehat{V})$ conditionally on data, $\widetilde{\mu}$ is a recentered null-compatible mean and $K(T) = o(\sqrt{T})$ is some positive "tuning" function increasing without bound as $T \longrightarrow \infty$. Basic recentering defines $\widetilde{\mu}_j = 0$ for $K(T)\widehat{\mu}_j \leq 1$. Setting $\widetilde{\mu}_j = 0$ amounts to selecting j as the index of a binding constraint. For $K(T)\widehat{\mu}_j > 1$, $\widetilde{\mu}_j$ is defined to ensure $K(T)\ \widetilde{\mu}_j \longrightarrow \infty$ as $T \longrightarrow \infty$, this being simply achieved by taking $\widetilde{\mu}_j = \widehat{\mu}$. Basic selection as stated here is a special case of the Andrews and Soares (2010) Generalized Moment Selection (GMS) procedure.[1]

[1]Indeed, this selection rule corresponds to use of moment selection function $\varphi_j^{(2)}$ considered by Andrews and Soares (2010, pp. 131-132) with due allowance for standardization of parameter estimates. See also Andrews and Barwick (2012, pp. 8-9) for various examples of the GMS selection rules.

Data-dependent selection of binding constraints reduces possible inefficiencies arising from fixing all the elements of $\widetilde{\mu}$ to be zero (least favorable). On the other hand, regardless of how $\widetilde{\mu}$ is constructed, simulation (or bootstrap) is still needed since the asymptotic distribution of the statistic used in this literature is generally non-standard. This applies even to test statistics which aggregate individual discrepancy values $\min(\widehat{\mu}_j, 0)$ in a simple manner. They include the extreme value form studied by Hansen (2005) and the sum

$$\sum_{j=1}^{p}[-\sqrt{T}\min(\widehat{\mu}_j, 0)] \tag{2.2}$$

lying within the very wide class of right-tailed tests studied by Andrews and Soares (2010).

2.2 The Smoothed Indicator Approach

Let $1\{.\}$ denote the indicator taking value unity if the statement inside the bracket is true and zero otherwise. The root cause of non-standard distribution of (2.2) is the discontinuity at the origin of the indicator $1\{x \leq 0\}$ underlying the negative-part function $\min(x, 0) = 1\{x \leq 0\}x$. To overcome this problem, the present paper investigates an indicator smoothing approach as follows.

First, we approximate the function $\min(x, 0)$ by $\Psi_T(x)x$ where $\{\Psi_T(x)\}$ is a sequence of non-negative and non-increasing functions each of which is continuously differentiable at the origin and converges pointwise (except possibly at the origin) as $T \longrightarrow \infty$ to the indicator function $1\{x \leq 0\}$. We refer to $\Psi_T(x)$ as an (origin-smoothed) indicator smoother or a smoothed indicator for $1\{x \leq 0\}$.

In this paper, we will focus on the class of smoothed indicators generated as $\Psi_T(x) = \Psi(K(T)x)$ for some fixed function Ψ and a "tuner" $K(T)$ of the type mentioned in Subsection 2.1. The functional form of Ψ includes decumulative distribution functions for continuous variates as well as discrete yet origin-smooth functions. We therefore replace the individual negative-part statistic $\sqrt{T}\min(\widehat{\mu}_j, 0)$ of (2.2) by $\sqrt{T}\Psi_T(\widehat{\mu}_j)\widehat{\mu}_j$. Subject to regularity conditions set out later, $\Psi_T(\widehat{\mu}_j) = o_p(1/\sqrt{T})$ for strictly positive μ_j and hence the term $\sqrt{T}\Psi_T(\widehat{\mu}_j)\widehat{\mu}_j$ vanishes asymptotically. For zero-valued μ_j, $\Psi_T(\widehat{\mu}_j)$ tends to $\Psi(0)$ in probability and $\sqrt{T}\Psi_T(\widehat{\mu}_j)\widehat{\mu}_j$ is asymptotically equivalent to $\Psi(0)\sqrt{T}\widehat{\mu}_j$.

Second, we consider a left-tailed test based on the statistic that replaces (2.2) with

$$\sum_{j=1}^{p}\left[\sqrt{T}\Psi_T(\widehat{\mu}_j)\widehat{\mu}_j - \Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj})\right] \tag{2.3}$$

where $\widehat{v}_{jj}$ is the jth diagonal element of $\widehat{V}$ and Λ_T is an adjustment term approximating the expectation of $[\Psi_T(\widehat{\mu}_j) - \Psi(0)]\sqrt{T}\widehat{\mu}_j$ evaluated at $\mu_j = 0$. This expectation is non-positive, though shrinking to zero in large samples.[2] Under suitable regularity conditions Λ_T, whose detailed construction is given in Section 3, is non-positive for all T but converges to zero in probability. Hence, under the null hypothesis the statistic (2.3) will be asymptotically either degenerate or equivalent in distribution to a normal variate and thus critical values for a test using (2.3) will not require simulation.

Besides indicator smoothing, it is also appropriate to view Ψ_T as a form of binding inequality selection akin to the aforementioned GMS procedure. The smoothed indicators in (2.3) essentially embed a data driven weighting scheme which automatically concentrates the statistic (2.3) onto those parameter estimates signaling binding inequalities. Indeed, consider the specific smoothed indicator constructed as $\Psi_T(x) = 1\{K(T)x \leq 1\}$. Such $\Psi_T(x)$ simply shifts the point of discontinuity away from the origin whilst still acting as a pure zero-one selector. Then the GMS based recentering described in Subsection 2.1 would amount to setting $\widetilde{\mu}_j = (1 - \Psi_T(\widehat{\mu}_j))\widehat{\mu}_j$. In this case, the statistic (2.3) is equal to

$$\sum_{j=1}^{p} \sqrt{T}(\widehat{\mu}_j - \widetilde{\mu}_j) + o_p(1). \tag{2.4}$$

Since both $\widehat{\mu}$ and $\widetilde{\mu}$ are available as a by-product of the mainstream tests of Subsection 2.1, one may as well perform a test on their difference. The asymptotic distribution of (2.4) does not itself require simulation and recentering, so there is no circularity of argument. Though (2.4) and the GMS test procedure are closely related, it is important to stress that the present test enforces data driven selection of binding inequalities through smoothed indicators within the test statistic itself rather than at the stage of critical value estimation. Therefore, the class of statistics (2.3) does not lie in the otherwise very wide class covered by the work of Andrews and Soares (2010).

It is worth noting that the approach to achieve asymptotic normality in this paper is distinct from alternative devices such as those of Dykstra (1991) and Menzel (2008) who demonstrate that even the QLR statistic can be asymptotically normal when p, the dimension of μ, is viewed as increasing with T to infinity. Recent papers by Lee and Whang (2009) and Lee, Song and Whang (2011) obtain asymptotic normality for a class of functional inequality test statistics. Their particular device (poissonization) requires μ to be infinitely dimensional at the outset. By contrast, in the framework of testing finite and fixed p inequalities, the present paper (and its preliminary versions (Chen and Szroeter (2006, 2009) and Chen (2009, Chapter 3)) where a prototype asymptotically normal test statistic appears) uses only large T asymptotics and an indicator smoothing device. The strategy adopted by this work in testing is akin to Horowitz (1992) who sought to resolve non-standard asymptotic behavior in estimation by replacing a discrete indicator function with a smoothed version. Therefore, the smoothing mechanism in-

[2]Note that $\Psi_T(\widehat{\mu}_j)\widehat{\mu}_j \leq \Psi(0)\widehat{\mu}_j$ for any T because the function $\Psi_T(x) = \Psi(K(T)x)$ is constructed to be non-negative and non-increasing in x.

vestigated by this paper to obtain standard asymptotic distribution results could also be of theoretical interest in its own right.

3 Smoothed Indicator Class and Test Procedure

We now formally set out regularity conditions on the smoothed indicator $\Psi_T(x)$, $x \in R$. We require that

$$\Psi_T(x) = \Psi(K(T)x) \tag{3.1}$$

where $\Psi(.)$ and $K(T)$ are functions satisfying the following assumptions:

[A1] *$\Psi(x)$ is a non-increasing function and $0 \leq \Psi(x) \leq 1$ for $x \in R$.*

[A2] *$\Psi(0) > 0$ and, throughout some open interval containing $x = 0$ and at all except possibly a finite number of points outside that interval, $\Psi(x)$ has a continuous first derivative $\psi(x)$ that is bounded absolutely by a finite positive constant. The left-hand limits of $\psi(y)$ as y approaches x exist at any $x \in R$.*

[A3] *$K(T)$ is positive and increasing in T.*

[A4] *$K(T) \longrightarrow \infty$ and $K(T)/\sqrt{T} \longrightarrow 0$ as $T \longrightarrow \infty$.*

[A5] *$\Psi(x) \longrightarrow 1$ as $x \longrightarrow -\infty$.*

[A6] *$\sqrt{T}\Psi(K(T)x) \longrightarrow 0$ as $T \longrightarrow \infty$ for $x > 0$.*

Assumptions [A1]-[A6] are very mild and satisfied by all the particular Ψ functions including step-at-unity, logistic and normal, discussed in Section 7.1 and used in the simulations of this paper. Assumption [A4] regulates the rate at which the "tuning" parameter $K(T)$ can grow and, in the context of Andrews and Soares (2010) discussed in Subsection 2.1, enables consistent selection of binding constraints. Forms of tuning are also used by Chernozhukov et al. (2007) and Linton et al. (2010). [A2] enables smoothing for asymptotic normality through zero-valued μ_j, whilst [A6] creates data-driven importance weighting in the sense that each $\widehat{\mu}_j$ corresponding to strictly positive μ_j is likely to contribute ever less to the value of the test statistic as T increases. In consequence, the statistic will be asymptotically dominated by those $\widehat{\mu}_j$ corresponding to zero or negative μ_j, detection of which is the very purpose of the test.

To implement the test, we have to construct the term Λ_T in (2.3) of Subsection 2.2. Though Assumptions [A2], [A4] and (3.1) above are given so that, for $\mu_j = 0$, $\sqrt{T}\Psi_T(\widehat{\mu}_j)\widehat{\mu}_j$ in (2.3) is asymptotically equivalent to $\Psi(0)\sqrt{T}\widehat{\mu}_j$, the difference $\sqrt{T}\Psi_T(\widehat{\mu}_j)\widehat{\mu}_j - \Psi(0)\sqrt{T}\widehat{\mu}_j$ remains non-positive in large samples. Whilst asymptotically negligible, this may be size-distorting in finite samples. To systematically offset that effect, the adjustment term Λ_T is constructed as follows to approximate the expectation of $[\Psi_T(\widehat{\mu}_j) - \Psi(0)]\sqrt{T}\widehat{\mu}_j$.

Under Assumption [A2], there are finite increasing values $a_1, ..., a_n$ for some $n \geq 1$ such that $\Psi(x)$ is continuously differentiable in intervals $(-\infty, a_1), (a_1, a_2), ..., (a_n, \infty)$. Because Ψ is bounded and non-increasing, its one-sided limits $\Psi(a_i^-) \equiv \lim_{x \longrightarrow a_i^-} \Psi(x)$ and $\Psi(a_i^+) \equiv \lim_{x \longrightarrow a_i^+} \Psi(x)$ for $i \in \{1, 2, ..., n\}$ exist. Let $\widetilde{\psi}(x)$, $x \in R$ be the "extended" derivative of Ψ defined as the left-hand limit of $\psi(x)$. Namely, $\widetilde{\psi}(x) \equiv \lim_{y \longrightarrow x^-} \psi(y)$. Then the algebraic form of Λ_T whose detailed derivation is given in Appendix A can be written as

$$\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj}) = \widehat{v}_{jj} \widetilde{\psi}(K(T)\widehat{\mu}_j)K(T)/\sqrt{T} - \sqrt{\widehat{v}_{jj}} \sum_{i=1}^{n} (\Psi(a_i^-) - \Psi(a_i^+))\phi(\frac{a_i \sqrt{T}}{\sqrt{\widehat{v}_{jj}}K(T)}) \qquad (3.2)$$

where ϕ is the standard normal density function.

For the simple choice $\Psi(x) = 1\{x \leq 1\}$ used to form the statistic (2.4), $\widetilde{\psi} = 0$ and there is a single discontinuity at $x = 1$ so the proxy simplifies to

$$\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj}) = -\sqrt{\widehat{v}_{jj}} \phi(\frac{\sqrt{T}}{\sqrt{\widehat{v}_{jj}}K(T)}). \qquad (3.3)$$

On the other hand, for everywhere continuously differentiable Ψ, $\widetilde{\psi}(x) = \psi(x)$ for $x \in R$ and $\Psi(a_i^-) = \Psi(a_i^+)$ for $i \in \{1, 2, ..., n\}$. Hence Λ_T for such case simplifies to

$$\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj}) = \widehat{v}_{jj} \psi(K(T)\widehat{\mu}_j)K(T)/\sqrt{T}. \qquad (3.4)$$

Note that since Ψ is non-increasing, for any T, $\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj})$ given by (3.2) is non-positive by construction. Besides, under Assumption [A4] $\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj})$ tends to zero in probability as T tends to infinity. Hence for those $\mu_j \neq 0$, the impact of adjusting $\sqrt{T}\Psi_T(\widehat{\mu}_j)\widehat{\mu}_j$ with the term $\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj})$ on test behavior is asymptotically negligible though the adjustment (3.2) is applied for each $j \in \{1, 2, .., p\}$.

Finally, we consider a further useful generalization by replacing each $\widehat{\mu}_j$ in (2.3) with $\widehat{\theta}_j \widehat{\mu}_j$ for any positive scalar $\widehat{\theta}_j$, which can be fixed known or estimated. Choosing $\widehat{\theta}_j$ to be inverse of the estimated asymptotic standard deviation of $\widehat{\mu}_j$ amounts to conducting the test on t-ratios. Other choices of $\widehat{\theta}_j$ are discussed in Appendix C which deals with estimator covariance singularity issues. With this enhancing feature, the adjustment term $\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj})$ is replaced by $\Lambda_T(\widehat{\theta}_j \widehat{\mu}_j, \widehat{\theta}_j^2 \widehat{v}_{jj})$. We now present the test procedure as follows.

Let $\widehat{\Psi}, \widehat{\Lambda}, e_p$ be the p dimensional column vectors and $\widehat{\Delta}$ be the diagonal matrix defined as

$$\widehat{\Psi} \equiv (\Psi(K(T)\widehat{\theta}_1 \widehat{\mu}_1), \Psi(K(T)\widehat{\theta}_2 \widehat{\mu}_2), ..., \Psi(K(T)\widehat{\theta}_p \widehat{\mu}_p))', \qquad (3.5)$$

$$\widehat{\Lambda} \equiv (\Lambda_T(\widehat{\theta}_1 \widehat{\mu}_1, \widehat{\theta}_1^2 \widehat{v}_{11}), \Lambda_T(\widehat{\theta}_2 \widehat{\mu}_2, \widehat{\theta}_2^2 \widehat{v}_{22}), ..., \Lambda_T(\widehat{\theta}_p \widehat{\mu}_p, \widehat{\theta}_p^2 \widehat{v}_{pp}))', \qquad (3.6)$$

$$e_p \equiv (1, 1, ..., 1)', \qquad (3.7)$$

$$\widehat{\Delta} \equiv diag(\widehat{\theta}_1, \widehat{\theta}_2, ..., \widehat{\theta}_p). \qquad (3.8)$$

Let

$$Q_1 \equiv \sqrt{T}\widehat{\Psi}'\widehat{\Delta}\widehat{\mu} - e_p'\widehat{\Lambda} \tag{3.9}$$

$$Q_2 \equiv \sqrt{\widehat{\Psi}'\widehat{\Delta}\widehat{V}\widehat{\Delta}\widehat{\Psi}}. \tag{3.10}$$

We define the test statistic as

$$Q = \begin{cases} \Phi(Q_1/Q_2) & \text{if } Q_2 > 0 \\ 1 & \text{if } Q_2 = 0 \end{cases} \tag{3.11}$$

where $\Phi(x)$ is the standard normal distribution function. For asymptotic significance level α, we reject H_0 if $Q < \alpha$. The test statistic Q is therefore a form of tail probability or p-value.

We now sketch the reasoning which validates the test. Formal theorems are given later. Intuitively, we should reject H_0 if Q_1 is too small. For those parameter points under H_0 for which the probability limit of Q_2 is nonzero, Q_2 will be strictly positive with probability approaching one. Then the ratio Q_1/Q_2 will exist and be asymptotically normal. By contrast, for all points under H_1, the value of Q_1 will go in probability to minus infinity. Therefore, in cases where Q_2 is positive, we propose to reject H_0 if Q_1/Q_2 is too small compared with the normal distribution.

Note that our assumptions on the smoothed indicators do not rule out discrete but origin-smooth Ψ functions such as the step-at-unity example of Section 7.1. For such a discrete function, $\widehat{\Psi}$ will be a null vector with probability approaching one when all μ_j, $j \in \{1, 2, ..., p\}$, are strictly positive. In this case, Q_2 is also zero by (3.10) with probability approaching one. Therefore, occurrence of the event $Q_2 = 0$ is possible and signals that we should not reject H_0. Note that it is not an adhoc choice to set $Q = 1$ when $Q_2 = 0$ occurs because the probability limit of $\Phi(Q_1/Q_2)$ is also one when all μ_j parameters are strictly positive and Ψ is an everywhere positive function.[3]

4 Distributional Assumptions and Asymptotic Null Distribution

We begin by stating the following high-level assumptions which enable us to derive some basic asymptotic properties of the test. Except for [D2], these assumptions are standard.

Define Δ as the diagonal matrix $\Delta \equiv diag(\theta_1, \theta_2, ..., \theta_p)$ where θ_j is strictly positive and its estimator $\widehat{\theta}_j$ is almost surely strictly positive for $j \in \{1, 2, ..., p\}$. Let $d(\mu)$ be defined as the p dimensional vector whose jth element equals 0, $\Psi(0)$, 1 when $\mu_j > 0$, $\mu_j = 0$, $\mu_j < 0$

[3] The case of Ψ being everywhere positive is more complicated because Q_2 can then be almost surely strictly positive. If all μ_j parameters are strictly positive, both numerator and denominator in the ratio Q_1/Q_2 tend to zero in probability. See Appendix B.4 for analysis of the asymptotic properties of the test statistic Q in that case.

respectively. For notational simplicity, we keep implicit the possible dependence of the true values of the parameters μ, V and Δ on the underlying data generating process.

We assume that, as T tends to infinity,

[D1] $\sqrt{T}(\widehat{\mu}-\mu) \xrightarrow{d} N(0,V)$ *where* V *is some finite positive semi-definite matrix.*

The variance V need not be invertible but must satisfy the following condition (whose verification is illustrated in Appendix C).

[D2] $V\Delta d(\mu) \neq 0$ *for non-zero* $d(\mu)$.

Assumption [D2] amounts to saying that the asymptotic distribution of$\sqrt{T}d(\mu)'\Delta(\widehat{\mu}-\mu)$ should not be degenerate.

[D3] $\widehat{V} \xrightarrow{p} V$ *for some almost surely positive semi-definite estimator* $\widehat{V}$.

[D4] $\widehat{\Delta} \xrightarrow{p} \Delta$.

Now let J denote the set $\{1,2,...,p\}$ and decompose this as $J = A \cup M \cup B$, where

$$A \equiv \{j \in J : \mu_j > 0\},\ M \equiv \{j \in J : \mu_j = 0\},\ B \equiv \{j \in J : \mu_j < 0\}.$$

Let $U(0,1)$ denote a scalar random variable that is uniformly distributed in the interval $[0,1]$. We now present the asymptotic null distribution of the test statistic.

Theorem 1 (Pointwise Asymptotic Null Distribution) *Given [A1], [A2], [A3], [A4], [A6] with [D1] - [D4], the following are true under* $H_0 : \mu_j \geq 0$ *for all* $j \in J$ *with limits taken along* $T \longrightarrow \infty$.
(1) If $M \neq \varnothing$, *then* $Q \xrightarrow{d} U(0,1)$.
(2) If $M = \varnothing$, *then* $Q \xrightarrow{p} 1$.

Part (1) of this theorem reflects the fact that, for any fixed data generating process whose μ value lies on the boundary of null hypothesis space, the distribution of the test statistic Q is asymptotically non-degenerate and given (3.11), the limiting distribution of the ratio Q_1/Q_2 is standard normal. This justifies the idea of smoothing for normality. Moreover, Q has the same limiting distribution at each boundary point. Part (2) says that, at any fixed data generating process whose μ value lies in the interior of null hypothesis space, the asymptotic distribution of Q is degenerate and Q will take value above α with probability tending to 1.

5 Asymptotic Test Size

5.1 Pointwise and Uniform Asymptotic Control of Test Size

Theorem 1 shows that the test statistic Q is not asymptotically pivotal since its limiting distribution and hence the asymptotic null rejection probability depend on the true value of μ. By definition, the *pointwise* asymptotic size of the test is the supremum of the asymptotic rejection probability viewed as a function of μ on the domain defined by H_0. So Theorem 1 implies that this size equals the nominal level α and hence the test is asymptotically exact in the pointwise sense. However, pointwise asymptotic exactness is a weak property. It is desirable to ensure the convergence of the test size to the nominal level holds uniformly over the null-restricted parameter and data distribution spaces. In this section we present results showing that the test size is asymptotically exact in the uniform sense.

To distinguish between pointwise and uniform modes of analysis, we need some additional notation. Note that parameters such as μ and V are functionals of the underlying data generating distribution. Suppose the data consist of i.i.d. vectors x_t $(t = 1, ..., T)$ drawn from a joint distribution G. We henceforth use the notation $P_G(.)$ to make explicit the dependence of probability on G. Let Γ denote the set of all possible G compatible with prior knowledge or presumed specification of the data generating process. Then Assumptions [D1] - [D4] amount to restrictions characterizing the class Γ. Let Γ_0 be the subset of Γ that satisfies the null hypothesis. In the present test procedure, "$Q < \alpha$" is synonymous with "Q rejects H_0". Hence, the rejection probability of the test is $P_G(Q < \alpha)$ and the finite sample test size is $\sup_{G\in\Gamma_0} P_G(Q < \alpha)$.

Though Theorem 1 implies that convergence of rejection probability is not uniform over $G \in \Gamma_0$, the test can be shown to be uniformly asymptotically level α (Lehmann and Romano (2005, p. 422)) in the sense that

$$\limsup_{T\longrightarrow\infty} \sup_{G\in\Gamma_0} P_G(Q < \alpha) \le \alpha. \tag{5.1}$$

Inequality (5.1) and Part (1) of Theorem 1 together imply the test size is asymptotically exact in the uniform sense that

$$\limsup_{T\longrightarrow\infty} \sup_{G\in\Gamma_0} P_G(Q < \alpha) = \alpha. \tag{5.2}$$

The property (5.2) is important for the asymptotic size to be a good approximation to the finite-sample size of the test.[4] Such uniformity property has been emphasized in recent literature (e.g., see Mikusheva (2007), Romano and Shaikh (2008), Andrews and Guggenberger (2009) and Andrews and Soares (2010)) particularly when limit behavior of the test statistic can be discontinuous. Accordingly, we establish the validity of (5.2) in Theorem 2.

[4] Note that the notion of asymptotic test size using $\limsup_{T\longrightarrow\infty} \sup_{G\in\Gamma_0} P_G(Q < \alpha)$ is stronger than its pointwise version $\sup_{G\in\Gamma_0} \limsup_{T\longrightarrow\infty} P_G(Q < \alpha)$. See Lehmann and Romano (2005, p. 422) for an illustrating example in which pointwise asymptotic size can be a very poor approximation to the finite sample test size.

Before presenting the formal regularity conditions ensuring (5.2), we explain here how (5.2) is possible despite asymptotic non-pivotality of the test statistic. First note that by (3.11),

$$P_G(Q < \alpha) \leq P_G(Q_1 - z_\alpha Q_2 < 0) \tag{5.3}$$

where z_α is the α quantile of the standard normal distribution. The transformed statistic $(Q_1 - z_\alpha Q_2)$ is still not asymptotically pivotal but it can be shown that, given any arbitrary sufficiently small (relative to model constants) positive scalar η, we have with probability at least $(1-\eta)$ for all sufficiently large T that

$$Q_1 - z_\alpha Q_2 \geq r_T' \sqrt{T}(\widehat{\mu} - \mu) - (z_\alpha c_2(\eta) + c_1(\eta))\sqrt{r_T' V r_T}$$

where r_T, μ and V are non-stochastic G-dependent quantities such that either $r_T = 0$ or $r_T' V r_T$ is bounded away from zero over $G \in \Gamma_0$, whilst $c_1(\eta)$ and $c_2(\eta)$ are non-stochastic functions that do not depend on G and $c_1(\eta) \longrightarrow 0$ and $c_2(\eta) \longrightarrow 1$ as $\eta \longrightarrow 0$. Therefore,

$$P_G(Q_1 - z_\alpha Q_2 < 0) \leq P_G(r_T' \sqrt{T}(\widehat{\mu} - \mu) < (z_\alpha c_2(\eta) + c_1(\eta))\sqrt{r_T' V r_T}) + \eta \tag{5.4}$$

whose right hand will tend, uniformly over G giving non-zero r_T, to $\Phi(z_\alpha c_2(\eta) + c_1(\eta)) + \eta$ which is also automatically a weak upper bound on (5.4) for the case $r_T = 0$. This uniformly valid probability bound therefore applies to (5.3) for arbitrarily small η hence implies that (5.1) holds. Equality is obtained by invoking Theorem 1 which says α is actually attained as the limit of $P_G(Q < \alpha)$ evaluated at any fixed $G \in \Gamma_0$ whose μ has at least a zero-valued element.

The explanation provided above is indicative but short of a formal proof. In the next subsection we present additional "uniform" assumptions, strengthening the existing "pointwise" assumptions [D1] - [D4] of Section 4, that are needed to make the argument rigorous. The full proof, along with examples to illustrate some of the assumptions, will be found in the Appendix B.

5.2 Uniform Asymptotic Exactness of Test Size

In this section we rigorously address the issue of asymptotic exactness of test size in the uniform sense given by (5.2). For this purpose, we strengthen Assumptions [D1] - [D4] by the following Assumptions [U1] - [U4] where objects such as $K(T)$ have already been defined in Assumptions [A1] - [A6]. Define the vector Y and the scalar δ_T as

$$Y \equiv \sqrt{T}(\widehat{\mu} - \mu), \qquad \delta_T \equiv \sqrt{K(T)/\sqrt{T}}.$$

Note that Assumption [A4] implies that $\delta_T \longrightarrow 0$ as $T \longrightarrow \infty$. For any matrix m, let $\|m\| \equiv \max\{|m_{ij}|\}$ where m_{ij} denotes the (i,j)-th element of m.

Assumption [U1] : *For any finite scalar value* $\eta > 0$,

$$\lim_{T\longrightarrow\infty} \inf_{G\in\Gamma_0} P_G(\delta_T \|Y\| < \eta,\ ||\widehat{V} - V_G|| \ < \eta) = 1.$$

Assumption [U2] : *Let* $\Phi(.)$ *denote the standard normal distribution function. Then given any finite scalar* c,

$$\lim_{T\longrightarrow\infty} \sup_{G\in\Gamma_0} \sup_{\beta:\beta' V_G \beta=1} |P_G(\beta' Y \leq c) - \Phi(c)| \ = 0. \tag{5.5}$$

To illustrate how the high-level Assumptions [U1] and [U2] may be verified, consider the leading example where $\widehat{\mu}$ and $\widehat{V}$ are the sample mean and variance of i.i.d. random vectors x_t, $(t = 1, 2, ..., T)$ with joint distribution G.[5] Then the simple but not necessarily the weakest primitive condition guaranteeing both Assumptions [U1] and [U2] is that the first four moments of every element of x_t exist and are bounded uniformly over $G \in \Gamma_0$. This condition allows the application of the Chebychev inequality to components of the right-hand side of the inequality

$$P_G(\delta_T \|Y\| < \eta,\ ||\widehat{V} - V_G|| \ < \eta) \geq P_G(\delta_T \|Y\| < \eta) + P_G(||\widehat{V} - V_G|| \ < \eta) - 1$$

to deduce that Assumption [U1] holds. To verify Assumption [U2] we first note that, by Lemma 4 proved in the Appendix, it is sufficient for (5.5) that

$$\lim_{T\longrightarrow\infty} |P_{G_T}(\beta_T' Y \leq c) - \Phi(c)| \ = 0 \tag{5.6}$$

for all non-stochastic sequences (G_T, β_T) satisfying $G_T \in \Gamma_0$ and $\beta_T' V_{G_T} \beta_T = 1$. By the i.i.d. assumption, $\beta_T' Y$ is $1/\sqrt{T}$ times the sum of T variates $\beta_T'(x_t - E_{G_T}(x_t))$ which are mutually i.i.d. with mean 0 and variance 1 for each T when $\beta_T' V_{G_T} \beta_T = 1$. This meets the requirements of the double array version of the classic Lindeberg-Feller central limit theorem thus establishing asymptotic unit normality of $\beta_T' Y$ hence verifying (5.6).

For the next assumption, recall that θ_j is the jth diagonal element of the matrix Δ. For notational simplicity, the general dependence of θ_j and Δ on G will be kept implicit.

Assumption [U3] : *(i) There are finite positive scalars* λ *and* λ' *such that* $\lambda' \leq \theta_j \leq \lambda$, $(j = 1, 2, ..., p)$ *uniformly over* $G \in \Gamma_0$. *(ii) For any finite scalar value* $\eta > 0$,

$$\lim_{T\longrightarrow\infty} \inf_{G\in\Gamma_0} P_G(\left\|\widehat{\Delta} - \Delta\right\| < \eta\delta_T) = 1.$$

Assumption [U3] holds automatically when Δ is numerically specified by the user hence $\widehat{\Delta} = \Delta$. It also allows θ_j to be $1/\sqrt{v_{jj}}$ where v_{jj} is the jth diagonal element of V_G provided that

[5] This simple average framework is used extensively in recent literature on inference for (unconditional) moment inequality models. See, e.g. Chernozhukov et al. (2007), Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Andrews and Barwick (2012) and references cited therein.

v_{jj} is bounded below by some constant, say $L > 0$, uniformly over $G \in \Gamma_0$.[6] In such case,

$$\left|\widehat{\theta}_j - \theta_j\right| \leq |\widehat{v}_{jj} - v_{jj}| \sqrt{2} L^{-3/2} \tag{5.7}$$

when $|\widehat{v}_{jj} - v_{jj}| < L/2$.[7] Hence in the sample mean example described after Assumption [U2], we can verify [U3]-(ii) by applying the Chebychev inequality to show that $P_G(|\widehat{v}_{jj} - v_{jj}| < \eta\delta_T)$ also tends to 1 uniformly over $G \in \Gamma_0$.

For any given positive scalar σ, let $d_\sigma(\mu)$ denote the p dimensional vector whose jth element equals $\Psi(0)$ when $0 \leq \mu_j \leq \sigma$ and equals 0 otherwise.

Assumption [U4] : *There are finite positive real scalars ω, ω' and σ such that the following hold uniformly over $G \in \Gamma_0$: (i) $\|V_G\| < \omega$. (ii) $d_\sigma(\mu)'\Delta V_G \Delta d_\sigma(\mu) > \omega'$ for all non-zero $d_\sigma(\mu)$.*

Assumption [U4]-(i) is simply a boundedness assumption which automatically holds when V_G is a correlation matrix. [U4]-(ii) holds automatically when the smallest eigenvalue of V_G is bounded away from zero over $G \in \Gamma_0$. Note that [U4]-(ii), essentially strengthening Assumption [D2], requires that the asymptotic variance of $\sqrt{T} d_\sigma(\mu)'\Delta(\widehat{\mu} - \mu)$ be bounded away from zero for all non-zero $d_\sigma(\mu)$. This is a high level assumption whose verification will be illustrated in examples of Appendix C.

We can now present the following theorem establishing asymptotic exactness of the test in the uniform sense.

Theorem 2 (Uniform Asymptotic Exactness of Test Size) *Given Assumptions [D1] - [D4], suppose Assumptions [U1] - [U4] also hold. Assume some $G \in \Gamma_0$ has μ value containing at least one zero-valued element. Then under Assumptions [A1], [A2], [A3], [A4], [A6] and given $0 < \alpha < 1/2$,*

$$\limsup_{T \longrightarrow \infty} \sup_{G \in \Gamma_0} P_G(Q < \alpha) = \alpha.$$

6 Asymptotic Power of the Test

In this section, we study the asymptotic power properties of the test. Proof of all results are presented in the Appendix. For notational simplicity, we suppress the dependence of probability and parameters on the underlying data generating distribution. We first show that the test is consistent against fixed alternative hypotheses.

[6] Assumption [U3]-(ii) is stronger than requiring consistency of $\widehat{\theta}_j$ as an estimator of θ_j. An alternative approach is to strengthen Assumption [U2] by taking Y to be $\sqrt{T}(\widehat{\Delta\mu} - \Delta\mu)$ rather than just $\sqrt{T}(\widehat{\mu} - \mu)$. But that would be implicitly assuming $\sqrt{T}(\widehat{\theta}_j - \theta_j)$ is asymptotically normal (or degenerate). Such an assumption is even stronger than [U3]-(ii) and quite unnecessary for our results.

[7] By mean value expansion, $\left|\widehat{\theta}_j - \theta_j\right| = |\widehat{v}_{jj} - v_{jj}| / (2|\overline{v}_{jj}|^{3/2})$ where $\overline{v}_{jj}$ lies between $\widehat{v}_{jj}$ and v_{jj}. Thus when $|\widehat{v}_{jj} - v_{jj}| < L/2$, inequality (5.7) follows by noting that $|\overline{v}_{jj} - v_{jj}| \leq |\widehat{v}_{jj} - v_{jj}|$.

Theorem 3 (Consistency) *Given [A1] - [A6] with [D1] - [D4], the following is true under* $H_1 : \mu_j < 0$ *for some* $j \in \{1, 2, ..., p\}$.

$$P(Q < \alpha) \longrightarrow 1 \quad as \quad T \longrightarrow \infty.$$

Besides consistency, we are also interested in the local behavior of the test. In order to derive a local power function, we consider a sequence of μ values in the alternative-hypothesis space tending at rate $T^{-1/2}$ to a value $\gamma \equiv (\gamma_1, \gamma_2, ..., \gamma_p)'$ on the boundary of the null-hypothesis space. Specifically, we represent the jth element of μ of such a local sequence as

$$\mu_j = \gamma_j + \frac{c_j}{\sqrt{T}} \tag{6.1}$$

where $\gamma_j \geq 0$ and c_j are constants such that $\gamma_j = 0$ and $c_j < 0$ hold simultaneously for at least one j. The sequence (6.1) is said to be *core* if $c_j < 0$ holds in every instance of $\gamma_j = 0$. A core local sequence corresponds to Neyman-Pitman drift in the original sense (McManus (1991)) whereby parameter values conflicting with the null hypothesis are imagined *ceteris paribus* to draw ever closer to compliance as T increases. In the easily-visualized case $p = 2$, all points on the boundary of null-restricted space are limits of core sequences. Non-core sequences can only converge to the origin, a single point compared to the continuum of the full boundary. We may now state :

Theorem 4 (Local Power) *Assume [A1], [A2], [A3], [A4], [A6] and [D1], [D3], [D4] hold with the elements* μ_j *of* μ *taking the T-dependent forms as specified by (6.1). Define*

$$\begin{aligned}
\tau &\equiv \sum_{j=1}^{p} 1\{\gamma_j = 0\}\theta_j c_j \\
\kappa &\equiv \sum_{i=1}^{p}\sum_{j=1}^{p} 1\{\gamma_i = 0\}1\{\gamma_j = 0\}\theta_i\theta_j v_{ij}
\end{aligned}$$

where v_{ij} *denotes the* (i, j)*-th element of variance matrix* V. *Assume* $\kappa > 0$. *Then, as* $T \longrightarrow \infty$,

$$P(Q < \alpha) \longrightarrow \Phi(z_\alpha - \kappa^{-1/2}\tau), \tag{6.2}$$

where z_α *is the* α *quantile of the standard normal distribution.*

Theorem 4 implies that the test has power exceeding size against all core sequences because the composite drift parameter τ is necessarily negative for such local scenarios. By contrast, tests based on LFC critical values can be biased against core local sequences tending to boundary points off the origin. This is easily seen for statistics such as EV and QLR which are continuous in their arguments. In such cases, local power under any core sequence (6.1) tends to rejection probability at the boundary point $\mu = (\gamma_1, \gamma_2, ..., \gamma_p)'$. Unless this point is the LFC itself, rejection probability there will be smaller than that at any LFC point by definition. Hence the LFC critical value based test is biased against core local alternatives. A similar argument is given in Hansen (2003, 2005).

Against non-core local sequences, our test can be biased because a trade-off comes into force between negative and positive c_j as Theorem 4 shows. Some degree of local bias is common in multivariate one-sided tests and exists even in GMS procedures using estimated rather than LFC test critical values, as noted by Andrews and Soares (2010, p.146, comment (vi)). However, the exact local direction at which a test exhibits strength or weakness may vary across tests. Therefore, different tests are complementary rather than competing. To obtain a formal result, we consider a local sequence converging to the origin, namely $\gamma_j = 0$ for $j \in \{1, 2, ..., p\}$. Let c denote the vector $(c_1, c_2, ..., c_p)'$. Under such a local scenario, the GMS procedure will asymptotically treat all inequalities as binding in the critical value calculation. Thus the asymptotic distribution of the statistic $S(\sqrt{T}\widehat{\mu}, \widehat{V})$ of Subsection 2.1 is the same as that of $S(Z + c, V)$ and the test rejection probability tends to

$$P(S(Z + c, V) > q_\alpha) \tag{6.3}$$

where q_α is the $(1-\alpha)$ quantile of $S(Z, V)$ under $Z \sim N(0, V)$. We now present a theorem showing that the test of this paper is locally most powerful for a non-empty subclass of directions. Let θ denote the vector of diagonal elements of the matrix Δ.

Theorem 5 *Suppose the variance matrix V is positive definite and $\gamma_j = 0$ for $j \in \{1, 2, ..., p\}$ in the local sequence (6.1). Then for every testing function $S(.,.)$ such that $P(S(Z, V) > q_\alpha) = \alpha$ under $Z \sim N(0, V)$, the asymptotic local power in (6.2) is at least α and is not smaller than (6.3) when $c = -\delta V \theta$ for any positive scalar δ.*

Depending on the off-diagonal elements of V, the local directions $-\delta V \theta$ can be for either core or non-core sequences.[8] Theorem 5 implies that along such local alternatives, the present test is not biased and its limiting local power is not dominated by those of existing tests based on GMS critical values. Note that the result of Theorem 5 does not require specification of particular functional forms of $S(.,.)$. It is achieved by indirectly exploiting the Neyman-Pearson lemma. Some special forms are used in Section 7 for numerical illustration.

7 Monte Carlo Simulation Studies

In this section we conduct a series of Monte Carlo simulations to study the finite sample performance of the test. All tables of simulation results are placed together at the end of the section.

[8]Note that the vector $-\delta V \theta$ necessarily contains at least one negative element since V is positive definite, θ is a positive vector and δ is a postive scalar.

7.1 The Specification of Smoothed Indicator

Our objective is to investigate how well the asymptotic theory of the test works in finite sample simulations. For this purpose, we choose Ψ functions which are simple, recognized and not contrived. It would be premature at this stage to undertake a more elaborate exercise to find an optimal combination of $\Psi(x)$ and $K(T)$.

For the specification of Ψ, the following functions are heuristic choices that are widely adopted in research on smoothed threshold crossing models.

$$\begin{aligned} \textit{Normal} \quad &: \quad \Psi_{Nor}(x) \equiv 1 - \Phi(x) \\ \textit{Logistic} \quad &: \quad \Psi_{Log}(x) \equiv (1 + \exp(x))^{-1} \end{aligned}$$

Besides Ψ_{Nor} and Ψ_{Log}, the following simple choice of Ψ, mentioned in Section 2.2, is also valid.

$$\textit{Step-at-unity} : \Psi_{Step}(x) \equiv 1\{x \leq 1\}$$

As regards the choice of $K(T)$, the following two specifications closely match tuning parameters used in recent literature on inference of moment inequality models (See e.g. Chernozhukov et al. (2007) and Andrews and Soares (2010)). These choices are

$$\begin{aligned} \textit{SIC} \quad &: \quad K_{SIC}(T) \equiv \sqrt{T/\log(T)} \\ \textit{LIL} \quad &: \quad K_{LIL}(T) \equiv \sqrt{T/(2\log\log(T))} \end{aligned}$$

The first name reflects a connection with the Schwarz Information Criterion (SIC) for model selection and the second with the Law of the Iterated Logarithm (LIL).

7.2 The Simulation Setup

The simulation experiments are designed as follows. We choose a nominal test size of $\alpha = 0.05$. We use $R = 10000$ replications for simulated rejection probabilities. In each replication, we generate i.i.d. observations $\{x_t\}_{t=1}^{T}$ with $T = 250$ according to the following scheme :

$$x_t = \mu + V^{1/2} w_t \tag{7.1}$$

where w_t is a p dimensional random vector whose elements are i.i.d. from distribution G_w.

We compute $\widehat{\mu}$ and $\widehat{V}$ as the sample average and sample variance of the generated data. We take the scalars $\theta_j = 1/\sqrt{v_{jj}}$ and $\widehat{\theta}_j = 1/\sqrt{\widehat{v}_{jj}}$ where v_{jj} and $\widehat{v}_{jj}$ are the jth diagonal elements of V and $\widehat{V}$ respectively. This simple simulation setup is also adopted by Andrews and Soares (2010) and Andrews and Barwick (2012) in simulation study of the GMS tests. For G_w, we consider three distributions: standard normal, logistic and $U(-1, 2)$, the uniform distribution on the interval $[-1, 2]$. All of these distributions are centered and scaled such that $E(w_{t,j}) = 0$ and $Var(w_{t,j}) = 1$ for $j \in \{1, 2, ..., p\}$. Standard normality of G_w is the benchmark. The logistic distribution has thicker tails than the normal whilst the support of a uniform distributed random variate is bounded. The latter two distributions are included to assess the test performance under finite sample non-normality of $\widehat{\mu}$. For comparison, we also conduct simulations using the following test statistics:

$$
\begin{aligned}
S_1 &= -\min\{\sqrt{T}\widehat{\theta}_1\widehat{\mu}_1, \sqrt{T}\widehat{\theta}_2\widehat{\mu}_2, ..., \sqrt{T}\widehat{\theta}_p\widehat{\mu}_p, 0\}, \\
S_2 &= \min_{\mu:\mu\geq 0} T(\widehat{\mu} - \mu)'\widehat{V}^{-1}(\widehat{\mu} - \mu), \\
S_3 &= \sum_{j=1}^{p}(\min\{\sqrt{T}\widehat{\theta}_j\widehat{\mu}_j, 0\})^2, \\
S_4 &= \sum_{j=1}^{p}[-\sqrt{T}\min(\widehat{\theta}_j\widehat{\mu}_j, 0)].
\end{aligned}
$$

The extreme value form S_1 is essentially Hansen (2005)'s test statistic appropriated for testing multiple non-negativity hypotheses. S_2 is the classic QLR test statistic. S_3 is the modified-method-of-moments (MMM) statistic considered in the literature of moment inequality models (see, e.g. Chernozhukov et al. (2007), Romano and Shaikh (2008), Andrews and Guggenberger (2009) and Andrews and Soares (2010)). S_4 is the raw sum-of-negative-part statistic which can be transformed by smoothing into the key component of the test of the present paper.

The critical values for tests based on S_1 to S_4 are estimated using bootstrap coupled with the GMS procedure of the elementwise t-test type as suggested by Andrews and Soares (2010) and Andrews and Barwick (2012). We use 10000 bootstrap repetitions for calculation of the GMS test critical values. The tuning parameter in the GMS procedure is set to be the SIC or LIL type (Andrews and Soares (2010, p. 131)). For ease of reference, let $S_j(SIC)$ and $S_j(LIL)$ denote the GMS test using statistic S_j with tuning SIC and LIL respectively. Furthermore, let $Q(\Psi, K)$ denote the present test implemented with its smoothed indicator specified by Ψ and K.

We consider simulation scenarios based on $p \in \{4, 6, 10\}$. For multivariate simulation design, we have to be more selective on the specifications of μ and V parameters of (7.1). Concerning the μ vector, we follow a design similar to that previously employed by Hansen (2005, p. 373) in simulation study of the test size performance. To be specific, μ is the p dimensional vector given by

$$\mu_1 = 0, \mu_j = \lambda(j-1)/(p-1) \quad \text{for } p \geq j \geq 2$$

where $\lambda \in \{0, 0.25, 0.5\}$. Note that the λ values are introduced to control the extent to which inequalities satisfying the null hypothesis are in fact non-binding. Regarding the variance matrix

V, we set V to be a Toeplitz matrix with elements $V_{i,j} = \rho^{j-i}$ for $j \geq i$, where $\rho \in \{0, -0.5, 0.5\}$. This greatly simplifies the specification for off-diagonal elements of V but still allows for presence of various degrees of both positive and negative correlations.

For power studies, we consider the μ vector given by

$$\mu = -\delta V\theta + \epsilon\widetilde{\mu} \tag{7.2}$$

where $\delta \in \{0.15, 0.1, 0.05\}$, V is the variance matrix given as above, $\theta = (\theta_1, \theta_2, ..., \theta_p)'$, $\epsilon \in \{0, 0.5, 0.8\}$ and $\widetilde{\mu}$ is the vector with $\widetilde{\mu}_j = \delta$ for $1 \leq j \leq p/2$ and $\widetilde{\mu}_j = -\delta$ for $p/2 < j \leq p$. For $\epsilon = 0$, the design (7.2) mimics the local direction as suggested by Theorem 5 under which the test $Q(\Psi, K)$ is expected to outperform other tests. When ϵ is non-zero, the local direction in favor of the present test is perturbed with another vector $\widetilde{\mu}$ containing mixture of positive and negative elements. Such $\widetilde{\mu}$ may incur power trade-off in light of Theorem 4 and thus the perturbation parameter ϵ controls the degree of deviation toward $\widetilde{\mu}$ and enables some sensitivity check of test power performance.

7.3 Simulation results

We report the simulated maximum null rejection probability (MNRP) and average power (AP) for each test. Given G_w, the maximization for the MNRP is over all H_0 compatible combinations of μ and ρ values whilst given both G_w and ϵ, the averaging for AP is over all H_1 compatible μ and ρ configurations. Table 1 lists the MNRP values in three block columns side by side for the three specifications of G_w. The AP values generated by three ϵ values are then listed separately for each G_w in Tables 2, 3 and 4.

In Table 1, the primary interest is how close the MNRP values are to the nominal 5% significance level, particularly in cases of over-rejecting. In that respect, we compare the percentage of values not exceeding 0.05, 0.055, 0.06, 0.065. These percentages are about 18, 51, 87, 96 for the 54 $Q(\Psi, K)$ values and 9, 52, 79, 94 for the 72 values of the GMS tests. Plainly, the $Q(\Psi, K)$ test is no more prone to over-rejection than the GMS tests. A common feature across all tests is that over-rejection tends to increase with p. However, only 2 out of 54 $Q(\Psi, K)$ entries and 4 out of 72 GMS entries exceed 0.065. These excesses amount to less than 5% of a table of 126 simulated entries.

We now examine the sensitivity of MNRP to the underlying data generating distribution G_w. For all tests, Table 1 exhibits little systematic difference attributable to the three different specifications of G_w. These figures suggest that the MNRP results are not sensitive to finite sample non-normality. Furthermore, for each test, regardless of G_w, Table 1 suggests that use of SIC type tuner in place of the LIL can yield better control of test size. This finding is consistent with the simulation studies of Andrews and Soares (2010, pp. 149-152) demonstrating that the SIC tuner tends to give better MNRP properties. Overall, $Q(\Psi_{Step}, K_{SIC})$ and $Q(\Psi_{Log}, K_{SIC})$ have better MNRP results among the class of $Q(\Psi, K)$ tests and their size performance is comparable to that of the four SIC tuned GMS tests.

We now turn to Tables 2, 3, 4 giving AP results of the tests. For the unperturbed direction ($\epsilon = 0$), Theorem 5 of Section 6 indicates that the $Q(\Psi, K)$ test is locally more powerful than the GMS tests considered in the simulations. Along such local direction, irrespective of the underlying G_w, the simulation results indicate that the $Q(\Psi, K)$ tests dominate the GMS tests in AP performance. The GMS QLR test (S_2) is not far behind. Hansen's test (S_1), which is arguably the most stable in terms of MNRP performance, has distinctly lower power. But it is still a good performer. For the perturbed directions ($\epsilon \in \{0.5, 0.8\}$), while the $Q(\Psi, K)$ tests still outperform the S_1 tests, they do not generally dominate other versions of the GMS tests but the AP differences are not large.

We comment on the comparative performance of the $Q(\Psi, K)$ tests with the S_4 tests. Their comparison is of particular interest since the present test essentially attempts to smooth the statistic S_4. The smoothed version is less costly in computation because its critical value is obtained without resampling. We compare $S_4(SIC)$ with $Q(\Psi_{Step}, K_{SIC})$ and $Q(\Psi_{Log}, K_{SIC})$. The simulation results suggest that the $Q(\Psi_{Step}, K_{SIC})$ and $Q(\Psi_{Log}, K_{SIC})$ tests have similar degree of size control as $S_4(SIC)$. Against the alternative hypothesis, $Q(\Psi_{Log}, K_{SIC})$ has slightly larger power than $S_4(SIC)$ in all 27 cases while $Q(\Psi_{Step}, K_{SIC})$ outperforms $S_4(SIC)$ in 18 out of the 27 cases. These findings suggest that implementational advantage of the present test based on smoothing does not appear to be achieved at the cost of test performance.

Perusing all the other entries in Tables 2, 3, 4, it seems that the different variants of the $Q(\Psi, K)$ test perform quite similarly to one another retaining power well in excess of 0.73 throughout. What these results illustrate is that the $Q(\Psi, K)$ test has identifiable directions of strength as indicated theoretically by this paper. Given the simulation results above, the $Q(\Psi_{Step}, K_{SIC})$ and $Q(\Psi_{Log}, K_{SIC})$ tests work at least as well as other $Q(\Psi, K)$ versions examined here but have better size performance. Hence while K_{SIC} is the preferred tuner, both Ψ_{Step} and Ψ_{Log} are the recommended smoothers.

Table 1 : Simulated Maximum Null Rejection Probability for $T = 250$

DGP G_w	$N(0,1)$			*Logistic*			$U(-1,2)$		
Number of inequalities	4	6	10	4	6	10	4	6	10
$Q(\Psi_{Step}, K_{SIC})$	.049	.056	.055	.052	.054	.056	.051	.052	.055
$Q(\Psi_{Log}, K_{SIC})$	.046	.053	.055	.046	.054	.057	.048	.052	.058
$Q(\Psi_{Nor}, K_{SIC})$	.050	.059	.061	.050	.058	.063	.050	.056	.063
$Q(\Psi_{Step}, K_{LIL})$	.051	.059	.059	.053	.056	.059	.051	.053	.057
$Q(\Psi_{Log}, K_{LIL})$	.049	.056	.057	.048	.057	.060	.048	.053	.059
$Q(\Psi_{Nor}, K_{LIL})$	.054	.062	.065	.052	.059	.066	.053	.058	.066
$S_1(SIC)$	.050	.052	.054	.049	.052	.053	.051	.052	.053
$S_2(SIC)$	.050	.054	.053	.052	.055	.054	.050	.050	.054
$S_3(SIC)$	.050	.056	.052	.050	.051	.057	.052	.052	.056
$S_4(SIC)$	.051	.058	.054	.053	.054	.057	.052	.055	.058
$S_1(LIL)$	.053	.055	.055	.051	.054	.056	.054	.054	.056
$S_2(LIL)$	.058	.061	.061	.059	.063	.063	.058	.058	.061
$S_3(LIL)$	.056	.061	.057	.055	.058	.065	.058	.058	.064
$S_4(LIL)$	.059	.068	.066	.060	.064	.070	.061	.065	.070

Table 2 : Simulated Average Power for $T = 250$, $G_w = N(0,1)$

	$\epsilon = 0$			$\epsilon = 0.5$			$\epsilon = 0.8$		
Number of inequalities	4	6	10	4	6	10	4	6	10
$Q(\Psi_{Step}, K_{SIC})$	.770	.837	.900	.773	.840	.904	.783	.849	.909
$Q(\Psi_{Log}, K_{SIC})$	.754	.827	.893	.783	.849	.910	.813	.872	.927
$Q(\Psi_{Nor}, K_{SIC})$	.741	.814	.882	.780	.845	.906	.817	.875	.928
$Q(\Psi_{Step}, K_{LIL})$	.752	.822	.886	.761	.830	.895	.780	.847	.906
$Q(\Psi_{Log}, K_{LIL})$	.748	.821	.888	.781	.847	.908	.815	.874	.928
$Q(\Psi_{Nor}, K_{LIL})$	.734	.807	.875	.778	.844	.903	.819	.876	.928
$S_1(SIC)$	.593	.626	.650	.699	.728	.761	.774	.803	.831
$S_2(SIC)$	.714	.781	.847	.784	.844	.901	.834	.887	.937
$S_3(SIC)$	.678	.735	.793	.750	.804	.858	.805	.854	.899
$S_4(SIC)$	.730	.794	.855	.767	.830	.886	.808	.864	.913
$S_1(LIL)$	.594	.626	.650	.700	.729	.762	.776	.805	.832
$S_2(LIL)$	.716	.782	.848	.785	.846	.903	.836	.889	.939
$S_3(LIL)$	.678	.736	.794	.751	.805	.860	.808	.856	.902
$S_4(LIL)$	.732	.795	.857	.769	.833	.889	.811	.868	.916

Table 3 : Simulated Average Power for $T = 250$, $G_w = Logistic$

	$\epsilon = 0$			$\epsilon = 0.5$			$\epsilon = 0.8$		
Number of inequalities	4	6	10	4	6	10	4	6	10
$Q(\Psi_{Step}, K_{SIC})$	.772	.839	.900	.774	.841	.903	.781	.850	.910
$Q(\Psi_{Log}, K_{SIC})$	.757	.828	.893	.785	.851	.910	.813	.875	.929
$Q(\Psi_{Nor}, K_{SIC})$	.744	.815	.882	.781	.847	.906	.817	.878	.930
$Q(\Psi_{Step}, K_{LIL})$	.753	.824	.886	.763	.831	.894	.779	.848	.908
$Q(\Psi_{Log}, K_{LIL})$	.751	.823	.888	.783	.849	.908	.815	.876	.930
$Q(\Psi_{Nor}, K_{LIL})$	.738	.808	.874	.780	.845	.904	.819	.878	.930
$S_1(SIC)$	.599	.629	.651	.697	.729	.762	.775	.803	.831
$S_2(SIC)$	.718	.782	.847	.784	.845	.901	.834	.889	.938
$S_3(SIC)$	.681	.737	.794	.750	.803	.858	.806	.855	.901
$S_4(SIC)$	.734	.795	.854	.768	.830	.886	.807	.866	.915
$S_1(LIL)$	.600	.629	.651	.699	.730	.763	.777	.805	.833
$S_2(LIL)$	.719	.784	.849	.786	.846	.903	.837	.891	.940
$S_3(LIL)$	.682	.738	.796	.751	.805	.861	.808	.857	.903
$S_4(LIL)$	.735	.797	.856	.771	.833	.889	.811	.869	.919

Table 4 : Simulated Average Power for $T = 250$, $G_w = U(-1, 2)$

	$\epsilon = 0$			$\epsilon = 0.5$			$\epsilon = 0.8$		
Number of inequalities	4	6	10	4	6	10	4	6	10
$Q(\Psi_{Step}, K_{SIC})$	.769	.837	.899	.775	.842	.902	.782	.849	.908
$Q(\Psi_{Log}, K_{SIC})$	.754	.826	.892	.785	.850	.910	.812	.874	.926
$Q(\Psi_{Nor}, K_{SIC})$	.741	.813	.880	.781	.846	.906	.817	.876	.927
$Q(\Psi_{Step}, K_{LIL})$	.752	.821	.885	.763	.832	.894	.779	.847	.907
$Q(\Psi_{Log}, K_{LIL})$	.749	.820	.886	.784	.848	.908	.815	.876	.927
$Q(\Psi_{Nor}, K_{LIL})$	.735	.806	.873	.780	.844	.903	.819	.878	.928
$S_1(SIC)$	.594	.623	.652	.698	.727	.758	.773	.801	.830
$S_2(SIC)$	.715	.778	.846	.784	.843	.900	.834	.887	.937
$S_3(SIC)$	.678	.733	.793	.749	.803	.858	.805	.854	.899
$S_4(SIC)$	.730	.793	.852	.768	.831	.886	.807	.866	.914
$S_1(LIL)$	.594	.623	.652	.699	.728	.759	.775	.803	.831
$S_2(LIL)$	.716	.780	.848	.785	.845	.902	.836	.889	.939
$S_3(LIL)$	.679	.734	.794	.751	.805	.860	.807	.857	.901
$S_4(LIL)$	.731	.794	.853	.770	.833	.889	.811	.869	.918

8 Conclusions

This paper develops a test of multiple inequality hypotheses whose implementation does not require computationally intensive procedures. The test is based on origin-smooth approximation of indicators underlying the sum-of-negative-part statistic. This yields a simply structured statistic whose asymptotic distribution, whenever non-degenerate, is normal under the null hypothesis. Hence test critical values can be fixed ex ante and are essentially based on the unit normal distribution. Moreover, the test is applicable under weak assumptions allowing for estimator covariance singularity.

We have proved that the size of the test is asymptotically exact in the uniform sense. The test is consistent against all fixed alternative hypotheses. We have derived a local power function and used it to demonstrate that the test is unbiased against a wide class of local alternatives. We have also provided a new theoretical result pinpointing directions of alternatives for which the test is locally most powerful.

We have performed simulations which illustrate the potential of the test to be of practical inferential value along with simplicity and speed. These simulations, carried out for a range of p values, also shed light on the choice of smoothed indicator. They suggest that when coupled with the SIC type tuner, both the logistic and the step-at-unity smoothers perform well in finite samples. These are the recommended choices for test implementation. The simulation study also compares the test of this paper with several different tests which estimate critical values using the GMS procedure. We find that the test appears to be a viable complement to the GMS critical value estimation methodology.

References

[1] Andrews, D. W. K. and P. J. Barwick (2012), "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure", *Econometrica*, forthcoming.

[2] Andrews, D. W. K. and P. Guggenberger (2009), "Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities", *Econometric Theory*, 25, 669-709.

[3] Andrews, D. W. K. and G. Soares (2010), “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection”, *Econometrica*, 78, 119-157.

[4] Chen, L-Y. (2009), Econometric Inference Involving Discrete Indicator Functions: Dynamic Discrete Choice and Multiple Inequality Tests, PhD dissertation, Department of Economics, University College London.

[5] Chen, L-Y, and J. Szroeter (2006), "Constraint Chaining: A New Technique for Testing Multiple One-Sided Hypotheses", working paper, University College London.

[6] Chen, L-Y, and J. Szroeter (2009), "Hypothesis testing of multiple inequalities: the method of constraint chaining", Cemmap working paper, CWP13/09, Institute for Fiscal Studies: London.

[7] Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models", *Econometrica*, 75, 1243-1284.

[8] Dykstra, R. (1991), "Asymptotic Normality for Chi-Bar-Square Distributions", *Canadian Journal of Statistics*, 19, 297-306.

[9] Gourieroux, C. and A. Monfort (1995), Statistics and Econometric Models, Vol. 2. Cambridge University Press.

[10] Hansen, P.R. (2003), "Asymptotic Tests of Composite Hypotheses", Economics WP 2003-09, Brown University.

[11] Hansen, P.R. (2005), "A Test for Superior Predictive Ability", *Journal of Business and Economic Statistics*, 23, 365-380.

[12] Horowitz, J. (1992), "A Maximum Score Estimator for the Binary Response Model", *Econometrica*, 60, 505-531.

[13] Imbens, G. W. and Manski, C. F. (2004), "Confidence Intervals for Partially Identified Parameters", *Econometrica*, 72, 1845-1857.

[14] Kodde, D.A. and F.C. Palm (1986), "Wald Criteria for Jointly Testing Equality and Inequality Restrictions", *Econometrica*, 54, 1243-1248.

[15] Lee S. and Y-J Whang (2009), "Nonparametric tests of conditional treatment effects," Cemmap working paper, CWP36/09, Institute for Fiscal Studies: London.

[16] Lee S., Song K. and Y-J Whang (2011), "Testing functional inequalities," Cemmap working paper, CWP12/11, Institute for Fiscal Studies: London.

[17] Lehmann, E. and Romano, J. P. (2005), Testing Statistical Hypotheses, 3rd ed. New York : Springer.

[18] Linton, O., Song, K. and Y-J. Whang (2010), "An improved bootstrap test of stochastic dominance", *Journal of Econometrics*, 154, 186-202.

[19] McManus, D. (1991), "Who Invented Local Power Analysis ?", *Econometric Theory*, 7, 265-268.

[20] Menzel, K. (2008), "Estimation and Inference with Many Moment Inequalities", Unpublished Working Paper, Department of Economics, MIT.

[21] Mikusheva, A. (2007), "Uniform Inference in Autoregressive Models", *Econometrica*, 75, 1411 - 1452.

[22] Perlman, M.D. (1969), "One-Sided Testing Problems in Multivariate Analysis", *Annals of Mathematical Statistics*, 40, 549-567.

[23] Robertson, T., F. T. Wright, and R. L. Dykstra (1988), Order Restricted Statistical Inference, New York : Wiley.

[24] Romano, J. P. and Shaikh, A.M. (2008), "Inference for identifiable parameters in partially identified econometric models", *Journal of Statistical Planning and Inference*, 138, 2786-2807.

[25] Rosen, A. (2008), "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146, 107-117.

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

[27] Stoye, J. (2009), "More on Confidence Intervals for Partially Identified Parameters", *Econometrica*, 77, 1299-1315.

[28] White, H. (2000), "A Reality Check for Data Snooping", *Econometrica*, 68, 1097-1126.

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

[30] Wolak, F. (1988), "Duality in Testing Multivariate Hypotheses", *Biometrika*, 75, 611-615.

[31] Wolak, F. (1989), "Testing Inequality Constraints in Linear Econometric Models", *Journal of Econometrics*, 41, 205-235.

[32] Wolak, F. (1991), "The Local Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models", *Econometrica*, 59, 981-995.

A Supplementary Derivation of $\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj})$

The term $\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj})$ acts as an approximation for the expectation of $[\Psi_T(\widehat{\mu}_j) - \Psi(0)]\sqrt{T}\widehat{\mu}_j$ evaluated at $\mu_j = 0$. Under regularity condition [D1], when $\mu_j = 0$, the distribution of $\sqrt{T}\widehat{\mu}_j$ for T sufficiently large is approximately normal with mean zero and variance v_{jj}. Let X denote any scalar random variable distributed as $N(0,c)$. Define $h_T \equiv K(T)/\sqrt{T}$. Given (3.1), $\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj})$ is thus constructed to approximate $E((\Psi(h_T X) - \Psi(0))X) = E(\Psi(h_T X)X)$ with $c = v_{jj}$. In what follows, we take as read the notation and definitions stated between equations (3.1) and (3.2).

Define $a_0 \equiv -\infty$ and $a_{n+1} \equiv \infty$. Let ϕ denote the standard normal density function. Note that

$$\begin{aligned}
&E(\Psi(h_T X)X) \\
= \;& \sum_{i=1}^{n+1} \int_{a_{i-1}/h_T}^{a_i/h_T} \Psi(h_T x) x \phi(x/\sqrt{c})/\sqrt{c}dx \\
= \;& \sqrt{c}\left[\sum_{i=1}^{n+1} \int_{a_{i-1}/h_T}^{a_i/h_T} h_T \psi(h_T x)\phi(x/\sqrt{c})dx - \sum_{i=1}^{n}(\Psi(a_i^-) - \Psi(a_i^+))\phi(\frac{a_i}{h_T\sqrt{c}})\right] \qquad (A.1) \\
= \;& c h_T E(\widetilde{\psi}(h_T X)) - \sqrt{c}\sum_{i=1}^{n}(\Psi(a_i^-) - \Psi(a_i^+))\phi(\frac{a_i}{h_T\sqrt{c}}) \qquad (A.2)
\end{aligned}$$

where (A.1) follows from integration by parts and re-arrangement of terms in the sum and (A.2) follows by using [A2] which implies $\widetilde{\psi}(x) = \psi(x)$ almost everywhere. Taking $c = v_{jj}$ and plugging in the parameter estimates, we hence construct $\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj})$ as

$$\Lambda_T(\widehat{\mu}_j, \widehat{v}_{jj}) \equiv \widehat{v}_{jj}\widetilde{\psi}(K(T)\widehat{\mu}_j)K(T)/\sqrt{T} - \sqrt{\widehat{v}_{jj}}\sum_{i=1}^{n}(\Psi(a_i^-) - \Psi(a_i^+))\phi(\frac{a_i\sqrt{T}}{\sqrt{\widehat{v}_{jj}}K(T)}). \qquad (A.3)$$

We now comment on the derivative term in the expression (A.3). Since h_T goes to zero as T increases, $E(\widetilde{\psi}(h_T X))$ tends to $\psi(0)$ by Assumption [A2] and the Dominated Convergence Theorem. The limit value $\psi(0)$ also coincides with the probability limit of $\widetilde{\psi}(K(T)\widehat{\mu}_j)$ for the case $\mu_j = 0$. Hence, we use $\widetilde{\psi}(K(T)\widehat{\mu}_j)$ instead of $E(\widetilde{\psi}(h_T X))$ to account for the slope effect,[9] thus allowing the derivative term to depend on the estimate $\widehat{\mu}_j$. This has the advantage that for non-zero valued μ_j, $\widetilde{\psi}(K(T)\widehat{\mu}_j)$ itself also tends to zero and hence yields faster convergence of Λ_T to zero when the function Ψ further has the properties of $\lim_{x\longrightarrow -\infty}\psi(x) = \lim_{x\longrightarrow\infty}\psi(x) = 0$. Specifications of Ψ satisfying these properties are numerous, including the logistic and the normal smoothers given in Section 7.1.

[9] By taking $X \sim N(0,c)$ with $c = \widehat{v}_{jj}$, $E(\widetilde{\psi}(h_T X))$ can be computed using numerical integral as

$$\int_{-\infty}^{\infty} \widetilde{\psi}(h_T x)\phi(x/\sqrt{\widehat{v}_{jj}})/\sqrt{\widehat{v}_{jj}}dx.$$

B Proofs of Theoretical Results

The section presents proofs of all theoretical results stated in the paper. Proofs of Theorems 1, 3, 4 and 5 (pointwise asymptotics and local power) along with preliminary Lemmas 1, 2 and 3 are presented in Subsections B.1 - B.7. Proofs of Lemma 4 providing a sufficient condition for Assumption [U2] and Theorem 2 (uniform asymptotics) are given separately in Subsections B.8 and B.9 of the Appendix.

Recall that J denotes the set $\{1, 2, ..., p\}$ and the sets A, M, and B are defined as

$$A \equiv \{j \in J : \mu_j > 0\}, \ M \equiv \{j \in J : \mu_j = 0\}, \ B \equiv \{j \in J : \mu_j < 0\}.$$

B.1 Probability Limits of the Smoothed Indicator

We first prove a lemma that states the probability limits of the smoothed indicator $\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)$, which will be referred to in the proofs of some theorems in this paper.

Lemma 1 (Probability Limits of the Smoothed Indicator)

Assume [D1] and [D4]. Then the following results are valid as $T \longrightarrow \infty$.

(1) *If* $j \in A$ *and [A1], [A3], [A6] hold, then* $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j) \xrightarrow{p} 0$.

(2) *If* $j \in M$ *and [A2], [A4] hold, then* $\Psi_T(\widehat{\theta}_j\widehat{\mu}_j) \xrightarrow{p} \Psi(0)$.

(3) *If* $j \in B$ *and [A1], [A3], [A5] hold, then* $\Psi_T(\widehat{\theta}_j\widehat{\mu}_j) \xrightarrow{p} 1$.

Proof. To show part (1), for $\varepsilon > 0$ and for $\eta > 0$, we want to find some $\overline{T}(\varepsilon, \eta) > 0$ such that for $T > \overline{T}(\varepsilon, \eta)$,

$$P(\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j) \leq \varepsilon) \geq 1 - \eta.$$

By [D1] and [D4], we have $\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{p} \theta_j\mu_j$, which is strictly positive for $j \in A$. Then there is a $T_1(\eta)$ such that for $T > T_1(\eta)$,

$$P(\theta_j\mu_j/2 \leq \widehat{\theta}_j\widehat{\mu}_j \leq 3\theta_j\mu_j/2) \geq 1 - \eta.$$

Therefore, by [A1] and [A3] we have

$$\begin{aligned} 1 - \eta &\leq P(\Psi_T(3\theta_j\mu_j/2) \leq \Psi_T(\widehat{\theta}_j\widehat{\mu}_j) \leq \Psi_T(\theta_j\mu_j/2)) \\ &\leq P(\Psi_T(\widehat{\theta}_j\widehat{\mu}_j) \leq \Psi_T(\theta_j\mu_j/2)) \\ &\leq P(\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j) \leq \sqrt{T}\Psi_T(\theta_j\mu_j/2)) \end{aligned}$$

where the first inequality follows because Ψ is a non-increasing function. [A6] implies that $\sqrt{T}\Psi_T(\theta_j\mu_j/2) \longrightarrow 0$ as $T \longrightarrow \infty$. Therefore, there is some $T_2(\varepsilon)$ such that for $T > T_2(\varepsilon)$,

$\sqrt{T}\Psi_T(\theta_j\mu_j/2) < \varepsilon$. Combining all these results, part (1) in this lemma follows by choosing $\overline{T}(\varepsilon,\eta) = \max(T_1(\eta), T_2(\varepsilon))$.

To show part (2), note that If $j \in M$, by [D1] and [D4], we have $\sqrt{T}\widehat{\theta}_j\widehat{\mu}_j = Op(1)$. By [A4], $K(T)/\sqrt{T} = o(1)$ so that $K(T)\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{p} 0$. By [A2], Ψ is continuous at origin. Therefore, part (2) follows from the application of the continuous mapping theorem.

To show part (3), for $\varepsilon > 0$ and for $\eta > 0$, we want to find some $\overline{T}(\varepsilon,\eta) > 0$ such that for $T > \overline{T}(\varepsilon,\eta)$,

$$P(1-\varepsilon \le \Psi_T(\widehat{\theta}_j\widehat{\mu}_j) \le 1+\varepsilon) \ge 1-\eta.$$

Following the proof given in part (1), we have that there is a $T_1(\eta)$ such that for $T > T_1(\eta)$

$$\begin{aligned} 1-\eta &\le P(\theta_j\mu_j/2 \le \widehat{\theta}_j\widehat{\mu}_j \le 3\theta_j\mu_j/2) \\ &\le P(\Psi_T(3\theta_j\mu_j/2) \le \Psi_T(\widehat{\theta}_j\widehat{\mu}_j) \le \Psi_T(\theta_j\mu_j/2)). \end{aligned}$$

Note that if $j \in B$, then $\theta_j\mu_j < 0$ and thus by [A5], $\Psi_T(\theta_j\mu_j/2) \longrightarrow 1$ and $\Psi_T(3\theta_j\mu_j/2) \longrightarrow 1$. Then there is some $T_3(\varepsilon)$ such that for $T > T_3(\varepsilon)$, $\Psi_T(\theta_j\mu_j/2) \le 1+\varepsilon$ and $\Psi_T(3\theta_j\mu_j/2) \ge 1-\varepsilon$. Therefore, part (3) follows by choosing $\overline{T}(\varepsilon,\eta) = \max(T_1(\eta), T_3(\varepsilon))$. ■

B.2 Asymptotic Properties of $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j$

Based on Lemma 1, we derive the asymptotic properties of the components corresponding to $j \in A$, $j \in M$, $j \in B$ of the sum $\sum_{j\in J}\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j$. The results are stated in the following lemma.

Lemma 2 (Asymptotic Properties of $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j$)

Let v_{jj} denote the jth diagonal element of V. Assume [D1] and [D4]. Then the following results are valid as $T \longrightarrow \infty$.

(i) *If $j \in A$ and [A1], [A3], [A6] hold, then $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{p} 0$.*

(ii) *If $j \in M$ and [A2], [A4] hold, then $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{d} N(0, (\Psi(0)\theta_j)^2 v_{jj})$.*

(iii) *If $j \in B$ and [A1], [A3], [A5] hold, then $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{p} -\infty$.*

Proof. Note that part (i) follows from [D1], [D4] and part (1) of Lemma 1. To show part (ii), by [D1] and [D4], if $j \in M$, we have that $\sqrt{T}\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{d} N(0, \theta_j^2 v_{jj})$. Therefore, part (ii) follows by applying part (2) of Lemma 1. To show part (iii), note that for $j \in B$,

$$\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j = \Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\sqrt{T}\widehat{\theta}_j(\widehat{\mu}_j - \mu_j) + \Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\sqrt{T}\widehat{\theta}_j\mu_j. \tag{B.1}$$

Therefore, part (iii) follows from the fact that by [D1], [D4] and part (3) of Lemma 1, the first term on the right hand side of (B.1) is $Op(1)$ and the second term goes to $-\infty$ in probability. ■

B.3 Asymptotic Properties of $\Lambda_T(\widehat{\theta}_j\widehat{\mu}_j, \widehat{\theta}_j^2\widehat{v}_{jj})$

The following lemma states the asymptotic properties of the adjustment term $\Lambda_T(\widehat{\theta}_j\widehat{\mu}_j, \widehat{\theta}_j^2\widehat{v}_{jj})$ defined by (3.2).

Lemma 3 (Asymptotic Properties of $\Lambda_T(\widehat{\theta}_j\widehat{\mu}_j, \widehat{\theta}_j^2\widehat{v}_{jj})$)

Assume [A1], [A2], [A4], [D3] and [D4]. Then for $j \in J$, $\Lambda_T(\widehat{\theta}_j\widehat{\mu}_j, \widehat{\theta}_j^2\widehat{v}_{jj}) \xrightarrow{p} 0$.

Proof. By [A1] and [A2] and the properties of standard normal density function, we find that

$$\left|\Lambda_T(\widehat{\theta}_j\widehat{\mu}_j, \widehat{\theta}_j^2\widehat{v}_{jj})\right| \leq \widehat{\theta}_j^2\widehat{v}_{jj}\frac{K(T)}{\sqrt{T}}\left[b_\Psi + \sqrt{2\widehat{\theta}_j^2\widehat{v}_{jj}\pi^{-1}}\frac{K(T)}{\sqrt{T}}\sum_{i=1}^{n} a_i^{-2}\right]$$

where b_Ψ denotes the finite positive bound on the derivative of Ψ given in Assumption [A2]. Note that [A2] also implies $a_i^2 > 0$ for each i. By [A4], [D3] and [D4], the right-hand side of the inequality above is $o_p(1)$ and thus Lemma 3 follows. ■

B.4 Proof of Theorem 1

Proof of part (1) :

By Lemma 3 and under H_0, the quantity Q_1 may be written as

$$Q_1 = \sum_{j\in A}\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j + \sum_{j\in M}\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j + o_p(1)$$

which, by part (i) of Lemma 2, is asymptotically equivalent in probability to merely

$$\sum_{j\in M}\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j.$$

which, by [D1], [D2], [D4] and part (2) of Lemma 1, is asymptotically normal with mean zero and strictly positive variance equal to $\Psi(0)^2\omega_M$ where $\omega_M \equiv d_M'\Delta V\Delta d_M$ in which d_M denotes the p dimensional vector whose jth element is unity for $j \in M$ but zero for $j \notin M$. Using similar arguments along with [D3], we also find that

$$Q_2 \equiv \sqrt{\widehat{\Psi}'\widehat{\Delta}\widehat{V}\widehat{\Delta}\widehat{\Psi}} \xrightarrow{p} \Psi(0)\omega_M^{1/2}.$$

From these results about Q_1 and Q_2 and the definition (3.11) of Q, we conclude that Q equals to $\Phi(Q_1/Q_2)$ with probability tending to 1 as $T \longrightarrow \infty$ and thus $Q \xrightarrow{d} U(0,1)$.

Proof of part (2) :

When M is empty yet H_0 holds, only the sums taken for $j \in A$ remain in the definitions of Q_1 and Q_2 hence the following analysis is confined to $j \in A$. We distinguish between smoothed indicators which are such that $\Psi_T(x) = 0$ for all T sufficiently large when $x > 0$ and smoothed indicators such that $\Psi_T(x)$ remains strictly positive for $x > 0$ for all T. In the former case, part (1) of Lemma 1 implies that $P(\Psi_T(\widehat{\theta}_j\widehat{\mu}_j) = 0) \longrightarrow 1$ for $j \in A$ and hence $P(Q_2 = 0) \longrightarrow 1$ and thus $P(Q = 1) \longrightarrow 1$.

Now we consider the latter case where $\Psi_T(x) > 0$ for $x > 0$ regardless of T. This happens for everywhere positive Ψ functions. Then the quantity $\widehat{\Upsilon}_j \equiv \widehat{\theta}_j\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)$ is almost surely strictly positive for all $j \in A$. By eigenvalue theory, for all T,

$$Q_2 \leq \sqrt{\widehat{\lambda}_{\max} \sum_{j \in A} \widehat{\Upsilon}_j^2} \leq \sqrt{p\widehat{\lambda}_{\max}} \max_{j \in A}\{\widehat{\Upsilon}_j\} \tag{B.2}$$

where $\widehat{\lambda}_{\max}$ is the largest eigenvalue of $\widehat{V}$. Note that (B.2) holds even if $Q_2 = 0$, which under current scenario could only happen because of singularity of $\widehat{V}$ and V. However, when $P(Q_2 = 0) \longrightarrow 1$, we have $P(Q = 1) \longrightarrow 1$ and thus part (2) of the theorem follows.

Note that for $j \in J$, equation (3.2) and Assumptions [A1] and [A2] imply that the term $\Lambda_T(\widehat{\theta}_j\widehat{\mu}_j, \widehat{\theta}_j^2\widehat{v}_{jj})$ is non-positive for all T. Hence, since all μ_j are positive by supposition, as $T \longrightarrow \infty$, by (3.9) we have that

$$Q_1 \geq \max_{j \in A}\{\widehat{\Upsilon}_j\} \min_{j \in A}\{\sqrt{T}\widehat{\mu}_j\}.$$

with probability tending to 1. Because the mapping from a positive semi-definite matrix to its maximum eigenvalue is continuous on the space of such matrices, by [D3] we have $\widehat{\lambda}_{\max} \xrightarrow{p} \lambda_{\max}$ where $\lambda_{\max}$ is the largest eigenvalue of V. By [D2], $0 < \lambda_{\max} < \infty$ and thus we have

$$Q_1/Q_2 \geq \min_{j \in A}\{\sqrt{T}\widehat{\mu}_j\}/\sqrt{p\widehat{\lambda}_{\max}}$$

with probability tending to 1 as $T \longrightarrow \infty$. Since $\sqrt{T}\widehat{\mu}_j$ goes to infinity as $T \longrightarrow \infty$ for $j \in A$, it follows that $Q = \Phi(Q_1/Q_2) \xrightarrow{p} 1$.

B.5 Proof of Theorem 3

Since rejection of H_0 occurs if $Q < \alpha$ for the test statistic (3.11), it suffices for consistency to show that under H_1, Q_2 goes in probability to some positive constant and Q_1 goes to minus infinity as $T \longrightarrow \infty$. By (3.5) and Lemma 1, the probability limit of $\widehat{\Psi}$ under H_1 is the p dimensional vector whose jth element is $[1\{\mu_j < 0\} + \Psi(0)1\{\mu_j = 0\}]$. Therefore, by [D3] and [D4]

$$Q_2 \equiv \sqrt{\widehat{\Psi}'\widehat{\Delta}\widehat{V}\widehat{\Delta}\widehat{\Psi}} \xrightarrow{p} \sqrt{d(\mu)'\Delta V\Delta d(\mu)},$$

which is strictly positive by the regularity condition [D2]. On the other hand, Lemma 2 implies that $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j$ is bounded in probability for $j \in J\backslash B$ but tends to negative infinity for $j \in B$. Furthermore, Lemma 3 implies that $\Lambda_T(\widehat{\theta}_j\widehat{\mu}_j, \widehat{\theta}_j^2\widehat{v}_{jj}) = o_p(1)$ for $j \in J$. Under H_1, B is non-empty and thus Q_1/Q_2 goes to $-\infty$ in probability and hence $P(Q < \alpha) \longrightarrow 1$ as $T \longrightarrow \infty$.

B.6 Proof of Theorem 4

Under the assumed form of local sequence (6.1), for all j we have

$$K(T)\widehat{\theta}_j\widehat{\mu}_j = (K(T)/\sqrt{T})\widehat{\theta}_j[\sqrt{T}(\widehat{\mu}_j - \mu_j) + c_j] + K(T)\widehat{\theta}_j\gamma_j$$

where $\gamma_j \geq 0$. In the case $\gamma_j = 0$, Assumptions [A4], [D1] and [D4] imply that $K(T)\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{p} 0$ as $T \longrightarrow \infty$. By [A2] and the continuous mapping theorem, this then implies that $\Psi(K(T)\widehat{\theta}_j\widehat{\mu}_j) \xrightarrow{p} \Psi(0)$. On the other hand, if $\gamma_j > 0$, (6.1) implies that there is some $\delta > 0$ such that $\mu_j > \gamma_j - \delta > 0$ for all T sufficiently large. So under [A1], [A3], [A6], [D1] and [D4], we have that $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{p} 0$ by using arguments closely matching the proof of part (1) of Lemma 1.

Therefore, from these results and by (6.1), [D1], [D4] and Lemma 3, Q_1 is asymptotically equivalent in probability to

$$\Psi(0)\sum_{j=1}^{p} 1\{\gamma_j = 0\}\theta_j[\sqrt{T}(\widehat{\mu}_j - \mu_j) + c_j]$$

and thus has an asymptotic normal distribution with mean $\Psi(0)\tau$ and variance $\Psi(0)^2\kappa$. Using similar arguments, it is straightforward to see that $Q_2 \xrightarrow{p} \Psi(0)\sqrt{\kappa}$. Therefore, $Q_1/Q_2 \xrightarrow{d} N(\kappa^{-1/2}\tau, 1)$ from which the assertion of Theorem 4 follows.

B.7 Proof of Theorem 5

We shall establish that for any non-zero vector c,

$$\Phi(z_\alpha + \sqrt{c'V^{-1}c}) \geq P(S(Z + c, V) > q_\alpha) \tag{B.3}$$

holds for every testing function $S(.,.)$ such that $P(S(Z, V) > q_\alpha) = \alpha$ under $Z \sim N(0, V)$. The theorem then follows by noting that the left-hand side of (B.3) when $c = -\delta V\theta$ coincides with the power function (6.2) under the local direction specified by the theorem.

To show (B.3), consider an imaginary situation where X is the observable random vector that is distributed as $Z + \mu_X$ where $Z \sim N(0, V)$. For given V, a simple application of the Neyman-Pearson lemma (Lehmann and Romano (2005, p. 60, Theorem 3.2.1)) implies that a most powerful test at level α of the simple null hypothesis $\mu_X = 0$ versus the simple alternative $\mu_X = c$ is to reject the null if and only if $-c'V^{-1}X/\sqrt{c'V^{-1}c} < z_\alpha$. Hence (B.3) holds by noting that such test has power equal to $\Phi(z_\alpha + \sqrt{c'V^{-1}c})$ which is therefore not smaller than

$P(S(Z+c,V) > q_\alpha)$, the power of another test at level α which rejects the null hypothesis $\mu_X = 0$ if and only if $S(X,V) > q_\alpha$.

B.8 Sufficient Condition for Assumption [U2]

The following lemma provides a sufficient condition for Assumption [U2] of Section 5. Recall that $Y \equiv \sqrt{T}(\widehat{\mu} - \mu)$.

Lemma 4 *Assumption [U2] holds provided that given any finite scalar c,*

$$\lim_{T \longrightarrow \infty} |P_{G_T}(\beta_T' Y \leq c) - \Phi(c)| \; = 0 \tag{B.4}$$

for any sequence (G_T, β_T) satisfying $G_T \in \Gamma_0$ and $\beta_T' V_{G_T} \beta_T = 1$.

Proof. Let

$$f_T(G,\beta) \equiv |P_G(\beta' Y \leq c) - \Phi(c)|.$$

Let S denote the set $\{(G,\beta) : G \in \Gamma_0, \beta \in \Sigma(G)\}$ where the set $\Sigma(G) \equiv \{\beta \in R^p : \beta' V_G \beta = 1\}$. Note that

$$\sup_{G \in \Gamma_0} \sup_{\beta \in \Sigma(G)} f_T(G,\beta) = \sup_{(G,\beta) \in S} f_T(G,\beta). \tag{B.5}$$

Since for any $\varepsilon > 0$, there is a pair $(G_T(\varepsilon), \beta_T(\varepsilon))$ in S such that

$$\sup_{(G,\beta) \in S} f_T(G,\beta) < f_T(G_T(\varepsilon), \beta_T(\varepsilon)) + \varepsilon,$$

Assumption (B.4) used with equality (B.5) implies

$$\lim_{T \longrightarrow \infty} \sup_{G \in \Gamma_0} \sup_{\beta \in \Sigma(G)} f_T(G,\beta) < \varepsilon.$$

Hence Assumption [U2] follows by noting that ε is arbitrary chosen and $f_T \geq 0$. ■

B.9 Proof of Theorem 2

We aim to establish the inequality

$$\limsup_{T \longrightarrow \infty} \sup_{G \in \Gamma_0} P_G(Q < \alpha) \leq \alpha. \tag{B.6}$$

Then Theorem 2 follows by combining together the results implied by (B.6) and Part (1) of Theorem 1.

Let z_α be the α quantile of the standard normal distribution. The test rejects the null hypothesis if and only if $Q_2 > 0$ and $Q_1 - z_\alpha Q_2 < 0$. Therefore,

$$P_G(\text{reject } H_0) \leq P_G(Q_1 - z_\alpha Q_2 < 0). \tag{B.7}$$

The strategy of the proof is to demonstrate that $P_G(Q_1 - z_\alpha Q_2 < 0)$ is asymptotically bounded by the nominal size α uniformly for all G satisfying the null hypothesis. That then validates (B.6) via (B.7). Note that $-z_\alpha > 0$ for $0 < \alpha < 1/2$ as used in this theorem. By (3.9), (3.10) and non-positivity of the Λ_T term, we have

$$\begin{aligned} Q_1 &\geq \sum_{j=1}^{p} \Psi(K(T)\widehat{\theta}_j\widehat{\mu}_j)\sqrt{T}\widehat{\theta}_j\widehat{\mu}_j \\ Q_2 &= \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p} \Psi(K(T)\widehat{\theta}_i\widehat{\mu}_i)\Psi(K(T)\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_i\widehat{\theta}_j\widehat{v}_{ij}} \end{aligned}$$

where $\widehat{v}_{ij}$ and v_{ij} are the (i,j) elements of $\widehat{V}$ and V_G, respectively. For notational simplicity, the dependence of μ and v_{ij} on G is kept implicit.

Now we give details of the proof. For ease of presentation, they are organized in the following headed subsections.

1. Lower Bound for the Difference $(Q_1 - z_\alpha Q_2)$

Let $\delta_T \equiv \sqrt{K(T)/\sqrt{T}}$. For any $\eta > 0$, define the set

$$R_T(\mu) \equiv \{j : 0 \leq K(T)\mu_j \leq 2\eta\delta_T\}.$$

We show that, with probability tending to 1 uniformly over $G \in \Gamma_0$ as $T \longrightarrow \infty$,

$$Q_1 - z_\alpha Q_2 \geq Q_{1,R_T} - z_\alpha Q_{2,R_T} \tag{B.8}$$

where

$$\begin{aligned} Q_{1,R_T} &\equiv \sum_{j \in R_T(\mu)} \Psi(K(T)\widehat{\theta}_j\widehat{\mu}_j)\sqrt{T}\widehat{\theta}_j\widehat{\mu}_j, \\ Q_{2,R_T} &\equiv \sqrt{\sum_{i \in R_T(\mu)}\sum_{j \in R_T(\mu)} \Psi(K(T)\widehat{\theta}_i\widehat{\mu}_i)\Psi(K(T)\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_i\widehat{\theta}_j\widehat{v}_{ij}}. \end{aligned}$$

We follow the convention that summation over an empty set yields value zero. Note that (B.8) automatically holds when $R_T(\mu) = \{1, 2, ..., p\}$. For $R_T(\mu)$ being a proper subset of $\{1, 2, ..., p\}$, we rely on the fact (proved in the next subsection) that, with probability tending to 1 uniformly over $G \in \Gamma_0$ as $T \longrightarrow \infty$,

$$K(T)\widehat{\mu}_j > \eta\delta_T \text{ for } j \notin R_T(\mu) \tag{B.9}$$

and, for $R_T(\mu)$ nonempty,

$$Q_{2,R_T} > \sqrt{\omega'/2} > 0 \tag{B.10}$$

where ω' is the constant defined in Assumption [U4]-(ii). Let m be any index such that $m \notin R_T(\mu)$ and $\widehat{\theta}_m\widehat{\mu}_m \leq \widehat{\theta}_j\widehat{\mu}_j$ for all $j \notin R_T(\mu)$. Since Ψ is non-negative, (B.9) implies

$$Q_1 \geq Q_{1,R_T} + \Psi(K(T)\widehat{\theta}_m\widehat{\mu}_m)\widehat{\theta}_m\eta\delta_T^{-1}. \tag{B.11}$$

Furthermore, by [A1] the function Ψ is non-increasing and $\Psi \leq 1$. Thus, (B.9) and (B.10) together imply

$$|Q_{2,R_T} - Q_2| \leq \left|Q_{2,R_T}^2 - Q_2^2\right| / Q_{2,R_T} \leq p^2\Psi(K(T)\widehat{\theta}_m\widehat{\mu}_m)\left\|\widehat{\Delta}\right\|^2\left\|\widehat{V}\right\|\sqrt{2/\omega'}. \tag{B.12}$$

Given that $-z_\alpha > 0$, when $R_T(\mu)$ is empty, (B.11) alone implies (B.8). With $R_T(\mu)$ non-empty, (B.11) and (B.12) together imply (B.8) provided

$$\widehat{\theta}_m\eta\delta_T^{-1} \geq -z_\alpha p^2\left\|\widehat{\Delta}\right\|^2\left\|\widehat{V}\right\|\sqrt{2/\omega'}. \tag{B.13}$$

We show that under the null hypothesis, (B.9), (B.10) and (B.13) will indeed hold for η small enough and T large enough (yielding δ_T small enough by Assumption [A4]) under the key event E_T^η described next.

2. The Key Event E_T^η and Lower Bound for the Difference $(Q_{1,R_T} - z_\alpha Q_{2,R_T})$

Let Y_j be the jth element of $Y \equiv \sqrt{T}(\widehat{\mu} - \mu)$. For $\eta > 0$, define the event

$$E_T^\eta \equiv \{\delta_T\|Y\| < \eta, ||\widehat{V} - V_G|| < \eta, \left\|\widehat{\Delta} - \Delta\right\| < \eta\delta_T\}$$

which holds with probability tending to 1 uniformly over $G \in \Gamma_0$ as $T \longrightarrow \infty$ by Assumptions [A4], [U1] and [U3]-(ii). Since $K(T)\widehat{\mu}_j = K(T)\mu_j + \delta_T^2 Y_j$, under the null hypothesis the event E_T^η implies the inequality (B.9). To show that the event E_T^η also implies (B.10) and (B.13), and then derive the key result (B.18) of this subsection, we first need to draw out the following inequalities (B.14) - (B.17).

Note that when $0 \leq K(T)\mu_j \leq 2\eta\delta_T$, we have that by Assumption [U3]-(i) and under the event E_T^η,

$$\sqrt{T}\widehat{\theta}_j\widehat{\mu}_j \geq \theta_j Y_j - 3\eta^2, \tag{B.14}$$

$$\left|K(T)\widehat{\theta}_j\widehat{\mu}_j\right| \leq 3\eta\delta_T(\lambda + \eta\delta_T). \tag{B.15}$$

By Assumption [A2], $\Psi(x)$ is differentiable on $|x| \leq 3\eta\delta_T(\lambda + \eta\delta_T)$ for η small enough and T large enough. Therefore, given $\Psi \leq 1$, the event E_T^η and inequalities (B.14) and (B.15) imply that

$$\Psi(K(T)\widehat{\theta}_j\widehat{\mu}_j)\sqrt{T}\widehat{\theta}_j\widehat{\mu}_j \geq \Psi(0)\theta_j Y_j - 3(\lambda b_\Psi(\lambda + \eta\delta_T) + 1)\eta^2$$

where b_Ψ denotes the bound on the derivative of $\Psi(x)$ defined in Assumption [A2]. Hence, when $\eta < 1$ and $\delta_T < 1$, we may certainly write

$$Q_{1,R_T} \geq \Psi(0) \sum_{j\in R_T(\mu)} \theta_j Y_j - C_1\eta \tag{B.16}$$

where C_1 is a fixed positive quantity given values of p, λ and b_Ψ. By Assumptions [U3]-(i) and [U4]-(i) and using similar arguments with $\eta < 1$ and $\delta_T < 1$, we can obtain a bound for Q_{2,R_T}^2 under the event E_T^η as the following

$$Q_{2,R_T}^2 \geq \Psi(0)^2 \sum_{i\in R_T(\mu)} \sum_{j\in R_T(\mu)} \theta_i\theta_j v_{ij} - C_2\eta \tag{B.17}$$

where C_2 is fixed and positive given values of p, λ, ω, b_Ψ and $\Psi(0)$.

We can choose η to satisfy $\eta < \min\{1, \omega'/(2C_2)\}$ and choose T such that $2\eta\delta_T/K(T) < \sigma$, where σ is the constant defined in Assumption [U4] by which the right-hand side of (B.17) is larger than $\omega'/2$ and hence inequality (B.10) is satisfied. Using Assumptions [U3]-(i) and [U4]-(i), under the event E_T^η, we see $\widehat{\theta}_m > \lambda' - \delta_T\eta$ whilst $\left\|\widehat{\Delta}\right\|^2 \left\|\widehat{V}\right\| \leq (\lambda + \delta_T\eta)^2(\omega + \eta)$. Since $\delta_T^{-1} \longrightarrow \infty$ by Assumption [A4], given $\eta > 0$, (B.13) will indeed hold for large enough T . Finally, let r_T denote the p dimensional vector whose jth element is θ_j if $j \in R_T(\mu)$ and zero, otherwise. Then given that $-z_\alpha > 0$ and with η small enough and T large enough, (B.16) and (B.17) together imply

$$Q_{1,R_T} - z_\alpha Q_{2,R_T} \geq \Psi(0)r_T'Y - C_1\eta - z_\alpha\sqrt{\Psi(0)^2 r_T' V_G r_T - C_2\eta}. \tag{B.18}$$

3. The Probability Bounds

We have shown above how occurrence of the event E_T^η implies the inequality (B.8) given η small enough and T large enough. Hence

$$\begin{aligned} P_G(Q_1 - z_\alpha Q_2 < 0) &\leq 1 - P_G(E_T^\eta) + P_G(Q_1 - z_\alpha Q_2 < 0, E_T^\eta) \\ &\leq 1 - P_G(E_T^\eta) + P_G(Q_{1,R_T} - z_\alpha Q_{2,R_T} < 0) \end{aligned} \tag{B.19}$$

where the last term of (B.19) is zero when $R_T(\mu)$ is empty. For non-empty $R_T(\mu)$, using (B.18) yields

$$P_G(Q_{1,R_T} - z_\alpha Q_{2,R_T} < 0) \leq P_G(r_T'Y - z_\alpha\sqrt{r_T'V_G r_T - C_2\eta/\Psi(0)^2} < C_1\eta/\Psi(0)). \tag{B.20}$$

The probability in the right-hand side of (B.20) may be written as

$$P_G(\beta_T' Y < z_\alpha \widetilde{C}_{2,R_T} + \eta \widetilde{C}_{1,R_T}) \tag{B.21}$$

where

$$\begin{aligned} \beta_T &\equiv r_T/\sqrt{r_T' V_G r_T}, \\ \widetilde{C}_{1,R_T} &\equiv C_1/(\Psi(0)\sqrt{r_T' V_G r_T}), \\ \widetilde{C}_{2,R_T} &\equiv \sqrt{r_T' V_G r_T - C_2\eta/\Psi(0)^2}/\sqrt{r_T' V_G r_T}. \end{aligned}$$

Note that by [U4]-(ii), we have that with T large enough, $0 \leq \widetilde{C}_{1,R_T} \leq C_1/(\Psi(0)\sqrt{\omega'})$ and $\sqrt{1 - C_2\eta/\omega'} \leq \widetilde{C}_{2,R_T} \leq 1$. Hence, given $z_\alpha < 0$ and η small enough, the probability (B.21) cannot exceed

$$P_G(\beta_T' Y < z_\alpha \sqrt{1 - C_2\eta/\omega'} + C_1\eta/(\Psi(0)\sqrt{\omega'})). \tag{B.22}$$

Given the fact that β_T is non-stochastic with $\beta_T' V_G \beta_T = 1$, Assumption [U2] implies that given η, for any $\xi > 0$, there is a threshold $T^*(\eta, \xi)$ such that for $T > T^*(\eta, \xi)$, the probability (B.22) will be smaller than

$$\Phi(z_\alpha \sqrt{1 - C_2\eta/\omega'} + C_1\eta/(\Psi(0)\sqrt{\omega'})) + \xi$$

uniformly over all G obeying the null hypothesis. On the other hand, by Assumptions [A4], [U1] and [U3]-(ii) applied to the event E_T^η, for any $\varepsilon > 0$, there is a threshold $T^{**}(\eta, \varepsilon)$ such that for $T > T^{**}(\eta, \varepsilon)$, $P_G(E_T^\eta) > 1 - \varepsilon$ uniformly over all G obeying the null hypothesis. Putting together these facts and (B.19), (B.20), (B.22), we have that for $T > \max\{T^*(\eta, \xi), T^{**}(\eta, \varepsilon)\}$,

$$\sup_{G \in \Gamma_0} P_G(Q_1 - z_\alpha Q_2 < 0) \leq \Phi(z_\alpha \sqrt{1 - C_2\eta/\omega'} + C_1\eta/(\Psi(0)\sqrt{\omega'})) + \xi + \varepsilon$$

from which by letting $T \longrightarrow \infty$ in accordance with $T > \max\{T^*(\eta, \xi), T^{**}(\eta, \varepsilon)\}$ as the scalars η, ξ and ε approach zero, it follows that $\limsup_{T \longrightarrow \infty} \sup_{G \in \Gamma_0} P_G(Q_1 - z_\alpha Q_2 < 0) \leq \alpha$.

C Covariance Singularity Examples

In this appendix section, we present three examples of estimator covariance singularity for which the high level assumptions [D2] and [U4]-(ii) are verified. Recall that G is the joint distribution from which the underlying individual data vector is randomly sampled. Γ is the set of all possible G compatible with presumed specification of the data generating process and Γ_0 is the subset of Γ that satisfies the null hypothesis. All parameter values such as μ and V depend on the point G of evaluation but we keep that implicit to avoid notational clutter.

In the first two examples, the econometric model is initially characterized by an r dimensional vector of parameters $\beta \equiv (\beta_1, \beta_2, ..., \beta_r)'$. The restrictions being tested are synthesized into the one-sided form $\mu \geq 0$ with $\mu = (\mu_1, \mu_2, ..., \mu_p)' = C\beta + b$ where C is a known $p \times r$ matrix and b is a known p dimensional vector of constants. We assume an asymptotically normal estimator $\widehat{\beta}$ is available with non-singular asymptotic variance matrix Ω. Since $V = C\Omega C'$, V value induced by any $G \in \Gamma$ is necessarily singular when $r < p$. In the third example, we consider a different scenario where singularity arises only for some specific V values.

Example 1: Triangle Restriction

For a Cobb-Douglas production function with capital and labor elasticity coefficients β_1 and β_2, the restrictions being tested $\beta_1 \geq 0$, $\beta_2 \geq 0$ and $\beta_1 + \beta_2 \leq 1$ (non-increasing returns to scale) form a triangle for the graph of (β_1, β_2). Here $r = 2$, $p = 3$ and

$$\mu = (\mu_1, \mu_2, \mu_3)' = (\beta_1, \beta_2, 1 - \beta_1 - \beta_2)'. \tag{C.1}$$

Verification of [D2] and [U4]-(ii) : Note that $V = C\Omega C'$ where Ω is the variance matrix of the asymptotic distribution of $\sqrt{T}(\widehat{\beta} - \beta)$ and

$$C' = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}, \quad C'\Delta d(\mu) = \begin{bmatrix} \theta_1 & 0 & -\theta_3 \\ 0 & \theta_2 & -\theta_3 \end{bmatrix} d(\mu).$$

We assume the primitive condition that the smallest eigenvalue of Ω is bounded away from zero over all $G \in \Gamma$. Assumption [D2] is true since $C'\Delta d(\mu)$ being zero for non-zero $d(\mu)$ would require all elements of $d(\mu)$ to be non-zero, in turn requiring all elements of μ given by (C.1) to be negative or zero, which is impossible. For Assumption [U4]-(ii), we note that for sufficiently small σ, the only non-zero values for $d_\sigma(\mu)$ possible under the null hypothesis are $\Psi(0)$ multiples of $(1,0,0)'$, $(0,1,0)'$, $(0,0,1)'$, $(1,1,0)'$, $(1,0,1)'$, $(0,1,1)'$, because it is not possible for more than two of the elements of μ to simultaneously lie between 0 and $\sigma < 1/3$ as $\mu_1 + \mu_2 + \mu_3 = 1$. Therefore, given Assumption [U3]-(i) and the primitive condition on Ω, Assumption [U4]-(ii) is satisfied here.

Example 2: Interval Restrictions with Fixed Known End-Points

Suppose the r dimensional parameter vector β is hypothesized to satisfy interval restrictions $l \leq \beta \leq u$, where l and u are numerically specified. In this case, $p = 2r$ and $\mu = ((\beta - l)', (u - \beta)')'$. An estimator $\widehat{\beta}$ is available such that $\sqrt{T}(\widehat{\beta} - \beta)$ is asymptotically normal with variance Ω whose smallest eigenvalue is assumed primitively to be bounded away from zero over all $G \in \Gamma$. Note that $V = C\Omega C'$ where $C' = [I_r, -I_r]$. Thus, $C'\Delta d(\mu)$ is the r dimensional vector whose jth element is

$$[1\{\beta_j < l_j\} + \Psi(0)1\{\beta_j = l_j\}]\theta_j - [1\{\beta_j > u_j\} + \Psi(0)1\{\beta_j = u_j\}]\theta_{j+r} \tag{C.2}$$

for $j \leq r$. We consider the following two cases of interval hypotheses.

Case I : All hypothesized intervals are non-degenerate

For Case I, the null hypothesis concerns only non-degenerate intervals in the sense that $u_j > l_j$ for all $j \leq r$.

Verification of [D2] and [U4]-(ii) for null hypothesis given by Case I : Note that under H_1, $\beta_j < l_j$ or $\beta_j > u_j$ for some $j \leq r$ and thus (C.2) is either θ_j or $-\theta_{j+r}$ for some $j \leq r$. Hence $C'\Delta d(\mu)$ is non-zero and Assumption [D2] holds under the alternative hypothesis. We need to further show that $C'\Delta d(\mu)$ is not equal to zero for non-zero $d(\mu)$ under the null hypothesis. But under H_0, (C.2) simplifies to

$$\Psi(0)\left[1\{\beta_j = l_j\}\theta_j - 1\{\beta_j = u_j\}\theta_{j+r}\right]. \tag{C.3}$$

for all $j \leq r$. Given that $u_j > l_j$ for all j, there is some j such that expression (C.3) equals either $\Psi(0)\theta_j$ or $-\Psi(0)\theta_{j+r}$ whenever $d(\mu)$ is non-zero under the null hypothesis. Hence, Assumption [D2] is verified.

We now verify the high level assumption [U4]-(ii). Under the null hypothesis, the jth element of $C'\Delta d_\sigma(\mu)$ is

$$\Psi(0)[1\{l_j + \sigma \geq \beta_j \geq l_j\}\theta_j - 1\{u_j \geq \beta_j \geq u_j - \sigma\}\theta_{j+r}]. \tag{C.4}$$

For $\sigma < \min_{j\in\{1,2,...,r\}}(u_j - l_j)/2$, if $d_\sigma(\mu)$ is a non-zero, then there is some j such that expression (C.4) equals either $\Psi(0)\theta_j$ or $-\Psi(0)\theta_{j+r}$ and thus $C'\Delta d_\sigma(\mu)$ is a non-zero vector of length which is bounded away from zero by Assumption [U3]-(i). Given the primitive eigenvalue assumption on Ω, this completes verification of Assumption [U4]-(ii).

Case II : At least one hypothesized interval is degenerate

For Case II, at least one interval is specified to be degenerate (i.e. $l_j = u_j$ for some $j \leq r$) in the null hypothesis. Let S_e denote the subset of $\{1, 2, ..., r\}$ such that $l_j = u_j$ holds for all $j \in S_e$ but $l_j < u_j$ for all $j \notin S_e$.

Verification of [D2] and [U4]-(ii) for null hypothesis given by Case II : Under H_1, Assumption [D2] holds by the same arguments as given in Case I. Under H_0, (C.3) becomes $\Psi(0)(\theta_j - \theta_{j+r})$ for all $j \in S_e$. In this case, Assumption [D2] still holds but the restriction that $\theta_j \neq \theta_{j+r}$ for at least one $j \in S_e$ has to be imposed. This extra restriction guarantees that $C'\Delta d(\mu)$ is not equal to zero for all non-zero $d(\mu)$ and thus [D2] is fulfilled.

We now verify the high level assumption [U4]-(ii). Note that [U4]-(ii) only concerns the null hypothesis under which (C.4) becomes $\Psi(0)(\theta_j - \theta_{j+r})$ for all $j \in S_e$. Therefore, provided that there is one $j \in S_e$ such that $|\theta_j - \theta_{j+r}|$ is bounded away from zero over all $G \in \Gamma_0$, then $C'\Delta d_\sigma(\mu)$ is also a non-zero vector of length which is bounded away from zero. Given the primitive condition on Ω, Assumption [U4]-(ii) is thus satisfied for any $\sigma > 0$.

We now comment on testing interval hypothesis of the Case II type within the framework of this paper. For validity of the test, it suffices to choose any single equality hypothesis indexed by $h \in S_e$ and specify $\theta_h \neq \theta_{h+r}$ at the outset. This single asymmetry requirement is the only operational difference compared with Case I. Moreover, since $v_{h,h} = v_{h+r,h+r}$ where $v_{h,h}$ denotes the h-th diagonal element of V, weighting inversely proportional to standard error is not ruled out. The user can indeed set

$$\theta_{h+r} = (1+\varepsilon)\theta_h \text{ with } \theta_h = 1/\sqrt{v_{h,h}},\ \varepsilon > -1 \text{ and } \varepsilon \neq 0. \tag{C.5}$$

Here ε is a non-stochastic quantity chosen by the user to control the degree of deviation from perfect standardization of the estimate $\widehat{\mu}_{h+r}$. The weighting scheme (C.5) ensures that the test has exact asymptotic size in the uniform sense and is consistent against all fixed alternatives. On the other hand, Theorem 4 suggests that the user can specify $\varepsilon < 0$ (or reverse) to attach more (or less) weight to detection of violation of H_0 in the direction of $\beta_h < l_h$.

Note that asymmetric weighting (C.5) adopted here can be viewed as "perturbing" both Q_1 and Q_2 from the values they would take under symmetry. One might think to perturb only Q_2 to ensure that singularity does not cause division by (near) zero. For example, one could perturb $\widehat{V}$ in the expression (3.10) defining Q_2 in a manner akin to Andrews and Barwick (2012) who adjust the QLR test statistic by perturbing $\widehat{V}$ with a diagonal matrix when the determinant of the correlation matrix induced by $\widehat{V}$ is smaller than some pre-specified threshold. This alternative approach can allow for symmetric weighting. However unperturbed Q_1 will asymptotically converge to zero and hence rejection probability will tend to zero under the null and local alternative scenarios where all non-degenerate interval inequalities are non-binding. By contrast, the procedure (C.5) perturbing both Q_1 and Q_2 in a balanced way ensures that the ratio Q_1/Q_2 stays asymptotically standard normal in the null even when the only binding constraints are the equality hypotheses. It thus enables non-zero test power to be retained in the aforementioned scenarios of local alternatives.

Example 3: Interval Restrictions with Unknown End-Points

In Example 2, testing the inequalities $l \leq \beta \leq u$ was performed on fixed known interval end-points. Suppose now that l and u are not known but are parameters which satisfy $l \leq u$ and can take a continuum of values including those which make $(u - l)$ arbitrarily close to zero as well as precisely zero. There is no point estimator for β but consistent estimators $\widehat{l}$ and $\widehat{u}$ are available having joint asymptotic normal distribution with variance matrix Ω. This, for the univariate case, is the scenario considered by Imbens and Manski (2004) and Stoye (2009). For clarity, we stay with the setup where β is a scalar. We consider testing $H_0 : l \leq \beta_0 \leq u$ for a numerically specified candidate value β_0 for β. We then take $\widehat{\mu} = (\beta_0 - \widehat{l}, \widehat{u} - \beta_0)'$ and $\mu = (\beta_0 - l, u - \beta_0)'$. The asymptotic distribution of $\sqrt{T}(\widehat{\mu} - \mu)$ is normal with variance

$$V = \begin{bmatrix} \Omega_{11} & -\Omega_{12} \\ -\Omega_{12} & \Omega_{22} \end{bmatrix}.$$

For any given l and u, there is no reason why V should be singular. However, Stoye (2009. p. 1304, Lemma 3) demonstrates that, if one insists on $P(\widehat{u} \geq \widehat{l}) = 1$ holding over the underlying data generating distribution space where the difference $(u-l)$ is bounded away from infinity and the elements Ω_{11} and Ω_{22} bounded away from zero and infinity, then V necessarily depends on $(u-l)$ in such a way that $\Omega_{12} - \Omega_{11} \longrightarrow 0$ and $\Omega_{22} - \Omega_{11} \longrightarrow 0$ as $u - l \longrightarrow 0$. Thus, singularity of V where $\Omega_{11} = \Omega_{22} = \Omega_{12}$ must be allowed for.

Verification of [D2] and [U4]-(ii) : For Assumption [D2], note that under the maintained assumption that $l \leq u$, the vector $d(\mu)$ can be non-zero only if it takes one of the following forms: $(1,0)'$, $(0,1)'$, $(\Psi(0),0)'$, $(0,\Psi(0))'$, $(\Psi(0),\Psi(0))'$. The first four of these cannot make $V\Delta d(\mu) = 0$. The last form can only occur when $l = \beta_0 = u$ in which case we have

$$V\Delta d(\mu) = \Psi(0)[\theta_1\Omega_{11} - \theta_2\Omega_{12}, -\theta_1\Omega_{12} + \theta_2\Omega_{22}]'. \tag{C.6}$$

Note that (C.6) is zero only if V is singular and $\theta_1/\theta_2 = \Omega_{12}/\Omega_{11} = \Omega_{22}/\Omega_{12}$. Singularity occurs in Stoye's scenario where the model allows for $\Omega_{11} = \Omega_{22} = \Omega_{12}$. Since the weights θ_1 and θ_2 are chosen by the user, we can use $\theta_1 = 1/\sqrt{\Omega_{11}}$ and $\theta_2 = (1+\varepsilon)/\sqrt{\Omega_{22}}$ where ε is a pre-specified non-stochastic and non-zero quantity satisfying $\varepsilon > -1$. Then Assumptions [D2] holds regardless of singularity of V. For Assumption [U4]-(ii), we only need to consider the null hypothesis. In this case, the possible forms of non-zero $d_\sigma(\mu)$ can take are $(\Psi(0),0)'$, $(0,\Psi(0))'$ and $(\Psi(0),\Psi(0))'$. It is easily seen that $d_\sigma(\mu)'\Delta V \Delta d_\sigma(\mu)$ equals $\Psi(0)^2$ for the first, $\Psi(0)^2(1+\varepsilon)^2$ for the second, and $\Psi(0)^2\left[\varepsilon^2 + 2(1+\varepsilon)(1 - \Omega_{12}/\sqrt{\Omega_{11}\Omega_{22}})\right]$ for the third form. Hence Assumption [U4]-(ii) holds.

In this example, the weights θ_1 and θ_2 are chosenly asymmetrically and setting ε to be greater (smaller) than zero amounts to attaching more (or less) weight to detection of violation of H_0 in the direction $u < \beta_0$. The ε-perturbation arguments adopted here are indeed based on those given in Case II of Example 2. The value of the perturbation parameter ε is a user's input to the test procedure. The choice does not affect validity of the results concerning asymptotic test size and consistency. Asymmetry does affect local power but, by the same device, offers the user an opportunity to input a subjective assessment of the relative importance of different directions of violation of the null hypothesis.