cemmap

Hypothesis testing of multiple inequalities: the method of constraint chaining

Le-Yu Chen
Jerzy Szroeter

Hypothesis Testing of Multiple Inequalities : The Method of Constraint Chaining[1]

Le-Yu Chen[a2] and Jerzy Szroeter[b]

[a]Institute of Economics, Academia Sinica
[b]Department of Economics, University College London

Revised: May 2009

Abstract

Econometric inequality hypotheses arise in diverse ways. Examples include concavity restrictions on technological and behavioural functions, monotonicity and dominance relations, one-sided constraints on conditional moments in GMM estimation, bounds on parameters which are only partially identified, and orderings of predictive performance measures for competing models. In this paper we set forth four key properties which tests of multiple inequality constraints should ideally satisfy. These are (1) (asymptotic) exactness, (2) (asymptotic) similarity on the boundary, (3) absence of nuisance parameters from the asymptotic null distribution of the test statistic, (4) low computational complexity and boostrapping cost. We observe that the predominant tests currently used in econometrics do not appear to enjoy all these properties simultaneously. We therefore ask the question : Does there exist any nontrivial test which, as a mathematical fact, satisfies the first three properties and, by any reasonable measure, satisfies the fourth ? Remarkably the answer is affirmative. The paper demonstrates this constructively. We introduce a method of test construction called chaining which begins by writing multiple inequalities as a single equality using zero-one indicator functions. We then smooth the indicator functions. The approximate equality thus obtained is the basis of a well-behaved test. This test may be considered as the baseline of a wider class of tests. A full asymptotic theory is provided for the baseline. Simulation results show that the finite-sample performance of the test matches the theory quite well.

KEYWORDS : Tests, multiple inequality constraints, exactness, similarity on the boundary, smoothing, chaining methods

JEL SUBJECT AREA : C1, C4

---

1 Introduction and related literature

The growing importance of research on testing econometric multiple inequality hypotheses cannot be disputed. Such hypotheses now arise in diverse contexts, both classical and at the frontier. This is illustrated by the following selection of six non-exclusive groupings in which we give literature examples where formal inference on multiple inequalities is called for. Additional references will be found within the papers listed. *Group 1* : Concavity and other restrictions characterizing consumer-producer optimality and technology. See Hazilla and Kopp (1986), Kodde and Palm (1987), Koebel, Falk and Laisney (2003), Reis (2006) and Wolak (2007). *Group 2* : Multiperiod inequalities due to liquidity-risk premiums and random walk effects in financial markets. See Richardson, Richardson and Smith (1992), Boudoukh, Richardson, Smith and Whitelaw (1999), Fisher, Willson and Xu (1998) and Fleming, Kirby and Ostdiek (2006). *Group 3* : Stochastic dominance and monotonicity restrictions. See Anderson (1996), Davidson and Duclos (2000), Barrett and Donald (2003), Post (2003), Linton, Maassoumi and Whang (2005), Linton, Song and Whang (2008) and Lee, Linton and Whang (2009). *Group 4* : One-sided constraints on (conditional) moments in GMM estimation. See Rosen (2008) and Andrews and Soares (2009). *Group 5* : Bounds on parameters which are only partially identified. See Guggenberger, Hahn and Kim (2008) and McAdams (2008). *Group 6* : Orderings of predictive performance measures for competing models. See White (2000), Hansen (2005), Hansen and Lunde (2005) and Martin, Reidy and Wright (2009). It will be noticed that the dates of the papers above fall within the last 20 or so years. This is a reflection of the fact that econometric theory for handling multiple inequalities has largely developed in that period of time and is still very much incomplete. The purpose of the present paper is to address four outstanding concerns in test design and to show constructively that there does exist a theoretically sound test which can simultaneously satisfy four key properties. That test is based on the novel concept of constraint chaining.

The articles cited above consider null hypotheses of the type where, for some finite or infinite subset S of the real line, the "true" values of parameters μ_r, $r \in S$, in a given econometric model are all nonnegative whilst the alternative is that at least one is negative. The subscript r may be simply an indexing argument but it could also be the value of an observable economic variable, in which case S may be a entire interval. This is the case in some papers of Group 3 above. In other groups, S is generally the set of integers

$$\{1, 2, \ldots, p\}$$

for some finite p. The baseline statistical properties we set forth shortly for tests of composite null hypotheses are relevant regardless of S. The paper conducts most of its specific analysis for finite p but that extends to the case where S is a continuum. We allow the vector

$$\mu \equiv (\mu_1, \mu_2, \ldots, \mu_p)'$$

to be functionally dependent on other parameters in the model. This framework is then sufficiently general to cover most finite multiple inequality contexts.

The difficulty of designing tests arises because simple one-sided testing of the single scalar parameter case ($p = 1$) does not carry over in an obvious way to the situation where $p \geq 2$. It is tempting to try to solve the problem by conducting a family of p individual tests each at reduced significance level in an attempt to keep the family-wide error rate (overall significance level) within a nominal value. However, since the actual overall significance level generally falls well short of the nominal, the family of tests is "inexact" and, being "conservative", is likely to have lower power than an efficient test which is designed to attain the nominal level. The family of tests (or "multiple comparisons") approach has useful features (Savin (1980, 1984), Hochberg and Tamhane (2007)) but we do not consider it in this paper since it does not enjoy the most important of the four key properties which set the standard in tests of composite null hypotheses. These are as follows.

Property 1 : *(Asymptotic) exactness.*
This property requires that the nominal test size is (asymptotically) achieved at some point of the null hypothesis (and not exceeded at any point).

Property 2 : *(Asymptotic) similarity on the boundary of a composite null hypothesis.*
Applied to the present context, this property requires that a test should have constant rejection probability at all points $\mu \geq 0$ where at least one element of μ is exactly zero. Theoretical work indicates that bias and inadmissibility of test, for example, can result from failure of Property 2. See Hansen (2003, 2005) and Linton, Song and Whang (2008).

Property 3 : *Absence of nuisance parameters from the (asymptotic) null distribution of the test statistic.*
The nuisance parameter (any parameter other than that subject to the null hypothesis) is usually a covariance matrix. When Property 3 fails, true test critical values cannot be evaluated even if the analytical form of the null distribution is known completely. Estimates of critical values can be obtained either by plugging consistent estimates of the nuisance parameters into the analytical formula (if available) for the critical values. Alternatively, one may directly estimate the critical values as quantiles of the bootstrap distribution of the test statistic. But the bootstrap samples must be adjusted to ensure that they mimic a population in which the null hypothesis holds (Politis, Romano and Wolf (1999, Section 1.8)). Using the bootstrap to eliminate nuisance parameters is quite separate from its potential in securing higher order improvement over first-order asymptotic approximation to the true finite-sample cdf of test statistics Indeed, the latter may be adversely affected by the former (Horowitz (2001)).

Property 4 : *Inexcessive computational complexity and boostrapping cost.*

Attempts to rescue tests from violation of Properties 1-3 may compromise Property 4.

In relation to the above properties, we proceed now to review the two predominant approaches used in econometrics for testing the validity of a set of inequality hypotheses. We initially assume that the exists an estimator $\widehat{\mu}$ based on sample size T such that

$$\sqrt{T}(\widehat{\mu} - \mu)$$

is (asymptotically) multivariate normal with mean 0 and covariance matrix V consistently estimated by $\widehat{V}$, both nonsingular. V and μ may depend on common parameters but this is generally kept implicit for notational simplicity.

The first test approach measures empirical discrepancy between unrestricted and inequality-restricted models using a quadratic form such as

$$\min_{\lambda:\lambda\geq 0} T(\widehat{\mu} - \lambda)'\widehat{V}^{-1}(\widehat{\mu} - \lambda). \tag{1}$$

This approach extends classical Wald, LM and LR equality tests to the context of multiple inequalities. The literature on this extension is vast and we therefore direct the reader to the key articles by Perlman (1969), Kodde and Palm (1986), Shapiro (1988), Wolak (1987, 1988, 1989a, 1989b, 1991), Dufour (1989) and the textbooks by Robertson, Wright and Dykstra (1988), Gourieroux and Monfort (1995, Chapter 27) and Silvapulle and Sen (2005, Chapters 3 and 4). The (asymptotic) distribution of (1) depends on the value of the DGP parameter μ from which data are drawn. For the case where V is known hence $\widehat{V} = V$, the work of Perlman (1969) and Wolak (1987, 1988) shows that the (asymptotic) cdf of (1) at the particular point $\mu = 0$ is of the so-called chi-bar squared distribution[3]

$$\sum_{r=0}^{p} w_{r,p}(V)F(x|r) \tag{2}$$

due to Kudo (1963), where $w_{r,p}(V)$ are nonnegative weights summing to 1 and $F(x|r)$ is the cdf of a central chi-square variate with r degrees of freedom. At boundary points of the composite null hypothesis $\mu \geq 0$, the cdf is also chi-bar but with p replaced by d, the number of binding constraints, and V replaced by a submatrix corresponding to those constraints[4]. Since the quantiles of this distribution vary with d, a test using quantiles as critical values calculated assuming $\mu = 0$ hence $d = p$ necessarily fails to have the desirable Property 2 (asymptotic similarity on the boundary) listed earlier. Nonetheless, in the

---

[3] The use of quadratic form statistic and chi-bar squared test is also related to the test problem of a simple null hypothesis against the one-sided alternative. See e.g., Kudo (1963), Nuesch (1966), Gourieroux, Holly and Monfort (1982), King and Smith (1986), Andrews (1998) and Feng and Wang (2007). See also Goldberg (1992) for discussions of the relation between the two testing problems in the context of the chi-bar squared approach.

[4] See equation (4.116) in Silvapulle and Sen (2004, pp 204).

case $\widehat{V} = V$, Property 1 (asymptotic exactness) still holds because Perlman (1969) proved that the particular value $\mu = 0$ is least favourable (LF) to the null hypothesis in the sense that the (asymptotic) rejection probability of the test based on the quadratic form statistic (1) is maximized over points satisfying $\mu \geq 0$ by the choice $\mu = 0$, provided that V is not functionally related to μ. Since the cdf of the test statistic at $\mu = 0$ is known, a critical value can in principle be calculated such that any nominal test size is attained. However, the weight $w_{r,p}(V)$ cannot be given a general algebraic form as it represents the probability that r elements out of a $p-$dimensional vector of correlated normal variables with covariance V take value in a cone determined by the constraints. Computation of these weights is non-trivial when p, the number of inequality constraints exceeds three. When p is large, numerical simulation of the weights or the tail probability of (2) is required. Silvapulle and Sen (2005, pp.78-81) provide a helpful recent review of available simulation techniques for the chi-bar squared distribution. Note that to ensure a reasonable estimate of the chi-bar squared distribution, iterations of constrained quadratic programming of (1) also have to be performed. This cycle of optimization and simulation could hardly be described as routine and has an adverse impact on Property 4.

There is unfortunately a further problem with the chi-bar squared test described above. When V is related to μ, which inevitably happens if μ is a nonlinear function of some underlying parameters, the choice $\mu = 0$ is no longer LF. In such cases, it is generally impossible to locate an LF value. One way out of this problem is to derive conservative bounds on the LF distribution. See, e.g., Kodde and Palm (1986). The bounds approach is not only popular, but it has even been proposed as necessary since, in most instances, it is quite difficult to obtain an empirically implementable asymptotically exact test, especially when the test is a NOS (nonlinear one-sided) hypothesis test. On the other hand, it runs the risk of inconclusive test results. Wolak (1991) discusses this problem in depth and points out that for tests with high dimensional inequality constraints, the bounds become very slack, making inconclusive test results more likely. The consequence is that a test based on the quadratic form statistic (1) using bounds of critical values will almost certainly not satisfy Property 1 (asymptotic exactness). Guggenberger, Hahn and Kim (2008) give an equivalence result which implies that the problem of testing specification of models containing moment inequality constraints involving partially identified parameters is equivalent to testing an appropriately-defined set of nonlinear multiple inequalities. Indeed Rosen (2008), using Wolak-style chi-bar theory in a partially identified setting, notes that his model specification test may be conservative. In a more general distributional setting also allowing partially identified parameters in moment inequalities, Andrews and Soares (2009, pp.22-23) warn that their GMS model specification test is conservative to unknown extent (hence may lose Property 1) in some cases. Finally, even when V is not related to μ, it is a nuisance parameter in the sense of Property 3 when its value is unknown. This causes the problems discussed in the paragraph following the statement of Property 3 and thus compromises Property 4.

The second predominant approach to testing the hypothesis $\mu \geq 0$, which

we call the extreme value (EV) method, is based on the statistic

$$\min\{\sqrt{T}\widehat{\mu}_1, \sqrt{T}\,\widehat{\mu}_2, \ldots, \sqrt{T}\widehat{\mu}_p\} \tag{3}$$

where $\widehat{\mu}_r$ denotes the r-th element of $\widehat{\mu}$. Recent heightened interest in this approach is due to White (2000). If precisely $d \geq 1$ elements of the true value of μ are zero (the rest being strictly positive), the asymptotic null distribution of (3) is that of the minimum element of a normal vector with mean 0 and covariance equal to an appropriate $d \times d$ submatrix of V. An analytic expression for the cdf of this distribution is not generally available, hence the cdf has to be obtained by simulation for each given V and d. That can be done to arbitrary accuracy but may impact on Property 4. Since the asymptotic null distribution of (3) depends on d, the test cannot satisfy Property 2. However, as long as V does not depend on μ, the cdf for the case $d = p$ is dominated by the cdfs for $d < p$. So the point $\mu = 0$ is least favourable hence Property 1 will hold. When V is related to μ, this conclusion is not sustained and we are faced with the problems discussed in connection with the chi-bar squared test. When V is not related to μ but its value is unknown, it is a nuisance parameter in the sense of Property 3 and the issues discussed in connection with that property come into play. But note that Property 4 is less affected than in the case of the chi-bar squared test whose statistic requires an inequality-constrained quadratic minimization coupled with V inversion.

The EV method of White has been modified in two ways by Hansen (2005) and Hansen and Lunde (2005). First, Hansen works with t-ratios of the $\widehat{\mu}_r$ rather than their raw values. Second, and more relevant to the four baseline properties we have set forth in this paper, White computes his critical values from an estimated null-hypothesis compliant cdf which corresponds to a population having mean $\mu = 0$, whereas Hansen's mean is data-driven having a particular element equal to zero only if the corresponding element of $\widehat{\mu}$ is negative or close to zero (the measure of closeness diminishing as sample size increases). Asymptotically, Hansen is consistently estimating the correct value of d, the true number of zeros to use in the asymptotic null distribution of the test statistic. In consequence, unlike White's test, Hansen's test will achieve the key Property 2. Nuisance parameter issues concerning V are the same as in White's test. The importance of estimating the number of binding inequality constraints (or "contact points" in the notion of Linton, Song, Whang (2008)) and the benefits of automatic data-driven components in test procedures has been further addressed in a number of more recent papers which include notably Linton, Song, Whang (2008), Andrews and Jia (2008), and Andrews and Soares (2009).

In the light of the above discussion, it is compelling to ask if there exists any approach which is theoretically guaranteed to satisfy Properties 1 through 3 as a mathematical fact and, by any reasonable measure, satisfies Property 4. For the last property, it is desired to develop a simple and reasonably effective test of multiple inequality constraints whenever the cost or complexity of alternative testing procedures is prohibitive. The test should be applicable under wide

assumptions and be easy to implement without need for bootstrapping or numerical simulation. Remarkably, the answer is affirmative. The rest of the present paper demonstrates this. We set out an approach to inequality test construction which is philosophically and operationally very different from the chi-bar squared and EV tests. The new idea is to chain multiple inequalities into a single equality by exploiting zero-one indicator functions. Thus testing multiple inequalities is reduced to testing a single equality. Then we smooth these functions to overcome the distribution problems of discrete functions of continuous variates. The smoothing embodies a data-driven importance weighting feature which at the outset automatically concentrates weight onto those parameter estimates having most inferential value. In respect of the role of smoothing, there is some affinity between our paper and the work of Horowitz (1992). The zero-one indicator in Horowitz's context represents a binary choice probability depending on model parameters. By working with a smoothed version of the indicator which approaches the original zero-one with increasing sample size, Horowitz finds that his smoothed maximum score estimation method results in parameter estimates which achieve root of sample size convergence rate and are asymptotically normal, quite unlike the original estimates obtained by Manski (1975, 1985) without smoothing which converge at slower rate (cubic root of sample size) and have very complicated nonstandard asymptotic distribution. Our paper is therefore a counterpart in testing of Horowitz's idea in estimation. Smoothed indicator functions invariably involve a function of sample size, called a "bandwidth" (Horowitz (1992, 2002)) or "tuning parameter" (see the works cited at the end of the previous paragraph), which controls the rate at which the smoothed indicator approaches the original. Since smoothed indicator functions which simply shift the discontinuity away from the origin, the size of that shift is the bandwidth. The work of Horowitz (2002) suggests that, with some additional restrictions on bandwidth, Edgeworth-type higher-order expansions are derivable for test statistics constructed using smoothed indicators, hence bootstrapping may improve test performance. This is a potential future bonus for the test of this paper but our present objective is restricted to setting out a complete first-order theory for an asymptotic test that enjoys Properties 1-4. In fact, we also present results of simulations which show that the test works quite well without bootstrapping.

At this junction, the related work by Andrews (1999, 2000, 2001, 2002) on hypothesis testing involving inequality restrictions should be noted since he raises issues which may at first sight appear to restrict the methods of the present paper. However, his work highlights reasons for nonstandard behaviour of test statistics other than those we have pointed out in our review above. For example, Andrews considers estimation and testing when the parameter estimates lie in a set defining a maintained hypothesis some of whose boundary points are included in the set defining the null hypothesis. In this case, the asymptotic distribution of the estimator cannot be symmetric under all points of the null hypothesis and thus conventional quadratic-form test statistics based on the estimator cannot be standard chi-square (see Andrews (1999, 2001)). Indeed, even bootstrap tests may not work (Andrews (2000)). This scenario

7

would apply, for example, to testing that a variance parameter is zero using a variance estimate which is naturally nonnegative. By contrast, the boundary of the null hypothesis set considered in this paper lies within the interior of the maintained hypothesis set. So the scenario envisaged by Andrews is not applicable within the setup of this paper.

The rest of the paper is organized as follows. Section 2 sets up the constraint chaining framework and the smoothing indicator functions. We note that, whilst the philosophy of our approach leads uniquely to a statistic which is of "quasi-linear" form, this can be thought of as a special case of a wider class of forms all using only products of parameter estimates and smoothed indicators. Section 3 states distributional assumptions and completes the construction of the chaining test statistic. Section 4 investigates the asymptotic properties of the chaining test. Section 5 discusses issues in choosing the smoothing indicator functions. Section 6 proposes a computable interpolating power function which reflects the intentional differential impact of importance weighting on power at local to origin points compared with medium to distant point. Section 7 studies the test's finite sample power using Monte Carlo simulations. Section 8 presents proofs of the stated theorems. Section 9 sketches, without proof, how the constraint chaining method extends to a test of

$$\mu(x) \geq 0, \ \ a \leq x \leq b$$

, where $\mu(x)$ is a function of an observable variate x . Section 10 concludes the paper.

2 The chaining framework and the smoothing indicators

Let $\mu = (\mu_1, \mu_2, ..., \mu_p)'$ be a column vector of (functions of) parameters appearing in an econometric model. The problem is to test :

$$H_0 : \mu_j \geq 0 \text{ for all } j \in \{1, 2, ..., p\} \text{ versus } H_1 : \mu_j < 0 \text{ for at least one } j. \quad (4)$$

Note that in this problem neither the number nor the specific values of j are known for which $\mu_j = 0$ under H_0 or $\mu_j < 0$ under H_1.

Let $D(x), x \in R$, be the indicator function where $D(x) = 1$ if $x \leq 0$ and $D(x) = 0$ if $x > 0$. Then we can chain the multiple inequalities with the indicators to rewrite the hypothesis under test as follows :

$$H_0 : \sum_{j=1}^{p} D(\mu_j)\mu_j = 0 \ \ \text{versus} \ \ \ H_1 : \sum_{j=1}^{p} D(\mu_j)\mu_j < 0 \quad \quad (5)$$

Evidently, testing multiple inequalities is equivalent to one-sided single equality testing. With parameter estimates $\widehat{\mu}$ in hand, one can consider the constraint

chaining test statistic

$$\sum_{j=1}^{p} D(\widehat{\mu}_j)\widehat{\mu}_j. \tag{6}$$

However, a complication arises from the discontinuity of the indicator function $D(x)$ at the origin. Such discontinuity will result in the above chaining statistic having an asymptotic distribution which is the same as that of a sum of correlated and truncated random variables. The operational problems this incurs will hence be of similar difficulty to those arising with the chi-bar squared test methods. To overcome this problem, instead of directly working with $D(x)$, we construct a sequence of functions $\{\Psi_T(x)\}$ each of which is continuous at the origin and converges to the indicator $D(x)$ pointwise (except possibly at the origin) as the sample size T goes to infinity. In particular, we require that this sequence have a functional structure

$$\Psi_T(x) = \Psi(K(T)x)$$

satisfying the following assumptions:

[A1] $\Psi(x)$ is non-negative, bounded, and non-increasing in $x \in R$
[A2] $\Psi(x)$ is continuous at $x = 0$ and $\Psi(0) > 0$
[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 $x > 0$

We shall refer to any $\Psi_T(x) = \Psi(K(T)x)$ that satisfies the above assumptions as an (origin-) smoothing indicator. The technique of improving asymptotic behaviour by replacing a discrete indicator function with a smoothed version was first notably used by Horowitz (1992) in parameter estimation. Our paper shows the idea is also effective in hypothesis testing. Note that our smoothness assumptions are not completely the same as those in his (1992) paper or the follow-up (2002) and his "bandwidth" parameter is actually the inverse counterpart of our tuning parameter $K(T)$. In our framework, the set of assumptions on $\Psi_T(x)$ is not very restrictive and the range of choice for a valid $\Psi(x)$ is wide. In particular, one can use $\Psi(x) = 1 - F(x)$ where $F(x)$ is the cdf of an arbitrary random variable with $F(0) < 1$. We shall discuss issues in choosing $\Psi(x)$, $K(T)$ and particular examples in Section 5.

All the assumptions on $\Psi_T(x)$ enable the smoothing indicator to approximate $D(x)$ pointwise at just such a rate that the smoothed version $\sum_{j=1}^{p} \Psi_T(\widehat{\mu}_j)\widehat{\mu}_j$ of $\sum_{j=1}^{p} D(\widehat{\mu}_j)\widehat{\mu}_j$ has the property of asymptotic normality even though the latter itself is not asymptotically normal. Moreover, the idea of imposing [A6] is to create data-driven importance weighting in the sense that each individual $\widehat{\mu}_j$

9

corresponding to a strictly positive μ_j is likely to contribute ever less to the value of the test statistic as sample size T increases. As a consequence, the test statistic will be asymptotically dominated by those $\widehat{\mu}_j$ corresponding to zero or negative μ_j, detection of which is indeed the purpose of the test. By such importance weighting, we are therefore able to design a simple consistent one-sided test of H_0 versus H_1.

Note that the idea of chaining constraints with indicators is quite general and also flexible in terms of formulation. In particular, for arbitrary strictly positive scalars θ_j, $j \in \{1, 2, ..., p\}$, it is clear that the original problem could also be posed more generally as the following.

$$H_0 : \sum_{j=1}^{p} D(\theta_j \mu_j) \theta_j \mu_j = 0 \quad \text{versus} \quad H_1 : \sum_{j=1}^{p} D(\theta_j \mu_j) \theta_j \mu_j < 0. \qquad (7)$$

One key choice of θ_j which can be motivated from standardization of parameter estimators is $\theta_j = (V_{jj})^{-1/2}$ where V_{jj} denotes the asymptotic variance of $\widehat{\mu}_j$[5]. Furthermore, we note that the unsmoothed and smoothed quasi-linear chaining test quantities

$$\sum_{j=1}^{p} D(\widehat{\mu}_j) \widehat{\mu}_j \ \text{and} \ \sum_{j=1}^{p} \Psi_T(\widehat{\mu}_j) \widehat{\mu}_j$$

may be thought of as special cases of a class of plausible test quantities generated by taking the unsmoothed

$$\eta \equiv (\eta_1, \eta_2, ..., \eta_p) \ \text{ where } \eta_j \equiv D(\widehat{\mu}_j) \widehat{\mu}_j$$

and the smoothed

$$\widetilde{\eta} \equiv (\widetilde{\eta}_1, \widetilde{\eta}_2, ..., \widetilde{\eta}_p) \ \text{ where } \widetilde{\eta}_j \equiv \Psi_T(\widehat{\mu}_j) \widehat{\mu}_j$$

in a suitable class of functions $g(\eta)$, $\eta \in R^p$ including at least the following

$$g(\eta) = \sum_{j=1}^{p} \eta_j, \ \ g(\eta) = \eta' U \eta, \ \ g(\eta) = \min\{\eta_1, \eta_2, ..., \eta_p\}$$

where U is a positive semi-definite matrix. This paper develops a complete test theory for the case of linear function. It is the benchmark. The quadratic case is considered in Chen (2009, Chapter 4)[6].

---

[5] See Hansen (2005) for motivating arguments of the use of standardized individual statistic for the test construction in the context of multiple inequality testing.

[6] Chen (2009, Chapter 4) shows that the smoothed indicator test based on quadratic form of $g(\widetilde{\eta})$ can also simultaneously satisfy the four properties discussed in Section 1. In the contrast, the extreme value form of $g(\widetilde{\eta})$ is not smooth in $\widetilde{\eta}$ and thus even used with smoothed indicator, such extreme value test still suffers from the problem of nuisance parameter from unknown covariance V as in White (2000) and Hansen (2005).

3 Distributional assumptions and the chaining test statistic

Let the true value and estimator of the parameter vector be denoted respectively by $\mu^* = (\mu_1^*, \mu_2^*, ..., \mu_p^*)'$ and $\widehat{\mu} = (\widehat{\mu}_1, \widehat{\mu}_2, ..., \widehat{\mu}_p)'$. In the simplest distributional set-up, $\widehat{\mu}$ is assumed to be $\sqrt{T}$ consistent and

[D1] $\quad \sqrt{T}(\widehat{\mu} - \mu^*) \xrightarrow{d} N(0, V)$

where V is some finite positive definite matrix. For notational ease, we keep implicit the possible dependence of V on the values of any underlying parameters. We assume there is available an estimator $\widehat{V}$, almost surely positive definite, such that

[D2] $\quad \widehat{V} \xrightarrow{p} V.$

Regarding the scalars introduced at the end of Section 2, we assume the availability of strictly positive estimators $\widehat{\theta}_j, j \in \{1, 2, ..., p\}$, such that

[D3] $\quad \widehat{\theta}_j \xrightarrow{p} \theta_j.$

To construct the chaining test statistic, we introduce the following objects. Let Δ and $\widehat{\Delta}$ be the p dimensional square matrices

$$\Delta \quad \equiv \quad diag(\theta_1, \theta_2, ..., \theta_p) \tag{8}$$
$$\widehat{\Delta} \quad \equiv \quad diag(\widehat{\theta}_1, \widehat{\theta}_2, ..., \widehat{\theta}_p) \tag{9}$$

Let $\widehat{\Psi}$ be the p dimensional column vector

$$\widehat{\Psi} \equiv (\Psi_T(\widehat{\theta}_1 \widehat{\mu}_1), \Psi_T(\widehat{\theta}_2 \widehat{\mu}_2), ..., \Psi_T(\widehat{\theta}_p \widehat{\mu}_p))'. \tag{10}$$

Define Q_1 and Q_2 as

$$Q_1 \quad \equiv \quad \sqrt{T} \widehat{\Psi}' \widehat{\Delta} \widehat{\mu} \tag{11}$$
$$Q_2 \quad \equiv \quad \sqrt{\widehat{\Psi}' \widehat{\Delta} \widehat{V} \widehat{\Delta} \widehat{\Psi}}. \tag{12}$$

We define a test statistic Q as follows :

$$Q = \left\{ \begin{array}{ll} \Phi(Q_1/Q_2) & \text{if } Q_2 > 0 \\ 1 & \text{if } Q_2 = 0 \end{array} \right\} \tag{13}$$

where $\Phi(x)$ is the standard normal distribution function. For asymptotic significance level α, we reject H_0 if $Q < \alpha$. The chaining statistic Q is therefore a form of tail probability or p-value. We now sketch the reasoning which motivates our test procedure. Formal theorems and rigorous proofs 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 as sample size grows. 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 as sample size grows. 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 smoothing indicators do not exclude discrete functions. Only continuity at origin is required. Therefore, a discrete function that is continuous at origin can be a candidate of the smoothing indicator and such an example is presented in Section 5. If all parameters are strictly positive and Ψ is a discrete function, then $\widehat{\Psi}_T$ is zero with probability approaching one as sample size grows. In this case, Q_2 is also zero with probability approaching one. Therefore, occurrence of the event that $Q_2 = 0$ signals that we should not reject H_0. Note that it is not an adhoc choice to set Q to be one for which the event $Q_2 = 0$ occurs because the probability limit of $\Phi(Q_1/Q_2)$ is also one when all parameters are strictly positive and Ψ is an everywhere continuous function[7].

4   Asymptotic properties of the chaining test

The first-order asymptotics of the chaining test can be justified by the following four theorems, whose proofs are given in Section 8. To facilitate the statement of those theorems, we introduce some convenient notation. Let S denote the set $\{1, 2, ..., p\}$. Decompose S as $S = A \cup M \cup W$, where

$$A \equiv \{j \in S : \mu_j^* > 0\}, \ M \equiv \{j \in S : \mu_j^* = 0\}, \ W \equiv \{j \in S : \mu_j^* < 0\}.$$

Let d_{MW} be p-dimensional selection vector whose jth element is unity or zero according as $j \in M \cup W$ or $j \notin M \cup W$. Let d_S be the p-dimensional vector in which all its elements are unity. Define the scalars, ω_{MW} and ω_S as

$$\omega_{MW} \equiv d'_{WM} \Delta V \Delta d_{WM} \tag{14}$$
$$\omega_S \equiv d'_S \Delta V \Delta d_S. \tag{15}$$

Note that these scalars are in fact the variances of the asymptotic distribution of the sum $\sum_j \sqrt{T} \theta_j (\widehat{\mu}_j - \mu_j^*)$ over $j \in M \cup W$ and $j \in S$ respectively.

Theorem 1 (Asymptotic Null Distribution)

Given [D1], [D2] and [D3], the following are true under $H_0 : \mu_j^* \geq 0$ for all $j \in \{1, 2, ..., p\}$.
(1) If $M \neq \varnothing$, then $Q \xrightarrow{d} U(0, 1)$.

---

[7]The case of Ψ being everywhere continuous is more complicated because Q_2 in this case is almost surely strictly positive. If all parameters are strictly positive, then both numerator and denominator in the "$t-$ratio", Q_1/Q_2 go to zero in probability. See Section 8 for derivation of asymptotic properties of the test statistic for such non-standard scenario.

(2) If $M = \varnothing$, then $Q \overset{p}{\longrightarrow} 1$.

Note that the asymptotic size of the test is the supremum of the asymptotic rejection probability with respect to those μ^* restricted to the set defined under H_0. Therefore, Theorem 1 implies that the asymptotic test size is equal to α. Furthermore, we observe that the size is attained uniformly over all points on M, the boundary of null-restricted parameter space and thus the test achieves asymptotic similarity on the boundary. In contrast to the chi-bar squared and the extreme value test approaches, the limiting null distribution of the chaining test, when non-degenerate, is uniform, which clearly fulfills ease of implementation as required in Property 4.

Theorem 2 (Consistency)

Given [D1], [D2], and [D3],   the following is true under $H_1 : \mu_j^* < 0$ for some $j \in \{1, 2, ..., p\}$:

$$P(Q < \alpha) \longrightarrow 1 \ \ \text{as} \ \ T \longrightarrow \infty.$$

Theorem 2 shows that the chaining test is a consistent test. Besides consistency, we are also interested in the local behavior of the test. In order to derive a local power function of the chaining test, we need to view the magnitudes of μ_j^* for $j \in W$ as becoming smaller when $T \longrightarrow \infty$ (See McManus 1991). Specifically, we imagine that

[L1]   For $j \in W$, $\mu_j^* = \frac{c_j}{\sqrt{T}}$, where c_j is a fixed negative scalar.

Then we have the following result.

Theorem 3 (Neyman-Pitman Local Power)

Assume [D1], [D2] and [D3] hold with the elements μ_j^* of μ^* taking the T-dependent forms [L1] for $j \in W$ but keeping fixed values for $j \in S \backslash W$ where W is non-empty. Define

$$c_W \equiv \sum_{j \in W} \theta_j c_j < 0.$$

Then as $T \longrightarrow \infty$

$$P(Q < \alpha) \longrightarrow \Phi(\Phi^{-1}(\alpha) - \omega_{MW}^{-1/2} c_W). \tag{16}$$

Theorem 3 shows that the asymptotic local power of the test is increasing in the absolute magnitude of each negative c_j. Further, this power is not diluted

by any fixed positive μ_j^*. On the other hand, suppose we extend the imagined condition [L1] to

[L2]   For $j \in A \cup W$, $\mu_j^* = \frac{c_j}{\sqrt{T}}$, where c_j is a fixed scalar taking negative value for j in W but positive value for j in A.

Then Theorem 3 would be replaced by the following.

Theorem 4 (Two-Sided Local Analysis)

Assume [D1], [D2] and [D3] hold with the elements μ_j^* of μ^* taking the T-dependent forms [L2] for $j \in A \cup W$, where A and W are both non-empty. Define

$$c_{AW} \equiv \sum_{j \in A \cup W} \theta_j c_j .$$

Then as $T \longrightarrow \infty$

$$P(Q < \alpha) \longrightarrow \Phi(\Phi^{-1}(\alpha) - \omega_S^{-1/2} c_{AW}). \tag{17}$$

In the case of two-sided local analysis, it is clear from Theorem 4 that the chaining test could be locally biased[8]. However, any parameter region of local bias will be shrinking as sample size increases, and its practical relevance will become negligible. An argument we present later in Section 7.2 using a specific origin-smoothed indicator function sheds further light on this matter. Moreover, the simulations we report there are encouraging of the view that local bias is not an issue of practical concern and rather a small price to pay for an otherwise effective test that is so easy to apply.

5   The choice of smoothing indicator

For fixed T, our proposed chaining approach implicitly replaces the testing problem (5) with the technically easier problem of testing

$$H_0 : \sum_{j=1}^p \Psi_T(\mu_j)\mu_j = 0 \quad \text{versus} \quad H_1 : \sum_{j=1}^p \Psi_T(\mu_j)\mu_j < 0 .$$

Though the former testing problem is equivalent to the original problem of testing multiple inequalities, the latter is only asymptotically equivalent to the

---

[8]Though Theorem 1 shows that the chaining test is asymptotically similar on the boundary, such property is only necessary rather than sufficient for the test to be asymptotically unbiased (See Theorem 5 in Hansen (2003)). However, as shown by Hansen (2003, Theorem 3 and Corollary 6), the LF-based test that is asymptotically non-similar on the boundary is asymptotically inadmissible when compared with another test which is asymptotically similar on the boundary. In other words, while both are asymptotically exact tests, the power of the former test against local alternatives is dominated by the latter.

former and hence for any fixed sample size T, the smoothed version of the test is indeed testing a deformed hypothesis. In principle, optimization exercises are desired to find choices of $\Psi(x)$ and $K(T)$ that minimize the discrepancy between the limiting distribution of the chaining statistic and its actual finite sample distribution. This is neither the same as nor incompatible with the objective of minimizing the discrepancy the original (unsmoothed) and the deformed (smoothed) hypotheses subject to the restrictions on the smoothing indicators imposed under [A1] ~[A6]. The whole problem is complicated by the fact that any such optimization has to consider all relevant values of sample size T simultaneouly since $\Psi(x)$,as distinct from $\Psi_T(x) = \Psi(K(T)x)$, has to be independent of T. Thus a formal objective function will need to be in the nature of a sum of terms over T. Because any choice of $\Psi(x)$ and $K(T)$ will in theory yield the same first order asymptotic properties as long as the conditions [A.1] to [A.6] are met, such optimization is unlikely to produce a tractable solution without some judicious approximation. In Appendix A of this paper, we give a preliminary analysis where we suggest that the "interpolating power function" to be introduced in the next section could play a useful role. Nonetheless, we leave a thorough analysis of this second-level problem for further research. In this paper, we propose a set of valid smoothing indicators by appealing to simplicity of functional form (an Occam's razor type of argument) and then implementing simulation exercises to assess practical size and power potential.

The simplest choice of $\Psi(x)$ that satisfies [A.1]-[A.2] is the following.

Step-at-unity : $\Psi(x) = 1$ if $x \leq 1$, and $\Psi(x) = 0$ if $x > 1$.

Another convenient choice of $\Psi(x)$ that is everywhere continuous is the following.

Logisitic[9] : $\Psi(x) = (1 + \exp(x))^{-1}$.

For the choice of $K(T)$,we propose to set $K(T) \equiv \sqrt{T/\log(T)}$, which is simple and complies fully with the conditions [A.3] ~[A.6] on smoothing indicators. To implement the chaining test in practice, we also argue for an adjustment to account for the potentially overwhelming effect of the total number of inferentially uninformative parameter estimates. To understand this point, note that we should reject H_0 even if only one of μ_j^*, $j \in \{1, 2, ..., p\}$ is negative. Hence, though the asymptotic properties of the test are not affected, it is still desirable that the smoothing indicator be made dependent on p, the total number of parameters in such a way that it is decreasing in p for $x > 0$. Then, in finite samples, the inferential value of a single negative $\widehat{\mu}_j$ is less likely to be overwhelmed by additional numbers of positive $\widehat{\mu}_j$ as p increases. To achieve this, an easy way is to replace $\Psi(x)$ with $\Psi(h(p)x)$ where $h(p)$ is a positive function increasing in p. For example, we can use the quadratic specification for $h(p)$ such that $h(p) \equiv p(p-1)$ for $p \geq 2$ and $h(1) \equiv 1$.

---

[9]Note that this $\Psi(x)$ is the complement of the logisitc distribution: $\Psi(x) = 1 - F(x)$, where $F(x)$ is the logistic cdf.

6 An interpolating computable theoretical power function

The results of Section 4 show that the power function (rejection probability) of the chaining test depends on which of the scenarios [L1] or [L2] is appropriate. Scenario [L2] applies if all μ_j^* are viewed as being small enough to be "local". Scenario [L1] applies if only the negative (and zero) μ_j^* are viewed as small enough to be local, but in that case the importance weighting via smoothing indicators comes into play to eliminate the "nuisance" positive ones. If negative μ_j^* cannot be considered small enough to be local, then local power theory does not apply and instead consistency of the chaining test simply gives an asymptotic power of unity.

But how do we judge what is small enough to be "local" ? In other words, how do we know which of the two local power functions or the power of unity is appropriate ? A way to resolve this issue is to use smoothing indicator functions which automatically interpolate between the two local powers and the unity. In other words, we let the "numbers do the talking" and avoid the adhoc suggestion that we should fix an arbitrary point value which explicitly distinguish "local" from "nonlocal".

Accordingly, our proposal is to use an "interpolating theoretical power function" defined by $P \equiv \Phi(\Phi^{-1}(\alpha) - \omega^{-1/2}c)$, where

$$c = \sqrt{T} \sum_{j \in AUW} \Psi_T(\theta_j \mu_j^*)\theta_j \mu_j^*, \tag{18}$$

$$\omega = d' \Delta V \Delta d, \tag{19}$$

$$d = (\Psi_T(\theta_1 \mu_1^*), \Psi_T(\theta_2 \mu_2^*), ..., \Psi_T(\theta_p \mu_p^*))'. \tag{20}$$

Note the limit of P as T tends to infinity is the power function (16) under the Neyman-Pitman local sequence [L1] yet is (17) under the two-sided local sequence [L2]. Moreover, for fixed ("nonlocal") and negative μ_j^*, (18) tends to minus infinity as T approaches infinity, thus correctly yielding asymptotic power of unity in such a "nonlocal" case. Note that the interpolating power P can also be computed for those cases of μ^* implied under H_0. If all μ_j^* are non-local and positive, then $\omega^{-1/2}c$ tends to infinity as T goes to infinity and the limit of P in such case is zero[10]. If all μ_j^* are non-local and non-negative with at least one μ_j^* being zero, then the limit of P in this case is equal to α, the size of the test. Therefore, in between these extreme scenarios the formula of P acts as an interpolation[11]. The algebraic form of the function Ψ used in interpolating power need not be the same as used in the construction of the test statistic since it serves for a different purpose. That purpose is to automatically interpolate

---

[10]Note that in this case, both c and ω go to zero. However, the ratio $\omega^{-1/2}c$ goes to infinity as a consequence of the properties of the smoothing indicators. Proof of this result can be based on the same arguments as that of part (2) of Theorem 1 given in Section 8.

[11]Indeed the power function P automatically interpolates the rejection probabilities obtained in the four theorems of Section 4.

between the three different scenarios. The Ψ function that is used in computing P should be continuous (such as the logistic specification) because interpolation is a form of continuous smoothing[12].

In the next section we make some illustrative computations of the interpolating power function and compare its graphs with those of power functions obtained by simulation.

7 Monte Carlo studies of finite sample test power

In this section we conduct a series of Monte Carlo simulations to study the finite sample power of the test statistic and to compare the finite sample performance of different choices of the smoothing indicators.

7.1 The simulation setup

The smoothing indicators $\Psi(h(p)K(T)x)$ are formed using the following specification :

(i) *Step-at-unity* : $\Psi(x) = 1\{x \leq 1\}$
   Logistic : $\Psi(x) = (1 + \exp(x))^{-1}$
(ii) $K(T) = \sqrt{T/\log(T)}$
(iii) No dimension adjustment : $h(p) = 1$
   Quadratic adjustment : $h(p) = p(p-1)$

Microeconomic and financial datasets are nowadays typically large, but we decided nonetheless to check the validity of our asymptotic theory by going down to a sample as small as 50. In fact, we generated R replications for each of two sample sizes and according to the following scheme :

(iv) Parameters : $r_1 \equiv \theta_1\mu_1$ and $r_2 \equiv \theta_2\mu_2$

(v) Two $(p = 2)$ jointly normal variates $\widehat{r}_1$ and $\widehat{r}_2$ such that

$$\left[\begin{array}{c} \widehat{r}_1 \\ \widehat{r}_2 \end{array} \right] \sim N(\left[\begin{array}{c} r_1 \\ r_2 \end{array} \right], \frac{V}{T}), \text{ where } V = \left[\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right] \text{ with known } \rho.$$

The variates in (v) are generated via

$$\widehat{r}_1 = r_1 + \frac{\varepsilon_1}{\sqrt{T}}$$

$$\widehat{r}_2 = r_2 + \rho(\widehat{r}_1 - r_1) + \frac{\sqrt{1-\rho^2}}{\sqrt{T}}\varepsilon_2$$

---

[12] The transition from power at scenarios of local points such as [L1] or [L2] to power at non-local large points is continuous. A further investigation of this issue would ideally be covered by a higher-order asymptotic theory and (Edgeworth) expansions in T. In the absence of that, we propose the novel technique of interpolation formula P based on everywhere continuous interpolating weights.

where ε_1 and ε_2 are independently drawn from the $N(0,1)$ distribution. The number of draws made was $R = 100000$. Each draw led to a particular value of our test statistic and to a particular decision (reject or accept). The simulated rejection probabilities enabled graphical presentations of a quality which suggests that the number of draws used was more than adequate for the accuracy of our general and specific conclusions. We focused on a nominal test size of $\alpha = 0.05$. We shall also refer to the simulated rejection probability as the (simulated) power of the test. Hence under H_1, the power of an unbiased test should be no smaller than α. Furthermore, the test size is not distorted if the power of the test under H_0 is not greater than α.

We considered several pre-fixed values of r_2 and for each such value we conducted simulations yielding an empirical power function calculated on a grid of 1000 values of r_1. In particular, we chose $r_2 \in \{-0.5, 0, 0.5\}$ because they illustrate the issues and depict an informative range of behaviour with visual clarity. Corresponding to each empirical power function we calculated an interpolating theoretical power function for comparison. Based on Section 6, the corresponding interpolating power in current simulation setup is

$$P \equiv \Phi(\Phi^{-1}(\alpha) - \omega^{-1/2}c), \tag{21}$$

where

$$c = \sqrt{T}(\Psi_T(r_1)r_1 + \Psi_T(r_2)r_2), \tag{22}$$
$$\omega = \Psi_T(r_1)^2 + \Psi_T(r_2)^2 + 2\rho\Psi_T(r_1)\Psi_T(r_2). \tag{23}$$

Since interpolation is a form of continuous smoothing, we use the continuous logistic function for interpolating power even when the step-at-unity function is used in constructing the test statistic.

7.2 Simulation results for large samples

To study how well the asymptotic properties of the test hold in finite samples, we first consider a very simple simulation configuration : $\Psi(x)$ is step-at-unity and there is no adjustment for p. Table 3.1[13] shows the simulated power curves in the left column and the corresponding interpolating theoretical power curves in the right column when r_2 is kept fixed and the sample size T is 1000. These curves are functions of r_1 varying from -1 to 1 when r_2 is kept fixed at the values -0.5 (dashed curve), 0 (continuous curve) and 0.5 (dotted curve). These choices allow r_1 to vary equally widely, small and large, in both positive and negative configurations. They are thus concisely informative about test rejection probability both under the null hypothesis and under the alternative. The top pair of panels in Table 1 show rejection probability when $\hat{r}_1$ and $\hat{r}_2$ are uncorrelated ($\rho = 0$), the middle panels when negatively correlated ($\rho = -0.8$), and the bottom panels when positively correlated ($\rho = 0.8$). The non-zero values ± 0.8 were chosen to represent serious correlation.

---

[13] All tables in this paper can be found at the end of Section 7.

In interpreting Table 3.1, we note that our proposed interpolating theoretical power formula describes very well the finite sample behavior of the test. All interpolating theoretical power curves fit remarkably well the corresponding curves of simulated power. Regarding the performance of the test, we first look to check whether rejection probability under the null anywhere exceeds nominal size $\alpha = 0.05$. The null hypothesis is effective on those fragments of the dotted ($r_2 = +0.5$) and continuous ($r_2 = 0$) curves lying to the right of $r_1 = 0$ (inclusive). The following observations apply to both simulated and interpolating theoretical curves in all panels of Table 3.1. It is clear that no point on those curve fragments is above 0.05 in height. The continuous curve soon climbs back at height 0.05 for values of r_1 exceeding 0. The dotted curve lies almost immediately at a height of zero and is barely discernible as it coincides with the horizontal axis. This happens because the dotted line represents the case where both r_1 and r_2 are strictly positive and so the importance weighting effect of the smoothing indicator wipes out the impact of these values with high probability in large samples.

A second check of Table 3.1 should look for high rejection probability at least equal to the nominal size $\alpha = 0.05$ under the alternative hypothesis. The alternative hypothesis is effective on all curves at points to the left of $r_1 = 0$ in all panels of Table 3.1. It is additionally effective on the dashed curves ($r_2 = -0.5$) at points to the right of $r_1 = 0$ (inclusive). Note that the dashed curve is always at height 1, thus indicating that the test is extremely successful even when only one parameter value is strictly negative. The dotted and continuous curves are at height no lower than the nominal size, and mostly at height 1, at all points strictly to the left of $r_1 = 0$. It is evident that the test is successful.

7.3   Simulation results for small samples

Table 3.2 shows the simulation results using the same configuration as Table 3.1 but a smaller sample size ($T = 50$). The findings are similar to those in Tables 3.1. Rejection probability under the null hypothesis is bounded by the nominal size $\alpha = 0.05$. The rejection probability under that region of the alternative hypothesis represented by the dashed curves is reduced compared with that of Table 3.1; however it remains at value 1 for the most part and everywhere still exceeds ten times the nominal size. The heights of the continuous curves stay above the nominal size in the alternative hypotheses region to the left of $r_1 = 0$ and increase towards 1 as becomes ever more negative. The height of the dotted curve in the lower left-hand panel falls below the nominal size in the tiny local alternative region $-0.04 \leq r_1 < 0$. As r_1 becomes ever more negative (to -1), the height increases progressively to 1. Overall the theoretical curves match the simulated well, though not as well as in the large sample case of Table 3.1.

The above simulation results show that the chaining test seems to perform quite well in finite samples even under a very simple choice of the smoothing indicator. In the next simulations, we consider the effect of making an adjustment for the number of parameters, p. Hence, we use the same simple configurations as above to study the effect of introducing the quadratic version of the adjust-

ment function $h(p)$.

7.4   Effect of adjustment for number of parameters

Tables 3.3 and 3.4 show the simulated difference in power between using a quadratic adjustment function $h(p) = p(p-1)$ and using no adjustment $h(p) = 1$. In computing this power, r_2 is kept fixed and the sample size T is set to be 1000 and 50, respectively. We find that the introduction of quadratic adjustment $h(p)$ has indeed raised the finite sample power. The power gain occurs over a much wider range of r_1 values for small samples than for large. The maximal gains achievable are greatest in small samples when the correlation between parameter estimators is nonnegative.

7.5   Effect of different choices of the $\Psi(x)$ function on power

Tables 3.5 and 3.6 shows the curves of difference of rejection probabilities of the logistic and the step-at-unity specifications of $\Psi(x)$ when r_2 is kept fixed and the sample size T is 1000 and 50 respectively. These tables suggest that the logistic specification generally incurs no greater probability of Type 1 error (rejecting the null hypothesis when both r_1 and r_2 are non-negative) than the step-at-unity. Also, the logistic is generally more powerful than the step-at-unity against the alternative hypothesis when r_2 is negative (dashed curves). The non-negativity of the continuous curves in the region $r_1 < 0$ indicates that the logistic smoothing indicator yields power at least as great as that of the step-at-unity when $r_2 = 0$.

The dotted curves concern the case $r_2 > 0$. They almost coincide with the horizontal axis where $r_1 < 0$ in Table 3.5, thus suggesting that the logistic and step-at-unity specifications have similar power in large samples when $r_2 > 0$. But in Table 3.6 the dotted curves go below the horizontal axis where $r_1 < 0$. Thus, in small samples, the logistic seems to be is less powerful than step-at-unity when $r_2 > 0$. This finding can be attributed to the fact that the step-at-unity indicator is capable of eliminating positive parameter estimates completely from the test statistic. In the contrast, the logistic does not completely eliminate these estimates and hence the power-reducing effect of such positives may have more impact in small samples compared to large where asymptotic equivalences between all smoothing indicators satisfying the theory assumptions [A1]-[A6].

Table 3.1 : Power function of step-at-unity Ψ under $h(p) = 1, T = 1000$
Vertical axis measures rejection probability.
Dotted, continuous, dashed curves represent r_2 values $+0.5, 0, -0.5$ respectively.
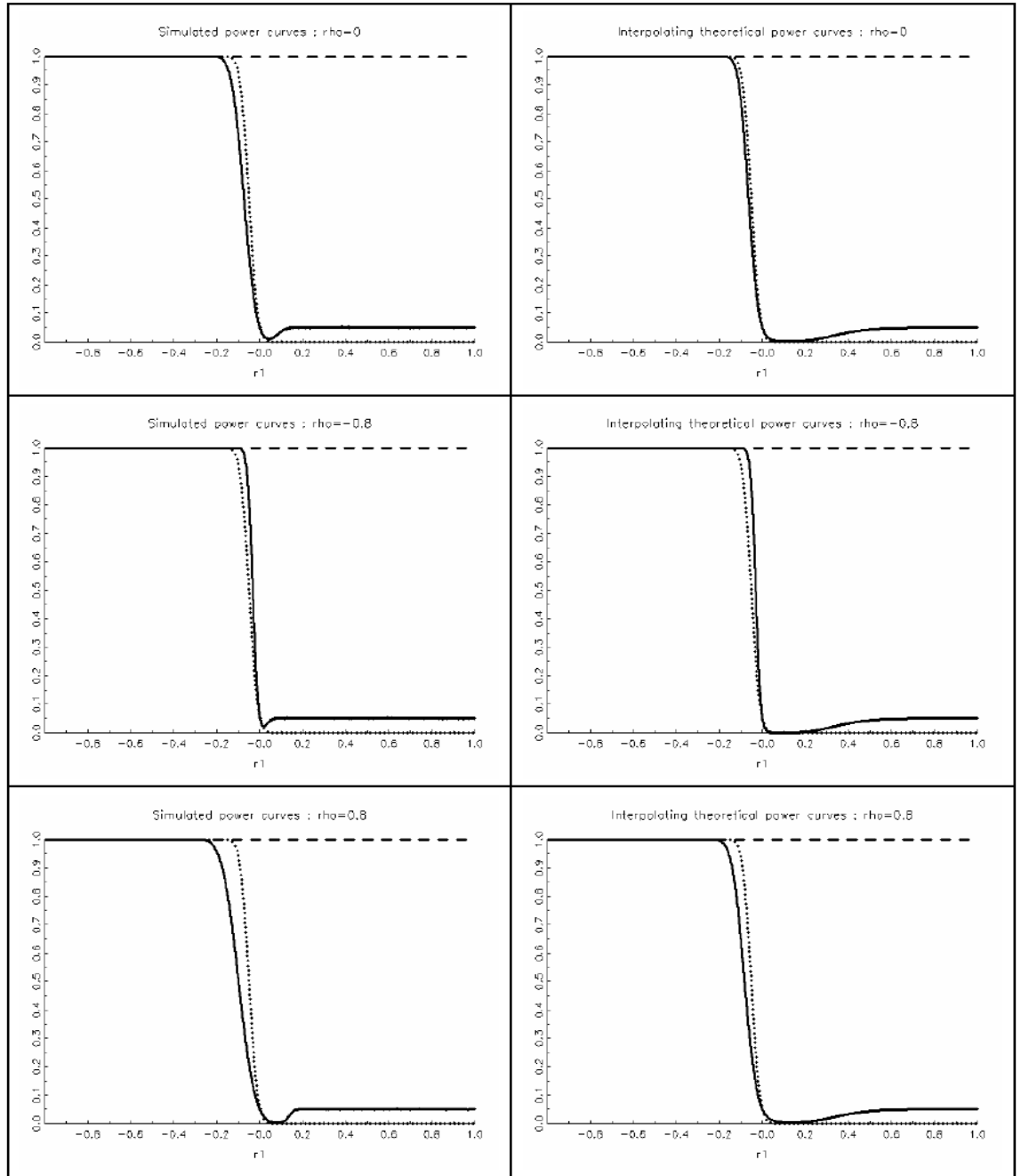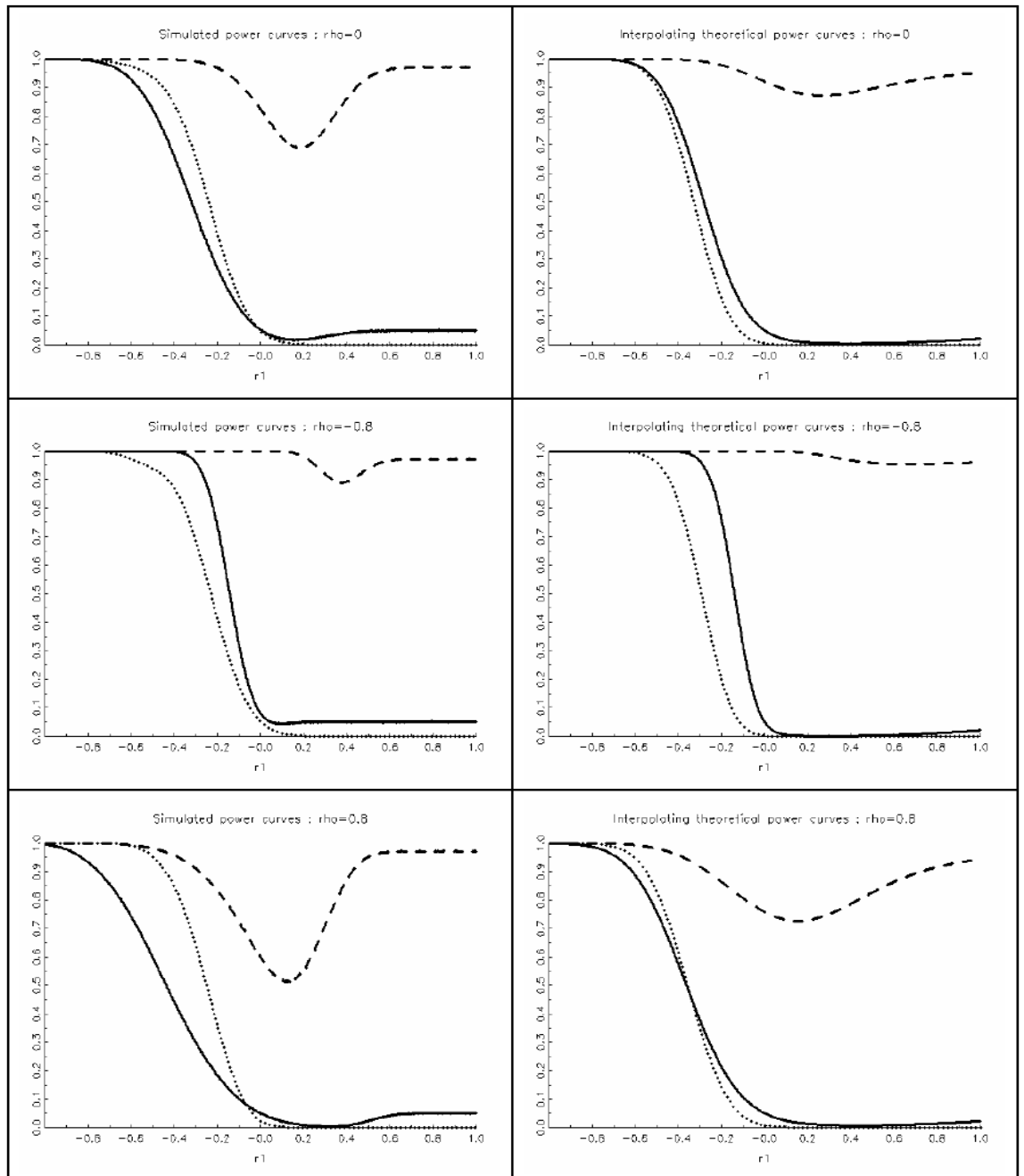
Table 3.2 : Power function of step-at-unity Ψ under $h(p) = 1, T = 50$
Vertical axis measures rejection probability.
Dotted, continuous, dashed curves represent r_2 values $+0.5, 0, -0.5$ respectively.

Table 3.3 : Power for $h(p) = p(p-1)$ minus power for $h(p) = 1$ ($T = 1000$)
Vertical axis measures difference of rejection probabilities.
Dotted, continuous, dashed curves represent r_2 values $+0.5, 0, -0.5$ respectively.
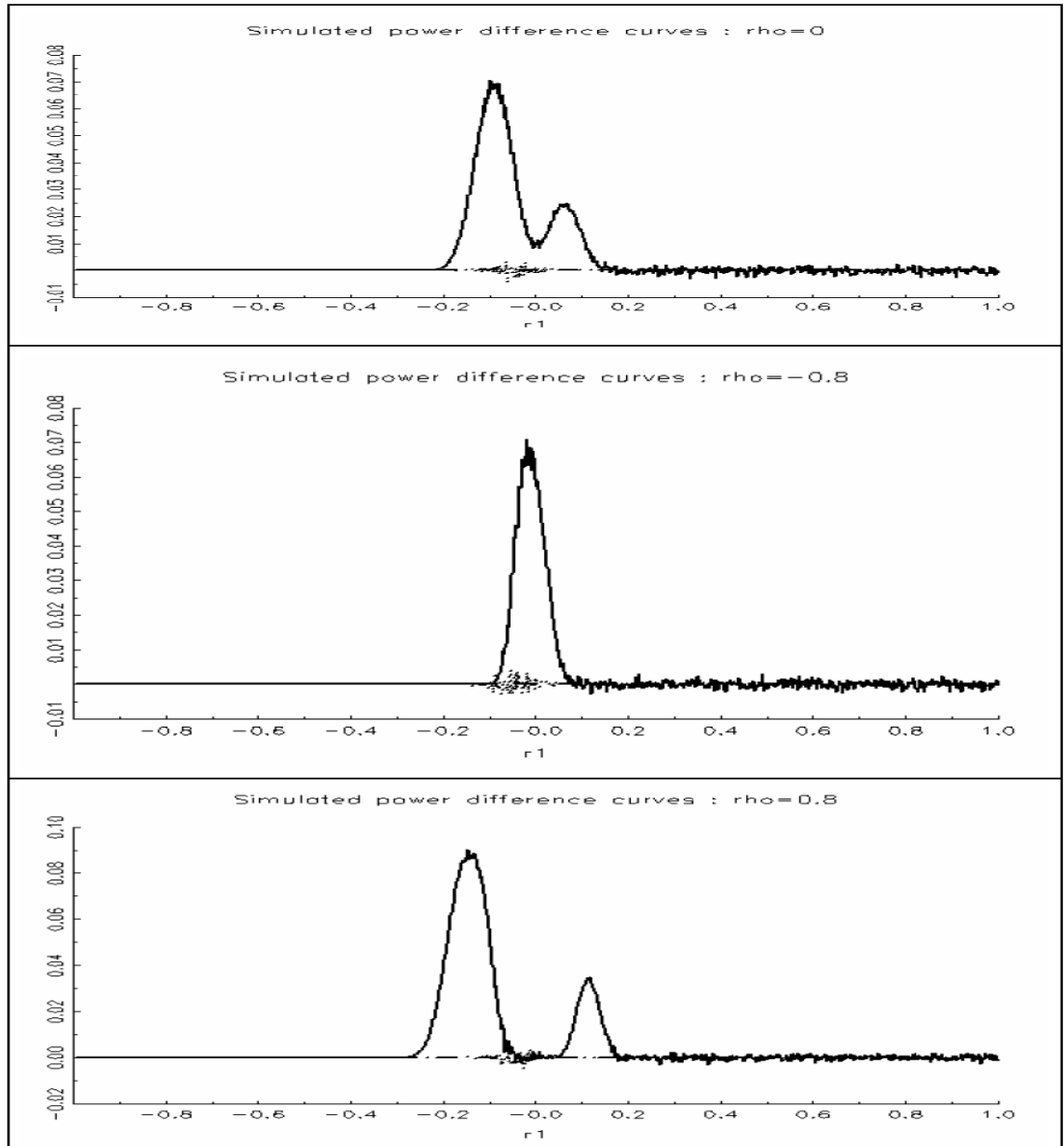
Table 3.4 : Power for $h(p) = p(p-1)$ minus power for $h(p) = 1$ ($T = 50$)
Vertical axis measures difference of rejection probabilities.
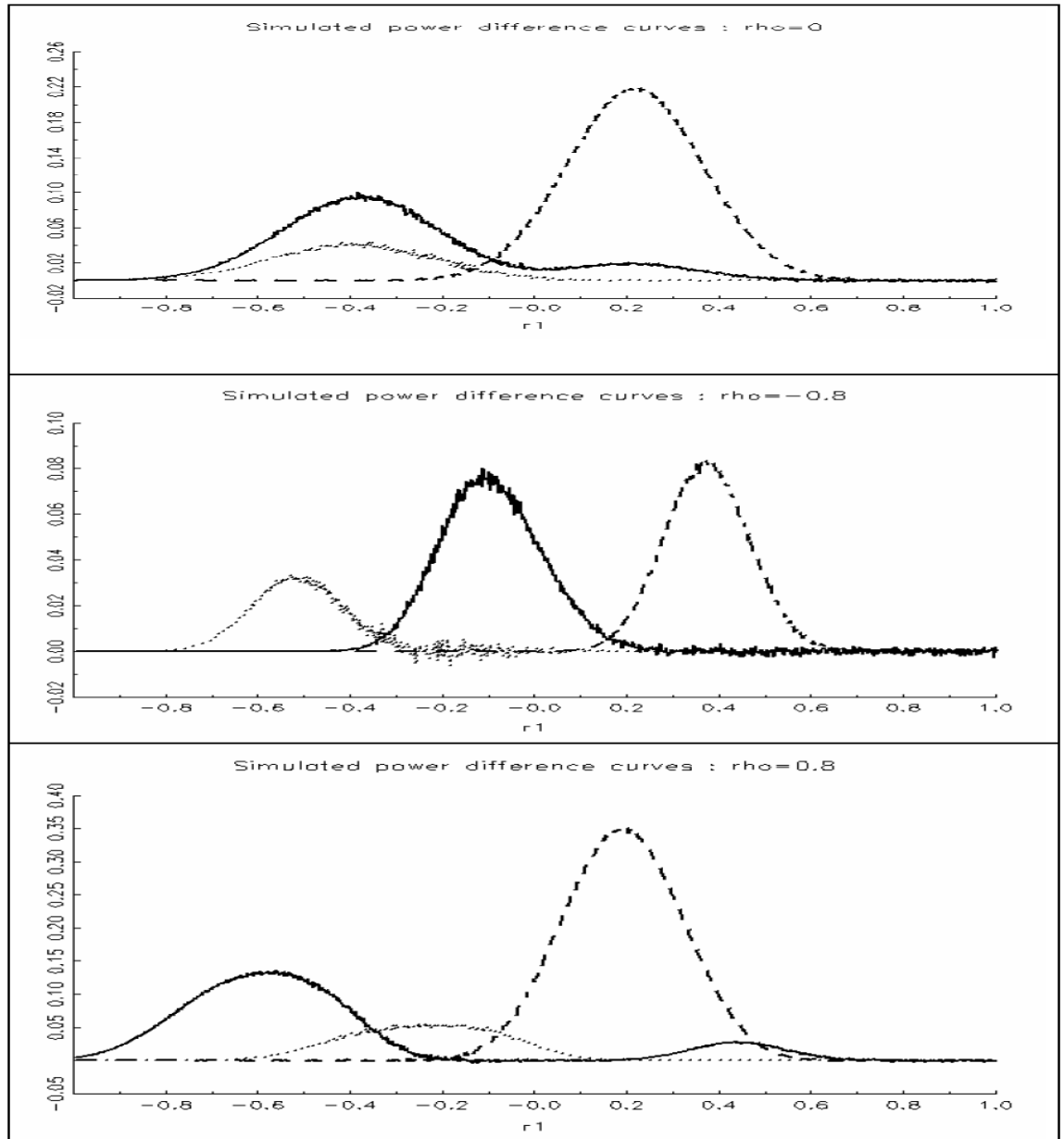Dotted, continuous, dashed curves represent r_2 values $+0.5, 0, -0.5$ respectively.

Table 3.5 : Logistic power minus step-at-unity power for $h(p) = p(p-1)$ $(T = 1000)$
Vertical axis measures difference of rejection probabilities.
Dotted, continuous, dashed curves represent r_2 values $+0.5, 0, -0.5$ respectively.

Table 3.6 : Logistic power minus step-at-unity power for $h(p) = p(p-1)$ $(T = 50)$
Vertical axis measures difference of rejection probabilities.
Dotted, continuous, dashed curves represent r_2 values $+0.5, 0, -0.5$ respectively.
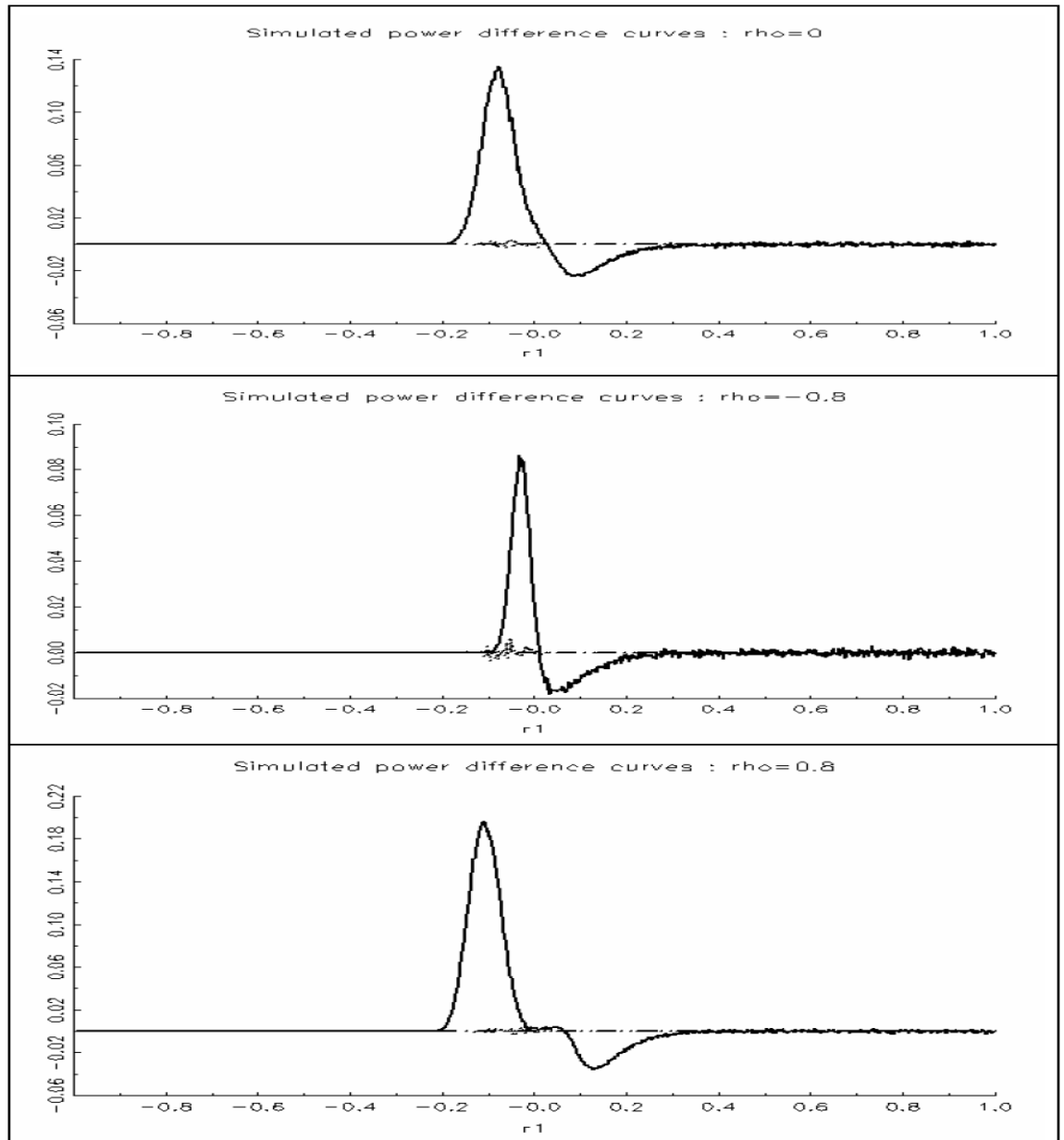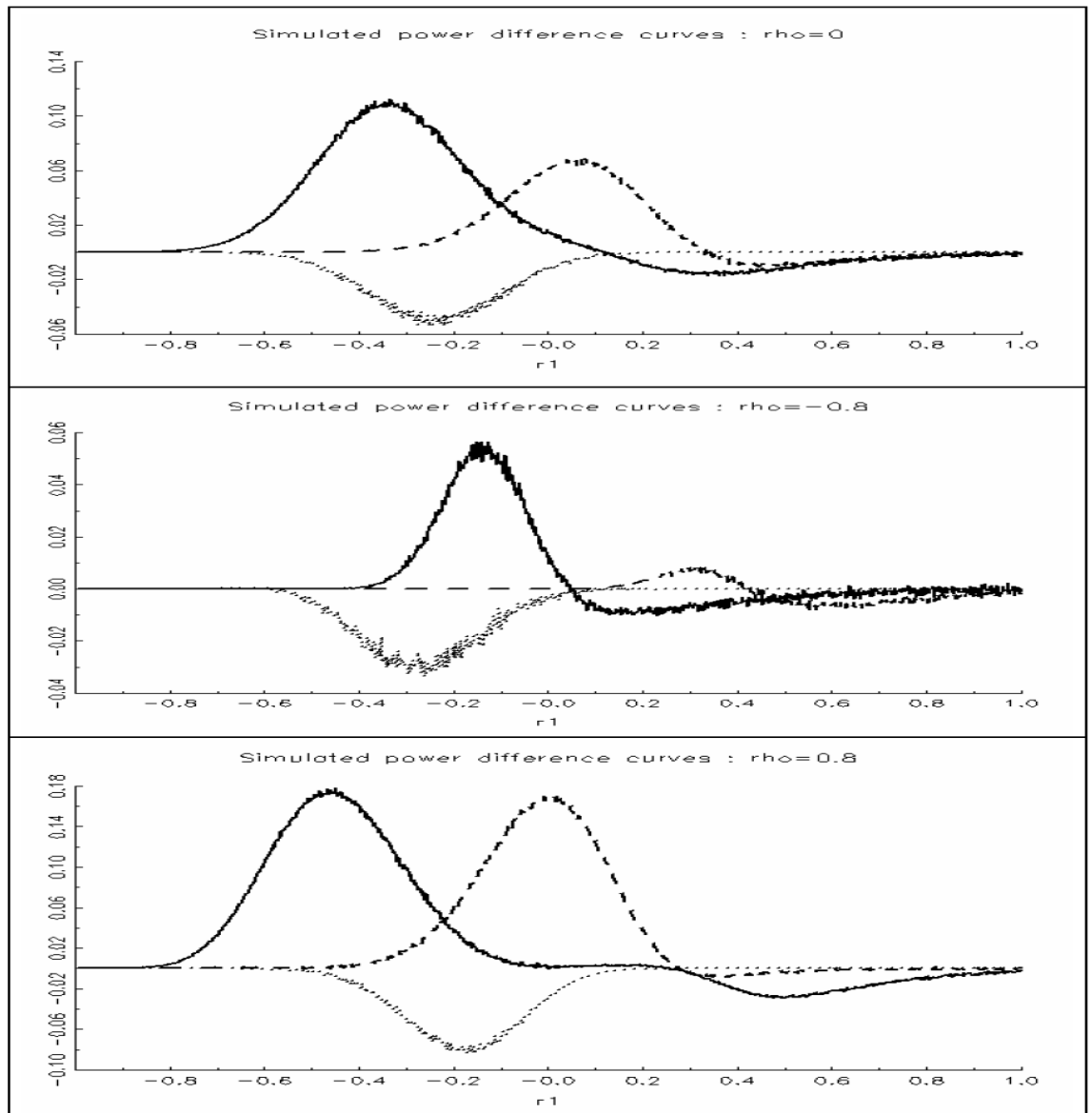
8 Proofs of theorems

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

$$A \equiv \{j \in S : \mu_j^* > 0\}, \ M \equiv \{j \in S : \mu_j^* = 0\}, \ W \equiv \{j \in S : \mu_j^* < 0\}.$$

8.1 Probability limits of the smoothing indicator

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

Lemma 5 (Probability Limits of the Smoothing Indicator)

Assume [A1] ~[A6], [D1] and [D3], then as $T \longrightarrow \infty$, the smoothing indicator $\Psi_T(\widehat{\theta}_j \widehat{\mu}_j) = \Psi(K(T)\widehat{\theta}_j \widehat{\mu}_j)$ has the following probability limits that depend on μ_j :

   (1)   *If $j \in A$, then $\sqrt{T}\Psi_T(\widehat{\theta}_j \widehat{\mu}_j) \overset{p}{\longrightarrow} 0$*

   (2)   *If $j \in M$, then $\Psi_T(\widehat{\theta}_j \widehat{\mu}_j) \overset{p}{\longrightarrow} \Psi(0)$*

   (3)   *If $j \in W$, then $\Psi_T(\widehat{\theta}_j \widehat{\mu}_j) \overset{p}{\longrightarrow} 1$*

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

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

By [D1] and [D3], we have $\widehat{\theta}_j \widehat{\mu}_j \overset{p}{\longrightarrow} \theta_j \mu_j$, which is strictly positive for $j \in A$. Then there is a $T_1(\eta)$ such that for $T > T_1$,

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

By [A1] and [A3], we have

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

where the third inequality follows because Ψ is a non-negative 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 $T > T_2$, $\sqrt{T}\Psi_T(\theta_j \mu_j/2) < \varepsilon$. Combining all these results, part (1) in this lemma follows by choosing $T^*(\varepsilon, \eta) = \max(T_1, T_2)$.

To show part (2), note that

$$
\begin{aligned}
\Psi_T(\widehat{\theta}_j\widehat{\mu}_j) &= \Psi(K(T)\widehat{\theta}_j\widehat{\mu}_j) \\
&= \Psi(\frac{K(T)}{\sqrt{T}}\sqrt{T}\widehat{\theta}_j\widehat{\mu}_j).
\end{aligned}
$$

If $j \in M$, by [D1] and [D3], we have $\sqrt{T}\widehat{\theta}_j\widehat{\mu}_j = O_p(1)$. By [A4] $\frac{K(T)}{\sqrt{T}} = o(1)$ so that $\frac{K(T)}{\sqrt{T}}\sqrt{T}\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{p} 0$. 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 $T^*(\varepsilon, \eta) > 0$ such that for $T \geq T^*(\varepsilon, \eta)$,

$$
P(1 - \varepsilon \leq \Psi_T(\widehat{\theta}_j\widehat{\mu}_j) \leq 1 + \varepsilon) \geq 1 - \eta. \tag{24}
$$

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

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

Note that if $j \in W$, then $\theta_j\mu_j < 0$ 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 $T > T_3$, $\Psi_T(\theta_j\mu_j/2) \leq 1 + \varepsilon$ and $\Psi_T(3\theta_j\mu_j/2) \geq 1 - \varepsilon$. Therefore, part (3) follows by choosing $T^*(\varepsilon, \eta) = \max(T_1, T_3)$. ∎

8.2   Asymptotic properties of $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j$

Based on Lemma 5, we derive the asymptotic properties of the components corresponding to $j \in A$, $j \in M$, $j \in W$ of the sum

$$
Q_1 = \sum_{j \in S} \sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j.
$$

The results are stated in the following lemma.

Lemma 6 (Asymptotic Properties of $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j$)

Let V_{jj} be the asymptotic variance of $\widehat{\mu}_j$. Assume [A1] ~[A6], [D1] and [D3], then as $T \longrightarrow \infty$, $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j$ has the following asymptotic properties that depend on μ_j :

(i)   *If $j \in A$, then $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{p} 0$*

(ii)   *If $j \in M$, 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 $j \in W$, 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], [D3] and part (1) of Lemma 5. To show part (ii), note that by [D1] and [D3], 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 5. To show part (iii), note that for $j \in W$,

$$\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j = \Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\sqrt{T}\widehat{\theta}_j(\widehat{\mu}_j - \mu_j^*) + \Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\sqrt{T}\widehat{\theta}_j\mu_j^*. \qquad (25)$$

Therefore, part (iii) follows from the fact that by [D1], [D3] and part (3) of Lemma 5, the first term on the right hand side of (25) is $Op(1)$ and the second term goes to $-\infty$ in probability. ■

8.3   Proof of Theorem 1

The proof of part (1) of Theorem 1 is obtained from the proof of Theorem 3 by setting $W = \varnothing$ and hence c_W vanishes.

Regarding part (2), we first consider the simpler situation where Ψ is chosen such that the event $\Psi_T(\widehat{\theta}_j\widehat{\mu}_j) = 0$ can occur with positive probability. Note that this is the case for Ψ being an origin-smoothed discrete function such as the step-at-unity smoothing indicator. In this scenario, part (1) of 5 implies that $P(\Psi_T(\widehat{\theta}_j\widehat{\mu}_j) = 0) \longrightarrow 1$ for $j \in A$ and hence $P(Q_2 = 0) \longrightarrow 1$. Now we consider the situation where Ψ is chosen such that the event $\Psi_T(\widehat{\theta}_j\widehat{\mu}_j) > 0$ occurs almost surely. Note that this is the case for Ψ being everywhere continuous such as the logistic smoothing indicator. In this scenario, Q_2 is almost surely positive and hence $Q = \Phi(Q_1/Q_2)$ almost surely. We write

$$\frac{Q_1}{Q_2} = \sum_{j \in A} \sqrt{T}\widehat{\mu}_j\widehat{\Upsilon}_j,$$

where

$$\widehat{\Upsilon}_j = \widehat{\theta}_j\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)/Q_2.$$

Note that each $\widehat{\Upsilon}_j$ is almost surely positive and each $\widehat{\mu}_j$ for $j \in A$ is strictly positive with probability tending to 1 as $T \longrightarrow \infty$. Therefore,

$$\frac{Q_1}{Q_2} \geq (\sum_{j \in A} \widehat{\Upsilon}_j)(\min_{j \in A}\{\sqrt{T}\widehat{\mu}_j\}) \geq \max_{j \in A}\{\widehat{\Upsilon}_j\}\min_{j \in A}\{\sqrt{T}\widehat{\mu}_j\}$$

with probability tending to 1 as $T \longrightarrow \infty$. Since

$$\max_{j \in A}\{\widehat{\Upsilon}_j\} \geq \sqrt{\frac{1}{p}\sum_{j \in A}\widehat{\Upsilon}_j^2}$$

and by eigenvalue theory

$$\sum_{j \in A} \widehat{\Upsilon}_j^2 = \frac{\widehat{\Psi}'\widehat{\Delta}\widehat{\Delta}\widehat{\Psi}}{\widehat{\Psi}'\widehat{\Delta}\widehat{V}\widehat{\Delta}\widehat{\Psi}} \geq \frac{1}{\lambda_{\max}},$$

29

where $\widehat{\lambda}_{\max}$ is the largest eigenvalue of $\widehat{V}$. Combining these inequalities, we have

$$\frac{Q_1}{Q_2} \geq \min_{j \in A}\{\sqrt{T}\widehat{\mu}_j\}\sqrt{\frac{1}{p\widehat{\lambda}_{\max}}}.$$

Because the mapping from a positive definite matrix to its maximum eigenvalue is continuous on the space of such matrices,

$$\widehat{\lambda}_{\max} \xrightarrow{p} \lambda_{\max} > 0,$$

where $\lambda_{\max}$ is the largest eigenvalue of V. Since for each $j \in A$,

$$\sqrt{T}\widehat{\mu}_j = \sqrt{T}(\widehat{\mu}_j - \mu_j^*) + \sqrt{T}\mu_j^*$$

in which by assumption [D1], the first term in the sum of the right-hand side is bounded in probability. The second term goes to infinity as $T \longrightarrow \infty$ since each μ_j^* is positive by supposition. Putting together all these results, we can conclude that Q_1/Q_2 goes to infinity in probability and hence

$$Q = \Phi(Q_1/Q_2) \xrightarrow{p} 1.$$

8.4 Proof of Theorem 2

Note that W is non-empty under H_1. By [D2], [D3] and Lemma 5, the probability limit of Q_2 is finite and strictly positive. Lemma 6 says that $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j$ is bounded in probability for $j \in S\backslash W$. Since $\sqrt{T}\Psi_T(\widehat{\theta}_j\widehat{\mu}_j)\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{p} -\infty$ for $j \in W$, Q_1/Q_2 goes to $-\infty$ in probability and hence $P(Q < \alpha) \longrightarrow 1$ as $T \longrightarrow \infty$ under H_1.

8.5 Proof of Theorem 3

Under [L1], for $j \in W$,

$$K(T)\widehat{\theta}_j\widehat{\mu}_j = (K(T)/\sqrt{T})\sqrt{T}\widehat{\theta}_j(\widehat{\mu}_j - \mu_j^*) + (K(T)/\sqrt{T})\widehat{\theta}_j c_j$$

and hence by [A4], [D1] and [D3], $K(T)\widehat{\theta}_j\widehat{\mu}_j \xrightarrow{p} 0$. Using [A2] and the continuous mapping theorem, $\Psi(K(T)\widehat{\theta}_j\widehat{\mu}_j) \xrightarrow{p} \Psi(0)$ for $j \in W$. This result and part (2) of Lemma 5 imply that $\Psi(K(T)\widehat{\theta}_j\widehat{\mu}_j) \xrightarrow{p} \Psi(0)$ for $j \in M \cup W$ and thus $\widehat{\Psi} \xrightarrow{p} \Psi(0)d_{MW}$. Therefore, under [D1], [D2], [D3] and [L1], Q_1 is asymptotically equivalent in probability to

$$\begin{aligned}
\Psi(0)d'_{MW}\Delta\sqrt{T}\widehat{\mu} &= \Psi(0)d'_{MW}\Delta\sqrt{T}\mu^* + \Psi(0)d'_{MW}\Delta\sqrt{T}(\widehat{\mu} - \mu^*) \\
&= \Psi(0)c_W + \Psi(0)d'_{MW}\Delta\sqrt{T}(\widehat{\mu} - \mu^*)
\end{aligned}$$

and thus has an asymptotic normal distribution with mean $\Psi(0)c_W$ and variance $\Psi(0)^2\omega_{MW}$. Using similar arguments, we find that $Q_2 \xrightarrow{p} \Psi(0)\omega_{MW}^{1/2}$. Therefore, Theorem 3 follows by noting that

$$\frac{Q_1}{Q_2} \xrightarrow{d} N(\omega_{MW}^{-1/2}c_W, 1).$$

30

8.6 Proof of Theorem 4

The proof of Theorem 4 proceeds along similar lines to that of Theorem 3 except that now μ_j^* is allowed to vary from both sides of the origin as specified in [L2]. In particular, $\Psi(K(T)\widehat{\theta}_j\widehat{\mu}_j) \xrightarrow{p} \Psi(0)$ for $j \in S$ and in this case Q_1 is asymptotically equivalent in probability to

$$\Psi(0)d'_{MW}\Delta\sqrt{T}\widehat{\mu} = \Psi(0)c_{AW} + \Psi(0)d'_S\Delta\sqrt{T}(\widehat{\mu} - \mu^*)$$

and thus has an asymptotic normal distribution with mean $\Psi(0)c_{AW}$ and variance $\Psi(0)^2\omega_S$. Furthermore, $Q_2 \xrightarrow{p} \Psi(0)\omega_S^{1/2}$. Therefore, Theorem 4 follows by noting that

$$\frac{Q_1}{Q_2} \xrightarrow{d} N(\omega_S^{-1/2}c_{AW}, 1).$$

9 Extension of the constraint chaining method to testing a continuum of inequalities

In this section we briefly sketch how the quasi-linear chaining test developed in the preceding sections could be extended to testing a continuum of inequalities such as those arising in stochastic dominance hypotheses. A full theory is outside the scope of this paper. Suppose instead of the test problem (4), we want to test

$$H_0 : \mu(x) \geq 0 \text{ for } x \in [a, b] \text{ versus } H_1 : \mu(x) < 0 \text{ for some } x \in [a, b] \qquad (26)$$

for the function $\mu(x)$, which is the parameter to be tested and a and b are given scalars. To test this hypothesis, we would use the following continuous analog for (11) and (12) as :

$$Q_1^c \equiv \sqrt{T}\int_a^b \Psi(K(T)\widehat{\mu}(x))\widehat{\mu}(x)dx$$

$$Q_2^c \equiv [\int_a^b \int_a^b \Psi(K(T)\widehat{\mu}(x))\Psi(K(T)\widehat{\mu}(y))\widehat{V}(x,y)dxdy]^{1/2},$$

where $\widehat{\mu}(x)$ is a consistent estimator of the function $\mu(x)$ such that

$$\sqrt{T}(\widehat{\mu}(x) - \mu(x)) \Longrightarrow G, \qquad (27)$$

where $\Longrightarrow$ denotes weak convergence as $T \to \infty$ and $G = \{G(x) : x \in [a, b]\}$ denotes some zero-mean Gaussian process with covariance kernel V consistently estimated (in an appropriate function norm) by $\widehat{V} = \{\widehat{V}(x, y) : (x, y) \in [a, b]^2\}$.

The continuous analogue of the key result of discrete chaining is that

$$Q_1^c \xrightarrow{d} \sqrt{T}\int_a^b 1\{\mu(x) = 0\}G(x)dx \qquad (28)$$

which implies that Q_1^c is asymptotically normal with mean zero and variance

$$\int_a^b \int_a^b 1\{\mu(x) = 0\}1\{\mu(y) = 0\}V(x,y)dxdy \tag{29}$$

the square root of which will be the probability limit of Q_2^c. Thus the ratio of Q_1^c to Q_2^c will be asymptotically unit normal under the null hypothesis whenever the above variance expression (29) is nonzero. Hence the test statistic Q defined as in Section 3, originally for the finite dimensional parameter scenario but now using Q_1^c and Q_2^c in place of the original Q_1 and Q_2, will continue to be valid in the present extension. Note that the asymptotic unit normality certainly holds at the LF point ($\mu(x) = 0$ for $x \in [a,b]$) of the null hypothesis where integral expression (29) simplifies to

$$\int_a^b \int_a^b V(x,y)dxdy$$

which is strictly positive for a non-degenerate Gaussian process.

To illustrate the continuous extension above, we use the simplest example of the stochastic dominance hypothesis in which it is desired to test whether a given known baseline cdf $H(x)$ first order stochastically dominates an unknown population cdf $F(x)$ on support $[a,b]$. Under this hypothesis, the function $\mu(x)$ defined as

$$\mu(x) \equiv F(x) - H(x), \ a \leq x \leq b,$$

satisfies the inequality

$$\mu(x) \geq 0 \text{ for } x \in [a,b].$$

This is precisely the null hypothesis of (26). We can use

$$\widehat{\mu}(x) = \widehat{F}(x) - H(x)$$

where $\widehat{F}(x)$ is an empirical cdf based on a random sample of observations. It is a known fact that root of sample size times the difference between an empirical cdf $\widehat{F}$ and the true population cdf F converges weakly to G_F, the $F-$brownian bridge process (van der Vaart (1998) pp 266). This limiting process forms the basis of derivation of $G(x)$ in (27).

In the context of testing (26), Linton, Song and Whang (2008) use a test statistic which, restricted to the present case of first-order stochastic dominance, simplifies to

$$\int_a^b [\min\{0, \sqrt{T}\widehat{\mu}(x)\}]^2 dx. \tag{30}$$

Under the null hypothesis, (30) converges in distribution to

$$\int_a^b 1\{\mu(x) = 0\}[\min\{0, \sqrt{T}\widehat{\mu}(x)\}]^2 dx. \tag{31}$$

Just as result (28) demonstrated for the chaining statistic, (31) also shows that the non-degenerate limiting null distribution of Linton-Song-Whang statistic

depends only on the "contact" set $\{x \in [a, b] : \mu(x) = 0\}$. A major objective of these authors is to design a bootstrap method in which test critical values are based on an estimate of this set rather than on the "least favorable" case which simply replaces $1\{\mu(x) = 0\}$ everywhere with the value 1. Their argument and solution is a continuous version of the discrete case advanced by Hansen (2005). In both versions the idea is to make the test asymptotically similar on the boundary of the null hypothesis to improve test performance. Asymptotic similarity on the boundary is also achieved by the chaining test of the present paper, but in a conceptually very different way. Rather than using estimates of contact points to adjust the critical values of test statistics *ex-post*, the construction of the chaining statistic *ex ante* reduces effects due to null hypothesis points outside the contact set by using the smoothed indicator function $\Psi(K(T)\widehat{\mu}(x))$. Thus fixed critical values can be used and bootstrapping adjustments are not needed.

10 Conclusions

This paper has set out four key properties which tests of multiple inequality constraints should satisfy. These are (1) (asymptotic) exactness, (2) (asymptotic) similarity on the boundary, (3) absence of nuisance parameters from the asymptotic null distribution of the test statistic, (4) low computational complexity and boostrapping cost. We noted that predominant tests currently used in econometrics do not appear to have all these properties simultaneously. In this paper, we demonstrated that it is possible to design a test procedure that simultaneously enjoys the four desired properties. We introduced a method of test construction called constraint chaining which begins by writing multiple inequalities as a single equality using zero-one indicator functions. We noted that substituting parameter estimates into this equality would lead to awkward distributions. To get round this, we harnessed an idea of Horowitz (1992) originally used to remedy a similar difficulty in parameter estimation theory rather than hypothesis testing. The idea is to first smooth the indicator functions. The approximate equality thus obtained is the basis of a well-behaved test. This test could also be considered as the baseline of an interesting wider class of tests.

We provided a full asymptotic theory for the baseline test. The smoothed indicator function in the test acts as an importance weight, enhancing the effect of inferentially critical parameter values and diminishing the effect of others. We therefore supplemented the asymptotic theory with a computable interpolating power function which captures the differential impact of importance weighting on power at local to origin points compared with medium and distant points. Simulation results show that the finite-sample performance of the test matches the theory quite well.

A number of topics for further research arise from the present paper :

(1) How far is test performance sensitive to the choice of smoothed indicator and the "bandwidth" or "tuning" parameter that determines its convergence to the discrete indicator ?. Is it feasible to pin down an optimum choice in some

metric ? Some preliminary ideas on this are discussed in Appendix A to this paper.

(2) Given the "smoothness" of the test statistic of this paper and the fact that it is asymptotically pivotal on the boundary, bootsrapping may enhance its performance. To investigate this fully, some work on the possibility of an Edgeworth-type expansion is called for. This may also give an alternative formal justification for the descriptive interpolating power function proposed in this paper.

(3) Even if the normality assumption made in the paper is dropped, the smoothing which eases the asymptotics still holds, as indicated in Appendix B to this paper. But, without normality, the asymptotic distribution of the test statistic will not be standard. However, it might be successfully obtainable by the bootstrap precisely because of the smoothing. So the research agenda item (2) above applies here too.

(4) The extension of the test to continuous functions sketched in the paper may benefit from a full and more general analysis based on empirical process theory.

An investigation of the finite-sample performance of this extension based on simulation at least and analytical derivation when tractable is desirable.

(5) We note that, whilst the philosophy of the approach of this paper leads uniquely to a statistic which is of "quasi-linear" form, this can be thought of as a special case of a wider class of forms all using only products of parameter estimates and smoothed indicators. A full investigation of this class is a matter for further research. Some findings are already available (Chen (2009, chapter 4)).

(6) Monte Carlo simulations involving comparison with other tests are of limited use without understanding the theoretical properties on which tests may differ ex ante and thus be responsible for experimental results ex post. Given that this paper has focussed upon such properties as a key motivation in the design of tests, a future research project comparing the practical performance of the chi-bar squared, extreme value and chaining tests is now worth pursuing.

A   A proposal of optimal choice of the smoothing indicator

Let $L(x|\mu^*)$ be the limiting distribution of the chaining statistic defined by $(13)^{14}$. Note that Theorems 1 and 2 imply that $L(x|\mu^*)$ depends on the value of μ^*. Let $F_T(x|\Psi, K)$ be the actual finite sample distribution of the test statistic Q implemented based on the smoothing indicator $\Psi_T(x) = \Psi(K(T)x)$ in which $\Psi(x)$ and $K(T)$ satisfy the regularity conditions $[A1] \sim [A6]$. We want to choose Ψ and K such that $F_T(x|\Psi, K)$ can be well approximated using $L(x|\mu^*)$ that has been derived in this paper. In other words, we consider the following criterion

---

^{14}For ease of exploration, in this appendix we assume $\theta_j = 1$ for all j.

for optimality of the smoothing indicator.

$$\sum_{T=1}^{\infty} \beta_T ||F_T(x|\Psi, K) - L(x|\mu^*)||_{2,\lambda}, \tag{32}$$

where $||f||_{2,\lambda}$ of a given function f is its λ-weighted L_2 norm defined as

$$||f||_{2,\lambda} \equiv [\int_0^1 f(x)^2 \lambda(x) dx]^{1/2},$$

and β_T is the weight reflecting how seriously the user takes discrepancies between F_T and L at sample size T. For instance, if it takes at least m observations to fulfill the rank condition of computing the OLS estimator $\widehat{\mu}$, then β_T can be set to zero for $T < m$. For hypothesis testing and p-values, the tails of distributions are more important than the central part and thus $\lambda(x)$ can be designed to reflect this concern.

Since $L(x|\mu^*)$ is also a function of μ^* and $F_T(x|\Psi, K)$ is unknown, the criterion (32) is infeasible. However, we can consider its empirical feasible version defined as

$$\sum_{T=1}^{\infty} \beta_T ||\widehat{F}_T(x|\Psi, K) - L(x|\widehat{\mu})||_{2,\lambda}, \tag{33}$$

where $\widehat{F}_T(x|\Psi, K)$ is the bootstrapped estimator of $F_T(x|\Psi, K)$ and $\widehat{\mu}$ is a consistent estimator of μ^*. Then the smoothing indicator (Ψ, K) is chosen to be the minimizer of the criterion (33) subject to the conditions [A1]$\sim$ [A6]. Though formulation of optimality of the smoothing indicator based on (33) is implementable, analysis of such optimization problem is a non-trivial task. The Monte Carlo simulation results discussed in Section 7 suggest that our developed interpolating power formula $P_T(\alpha|\Psi, K, \mu^*)$ matches well the finite sample simulated rejection probability. Thus we can consider a simplified criterion as follows by replacing $\widehat{F}_T(\alpha|\Psi, K)$ with $P_T(\alpha|\Psi, K, \widehat{\mu})$ for the test with significance level α[15]

$$\sum_{T=1}^{\infty} \beta_T |P_T(\alpha|\Psi, K, \widehat{\mu}) - L(\alpha|\widehat{\mu}))|. \tag{34}$$

Since P_T has a simple closed-form formula, it is easier to analyze the optimization problem based on the criterion (34). Note that

$$|\widehat{F}_T(\alpha|\Psi, K) - L(\alpha|\widehat{\mu})| \leq |\widehat{F}_T(\alpha|\Psi, K) - P_T(\alpha|\Psi, K, \widehat{\mu})| + |P_T(\alpha|\Psi, K, \widehat{\mu}) - L(x|\widehat{\mu})|.$$

Thus we require that the approximation error $|\widehat{F}_T(\alpha|\Psi, K) - P_T(\alpha|\Psi, K, \widehat{\mu})|$ is of small order of magnitude so that (34) can be a useful approximating criterion for optimality. We will investigate this issue in our further research work of this project.

---

[15]This amounts to taking $\lambda(x) = 1$ if $x = \alpha$ and $\lambda(x) = 0$, otherwise.

B The case of asymptotically non-normal estimators

The regularity condition [D1] assumes asymptotic normality of $\widehat{\mu}$:

$$\sqrt{T}(\widehat{\mu} - \mu^*) \xrightarrow{d} N(0, V).$$

If the normality assumption [D1] is dropped, can the constraint chaining method survive ? That will depend on what replaces [D1]. If [D1] is only required to hold under the null hypothesis, but under the alternative hypothesis $\widehat{\mu}$ still converges in probability to the true μ^*, then the asymptotic null distribution of the chaining test statistic remains as before and the test is still consistent. But local power properties cannot then be derived.

Suppose, more drastically, that [D1] is replaced by

$$\tau(T)(\widehat{\mu} - \mu^*) \xrightarrow{d} ND,$$

where $\tau(T)$ is a scalar increasing function of T (or diagonal matrix of increasing functions of T), and ND denotes some nondegenerate distribution which need not be normal. For example, in time series models, the presence of unit roots can cause certain parameter estimates to require normalisation by $\tau(T) = T^{3/2}$ rather than $T^{1/2}$ in order to get an asymptotically nondegenerate distribution. Moreover, the limit distribution in unit root models is often non-normal. In such cases, the numerator of the smoothing statistic would need to be adjusted to

$$Q_1 \equiv \widehat{\Psi}' \tau(T) \widehat{\Delta} \widehat{\mu}.$$

With appropriate modifications of the regular conditions [A1] to [A6] to reflect the change of $T^{1/2}$ to $\tau(T)$, the following key feature of the smoothing indicator still holds :

$$\widehat{\Psi} \xrightarrow{p} d_M \text{ under } H_0,$$

where d_M is a p-dimensional vector in which its j-th element is untiy if $\mu_j^* = 0$ and is zero, otherwise. Thus, under the null hypothesis, Q_1 would be asymptotically distributed as

$$d_M' \Delta X,$$

where $X \sim ND$.

It is at this point that the particular distribution of ND would have to be inserted to enable the next step of getting p-values. The required task of getting p-values will vary according to application and may be very difficult. Nonetheless, the chaining method has enabled us to get this far.

References

[1] Anderson, G. (1996), "Nonparametric Tests of Stochastic Dominance in Income Distributions", *Econometrica*, 64, 1183-1193.

[2] Andrews, D. (1998), "Hypothesis Testing with a Restricted Parameter Space", *Journal of Econometrics*, 84, 155-199.

[3] Andrews, D. (1999), "Estimation When a Parameter is on a Boundary", *Econometrica*, 67, 1341-83.

[4] Andrews, D. (2000), "Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space", *Econometrica*, 68, 399-405.

[5] Andrews, D. (2001), "Testing When a Parameter is on the Boundary of the Maintained Hypothesis", *Econometrica*, 69, 683-734.

[6] Andrews, D. (2002), "Generalized Method of Moments Estimation When a Parameter is on a Boundary", *Journal of Business and Economic Statistics*, 20, 530-44.

[7] Andrews, D. and P. Jia. (2008), "Inference for Parameters defined by Moment Inequalities : A Recommended Moment Selection Procedure", Cowles Foundation DP 1676. Yale University.

[8] Andrews, D. and G. Soares (2009), "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection", revision of Cowles Foundation DP 1631 (2007).

[9] Barrett, G. and S. Donald (2003), "Consistent Tests for Stochastic Dominance", *Econometrica*, 71, 71-104.

[10] Boudoukh, J., Richardson, M., Smith, T. and R. Whitelaw (1999), "Ex-Ante Bond Returns and the Liquidity Preference Hypothesis", *Journal of Finance*, 54, 1153-67.

[11] 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.

[12] Dagenais, M.G. and J-M. Dufour (1991), "Invariance, Nonlinear Models and Asymptotic Tests", *Econometrica*, 59, 1601-1615.

[13] Davidson, R. and J..Y. Duclos (2000), "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality", *Econometrica*, 68, 1435-1464.

[14] Dufour, J-M. (1989), "Nonlinear Hypotheses, Inequality Restrictions, and Non-Nested Hypotheses: Exact Simultaneous Test in Linear Regressions", *Econometrica*, 57, 335-355.

[15] Feng, Y. and J. Wang (2007), "Likelihood Ratio Test Against Simple Stochastic Ordering Among Several Multinomial Populations", *Journal of Statistical Planning and Inference*, 37, 1362-1374.

[16] Fisher, G., Willson, D. and K. Xu (1998), "An Empirical Analysis of Term Premiums Using Significance Tests for Stochastic Dominance", *Economics Letters*, 60, 195-203.

[17] Fleming, J., Kirby, C. and B. Ostdiek (2006), "Bootstrap Tests of Multiple Inequality Constraints on Variance Ratios", *Economics Letters*, 91, 343-348.

[18] Goldberg, A. S. (1992), "One-Sided and Inequality Tests for a Pair of Means", in *Contributions to Consumer Demand and Econometrics*, eds. R. Bewley and T. V. Hoa, New York: St. Martin's Press, 140-162.

[19] Gourieroux, C., A. Holly and A. Monfort (1982), "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters", *Econometrica*, 50, 63-80.

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

[21] Guggenberg, P., J. Hahn, and K. Kim (2008), "Specification Testing under Moment Inequalities", *Economics Letters*, 99, 375-378.

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

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

[24] Hansen, P.R. and A. Lunde. (2005), "A Forecast Comparison of Volatility Models: Does Anything Beat a Garch(1,1) ?", *Journal of Applied Econometrics*, 20, 873-889.

[25] Hazilla, M. and R.J. Kopp (1986), "Testing for Separable Functional Structure Using Temporary Equilibrium Models", *Journal of Econometrics*, 33, 119-141.

[26] Hochberg, Y. and Tamhane, A.C. (2007), Multiple Comparison Procedures, New York : Wiley.

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

[28] Horowitz, J. (2001) "The Bootstrap", in *Handbook of Econometrics*, Vol. 5, ed. by Z. Griliches and M. D. Intriligator , Amsterdam: North Holland, 3159-3228.

[29] Horowitz, J. (2002), "Bootstrap Critical Values for Tests Based on the Smoothed Maximum Score Estimator", *Journal of Econometrics*, 111, 141-167.

[30] King, M. L. and Smith, M.D. (1986), "Joint One-Sided Tests of Linear Regression Coefficients", *Journal of Econometrics*, 32, 367-387.

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

[32] Kodde, D.A. and F.C. Palm (1987), "A Parametric Test of the Negativity of the Substitution Matrix", *Journal of Applied Econometrics*, 54, 1243-1248.

[33] Koebel, B., Falk. M. and F. Laisney (2003), "Imposing and Testing Curvature Conditions on a Box-Cox Cost Function", *Journal of Business and Economic Statistics*, 21, 319-335.

[34] Kudo, A. (1963), "A Multivariate Analogue of the One-Sided Test", *Biometrika*, 50, 403-418.

[35] Lee, S., Linton, O. and Whang. J. (2009), "Testing for Stochastic Monotonicity", *Econometrica.* 77, 585-602.

[36] Linton, O., Maasoumi, E. and Y-J. Whang (2005), " Consistent Testing for Stochastic Dominance under General Sampling Schemes", *Review of Economic Studies,* 72, 735-765.

[37] Linton, O., Song, K. and Y-J. Whang (2008), " Bootstrap Tests of Stochastic Dominance with Asymptotic Similarity on the Boundary", Cemmap Working Paper CWP08/08. Institute for Fiscal Studies: London.

[38] Manski, C. F. (1975), "Maximum Score Estimation of the Stochastic Utility Model of Choice", *Journal of Econometrics*, 3, 205-228.

[39] Manski, C. F. (1985), "Semiparametric Analysis of Discrete Responses: Asymptotic Properties of the Maximum Score Estimator", *Journal of Econometrics*, 27,313-334.

[40] Martin, G.M., Reidy, A. and J. Wright (2009), "Does the Option Market Produce Superior Forecasts of Noise-Corrected Volatility Measures ? ", *Journal of Applied Econometrics*, 24, 77-104.

[41] McAdams, D. (2008), "Partial Identification and Testable Restrictions in Multi-Unit Auctions", *Journal of Econometrics*, 146, 74-85.

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

[43] Nuesch, P.E. (1966), "On the Problem of Testing Location in Multivariate Populations for Restricted Alternatives", *Annals of Mathematical Statistics*, 37, 113-119.

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

[45] Post, T. (2003), "Empirical Tests for Stochastic Dominance Efficiency", *Journal of Finance*, 5, 1905-21.

[46] Reis, R. (2006), "Inattentive Consumers", *Journal of Monetary Economics*, 53, 1761-1800.

[47] Richardson, M.. Richardson, P.A. and T. Smith (1992), "The Monotonicity of the Term Premium : Another Look" , *Journal of Financial Economics*, 31, 97-105.

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

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

[50] Savin, N. E. (1980), "The Bonferroni and Scheffe Multiple Comparison Procedures", *Review of Economic Studies*, 47, 255-273.

[51] Savin, N. E. (1984), "Multiple hypothesis testing", in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. D. Intriligator , Amsterdam: North Holland, 827-879.

[52] Shapiro, A. (1988), "Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis", *International Statistical Review*, 56, 49-62.

[53] Silvapulle, M.J. and P.K. Sen (2004), Constrained Statistical Inference, New York : Wiley.

[54] van Der Vaart, A. W. (1998), Asymptotic Statistics, Cambridge University Press.

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

[56] 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.

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

[58] Wolak, F. (1989a), "Local and Global Testing of Linear and Nonlinear Inequality Constraints in Nonlinear Econometric Models", *Econometric Theory*, 5, 1-35.

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

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

[61] Wolak, F. (2007), "Quantifying the Supply-Side Benefits from Forward Contracting in Wholesale Electricity Markets", *Journal of Applied Econometrics*, 22, 1179-1207.