

# Bayesian and frequentist tests of sign equality and other nonlinear inequalities

David M. Kaplan\*

Department of Economics, University of Missouri

July 14, 2015

## Abstract

Testing whether two parameters have the same sign is a nonstandard problem due to the non-convex shape of the parameter subspace satisfying the composite null hypothesis, which is a nonlinear inequality constraint. We describe a simple example where the ordering of likelihood ratio (LR), Wald, and Bayesian sign equality tests reverses the “usual” ordering: the Wald rejection region is a subset of LR’s, as is the Bayesian rejection region (either asymptotically or with an uninformative prior). Under general conditions, we show that non-convexity of the null hypothesis subspace is a necessary but not sufficient condition for this asymptotic frequentist/Bayesian ordering. Since linear inequalities only generate convex regions, a corollary is that frequentist tests are more conservative than Bayesian tests in that setting. We also examine a nearly similar-on-the-boundary, unbiased test of sign equality. Rather than claim moral superiority of one statistical framework or test, we wish to clarify the regrettably ineluctable tradeoffs.

*JEL classification:* C11, C12

*Keywords:* convexity, likelihood ratio, limit experiment, nonstandard inference, unbiased test, Wald

## 1 Introduction

Economic theory often concerns parameters’ signs, such as the negative cross-price elasticity of complements. More generally, theory often predicts inequality constraints

---

\*Email: kaplandm@missouri.edu. Mail: Department of Economics, University of Missouri, 909 University Ave, 118 Professional Bldg, Columbia, MO 65211-6040, United States. Thanks to Patrik Guggenberger for helpful comments and references.

within a parameter space. We explore differences between Bayesian and frequentist testing of such theoretical predictions, first in general, then in the case of testing sign equality.

Under general conditions, we characterize the role of convexity of the null hypothesis,  $H_0$ . More properly, we mean “convexity of the parameter subspace where  $H_0$  is satisfied,” but we use “convexity of  $H_0$ ” as shorthand throughout. We show that convexity of  $H_0$  is a sufficient but not necessary condition for frequentist testing to be more conservative than the Bayesian test minimizing posterior expected loss, either with an uninformative prior or asymptotically (assuming the prior’s influence disappears). “More conservative” means that the Bayesian test rejects with probability above the nominal level,  $\alpha$ , given any true parameter value on the boundary of  $H_0$ . Equivalently, non-convexity of  $H_0$  is a necessary but not sufficient condition for the Bayesian test to be more conservative.<sup>1</sup> Since all *linear* inequality constraints lead to convex  $H_0$ , an immediate corollary is that frequentist tests are more conservative than Bayesian tests of linear inequality constraints. This corollary seems to reflect common belief, and Kline (2011) discusses such a result for a certain type of linear inequality, but a general, formal proof appears absent in the literature.

Unfortunately, irrespective of the ordering, this Bayesian/frequentist difference means no test can achieve both Bayesian “coherence” and frequentist “calibration” in these settings, both of which are usually desirable properties. In decision-theoretic terms, no decision rule can minimize both posterior expected loss and maximum (over the parameter space) expected loss, using the common (implicit) loss function taking value  $1 - \alpha$  for type I error,  $\alpha$  for type II error, and zero otherwise. The frequentist test does not minimize posterior expected loss, and the Bayesian test can have up to

---

<sup>1</sup>In Lindley’s (1957) paradox, the frequentist test rejects while the Bayesian test does not, but the mechanism is a prior with probability mass on the true point null hypothesis value  $\theta = \theta_0$  and nowhere else.

100% type I or II error rate.

One example of a nonlinear inequality constraint null hypothesis arises in testing whether two scalar parameters,  $\theta_1$  and  $\theta_2$ , have the same sign, i.e.,  $H_0 : \theta_1\theta_2 \geq 0$ . We examine this example in more detail. This null hypothesis includes whether a parameter's sign is stable over time or geography, where the subscripts on  $\theta_1$  and  $\theta_2$  indicate different time periods or places. Another application is the hypothesis that a binary treatment  $T$  attenuates the effect of regressor  $X$ , i.e., that the sign of the coefficient on  $X$  ( $\theta_1$ ) is opposite that of the interaction term ( $\theta_2$ ). If  $\theta_1$  and  $\theta_2$  are point-identified reduced form parameters known to bound a structural parameter, but it is unknown whether  $\theta_1 \leq \theta_2$  or  $\theta_2 \leq \theta_1$ , then the inclusion of the value zero in the identified set is equivalent to  $H_0 : \theta_1\theta_2 \leq 0$ . More generally, sign equality testing can be applied to moment inequalities in the usual way (e.g., Andrews, 2012), but where the two moments interact with each other, so  $m_1m_2 \geq 0$  rather than  $m_1 \geq 0, m_2 \geq 0$ . We compare existing tests and a new test with improved frequentist properties.

The ordering among Wald, likelihood ratio (LR), and Lagrange multiplier (LM) tests for linear equality (Berndt and Savin, 1977) and inequality (Gouriéroux et al., 1982, eqn. (8)) constraint testing is usually given in terms of test statistic size,  $\text{Wald} \geq \text{LR} \geq \text{LM}$ . In terms of rejection regions,  $\text{Wald} \supseteq \text{LR} \supseteq \text{LM}$ . With a simple sign equality testing DGP, we show  $\text{Wald} \subseteq \text{LR} = \text{LM}$ .

The Bayesian sign equality test's rejection region is usually a subset of the frequentist rejection region, although the opposite is true when the correlation between estimators  $\hat{\theta}_1$  and  $\hat{\theta}_2$  is quite negative. This accords with our general result that a non-convex  $H_0$  is necessary but not sufficient for the Bayesian test to be more conservative.

Our general result also explains the contrast between sign equality testing and other settings where frequentist testing is more conservative. For example, for testing

linear inequality constraints of the form  $H_0 : \theta \geq 0$  with  $\theta \in \mathbb{R}^d$ , Kline (2011) finds frequentist testing to be more conservative (e.g., his Figure 1). As another example, frequentist inference is more conservative under set identification: asymptotically, the frequentist confidence sets (Imbens and Manski, 2004; Stoye, 2009) are strictly larger than the estimated identified set rather than strictly smaller like the Bayesian credible sets, as shown by Moon and Schorfheide (2012, Cor. 1).<sup>2</sup> Our setup is not directly comparable to theirs since a Bayesian credible set cannot be inverted into a test (as the frequentist framework permits). Our setup accommodates null hypothesis inequalities on the point-identified (“reduced form”) parameters corresponding to “the identified set contains zero,” which is different than “the set-identified parameter equals zero”: with a continuous posterior distribution, the probability of the latter is always zero, while that of the former can approach 100%.

Although the Bayesian sign equality test cannot be improved from a Bayesian perspective, the LR test is biased and conservative from a frequentist perspective. Under general conditions, we propose a new, nearly unbiased test with exact asymptotic size. This test can be derived primarily using the general setup of Chiburis (2008), with the addition of some constraints and analytic results. We also compare the test resulting from the discrete approximation method in Moreira and Moreira (2013, p. 19). Despite its bias, the LR test may still be preferred for possessing a type of “rejection monotonicity” similar to that in Section 4.2 of Kline (2011).

The existence of a nearly unbiased (and thus similar-on-the-boundary) sign equality test with good power highlights another important difference between  $H_0 : \theta_1 \theta_2 \geq 0$  and the more commonly studied  $H_0 : \theta_1 \geq 0, \theta_2 \geq 0$ . As Andrews (2012) says in the title, “Similar-on-the-boundary tests for moment inequalities exist, but have poor

---

<sup>2</sup>There seems to be a typo in the statement of Corollary 1(ii), switching the frequentist and Bayesian sets from their correct places seen in the Supplemental Material proof.

power.” Specifically, he considers  $H_0 : \theta \geq 0$ ,  $H_1 : \theta \not\geq 0$ ,  $\theta \in \mathbb{R}^d$ , as in his (1.1), under a single draw of  $X \sim N(\theta, \Sigma)$ . With  $\theta = (\theta_1, \theta_2)$ , his Theorem 3 states that a similar-on-the-boundary test will have rejection probability  $\alpha$  for all  $\{\theta : \theta_1\theta_2 = 0\}$ , which is problematic with  $H_0 : \theta \geq 0$  since, for example, the point  $(0, -100)$  is very far from  $H_0$ . However, for sign equality, all points in  $\{\theta : \theta_1\theta_2 = 0\}$  are on the boundary; none are far from  $H_0$ . Consequently, Theorem 3 in Andrews (2012) does not imply poor power for similar-on-the-boundary tests of sign equality.

In higher dimensions, though, Theorem 3 of Andrews (2012) has bite even for sign equality. Let  $H_0 : \theta \geq 0$  or  $\theta \leq 0$ ,  $\theta \in \mathbb{R}^3$ . The boundary of  $\{\theta : \theta \geq 0\}$  is a subset of the boundary of the sign equality  $H_0$ , so any similar-on-the-boundary sign equality test must also be similar on the boundary of  $\{\theta : \theta \geq 0\}$ . Consequently, Theorem 3 of Andrews (2012) applies: any similar-on-the-boundary sign equality test must have rejection probability  $\alpha$  in set  $\{\theta : \theta_1 = 0, \theta_2 < 0, \theta_3 > 0\}$  (among others), which is not on the boundary.

The curse of dimensionality hinders the computational search for a similar-on-the-boundary frequentist test, although casting the search for an optimal test as a linear program as in Chiburis (2008) greatly speeds computation. Computationally, the Bayesian approach is much simpler, e.g., with a Bayesian bootstrap (Rubin, 1981). The important caveat is that the Bayesian test can be a poor decision rule from a minimax risk perspective: it can have up to 100% type I or II error rate in some cases.

A natural question is whether any method can attain a mix of (some) Bayesian and frequentist properties, similar to the bet-proof confidence sets in Müller and Norets (2014). An application of their method to testing sign equality would start with a one-sided bet-proof confidence interval for  $\theta_1\theta_2$ , rejecting  $H_0 : \theta_1\theta_2 \geq 0$  if the interval is strictly negative. Since the interval must be a superset of a Bayesian

credible interval with respect to some prior, the test is more conservative than the corresponding Bayesian test. Given our result that the Bayesian sign equality test is (usually) more conservative than the frequentist test, this initially suggests the bet-proof approach simply sides with the Bayesian perspective here. However, our results assume an asymptotically normal posterior, which excludes cases like Lindley’s (1957) paradox that has prior probability mass on the true value. We leave investigation of such methods to future work.

In the literature, Wolak (1991) discusses frequentist tests of nonlinear inequality constraints in nonlinear models. His setting is more general than our sign equality example, the result being that it is difficult even to find any test with asymptotically exact size (though his Lemma 1 helps). Kodde and Palm (1986) also discuss a Wald test of nonlinear inequalities; more recently, Donald and Hsu (2011) propose a simulation method to find frequentist critical values without relying on the least favorable null. This paper focuses instead on Bayesian/frequentist differences and on finding an unbiased sign equality test. Inequality constraint testing also applies to specification testing of moment inequality models as in Guggenberger et al. (2008), among others.

Section 2 characterizes the role of null hypothesis convexity in determining the ordering of Bayesian and frequentist hypothesis tests. Section 3 compares existing sign equality tests. Section 4 describes the new test. Proofs are collected in Appendix A. Acronyms used include Lagrange multiplier (LM), likelihood ratio (LR), maximum likelihood (ML) estimator (MLE), and rejection probability (RP), and  $\Phi(\cdot)$  is the standard normal CDF with quantiles  $z_\alpha \equiv \Phi^{-1}(\alpha)$ .

## 2 Convex and non-convex null hypotheses

We first discuss testing in one dimension for intuition and then provide general results.

## 2.1 One dimension

Let  $\theta \in \mathbb{R}$  be the parameter of interest. For illustration, consider a Gaussian shift experiment where the sampling and posterior distributions are, respectively,

$$\hat{\theta} \sim N(\theta, 1), \quad \theta \sim N(\hat{\theta}, 1). \quad (1)$$

This may be seen either as a scaled finite-sample problem with a normal likelihood or as a limit experiment. Let  $\alpha$  denote the test level, which for the Bayesian test means rejecting if and only if  $P(H_0) \leq \alpha$ .

The Bayesian/frequentist differences below may be framed in decision-theoretic terms. The Bayesian test examined here is a generalized Bayes rule that minimizes posterior expected loss when the loss function takes value  $1 - \alpha$  for type I error,  $\alpha$  for type II error, and zero otherwise. The unbiased frequentist test minimizes maximum (over  $\theta$ ) expected loss under the same loss function, as is often true of unbiased tests (e.g., Lehmann and Romano, 2005, Problem 1.10). Ideally, a single decision rule satisfies both properties.

If  $H_0 : \theta \leq c_0$ , then the frequentist and Bayesian tests are equivalent: both reject if and only if  $\hat{\theta} > c_0 + z_{1-\alpha}$ . Given the least favorable null of  $\theta = c_0$ ,  $P(\hat{\theta} > c_0 + z_{1-\alpha}) = 1 - \Phi(z_{1-\alpha}) = \alpha$ . Given  $\hat{\theta} = c_0 + z_{1-\alpha}$ ,  $P(H_0) = P(\theta \leq c_0) = \Phi(-z_{1-\alpha}) = \alpha$ .

Any convex one-dimensional set is an interval  $(-\infty, b]$ ,  $[a, b]$ , or  $[a, \infty)$ , or the corresponding open intervals that are practically equivalent (for our purposes). These are also the only possible sets generated by linear inequality constraints. The null hypothesis  $H_0 : \theta \leq c_0$  may be expressed as  $H_0 : \theta \in \Theta_0$  with  $\Theta_0 = (-\infty, c_0]$ . By symmetry, the frequentist/Bayesian equivalence also holds for  $\Theta_0 = [c_0, \infty)$ . The only remaining case is  $\Theta_0 = [a, c_0]$ .

Given  $\Theta_0 = [a, c_0]$ , the test of  $H_0 : \theta \leq c_0$  still rejects with probability  $\alpha$  if  $\theta = c_0$ .

In contrast, given  $\hat{\theta} = c_0 + z_{1-\alpha}$ ,  $P(\theta \in \Theta_0) < P(\theta \leq c_0) = \alpha$ , so the Bayesian test will now reject some  $\hat{\theta} < c_0 + z_{1-\alpha}$ . Without loss of generality,  $[a, c_0]$  may be centered to  $[-h, h]$ .

**Proposition 1.** *Given (1) and  $H_0 : \theta \in [-h, h]$ , the frequentist and Bayesian rejection regions are respectively  $(-\infty, -c_f] \cup [c_f, \infty)$  and  $(-\infty, -c_B] \cup [c_B, \infty)$ , where*

$$\Phi(h - c_f) + \Phi(-c_f - h) = \alpha = \Phi(h - c_B) - \Phi(-h - c_B) \quad (2)$$

and  $c_B < h + z_{1-\alpha} < c_f$ . If  $h < z_{0.5+\alpha/2}$ , then the second equality in (2) becomes an inequality and  $c_B = 0$ , i.e., the Bayesian test always rejects.

In Proposition 1, the frequentist test is too conservative by Bayesian standards, and the Bayesian test has up to 100% type I error rate, which is bad from a minimax perspective.

One non-convex example is the *nonlinear* inequality constraint test of

$$H_0 : \theta^2 > h^2 \text{ against } H_1 : \theta^2 \leq h^2 \quad (3)$$

for  $h \geq 0$ . Proposition 2 states that the Bayesian test is more conservative.

**Proposition 2.** *Given (1) and (3), the frequentist and Bayesian rejection regions are respectively  $(-c_f, c_f)$  and  $(-c_B, c_B)$ , where*

$$\Phi(c_f - h) - \Phi(-c_f - h) = \alpha = \Phi(c_B - h) + \Phi(-c_B - h) \quad (4)$$

and  $c_B < h + z_\alpha < c_f$ . If  $h < z_{1-\alpha/2}$ , then the second inequality in (4) becomes an inequality and  $c_B = 0$ , i.e., the Bayesian test never rejects.

In Proposition 2, the frequentist test rejects too often when judged by posterior probability, while the Bayesian test has up to 100% type II error rate, which is bad from a minimax perspective.

However, there are non-convex  $H_0$  besides (3) for which the frequentist test is still more conservative than the Bayesian test. Consider  $H_0 : \theta \in \{0, \epsilon\}$ , a set of two points. The Bayesian test always rejects since  $P(H_0) = 0$  given any  $\hat{\theta}$ , while the frequentist test does not always reject. In light of this, Corollary 3 summarizes Propositions 1 and 2.

**Corollary 3.** *Given (1), non-convexity of  $H_0$  is a necessary but not sufficient condition for Bayesian testing to be more conservative than frequentist testing.*

## 2.2 General results

We now present formal results under more general assumptions.

**Assumption 1.** As the sample size  $n \rightarrow \infty$ , the asymptotic distribution of  $\hat{\theta}$ , an estimator of interior point  $\theta \in \Theta \subseteq \mathbb{R}^k$ , is  $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$ , and there exists a consistent estimator  $\hat{V} \xrightarrow{p} V$ . The asymptotic Bayesian posterior distribution is the same, but interpreting  $\theta$  as random and  $\hat{\theta}$  as fixed. The  $\sqrt{n}$  rate may be replaced by any rate  $a_n \rightarrow \infty$ .

Lower-level conditions sufficient for the sampling and posterior distributions' asymptotic equivalence in Assumption 1, including semiparametric models like GMM and quantile regression, are given and discussed in Hahn (1997, Thm. G and footnote 13), Kwan (1999, Thm. 2), Kim (2002, Prop. 1), and Sims (2010, Sec. III.2), among others. Results for posterior asymptotic normality in parametric models date back to 1847 (per Johnson, 1970), with many more in the 20th century.

Simpler to imagine than Assumption 1 is the limit experiment, a single draw of  $X \sim N(\theta, V)$  with known  $V$ , and an improper uninformative prior to yield posterior  $\theta \sim N(X, V)$ .

**Assumption 2.** Let  $\alpha \in (0, 1/2)$ . Define the Bayesian test to reject if and only if  $P(H_0) \leq \alpha$ , using the asymptotic posterior in Assumption 1. With respect to the sampling distribution in Assumption 1, the frequentist test maintains asymptotic type I error rate no greater than  $\alpha$  given any  $\theta$  satisfying  $H_0$ , with asymptotically exact  $\alpha$  rejection probability at one or more  $\theta$  satisfying  $H_0$ .

Lemma 4 corresponds to the discussion of  $H_0 : \theta \leq c_0$  in one dimension, and Theorem 5 gives the main result.

**Lemma 4.** *Let  $\theta \in \mathbb{R}^k$ , and let  $H_0 : c'\theta \leq c_0$  for vector of constants  $c \in \mathbb{R}^k$  and  $c_0 \in \mathbb{R}$ . Equivalently,  $H_0 : \theta \in \Theta_0$ , where  $\Theta_0$  is the half-space  $\{t : c't \leq c_0\}$ . Then, the Bayesian and frequentist tests in Assumption 2 are asymptotically equivalent, rejecting if and only if  $\sqrt{n}(c'\hat{\theta} - c_0)/\sqrt{c'\hat{V}c} > z_{1-\alpha}$ . The asymptotic rejection probability is exactly  $\alpha$  for all  $\theta$  solving  $c'\theta = c_0$ , and the asymptotic posterior probability of  $H_0$  given any  $\hat{\theta}$  along the boundary of the rejection region is exactly  $\alpha$ .*

**Theorem 5.** *Let  $H_0 : \theta \in \Theta_0 \subset \mathbb{R}^k$ , where  $\Theta_0$  is convex, and  $H_1 : \theta \in \Theta_0^c \equiv \mathbb{R}^k \setminus \Theta_0$ . Under Assumptions 1 and 2, frequentist testing of  $H_0$  against  $H_1$  is asymptotically “more conservative” than Bayesian testing: the (frequentist) rejection probability of the Bayesian test given any  $\theta$  on the boundary of  $\Theta_0$  is at least as big as  $\alpha$ .*

An immediate implication of Theorem 5 is that non-convexity is a necessary but not sufficient condition for Bayesian testing to be more conservative than frequentist testing. Beyond Proposition 2, testing sign equality is another example where Bayesian testing can be more conservative.

The fixed  $\Theta_0$  in Theorem 5 could be modified to shrink with  $n$  so that smooth (differentiable)  $\Theta_0$  do not degenerate to half-spaces in the limit. For example, if  $\Theta_0 = \{\theta = (\theta_1, \theta_2) : \|\theta\|_2 \leq 1\}$ , then as  $n \rightarrow \infty$ , restricting attention to a  $n^{-1/2}$

neighborhood around  $(0, 1)$ , the null simplifies to  $H_0 : \theta_2 \leq 1$ . Instead,  $\Theta_0 = \{\theta : \|\theta\|_2 \leq 1/\sqrt{n}\}$  lets our asymptotic approximation capture a smooth, non-degenerate (strictly convex)  $\Theta_0$ . Examples like  $\Theta_0 = \{\theta : \theta_1 \geq 0, \theta_2 \geq 0\}$  do not degenerate either way.

Corollary 6 states the implication of Theorem 5 for linear inequality testing.

**Corollary 6.** *For any (joint) linear inequality constraint null hypothesis, under Assumptions 1 and 2, frequentist testing is asymptotically “more conservative” than Bayesian testing in the sense given in Theorem 5.*

For the important special case of  $H_0 : \theta \geq 0$  (elementwise) and  $H_1 : \theta \not\geq 0$ , with  $\theta \in \mathbb{R}^d$ , Kline (2011, p. 3136) explains the possible divergence of Bayesian and frequentist conclusions when the dimensionality  $d$  grows. He gives the example of  $\hat{\theta} = 0$ , where the Bayesian  $P(H_0) \approx 0$  while the frequentist  $p$ -value is near one. If instead  $H_0 : \theta \not\geq 0$  and  $H_1 : \theta \geq 0$ , then  $P(H_0) \approx 1$  can occur inside a frequentist rejection region for a test of level  $\alpha \approx 0$ . Theorem 11 states a less extreme version of this phenomenon for testing sign equality.

### 3 Comparison of existing sign equality tests

In Section 3.1, we show that either Bayesian or frequentist testing may be more conservative with the null hypothesis of sign equality. In Section 3.2, we compare a variety of sign equality tests in a simple example where the “usual” test ordering is reversed.

### 3.1 Bayesian/frequentist ordering

Let  $\theta_1 \in \mathbb{R}$  and  $\theta_2 \in \mathbb{R}$  be the parameters of interest. The null hypothesis is  $H_0 : \theta_1\theta_2 \geq 0$ , i.e., that  $\theta_1$  and  $\theta_2$  have the same sign (positive or negative). Similar to the Gaussian shift experiment in Section 2.1, for illustration, let the sampling and posterior distributions be

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} \sim N \left( \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N \left( \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

If  $\rho = 1$ , then a non-convex, one-dimensional null hypothesis results. Observing  $\hat{\theta}_2 - \hat{\theta}_1 = b$  implies  $\theta_2 - \theta_1 = b$ , reducing the problem to one parameter with  $H_0 : \theta_1(\theta_1 + b) \geq 0$ . If  $b > 0$ , then  $H_0 : \theta_1 \notin (-b, 0)$ ; if  $b \leq 0$ , then  $H_0 : \theta_1 \notin (0, -b)$ . Proposition 2 says the Bayesian test is more conservative in this case.

If  $\rho = -1$ , then a convex, one-dimensional null hypothesis results. Observing  $\hat{\theta}_1 + \hat{\theta}_2 = a$  implies  $\theta_1 + \theta_2 = a$ , reducing the problem to  $H_0 : \theta_1(a - \theta_2) \geq 0$ . If  $a > 0$ , then  $H_0 : \theta_1 \in [0, a]$ ; if  $a \leq 0$ , then  $H_0 : \theta_1 \in [a, 0]$ . Proposition 1 says the frequentist test is more conservative in this case.

As shown in Theorem 5, non-convexity is not sufficient for the Bayesian test to be more conservative. For testing sign equality, the Bayesian/frequentist ordering depends on  $\rho$ .

The frequentist rejection regions for  $\rho = -1$  and  $\rho = 1$  are seen in the top-left and bottom-right graphs in Figure 4 for  $\alpha = 0.05$ . Interestingly, the Bayesian test when  $\rho = 1$  is equivalent to the frequentist test when  $\rho = -1$ , and the Bayesian test when  $\rho = -1$  is equivalent to the frequentist test when  $\rho = 1$  (unless  $\hat{\theta}_1 = \hat{\theta}_2$ ). For example, when  $\rho = -1$ , if  $a \equiv \hat{\theta}_1 + \hat{\theta}_2 < 0$ , the posterior  $P(H_0)$  is  $\Phi(-\hat{\theta}_1) - \Phi(-\hat{\theta}_1 + a) = \Phi(-\hat{\theta}_1) - \Phi(\hat{\theta}_2)$ , and similarly  $\Phi(\hat{\theta}_2) - \Phi(-\hat{\theta}_1)$  if  $a \geq 0$ , so for any  $a$ ,  $P(H_0) = \left| \Phi(\hat{\theta}_2) - \Phi(-\hat{\theta}_1) \right|$ . With  $\rho = 1$ , if  $b \equiv \hat{\theta}_2 - \hat{\theta}_1 > 0$  and  $\hat{\theta}_1 \geq -b/2$ , then the

frequentist test rejects when

$$\alpha \geq \Phi(\hat{\theta}_1) - \Phi\left(\hat{\theta}_1 - 2[\hat{\theta}_1 - (-b/2)]\right) = \Phi(\hat{\theta}_1) - \Phi(-\hat{\theta}_2) = \Phi(\hat{\theta}_2) - \Phi(-\hat{\theta}_1),$$

identical to the Bayesian rejection criterion when  $\rho = -1$ . Similar analysis shows the equivalence in the other cases.

### 3.2 Ordering in simple example

We compare various tests of  $H_0 : \theta_1\theta_2 \geq 0$  in a simple example:

$$y_{i1} \stackrel{iid}{\sim} N(\theta_1, 1), 1 \leq i \leq n; \quad y_{k2} \stackrel{iid}{\sim} N(\theta_2, 1), 1 \leq k \leq n; \quad y_{i1} \perp y_{k2}, \forall i, k. \quad (5)$$

Figure 1 depicts the rejection regions given in Lemmas 7–10.

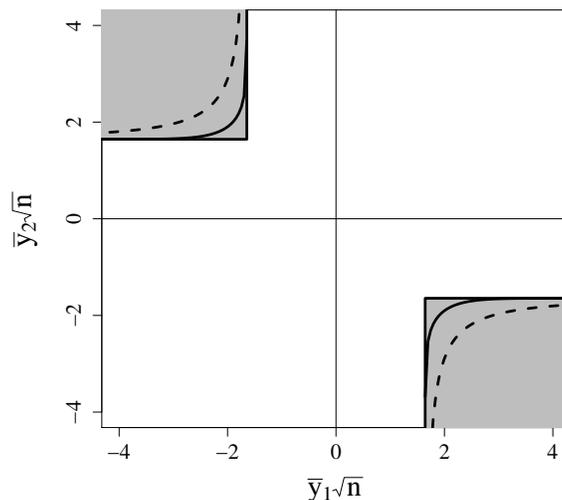


Figure 1: Rejection regions for different sign equality tests under model (5),  $\alpha = 0.05$ . The shaded area is the LR region from Lemma 8, which is also the rejection region for LM and LR-type GMM tests. The solid curved line is the boundary of the Bayesian region from Lemma 7. The dashed line is the boundary of the Wald region from Lemma 9. The horizontal and vertical axes show  $\sqrt{n}\bar{y}_1$  and  $\sqrt{n}\bar{y}_2$ , respectively.

With normal priors on  $\theta_1$  and  $\theta_2$ , specifically

$$\theta_1 \sim N(m_1, \tau_1^2), \quad \theta_2 \sim N(m_2, \tau_2^2), \quad (6)$$

the posteriors and asymptotic posteriors are

$$\theta_1 | \bar{y}_1 \sim N\left(\frac{\tau_1^2 \bar{y}_1 + m_1/n}{\tau_1^2 + 1/n}, \frac{\tau_1^2}{n\tau_1^2 + 1}\right), \quad \theta_2 | \bar{y}_2 \sim N\left(\frac{\tau_2^2 \bar{y}_2 + m_2/n}{\tau_2^2 + 1/n}, \frac{\tau_2^2}{n\tau_2^2 + 1}\right), \quad (7)$$

$$\sqrt{n}(\theta_1 - \bar{y}_1) \xrightarrow{d} N(0, 1), \quad \sqrt{n}(\theta_2 - \bar{y}_2) \xrightarrow{d} N(0, 1), \quad (8)$$

assuming  $\tau_1^2, \tau_2^2 \neq 0$ . Equation (8) is also the limit of (7) as  $\tau_1^2, \tau_2^2 \rightarrow \infty$ , i.e., the finite-sample posterior given improper uninformative priors. Equation (8) is also obtained by nonparametric Bayesian methods that do not assume a correctly specified likelihood; see, for example, Theorem 4.1 of Lo (1987) for the special case of Rubin's (1981) Bayesian bootstrap and Theorem 2.1 of Weng (1989), which includes general Dirichlet process priors. See also Kline (2011) for discussion of these different interpretations. Lemma 7 characterizes the asymptotic Bayesian test.

**Lemma 7.** *For the model in (5) and priors in (6) with  $\tau_1, \tau_2 \neq 0$ , the region where the asymptotic posterior probability of  $H_0$  is  $\alpha$  or less is*

$$\{(\bar{y}_1, \bar{y}_2) : [1 - \Phi(-\sqrt{n}\bar{y}_2)] [1 - \Phi(-\sqrt{n}\bar{y}_1)] + \Phi(-\sqrt{n}\bar{y}_1)\Phi(-\sqrt{n}\bar{y}_2) \leq \alpha\}.$$

From (5), iso-likelihood curves are concentric circles around  $(\bar{y}_1, \bar{y}_2)$ , so the likelihood ratio (LR) is equivalent to the Euclidean distance from  $(\bar{y}_1, \bar{y}_2)$  to the  $H_0$  region. The critical value is pinned down by considering  $\theta_2 \rightarrow \pm\infty$  (the least favorable null), and Lemma 8 results.

**Lemma 8.** *Under (5), with  $z_{1-\alpha} \equiv \Phi^{-1}(1 - \alpha)$ , the level- $\alpha$  LR rejection region is*

$$\{(\bar{y}_1, \bar{y}_2) : \max\{\min\{\bar{y}_1, -\bar{y}_2\}, \min\{-\bar{y}_1, \bar{y}_2\}\} \geq z_{1-\alpha}/\sqrt{n}\}.$$

Arguably the most natural<sup>3</sup> Wald statistic for  $H_0 : \theta_1\theta_2 \geq 0$  is

$$\widehat{W}_n = \frac{\sqrt{n}\widehat{\theta}_1\widehat{\theta}_2}{\sqrt{\widehat{\theta}_2^2 + \widehat{\theta}_1^2}}, \quad (9)$$

where the denominator is from the Delta method when  $(\theta_1, \theta_2) \neq (0, 0)$ . Using  $\widehat{W}_n$  is equivalent to using  $W_n$  on page 984 of Wolak (1991) or  $D$  in (2.16) of Kodde and Palm (1986), where  $W_n = \widehat{W}_n^2$  if  $\widehat{W}_n < 0$  and  $W_n = 0$  if  $\widehat{W}_n \geq 0$ .

**Lemma 9.** *For model (5), the rejection region for a level- $\alpha$  Wald test using (9) is*

$$\left\{ (\bar{y}_1, \bar{y}_2) : \frac{\sqrt{n}\bar{y}_1\bar{y}_2}{\sqrt{\bar{y}_1^2 + \bar{y}_2^2}} \leq z_\alpha \right\}.$$

Given (5), the Lagrange multiplier (LM) test (or “score test”) and LR-type GMM test have the same test statistic,  $\inf_{\theta \in H_0} n\|\bar{y} - \theta\|_2^2$ , where  $\|x\|_2 = \sqrt{x_1^2 + x_2^2}$  is the Euclidean ( $L_2$ ) vector norm. This is equivalent to the LR statistic and thus generates the same rejection region as LR, even in finite samples.

**Lemma 10.** *For model (5), the rejection region for a level- $\alpha$  LM test and that for a level- $\alpha$  LR-type GMM test are identical to the LR region in Lemma 8.*

The Wald test is strictly worse than the LR (and others) by frequentist criteria. The Wald rejection region is a subset of LR’s, so LR has greater power against any alternative while still controlling size.

The Bayesian test is similarly worse than LR by frequentist criteria, but the LR test is worse by Bayesian criteria. This is the same incompatibility seen in Section 2.

Even the LR test, which is uniformly most powerful for testing a single parameter’s sign, is a conservative sign equality test by frequentist standards. Theorem 11 states that all tests considered have rejection probability strictly below  $\alpha$  when  $\theta_1 = \theta_2 = 0$ .

---

<sup>3</sup>We ignore the multiplicity of Wald tests from equivalent null hypotheses  $H_0 : (\theta_1\theta_2)^r \geq 0$  for odd  $r$ .

**Theorem 11.** *For model (5) and test level  $\alpha \in (0, 0.5)$ , if  $\theta_1 = \theta_2 = 0$ , then the asymptotic rejection probabilities of the Bayesian, LR, Wald, LM, and LR-type GMM tests from Lemmas 7–10 satisfy*

$$\alpha > RP_{LR} = RP_{LR-GMM} = RP_{LM} > \max\{RP_{Bayes}, RP_{Wald}\}.$$

The Bayesian/frequentist ordering in Theorem 11 is possible due to the non-convexity of the sign equality null hypothesis, according to Theorem 5. Also, the LR/Wald ordering in Theorem 11 reverses the ordering for linear equality (Berndt and Savin, 1977) and linear inequality (Gouriéroux et al., 1982, eqn. (8)) constraint testing.

Changing a test’s critical value to solve the issue highlighted in Theorem 11 would introduce size distortion; the rejection region shape itself must change. We investigate this question of shape in Section 4, wherein we propose a new, nearly unbiased test under general conditions.

## 4 New sign equality test

We propose a new, nonrandomized, frequentist sign equality test that is asymptotically nearly unbiased (up to simulation/discretization error), unlike any existing test. We do not propose a new Bayesian test since it already minimizes posterior expected loss. Code for both deriving the test and running the test is available either online or upon request.

## 4.1 Setup and search methods

We observe a single draw of  $(\hat{t}_1, \hat{t}_2)'$  from

$$\begin{pmatrix} \hat{t}_1 \\ \hat{t}_2 \end{pmatrix} \sim N \left( \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad (10)$$

where  $d_1$  and  $d_2$  are unknown but  $\rho$  is known. We wish to test

$$H_0 : d_1 d_2 \geq 0, \quad H_1 : d_1 d_2 < 0. \quad (11)$$

This is a general limit experiment applicable to a variety of settings. The  $\hat{t}_1$  and  $\hat{t}_2$  could be sample moments, as in Andrews (2012) and others. Alternatively, they could be asymptotically bivariate normal  $t$ -statistics whose correlation has a consistent estimator,  $\hat{\rho}$ , with (10) providing the limit distribution under drifting (to zero) sequences of parameters of interest  $\theta_1$  and  $\theta_2$ , so that  $d_1 d_2 \geq 0 \iff \theta_1 \theta_2 \geq 0$ . Sampling dependence is not restricted, and nonparametric rates of convergence are allowed.

The primary difficulty is structuring a search over candidate tests. We use numerical integration to compute rejection probabilities for different  $(d_1, d_2)$ , given a candidate rejection region approximated by small squares on the  $(\hat{t}_1, \hat{t}_2)$  plane.<sup>4</sup> Without structure, the problem grows exponentially: with only  $10 \times 10 = 100$  squares in the  $t$ -statistic plane, there are  $2^{100} \approx 10^{30}$  candidates. If  $10^9$  candidates can be assessed each second, then an exhaustive search requires over 40 trillion years.

Chiburis (2008) casts this search as a linear program (LP) by allowing randomized tests and using weighted average power (WAP) as the objective function. Squares are

---

<sup>4</sup>The approach of Gouriéroux et al. (1982, p. 69) for linear inequality constraint testing and Wolak (1991, p. 987) for (locally linearized) nonlinear inequalities is to apply an affine transformation  $A$  so that, in this case,  $A(\hat{t}_1, \hat{t}_2)'$  is standard bivariate normal. However, computationally, it is easier to evaluate probabilities of correlated normals over rectangles than independent normals over parallelograms.

labeled  $j = 1, 2, \dots$ , and the test rejects with probability  $0 \leq x_j \leq 1$  if the observed  $(\hat{t}_1, \hat{t}_2)$  is inside the square. Given (10) and  $x_j$ , we can compute the contribution to WAP of square  $j$ ; denote this  $r_j$ , and let  $r = (r_1, \dots, r_J)'$ . Let  $a_{ij}$  be the contribution of  $x_j$  to rejection probability given  $d_1 = 0$  and any  $d_2 = v_i, v_i \leq 0$ . For grid  $\{v_i\}_{i=1}^n$ , size is controlled if  $Ax \leq \alpha$ , where  $a_{ij}$  is the row  $i$ , column  $j$  element of matrix  $A$ . For  $\epsilon > 0$ , a nearly similar-on-the-boundary constraint is  $Ax \geq \alpha - \epsilon$ . Thus, the linear program chooses  $x = (x_1, \dots, x_J)'$  to maximize  $r'x$  subject to  $Ax \leq \alpha$  and  $Ax \geq \alpha - \epsilon$ , i.e., maximize WAP subject to a nearly similar-on-the-boundary constraint.

Allowing randomized tests is critical. Solving the integer LP restricting  $x_j \in \{0, 1\}$  is intractable with the degree of precision we use below. However, converting the randomized LP solution to a nonrandomized test by simple thresholding often makes negligible difference.

Moreira and Moreira (2013, p. 19) propose a method that can solve a discrete approximation of the LP given an iid sample  $\{Y_j\}_{j=1}^J$  from known density  $m(\cdot)$ .<sup>5</sup> Here,  $x_j$  is the probability of randomized rejection of  $Y_j$ , rather than rectangle  $j$ .

For WAP, we use the improper weighting function that equals one for every point in  $H_1$ , i.e.,  $\pi(\theta_1, \theta_2) = 1\{\theta_1\theta_2 < 0\}$  in the notation of Chiburis (2008, eqn. (6)). Using the finite grid of point alternatives

$$\{(\theta_1, \theta_2) : \theta_1 \in \{\pm 0.5, \pm 1.0, \dots, \pm 3.0\}, \theta_2 \in \{\pm 0.5, \dots, \pm 3.0\}, \theta_1\theta_2 < 0\}$$

yields a similar but noticeably different test, especially without Constraint 3 below.

We include additional constraints:

1. By symmetry, the point  $(\hat{t}_1, \hat{t}_2)$ , where  $\hat{t}_1 > 0$  and  $-\hat{t}_1 < \hat{t}_2 < 0$ , should have the same  $x_j$  as the points  $(-\hat{t}_1, -\hat{t}_2)$ ,  $(\hat{t}_2, \hat{t}_1)$ , and  $(-\hat{t}_2, -\hat{t}_1)$ . This ensures invariance

---

<sup>5</sup>Alternatively, a low-discrepancy sequence like the Halton sequence could be used in place of pseudorandom draws of  $Y_j$  to improve the convergence rate, but for the present case, a good approximation of the test is achieved in only a few seconds anyway.

among tests of  $H_0 : d_1 d_2 \geq 0$ ,  $H_0 : d_2 d_1 \geq 0$ , and  $H_0 : (-d_1)(-d_2) \geq 0$ .

2. Away from the origin, we use a one-sided test (i.e., the LR-type rejection region); for example, letting  $z = \Phi^{-1}(1 - \alpha)$ , if  $\hat{t}_1 > 2z$ , then the test rejects if  $\hat{t}_2 < -z$ .
3. Optional: any square in Quadrant IV (II) must have  $x_j$  at least as great as a) the square diagonally up and left (down and right) of it and b) the adjacent square in the direction of the closest part of  $\Theta_0$  (e.g., to the left for a square at  $(2, -4)$ ).

We used additional structure to derive tests in a non-LP framework. As seen in Figure 4 for  $\alpha = 0.05$ , the rejection regions for different  $\rho$  nest within each other: the region for  $\rho$  is a subset of that for  $\rho'$  if  $\rho < \rho'$ . There is also monotonicity in  $\alpha$ , as seen in Figure 5.

Regardless of derivation, the monotonicity in  $\rho$  means that tests can be computed and stored ahead of time (for  $\alpha = 0.05$ , etc.), running in under a second in practice. We use  $\rho = -1, -0.95, \dots, 0.95$ . In practice, using the test for the grid  $\rho$  just below the actual  $\hat{\rho}$  errs slightly on the conservative side.

## 4.2 Properties of new tests

The LP methods of Chiburis (2008) and Moreira and Moreira (2013) and our non-LP method yield comparable tests, shown for  $\rho = 0$  in Figure 2. The runtime for Chiburis (2008) was under two minutes on a personal computer (8GB RAM, 3.2GHz i5 processor), and that for Moreira and Moreira (2013) was a few seconds; the non-LP method took much longer, including multiple starts. The bottom row of the figure shows that the two LP methods differ by a noticeable but small amount near the origin, without Constraint 3. The top row shows that, in turn, our non-LP method

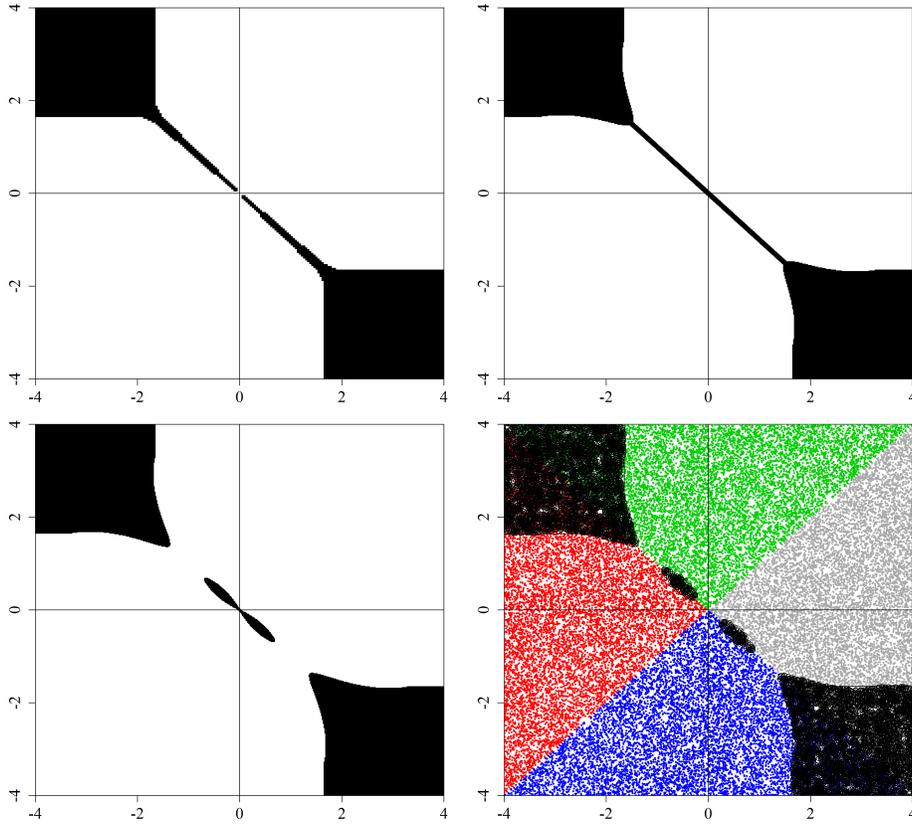


Figure 2: New rejection regions,  $\rho = 0$ ,  $\alpha = 0.05$ , Constraints 1 and 2, equal-weighted WAP, threshold  $x_j = 0.5$  used to convert randomized to nonrandomized tests. Top left: non-LP search, with additional Constraint 3 and constrained to be a proper superset of the LR-type region. Top right: Chiburis (2008) LP method, block width  $z_{1-\alpha}/100$ ,  $v_i = 0, -0.5, \dots, -7$ , with additional Constraint 3. Bottom left: same as top right but without Constraint 3. Bottom right: Moreira and Moreira (2013) discrete approximation LP,  $v_i = 0, -0.25, \dots, -4$ ,  $Y_j \sim \text{Unif}(-6, 6)$ ,  $J = 10^5$ ; the colored/gray dots are the  $Y_j$ , with black circles around those in the rejection region.

and Chiburis (2008) differ by a noticeable but small amount, with Constraint 3. The rejection probabilities of the nonrandomized<sup>6</sup> (thresholded) Chiburis (2008) test along the boundary ranges from 0.0480 to 0.0500, while the boundary rejection probabilities of the non-LP test are within  $[0.0499, 0.0500]$ . Due to the general similarity, non-LP results are the focus hereafter.

<sup>6</sup>It is easy to shrink the range for the *randomized* Chiburis (2008) test, but such improvement can actually make the nonrandomized version worse.

The rejection region shown in Figure 2 (top left) for  $\rho = 0$  and  $\alpha = 0.05$  corresponds to the power function in Figure 3, which also shows the power function of the LR test. As seen, the new test achieves RPs between 0.0499 and 0.0500 at all (grid) points along the boundary of  $H_0$ , whereas the LR-type test with the shape from Figure 1 has RPs ranging from 0.005 (at  $\theta_1 = \theta_2 = 0$ ) to 0.05 (beyond the graph's limits). Similarly, for  $\alpha = 0.1$ , the new test's RPs range between 0.0996 and 0.1000. Figure 3 also shows the nearly unbiasedness and asymptotic power improvement over the LR-type test.

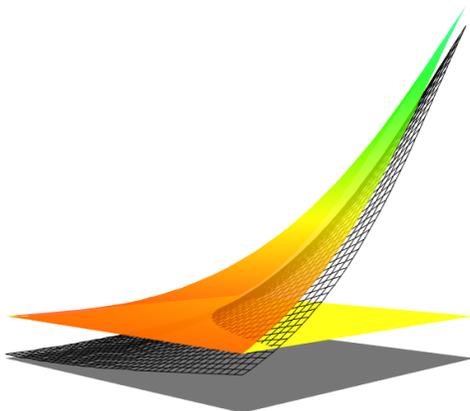


Figure 3: Asymptotic power functions of new (color/solid) and LR-type (black/wire mesh) tests,  $0 \leq \theta_1 \leq 2/\sqrt{n}$ ,  $0 \geq \theta_2 \geq -2/\sqrt{n}$ ,  $\rho = 0$ ,  $\alpha = 0.05$ . The parallel flat surfaces are zero and  $\alpha$  rejection probabilities.

Figure 4 contains additional examples of rejection regions for various  $\rho$ .

A decision-theoretic framework also distinguishes the new and LR-type tests in terms of minimax risk. We use the loss function that takes value  $\alpha$  for type II error,  $1 - \alpha$  for type I error, and zero otherwise. Risk (i.e., expected loss) under  $H_0$  is  $1 - \alpha$  times the type I error rate, and under  $H_1$  is  $\alpha$  times the type II error rate. An unbiased level- $\alpha$  test is a minimax risk decision rule, achieving  $\alpha(1 - \alpha)$  max risk. Max risk of the LR-type level- $\alpha$  test is  $\alpha(1 - 2\alpha^2)$  at  $(\theta_1, \theta_2) = \lim_{\epsilon \rightarrow 0^+} (\epsilon, -\epsilon)$ . For

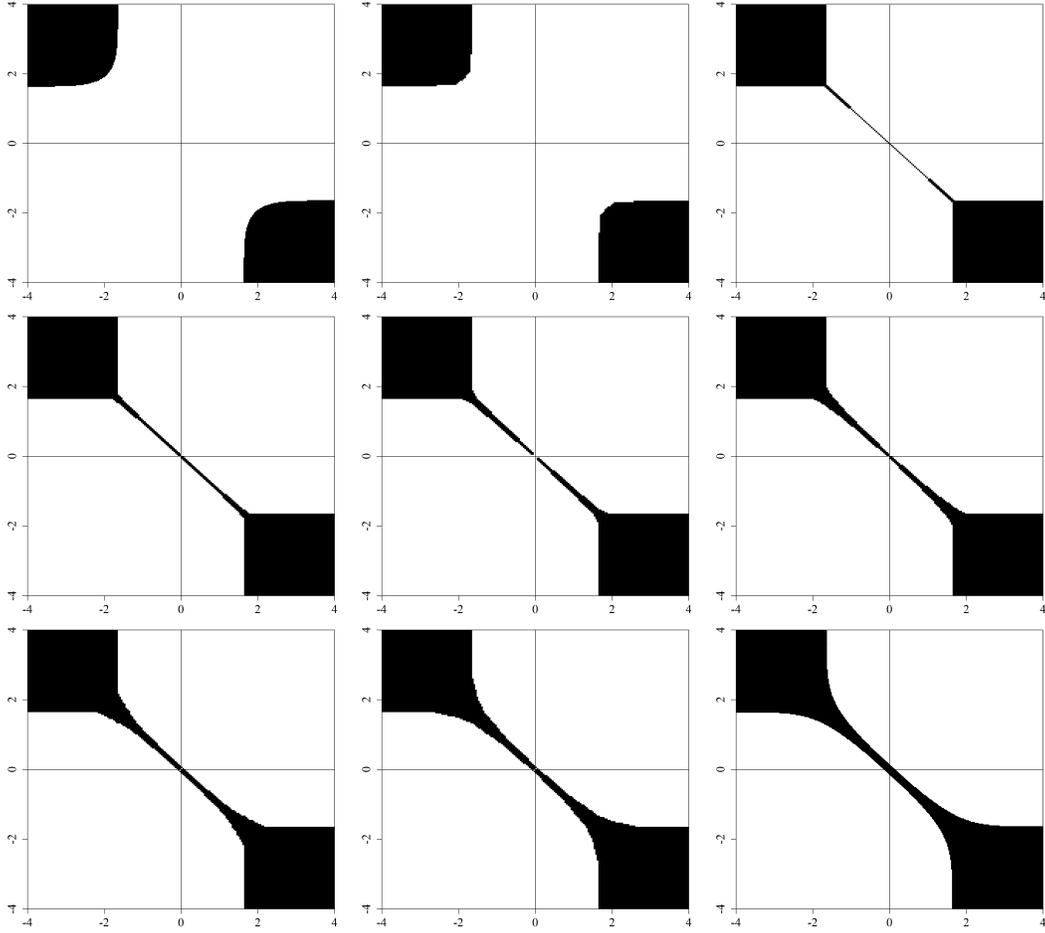


Figure 4: Rejection regions for new test,  $\alpha = 0.05$ ,  $\hat{t}_1$  on horizontal axis,  $\hat{t}_2$  on vertical axis. From left to right within rows, top to bottom:  $\rho = -1, -0.9, -0.5, -0.2, 0, 0.2, 0.5, 0.9, 1$ .

example, for  $\alpha = 0.05$ , the LR-type level- $\alpha$  test has max risk of 0.04975, whereas the new test's max risk is 5% lower, (nearly) 0.0475; if  $\alpha = 0.1$ , the max risk improvement is 8%.

Although the new test improves upon many properties of the LR-type test, it loses a type of rejection monotonicity similar to that in Section 4.2 of Kline (2011), so there is a trade-off. Specifically, the new test may reject  $(\hat{t}_1, \hat{t}_2)$  but not  $(\hat{t}'_1, \hat{t}'_2)$  even if  $|\hat{t}'_1| \geq |\hat{t}_1|$  and  $|\hat{t}'_2| \geq |\hat{t}_2|$ . The rationale for this type of rejection monotonicity is to avoid any samples where  $H_0$  is accepted but would have been rejected if  $(\hat{t}_1, \hat{t}_2)$

were closer to  $H_0$ . A Bayesian perspective finds a violation of this property to be egregious, whereas a strict frequentist perspective does not. Imposing such a rejection monotonicity constraint leads back to the rectangular LR-type region. Any level- $\alpha$  region is pinned down away from the origin at  $(z_{1-\alpha}, -\infty)$ . If it were to contain point  $(t_1, t_2)$  with  $t_1 < z_{1-\alpha}$ , rejection monotonicity would require the inclusion of  $(t_1, c)$  for all  $c < t_2$ , but this would violate the size constraint. Thus, the LR-type rejection region is the largest possible subject to size control and this type of rejection monotonicity.

This type of rejection monotonicity differs from power monotonicity. Comparing points with  $|\theta'_1| \geq |\theta_1|$  and  $|\theta'_2| \geq |\theta_2|$ , both the LR-type and new tests satisfy power monotonicity, as seen in Figure 3.

Rejection monotonicity in  $\alpha$  is yet a different property and is achieved by the new test. A level- $\alpha$  test rejects if (but not only if) a level- $\alpha'$  test rejects, for any  $\alpha > \alpha'$ . Figure 5 shows that the new test's  $\alpha = 0.05$  rejection region is a subset of the  $\alpha = 0.1$  rejection region when  $\rho = 0$ .

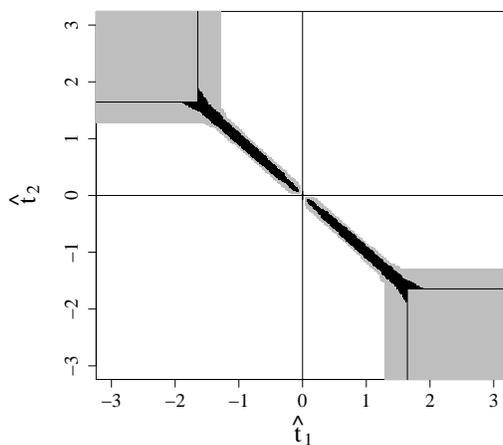


Figure 5: Overlay of  $\alpha = 0.05$  rejection region on  $\alpha = 0.10$  region, showing monotonicity in  $\alpha$ ;  $\rho = 0$ .

One open question is whether the new sign equality test may be represented (or

approximated) more elegantly as an LR-type test with data-dependent critical values.

## 5 Conclusion

We have explored the difference between Bayesian and frequentist inequality constraint tests, providing a formal result on the role of convexity of the null hypothesis parameter subspace. The Bayesian/frequentist test ordering reversal that can occur with non-convexity is shown to arise when testing sign equality of two parameters, unless the estimators are strongly negatively correlated. In either case, no test can achieve both Bayesian and frequentist optimality. We also contribute a new sign equality test that is nearly unbiased (unlike existing tests) and has strictly better power than the LR-type test.

Investigation of approaches like Müller and Norets (2014) applied to nonlinear inequality testing and extension of our results to cases without asymptotic (sampling and/or posterior) normality remain for future work.

## References

- Andrews, D. W. K. (2012). Similar-on-the-boundary tests for moment inequalities exist, but have poor power. Discussion paper 1815R, Cowles Foundation for Research in Economics at Yale University, New Haven, CT.
- Berndt, E. R. and Savin, N. E. (1977). Conflict among criteria for testing hypotheses in the multivariate linear regression model. *Econometrica*, 45(5):1263–1277.
- Chiburis, R. C. (2008). Approximately most powerful tests for moment inequalities. Job market paper, Princeton University, Princeton, NJ.
- Donald, S. G. and Hsu, Y.-C. (2011). A new test for linear inequality constraints when the variance–covariance matrix depends on the unknown parameters. *Economics Letters*, 113(3):241–243.
- Gouriéroux, C., Holly, A., and Monfort, A. (1982). Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica*, 50(1):63–80.
- Guggenberger, P., Hahn, J., and Kim, K. (2008). Specification testing under moment inequalities. *Economics Letters*, 99(2):375–378.

- Hahn, J. (1997). Bayesian bootstrap of the quantile regression estimator: a large sample study. *International Economic Review*, 38(4):795–808.
- Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857.
- Johnson, R. A. (1970). Asymptotic expansions associated with posterior distributions. *The Annals of Mathematical Statistics*, 41(3):851–864.
- Kim, J.-Y. (2002). Limited information likelihood and Bayesian analysis. *Journal of Econometrics*, 107(1):175–193.
- Kline, B. (2011). The Bayesian and frequentist approaches to testing a one-sided hypothesis about a multivariate mean. *Journal of Statistical Planning and Inference*, 141(9):3131–3141.
- Kodde, D. A. and Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*, 54(5):1243–1248.
- Kwan, Y. K. (1999). Asymptotic Bayesian analysis based on a limited information estimator. *Journal of Econometrics*, 88(1):99–121.
- Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer.
- Lindley, D. V. (1957). A statistical paradox. *Biometrika*, pages 187–192.
- Lo, A. Y. (1987). A large sample study of the Bayesian bootstrap. *The Annals of Statistics*, pages 360–375.
- Moon, H. R. and Schorfheide, F. (2012). Bayesian and frequentist inference in partially identified models. *Econometrica*, 80(2):755–782.
- Moreira, H. and Moreira, M. J. (2013). Contributions to the theory of optimal tests. *Ensaaios Econômicos 747*, FGV/EPGE, Rio de Janeiro.
- Müller, U. K. and Norets, A. (2014). Credibility of confidence sets in nonstandard econometric problems. Working paper, available at <http://www.princeton.edu/~umueller>.
- Rubin, D. B. (1981). The Bayesian bootstrap. *The Annals of Statistics*, 9(1):130–134.
- Sims, C. A. (2010). Understanding non-Bayesians. Unpublished book chapter, available at <http://sims.princeton.edu/yftp/UndrstndgNnBsns/GewekeBookChppter.pdf>.
- Stoye, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica*, 77(4):1299–1315.
- Weng, C.-S. (1989). On a second-order asymptotic property of the Bayesian bootstrap mean. *The Annals of Statistics*, pages 705–710.
- Wolak, F. A. (1991). The local nature of hypothesis tests involving inequality constraints in nonlinear models. *Econometrica*, 59(4):981–995.

## A Mathematical proofs

### A.1 Proof of Proposition 1

The least favorable null is  $\theta = h$  (or  $\theta = -h$ ), so

$$\alpha = P(\hat{\theta} < -c_f \mid \theta = h) + P(\hat{\theta} > c_f \mid \theta = h) > P(\hat{\theta} > c_f \mid \theta = h) = 1 - \Phi(c_f - h),$$

and  $c_f - h > z_{1-\alpha}$ . For the Bayesian test,

$$\alpha = P(-h < \theta < h \mid \hat{\theta} = c_B) = \Phi(h - c_B) - \Phi(-h - c_B) < \Phi(h - c_B),$$

so  $h - c_B > z_\alpha$  and  $c_B < h + z_{1-\alpha} < c_f$ . If  $P(H_0 \mid \hat{\theta} = 0) = \Phi(h) - \Phi(-h) \leq \alpha$ , then  $c_B = 0$  and the Bayesian test always rejects.  $\square$

### A.2 Proof of Proposition 2

Denote the frequentist and Bayesian rejection regions  $[-c_f, c_f]$  and  $[-c_B, c_B]$ . The value  $c_f$  is determined by<sup>7</sup>

$$\alpha = P(\hat{\theta} \in [-c_f, c_f] \mid \theta = h) = \Phi(c_f - h) - \Phi(-c_f - h) < \Phi(c_f - h),$$

so  $c_f - h > z_\alpha$ . In contrast,  $c_B$  is determined by

$$\begin{aligned} \alpha &= P(H_0 \mid \hat{\theta} = c_B) = P(\theta > h \mid \hat{\theta} = c_B) + P(\theta < -h \mid \hat{\theta} = c_B) \\ &> P(\theta > h \mid \hat{\theta} = c_B) = 1 - \Phi(h - c_B) = \Phi(c_B - h), \end{aligned}$$

so  $c_B - h < z_\alpha$ . Altogether,  $c_B < h + z_\alpha < c_f$ , so the Bayesian rejection region is a subset of the frequentist region.  $\square$

### A.3 Proof of Lemma 4

Even though  $\theta$  is  $k$ -dimensional, the test is essentially one-dimensional, where  $c'\hat{\theta}$  has a normal sampling distribution and  $c'\theta$  has an equivalent normal posterior. The  $\sqrt{n}$  below may be replaced by any  $a_n \rightarrow \infty$  rate.

For the frequentist test, from the sampling distribution in Assumption 1 and the continuous mapping theorem,

$$\sqrt{n}(c'\hat{\theta} - c'\theta) / \sqrt{c'\hat{V}c} \xrightarrow{d} N(0, 1).$$

---

<sup>7</sup>Technically,  $\theta = h$  is not in  $H_0$ , but this yields the supremum RP over  $H_0$  due to continuity of the RP in  $\theta$ .

The least favorable null is any  $\theta$  satisfying  $c'\theta = c_0$ , on the boundary of  $\Theta_0$ . If  $c'\theta = c_0$ , then

$$\sqrt{n}(c'\hat{\theta} - c_0)/\sqrt{c'\hat{V}c} \xrightarrow{d} N(0, 1),$$

so the test in Lemma 4 has exactly  $\alpha$  rejection probability. For any other  $\theta \in \Theta_0$ ,  $c'\theta < c_0$ , so  $c'\theta - c_0 < 0$ , and the test statistic diverges to  $-\infty$ , leading to zero rejection probability.

For the Bayesian test, we can check  $P(H_0) = \alpha$  everywhere on the boundary of the rejection region from Lemma 4. The boundary is equivalent to  $c'\hat{\theta} = c_0 + n^{-1/2}z_{1-\alpha}\sqrt{c'\hat{V}c}$ , so

$$\begin{aligned} P\left(c'\theta \leq c_0 \mid c'\hat{\theta} = c_0 + n^{-1/2}z_{1-\alpha}\sqrt{c'\hat{V}c}\right) &= P\left(\sqrt{n}(c'\theta - c'\hat{\theta} + n^{-1/2}z_{1-\alpha}\sqrt{c'\hat{V}c}) \leq 0\right) \\ &= P\left(\sqrt{n}(c'\theta - c'\hat{\theta})/\sqrt{c'\hat{V}c} \leq -z_{1-\alpha}\right) \\ &\rightarrow \Phi(-z_{1-\alpha}) = \alpha, \end{aligned}$$

using the asymptotic posterior from Assumption 1. Similar calculations show the probability to be below  $\alpha$  for larger  $c'\hat{\theta} - c_0$  and above  $\alpha$  for smaller  $c'\hat{\theta} - c_0$ .  $\square$

## A.4 Proof of Theorem 5

Let  $\theta_b$  be any point on the boundary of  $\Theta_0$ . Since  $\Theta_0$  is convex, there exists at least one hyperplane containing  $\theta_b$  such that  $\Theta_0$  is entirely on one side of the hyperplane. Let  $\Theta_{0,b}$  be the subspace containing all points on the  $\Theta_0$  side of this hyperplane, so  $\Theta_0 \subseteq \Theta_{0,b}$ . Define  $c \in \mathbb{R}^k$  and  $c_0 \in \mathbb{R}$  such that  $\Theta_{0,b} = \{\theta : c'\theta \leq c_0\}$ .

If  $\Theta_0 = \Theta_{0,b}$ , then Lemma 4 applies, which gives the desired conclusion.

Even if  $\Theta_0 \subset \Theta_{0,b}$  strictly, Lemma 4 says that the test defined therein has exact  $\alpha$  rejection probability if  $\theta = \theta_b$  since  $c'\theta_b = c_0$  by construction. We now show that each point on the boundary of that test's rejection region is in the interior of the Bayesian rejection region for  $H_0 : \theta \in \Theta_0$ . Consequently, the Bayesian rejection region for  $H_0 : \theta \in \Theta_0$  is a superset of the rejection region for  $H_0 : \theta \in \Theta_{0,b}$ , so the (frequentist) rejection probability (given  $\theta_b$ ) of the former Bayesian test is larger than  $\alpha$  (the rejection probability of the latter test).

Consider any point  $\hat{\theta}_b$  on the boundary of the rejection region for  $H_0 : \theta \in \Theta_{0,b}$ . From Lemma 4, the asymptotic posterior probability  $P(\theta \in \Theta_{0,b} \mid \hat{\theta}_b) = \alpha$  exactly. Since  $\Theta_0 \subset \Theta_{0,b}$ ,  $P(\theta \in \Theta_0 \mid \hat{\theta}_b) < \alpha$ , so  $\hat{\theta}_b$  is an interior point of the Bayesian rejection region.  $\square$

## A.5 Proof of Corollary 6

Any individual linear inequality constraint generates a convex subset of the parameter space (specifically, a half-space). The intersection of any number of convex sets is convex. Thus, any (joint) linear inequality constraint null hypothesis is equivalent to  $H_0 : \theta \in \Theta_0$  for some convex  $\Theta_0 \subset \mathbb{R}^k$ , so the assumptions of Theorem 5 are satisfied.  $\square$

## A.6 Proof of Lemma 7

Using the assumed independence and the asymptotic posteriors in (8),

$$P_a(H_0 \mid \bar{y}_1, \bar{y}_2) = [1 - \Phi(-\sqrt{n}\bar{y}_2)] [1 - \Phi(-\sqrt{n}\bar{y}_1)] + \Phi(-\sqrt{n}\bar{y}_1)\Phi(-\sqrt{n}\bar{y}_2). \quad \square \quad (12)$$

## A.7 Proof of Lemma 8

Given (5), the unrestricted MLE is  $(\hat{\theta}_1, \hat{\theta}_2) = (\bar{y}_1, \bar{y}_2)$ , the sample averages. If this satisfies  $H_0$ , then the likelihood ratio (LR) is one. If not, the restricted (to  $H_0$ ) MLE sets either  $\hat{\theta}_1$  or  $\hat{\theta}_2$  to zero since  $\bar{y}_1 \perp \bar{y}_2$ . Since  $\text{Var}(\bar{y}_1) = \text{Var}(\bar{y}_2)$ , the LR is a function of the Euclidean distance from  $(\bar{y}_1, \bar{y}_2)$  to  $H_0$ , which forms L-shaped iso-LR curves in quadrants II and IV of the  $(\bar{y}_1, \bar{y}_2)$  plane as in Figure 1.

By the sufficiency of the sample average  $\bar{y}_1$  for  $\theta_1$  and  $\bar{y}_2$  for  $\theta_2$ , we can use the likelihoods for  $\bar{y}_1 \sim N(\theta_1, 1/n)$  and  $\bar{y}_2 \sim N(\theta_2, 1/n)$ . Using  $\bar{y}_1 \perp \bar{y}_2$ , the joint likelihood factors into

$$\mathcal{L}(\theta_1, \theta_2 \mid \bar{y}_1, \bar{y}_2) = \mathcal{L}(\theta_1, \theta_2 \mid \bar{y}_1)\mathcal{L}(\theta_1, \theta_2 \mid \bar{y}_2) = \mathcal{L}(\theta_1 \mid \bar{y}_1)\mathcal{L}(\theta_2 \mid \bar{y}_2).$$

For  $j = 1, 2$ ,

$$\begin{aligned} \mathcal{L}(\theta_j \mid \bar{y}_j) &= (2\pi/n)^{-1/2} \exp\{-n(\bar{y}_j - \theta_j)^2/2\}, \\ \frac{\partial}{\partial \theta_j} \mathcal{L}(\theta_j \mid \bar{y}_j) &= (2\pi/n)^{-1/2} \exp\{-n(\bar{y}_j - \theta_j)^2/2\} n(\bar{y}_j - \theta_j), \end{aligned} \quad (13)$$

and the unrestricted maximum likelihood estimator (MLE) is  $(\hat{\theta}_1, \hat{\theta}_2) = (\bar{y}_1, \bar{y}_2)$ .

If the unrestricted MLE does not satisfy  $H_0$ , then the restricted (to  $H_0$ ) MLE is either  $(0, \bar{y}_2)$  or  $(\bar{y}_1, 0)$ . For example, if  $\bar{y}_1 > 0$  and  $\bar{y}_2 < 0$ , then by (13),  $\mathcal{L}(0 \mid \bar{y}_1) > \mathcal{L}(\theta_1 \mid \bar{y}_1)$  for any  $\theta_1 < 0$ , and  $\mathcal{L}(0 \mid \bar{y}_2) > \mathcal{L}(\theta_2 \mid \bar{y}_2)$  for any  $\theta_2 > 0$ . Since (13) is the same for  $j = 1$  and  $j = 2$ , the restricted MLE is  $(0, \bar{y}_2)$  if  $|\bar{y}_1| < |\bar{y}_2|$ , otherwise  $(\bar{y}_1, 0)$ .

The likelihood ratio if the MLE is in quadrant II or IV and  $|\bar{y}_1| < |\bar{y}_2|$  is

$$\frac{\mathcal{L}(0 \mid \bar{y}_1)\mathcal{L}(\hat{\theta}_2 \mid \bar{y}_2)}{\mathcal{L}(\hat{\theta}_1 \mid \bar{y}_1)\mathcal{L}(\hat{\theta}_2 \mid \bar{y}_2)} = \frac{\mathcal{L}(0 \mid \bar{y}_1)}{\mathcal{L}(\hat{\theta}_1 \mid \bar{y}_1)} = \exp\{-n\bar{y}_1^2/2\},$$

and similarly  $\exp\{-n\bar{y}_2^2/2\}$  if  $|\bar{y}_2| < |\bar{y}_1|$ . For critical value  $c$ , the LR rejection region is

$$\{(\bar{y}_1, \bar{y}_2) : \max\{\min\{\bar{y}_1, -\bar{y}_2\}, \min\{-\bar{y}_1, \bar{y}_2\}\} \geq c\}.$$

It remains to derive the critical value,  $c$ . The rejection probability (RP) is

$$\begin{aligned} & P(\max\{\min\{\bar{y}_1, -\bar{y}_2\}, \min\{-\bar{y}_1, \bar{y}_2\}\} > c) \\ &= P(\bar{y}_1 > c, \bar{y}_2 < -c) + P(\bar{y}_1 < -c, \bar{y}_2 > c) \\ &= P(\bar{y}_1 > c)P(\bar{y}_2 < -c) + P(\bar{y}_1 < -c)P(\bar{y}_2 > c). \end{aligned}$$

Under  $H_0$ , this is maximized at  $(\theta_1, \theta_2) = (0, m)$  as  $m \rightarrow -\infty$ , or symmetrically at  $(m, 0)$  as  $m \rightarrow \infty$ . (The case  $\theta_1 = 0$  and  $\theta_2 \rightarrow -\infty$  for fixed  $n$  produces the same results as  $\theta_2 < 0$  fixed as  $n \rightarrow \infty$ .) This is the least favorable null. In the case of  $(0, m)$ ,

$$P(\bar{y}_2 > c) = P(\sqrt{n}(\bar{y}_2 - m) > \sqrt{n}(c - m)) = 1 - \Phi(\sqrt{n}(c - m)),$$

where  $\Phi(\cdot)$  is the standard normal CDF. As  $m \rightarrow -\infty$ ,  $c - m \rightarrow \infty$ , so  $\Phi(\sqrt{n}(c - m)) \rightarrow 1$  and  $P(\bar{y}_2 > c) \rightarrow 0$ . Similarly,  $P(\bar{y}_2 < -c) \rightarrow 1$ . The RP thus reduces to  $P(\bar{y}_1 > c)$ , where  $\bar{y}_1 \sim N(0, 1/n)$ . Setting the RP equal to  $\alpha$  yields  $c = z_{1-\alpha}/\sqrt{n}$ , where  $z_{1-\alpha} \equiv \Phi^{-1}(1 - \alpha)$ . For  $\theta_1 = 0$  and general  $\theta_2 = m \leq 0$ , the RP with this  $c$  is

$$\begin{aligned} & P(\bar{y}_1 > c)P(\bar{y}_2 < -c) + P(\bar{y}_1 < -c)P(\bar{y}_2 > c) \\ &= \alpha\Phi(\sqrt{n}(-c - m)) + \alpha[1 - \Phi(\sqrt{n}(c - m))] \\ &= \alpha\Phi(-z_{1-\alpha} - m\sqrt{n}) + \alpha\Phi(-z_{1-\alpha} + m\sqrt{n}) \\ &= \alpha[\Phi(z_\alpha - m\sqrt{n}) + \Phi(z_\alpha + m\sqrt{n})]. \end{aligned}$$

The derivative of the term in square brackets with respect to  $m$  is  $\sqrt{n}[\phi(z_\alpha + m\sqrt{n}) - \phi(z_\alpha - m\sqrt{n})] \leq 0$  for  $m \leq 0$ , so the global maximum is at  $m \rightarrow -\infty$ . That is, the rejection probability will be strictly smaller than  $\alpha$  with  $c = z_{1-\alpha}/\sqrt{n}$  when  $\theta_1 = 0$  and  $\theta_2 \in (-\infty, 0]$ . By symmetry, a similar conclusion is reached for  $\theta_2 = 0$  and  $\theta_1 \in [0, \infty)$ .  $\square$

## A.8 Proof of Lemma 9

When  $\theta_1 = 0$  and  $\theta_2 < 0$  is fixed as  $n \rightarrow \infty$  (the least favorable null), the probability limit of the denominator of  $\widehat{W}_n$  is  $|\theta_2|$ , so  $\widehat{W}_n \xrightarrow{d} N(0, 1)$ . (The same distribution arises if  $\theta_2 \rightarrow -\infty$  with fixed  $n$ .) Rejecting when  $\widehat{W}_n < z_\alpha$  thus has  $\alpha$  probability. A similar argument when  $\theta_1 > 0$  and  $\theta_2 = 0$  yields the same  $z_\alpha$  critical value, and the rejection probability with this critical value is below  $\alpha$  when  $\theta_1 = \theta_2 = 0$ .  $\square$

## A.9 Proof of Lemma 10

For the LM test, the log likelihood and score are, using (13),

$$\begin{aligned}\ln \mathcal{L}(\theta_1, \theta_2 \mid \bar{y}_1, \bar{y}_2) &= \ln \mathcal{L}(\theta_1 \mid \bar{y}_1) + \ln \mathcal{L}(\theta_2 \mid \bar{y}_2) \\ &= -\ln(2\pi/n) - n(\bar{y}_1 - \theta_1)^2/2 - n(\bar{y}_2 - \theta_2)^2/2,\end{aligned}\quad (14)$$

$$\frac{\partial}{\partial \theta_1} \ln \mathcal{L}(\theta_1, \theta_2 \mid \bar{y}_1, \bar{y}_2) = n(\bar{y}_1 - \theta_1), \quad \frac{\partial}{\partial \theta_2} \ln \mathcal{L}(\theta_1, \theta_2 \mid \bar{y}_1, \bar{y}_2) = n(\bar{y}_2 - \theta_2). \quad (15)$$

The Fisher information and score

$$\mathcal{I}(\theta_1, \theta_2) = \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix}, \quad (16)$$

The LM test statistic is thus  $S'\mathcal{I}^{-1}S = n(\bar{y}_1 - \theta_1)^2 + n(\bar{y}_2 - \theta_2)^2 = n\|\bar{y} - \theta\|_2^2$ , where  $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \dots}$  is the Euclidean ( $L_2$ ) vector norm. Since  $H_0$  is not a point, the smallest value  $\inf_{\theta \in H_0} n\|\bar{y} - \theta\|_2^2$  is taken, as usual; i.e., the Euclidean distance from  $(\bar{y}_1, \bar{y}_2)$  to  $H_0$ . Consequently, the shape of the LM rejection region is identical to that of LR, so they must be completely identical for LM to control size.

For the LR-type GMM test, the moment conditions are  $\mathbb{E}(y_1 - \theta_1) = 0$  and  $\mathbb{E}(y_2 - \theta_2) = 0$ . Since the optimal weighting matrix is the identity matrix  $I_2$ , and since the criterion function at  $(\hat{\theta}_1, \hat{\theta}_2)$  is zero since the model is exactly identified, the test statistic is  $n$  times the quadratic form of the sample moments  $n^{-1} \sum_{i=1}^n (y_{ik} - \theta_k)$  (again, then minimized over  $H_0$ ):

$$J(\theta) = n(\bar{y}_1 - \theta_1, \bar{y}_2 - \theta_2)I_2(\bar{y}_1 - \theta_1, \bar{y}_2 - \theta_2)' = n\|\bar{y} - \theta\|_2^2, \quad (17)$$

matching LM exactly. □

## A.10 Proof of Theorem 11

We first show that the (asymptotic) Bayesian and Wald tests exclude the boundary of the LR rejection region. Without loss of generality (by symmetry), consider the LR boundary point  $(z_{1-\alpha}/\sqrt{n}, z_\gamma/\sqrt{n})$  for some  $z_\alpha > z_\gamma > -\infty$ . Using (12),

$$\begin{aligned}P_a(H_0 \mid z_\alpha/\sqrt{n}, z_\gamma/\sqrt{n}) &= [1 - \Phi(-\sqrt{n}z_\gamma/\sqrt{n})][1 - \Phi(-\sqrt{n}z_{1-\alpha}/\sqrt{n})] \\ &\quad + \Phi(-\sqrt{n}z_{1-\alpha}/\sqrt{n})\Phi(-\sqrt{n}z_\gamma/\sqrt{n}) \\ &= \gamma(1 - \alpha) + \alpha(1 - \gamma) = \alpha + \gamma(1 - 2\alpha) > \alpha\end{aligned}$$

if  $\alpha < 1/2$ . At the same LR boundary point, the Wald statistic value is

$$\widehat{W}_n = \frac{\sqrt{n}(z_{1-\alpha}/\sqrt{n})(z_\gamma/\sqrt{n})}{\sqrt{(z_{1-\alpha}/\sqrt{n})^2 + (z_\gamma/\sqrt{n})^2}} = z_\alpha \frac{z_{1-\gamma}}{\sqrt{z_{1-\alpha}^2 + z_{1-\gamma}^2}} > z_\alpha$$

if  $\alpha < 1/2$ . Thus, the Bayesian and Wald rejection regions are strict subsets of the LR region.

It remains to calculate the RP of the LR test when  $\theta_1 = \theta_2 = 0$ . Let  $Z_1 \sim N(0, 1)$ ,  $Z_2 \sim N(0, 1)$ ,  $Z_1 \perp Z_2$ . The RP is

$$\begin{aligned} &P(\mathcal{Z}_1 < z_\alpha, \mathcal{Z}_2 > z_{1-\alpha}) + P(\mathcal{Z}_1 > z_{1-\alpha}, \mathcal{Z}_2 < z_\alpha) \\ &= \Phi(z_\alpha)[1 - \Phi(z_{1-\alpha})] + [1 - \Phi(z_{1-\alpha})]\Phi(z_\alpha) = 2\alpha^2. \end{aligned}$$

For any  $\alpha \in (0, 0.5)$ ,  $2\alpha^2 < \alpha$ , so the tests are all conservative. If  $\alpha = 0.5$ , then the LR rejection region includes all of quadrants II and IV, and the Bayesian and Wald rejection regions are identical.  $\square$