Parametric Optimal Testing: Theory and Applications

Ping Yu*
University of Auckland

First Version: November 2008
This Version: August 2009
(Comments Welcome)

Abstract

This paper puts forward a unified framework for asymptotically optimal tests in parametric models and applies the new theory to two tests. The first test is a classical test in locally asymptotically normal (LAN) models, but the assumptions are weakened and a lacuna in the literature is filled. The second test concerns the location of the threshold point. The main result here is that the optimal test in the weighted average power sense is based on the posterior odds which depends on the prior on the local parameter space and is not unique. Furthermore, the likelihood ratio test is not asymptotically equivalent to the posterior odds and there is a discrete component in its asymptotic distribution. Since the asymptotic distribution of the posterior odds is not pivotal to the true value under the null hypothesis, a parametric bootstrap is used to find asymptotic critical values. The results in the second test are very different from those in classical LAN models.

KEYWORDS: optimal test, threshold regression, posterior odds, Bayes factor, nonstandard testing problem, integrated likelihood ratio, limits of experiments
JEL-CLASSIFICATION: C11, C12, C21.

*Email: p.yu@auckland.ac.nz. I would like to thank professor Adrian Raftery for introducing me the optimality property of Bayes factors. I also want to thank Professor Van der Vaart for helping me solving a technical difficulty. This paper benefits heavily from helpful comments of Bruce Hansen and careful modifications of Peter Phillips for which I want to express my gratitude. All remaining errors are mine.

1 Introduction

Optimal testing is part of classical statistical theory and has important applications throughout econometrics. In the frequentist literature, the Neyman-Pearson lemma is the cornerstone for considering optimality, but only applies for simple hypotheses. When the hypotheses are composite, the weighted average power (WAP) criterion is often used. For example, Andrews and Ploberger (1994) consider parametric optimal tests using WAP when a nuisance parameter is present only under the alternative. In the Bayes literature, Bayes factors are the usual choice for testing. It is well known that Bayes factors have some optimality properties. For example, Jeffreys (1961) shows the Bayes factor minimizes the sum of the Type I and Type II error probabilities averaged over the prior.

It is standard to apply the decision theory to the testing problem in finite samples; see, e.g., Section 8.1 of Chamberlain (2007). But the decision theory is not applied to optimal testing in large samples yet. This paper tries to find the optimal tests in the asymptotic WAP sense by using the decision theory framework. Such a unified framework is given in Section 2. This framework is very powerful. It helps us to sort out the key problems we need to answer when we try to find a test with the maximum WAP. The basic result is that the posterior odds procedure is optimal in the WAP sense as long as it is asymptotically pivotal to the local parameters under the null hypothesis.

There is considerable literature related to the topic discussed in this paper. Andrews (1994) builds the large sample correspondence between classical hypothesis tests and Bayesian posterior odds tests in locally asymptotically normal (LAN) models, but no optimality results are covered. See also his Section 7 for the interaction of frequentist and Bayesian literature in hypothesis testing. Andrews and Ploberger (1994) discuss optimal tests when a nuisance parameter can not be identified under the null, but their Theorem 2 actually claims that the optimal tests there have the smallest weighted average type I and type II error probabilities, not the maximum WAP. A classical test in LAN models is reconsidered in Section 3, but the assumptions are weakened and the lacuna in Andrews and Ploberger (1994) is filled.

A more interesting test about the location of the threshold point in threshold regression is discussed in Section 4. The threshold regression model is specified as

$$y = \begin{cases} x'\beta_1 + \sigma_1 e, & q \le \gamma; \\ x'\beta_2 + \sigma_2 e, & q > \gamma. \end{cases} \tag{1}$$

where q is the threshold variable used to split the sample, $x \in \mathbb{R}^{\ell}$, $\beta \equiv (\beta_1', \beta_2')' \in \mathbb{R}^{2\ell}$ and $\sigma \equiv (\sigma_1, \sigma_2)'$ are threshold parameters on mean and variance in the two regimes, $E[e|x,q] = 0$, $E[e^2] = 1$ is a normalization of the error variance and adopts conditional heteroskedasticity, and all the other variables are understood as in the linear regression framework. In the parametric model, the density of e conditional on (x,q) is $f_{e|x,q}(e|x,q;\eta)$, $\eta \in \mathbb{R}^{d_\eta}$ is some nuisance parameter affecting the shape of this error distribution only, the joint distribution of (x,q) is $f_{x,q}(x,q)$, the marginal density of q is $f_q(q)$, and the unknown parameter is $\theta = (\gamma, \beta_1', \beta_2', \sigma_1, \sigma_2, \eta')' \equiv (\gamma, \underline{\theta})$. Threshold regression models of the type (1) have many applications in economics; see the introduction of Yu (2007) or Lee and Seo (2008) for a summary.

The hypotheses of interest are

$$\begin{aligned} H_0 &: \gamma = \gamma_0, \\ H_1 &: \gamma \neq \gamma_0. \end{aligned} \tag{2}$$

This test is never considered independently although it appears as an intermediate step in the confidence interval construction literature such as Hansen (2000). The optimal tests depend on the weighting scheme used, so are not unique and depend on user preferences. Furthermore, the likelihood ratio test is not asymptotically equivalent to the posterior odds as in the usual LAN model discussed in Andrews (1994), and there is a discrete component in its asymptotic distribution. So the results in this test are very different from those in classical LAN models. Since the asymptotic distribution of the posterior odds test is not pivotal to the true value under H_0, a parametric bootstrap procedure is used to find critical values. Section 5 concludes, and all assumptions, proofs, lemmas and algorithms are given in Appendices A, B, C and D, respectively.

Before closing this introduction, it should be pointed out that the framework is essentially frequentist. Decision theory is used to attack the optimal testing problem and some Bayes procedures are used in deriving the test. But the randomness is confined to the data and does not include parameters. Throughout the paper, the data are assumed to be randomly sampled to simplify the theory, but the central ideas and methods of the paper may readily be applied to more general data generating processes. A word on notation: the letter c is used as a generic positive constant, which need not be the same in each occurrence, $\widetilde{\pi}$ is pi (more usually π), and $\overset{\theta_0}{\rightsquigarrow}$, $\overset{h_0}{\rightsquigarrow}$ signify weak convergence under θ_0 and h_0 respectively. The code for figures and tables is available at

http://homes.eco.auckland.ac.nz/pyu013/research.html

2 A Unified Framework for Asymptotically Optimal Tests

Suppose the model under consideration is P_θ with density $f(\cdot|\theta)$ for some $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^k$ is the parameter space, and θ_0 is uniquely identified in the interior of Θ. The data $W_n = (w_1, \cdots, w_n)$ are randomly sampled with support $\mathbb{W} \subset \mathbb{R}^K$. This section seeks to provide a unified framework for the following testing problem:

$$
\begin{aligned}
H_0 &: \theta \in \Theta_0, \\
H_1 &: \theta \in \Theta_1,
\end{aligned}
$$

where $\Theta_1 = \Theta_0^c$. This problem is formulated in a decision framework. The actions are to accept H_0 $(d = 0)$ or to reject H_0 $(d = 1)$. A conventional loss function is

$$
L(\theta, d) = \begin{cases} 0, & \text{if } \theta \in \Theta_0, d = 0, \\ b, & \text{if } \theta \in \Theta_0, d = 1, \\ 1, & \text{if } \theta \in \Theta_1, d = 0, \\ 0, & \text{if } \theta \in \Theta_1, d = 1, \end{cases}
$$

where $b \geq 0$ reflects the importance of the type I error relative to the type II error and will be specified below.

2.1 Weighted Average Power Envelope

In the frequentist literature, much of the focus centers on the power function on the local parameter space instead of the original parameter space. This is because the power generally converges to 1 for any fixed $\theta \in \Theta_1$, which makes the optimal testing problem trivial.

Suppose the appropriate normalization rate for θ is φ_n which converges to zero as n goes to infinity, then the local parameter space is $H_n = \varphi_n^{-1}(\Theta - \theta_0)$, which converges to $\mathbb{R}^k$ as n gets large since θ_0 is in the interior of Θ.[1] Instead of imposing a prior on Θ, we put a prior distribution $\Pi(h)$ on the local parameter space H_n. This specification is not new, e.g., Andrews (1994) and Andrews and Ploberger (1994) take this perspective. For a fixed prior $\Pi(h)$, the corresponding prior $\Pi(\theta)$ on the original parameter space Θ changes. If $\pi(h)$ is the prior density on H_n, then the prior density on Θ is $\pi(\theta) = \varphi_n^{-1}\pi\left(\varphi_n^{-1}(\theta-\theta_0)\right)$ which concentrates around θ_0 when n gets large. Here, there is some abuse of notation from using $\pi(\cdot)$ and $\Pi(\cdot)$ for the priors on Θ and H_n. Specificaton of $\pi(h)$ depends on user objectives. From the discussion below, $\pi(h)$ plays a similar role to the loss function in estimation problems.

When the null hypothesis is simple, θ_0 is known, and H_n is known, so a prior can be placed on H_n. But when the null hypothesis is composite, part of θ_0 is unknown and H_n is partly unknown, so the prior can not be placed on H_n.[2] In this case, the optimal test can not be found for an arbitrary prior. For many problems, however, the optimal test is invariant to the nuisance unknown part of θ_0 for some special class of reasonable priors, as discussed in the examples below.

For the given prior $\pi(h)$, the average (Bayes) risk for the decision $d(W_n,\pi)$, abbreviated as d_n, is

$$\begin{aligned} B_n(d_n,\pi) &= \int_\Theta\int_{\mathbb{W}^n} L(\theta,d_n)\, dP_\theta^n(W_n)\, d\Pi(\theta) \\ &= \int_{H_n}\int_{\mathbb{W}^n} L(\theta_0+\varphi_n h,d_n)\, dP_{\theta_0+\varphi_n h}^n(W_n)\, d\Pi(\theta_0+\varphi_n h) \\ &= |\varphi_n|\int_{H_n}\left[\int_{\mathbb{W}^n} L(\theta_0+\varphi_n h,d_n)\, dP_h^n(W_n)\right]\pi(h)dh \\ &= |\varphi_n|\int_{\mathbb{W}^n}\left[\int_{H_n} L(\theta_0+\varphi_n h,d_n)\prod_{i=1}^n f_h(w_i)\,\pi(h)dh\right]dW_n \end{aligned}$$

where $|\varphi_n|$ is the determinant of the matrix φ_n, $P_h = P_{\theta_0+\varphi_n h}$, $f_h(\cdot)$ is the density function corresponding to $P_h(\cdot)$, and the last equality is from Fubini's theorem. Since $B_n(d_n,\pi)$ will converge to zero when n goes to infinity, $B_n(d_n,\pi)$ will be used for $|\varphi_n|^{-1}B_n(d_n,\pi)$ in the following discussion. Interestingly, the effect of the prior on the decision procedure does not disappear even asymptotically. This is basically because this prior is imposed on the local parameter space, not on the original parameter space. When n gets larger, $\pi(\theta)$ actually concentrates around θ_0. If a fixed prior $\pi(\theta)$ is used, then $B_n(d_n,\pi)$ will asymptotically only depend on $\pi(\theta_0)$, not the whole function $\pi(\cdot)$.[3] The question left is to find the best d_n to minimize $B_n(d_n,\pi)$.

From the last equality in $B_n(d_n,\pi)$, the expected posterior loss can be minimized for each W_n to minimize

[1] The matrix φ_n is selected by letting $nH^2(P_{\theta_0},P_{\theta_n}) \asymp 1$, where $H^2(P_0,P_1) \equiv \frac{1}{2}\int\left[\sqrt{p_1(x)}-\sqrt{p_2(x)}\right]^2 d\mu(x)$ is the Hellinger distance between P_0 and P_1 with p_i the density of P_i with respect to any measure μ dominating P_0 and P_1, $\theta_n = \theta_0+\varphi_n h$ for some fixed h, and $nH^2(P_{\theta_0},P_{\theta_n}) \asymp 1$ means $nH^2(P_{\theta_0},P_{\theta_n})/1$ is equal to some constant c. See Section 13.1 of Lehmann and Romano (2005) for more details about the selection of φ_n. Usually, φ_n is known from the estimation problem and makes P_{θ_n} contiguous with respect to P_{θ_0}. In the examples considered below, P_{θ_n} is indeed contiguous with respect to P_{θ_0}. But $nH^2(P_{\theta_0},P_{\theta_n}) \asymp 1$ is not equivalent to the contiguity between P_{θ_n} and P_{θ_0}; see Oosterhoff and van Zwet (1979) for a sharp result on the relationship between contiguity and the Hellinger distance.

[2] In this paper, we assume θ_0 "uniquely" exists in the "interior" of Θ, so H_n can be potentially constructed although we do not know it exactly when θ_0 is not completely specified in H_0. This assumption excludes partially identified models and the case where θ_0 is on the boundary of Θ which is considered, for example, in Andrews (2001).

[3] This relies on $\varphi_n \to 0$ as $n\to\infty$. If $\varphi_n = 1$ for some part of the parameter θ such as γ in the specification test of threshold regression in Hansen (1996), the whole prior function on the original parameter space of that component of parameters (not only $\pi(\theta_0)$) matters. The parameters with $\varphi_n = 1$ are essentially localized parameters.

$B_n(d_n, \pi)$:

$$\begin{aligned}
& \int_{H_n} L(\theta_0 + \varphi_n h, d_n) \prod_{i=1}^{n} f_h(w_i) \pi(h) dh \\
= \;& (1 - d_n) \int_{H_n} L(\theta_0 + \varphi_n h, 0) \prod_{i=1}^{n} f_h(w_i) \pi(h) dh + d_n \left(\int_{H_n} L(\theta_0 + \varphi_n h, 1) \prod_{i=1}^{n} f_h(w_i) \pi(h) dh \right) \\
= \;& (1 - d_n) \int_{H_{1n}} \prod_{i=1}^{n} f_h(w_i) \pi(h) dh + d_n \left[b \int_{H_{0n}} \prod_{i=1}^{n} f_h(w_i) \pi(h) dh \right]
\end{aligned}$$

where $H_{0n} = \varphi_n^{-1}(\Theta_0 - \theta_0)$, and $H_{1n} = \varphi_n^{-1}(\Theta_1 - \theta_0)$. So

$$d(W_n, \pi) = \begin{cases} 1, & \text{if } PO(W_n, \pi) > b, \\ 0, & \text{if } PO(W_n, \pi) < b, \\ 0 \text{ or } 1, & \text{if } PO(W_n, \pi) = b \end{cases}$$

where

$$PO(W_n, \pi) = \frac{\int_{H_{1n}} \prod_{i=1}^{n} f_h(w_i) \pi(h) dh}{\int_{H_{0n}} \prod_{i=1}^{n} f_h(w_i) \pi(h) dh} = \frac{\int_{H_{1n}} L_n(\theta_0 + \varphi_n h) \pi(h) dh}{\int_{H_{0n}} L_n(\theta_0 + \varphi_n h) \pi(h) dh} \tag{3}$$

is the posterior odds which measures the ratio of the posterior probability of H_1 and H_0, and $L_n(\theta) = \prod_{i=1}^{n} f(w_i|\theta)$ is the likelihood function.[4]

Under d_n, the average risk $B_n(d_n, \pi)$ is

$$b \int_{H_{0n}} P_h^n(d_n = 1) \pi(h) dh + \int_{H_{1n}} (1 - P_h^n(d_n = 1)) \pi(h) dh,$$

which is the prior-weighted sum of type I and type II error probabilities. The Neyman-Pearson framework of testing is actually to select a specific b such that

$$\sup_{h \in H_{0n}} P_h^n(d_n = 1) \leq \alpha. \tag{4}$$

If b can be selected such that $P_h^n(d_n = 1) = \alpha$ for all $h \in H_{0n}$, where α is the prespecified significance level, then minimizing the average risk is equivalent to maximizing the WAP on H_{1n}:

$$WAP(d_n, \pi) = \int_{H_{1n}} P_{h_0}^n(d_n = 1) \pi(h_0) dh_0.$$[5]

Here, we use h_0 instead of h to distinguish the alternative h and the h in the integral of $PO(W_n, \pi)$. If a uniformly most powerful (UMP) test exists, then it maximizes WAP for all choices of π. Usually, b is hard to find for a fixed n, but can be selected when n goes to infinity.[6]

[4] $L_n(\theta)$ is the joint density of the data when treated as a function of W_n. In the following discussion, $L_n(\theta)$ is used for both the joint density and the likelihood function without introducing new notations to distinguish them.

[5] That $\pi(h) = 0$ for $h \in H_{0n}$ will also make the first term in $B_n(d_n, \pi)$ a constant, but it will induce $PO(W_n, \pi) = \infty$, which will further induce rejection for all W_n and b and make the testing problem trivial. This is understandable. Since only the type II error is taken care of by using this weighting scheme, rejection is always the best choice.

[6] There is another chance to make the first term of $B_n(d_n, \pi)$ vanish asymptotically: $\lim_{n \to \infty} \sup_{h \in H_{0n}} P_h^n(d_n = 1) = 0$, but this will make b go to ∞; that is, there is no rejection.

Note that the specification of α is related to punishment arising from the type I error. As α gets smaller, a larger b is required to satisfy (4), so punishment on the type I error gets more severe. Consequently, $P_{h_0}^n\left(PO\left(W_n,\pi\right)>b\right)$ is smaller, and the type II error is larger, so there is a trade-off between type I and type II errors.

2.2 Likelihood Ratio Process

For any testing in large samples, we must provide asymptotic critical values and give out local powers besides finding a test statistic. In a parametric framework, these two questions can be answered by utilizing the likelihood ratio process.

To find asymptotic critical values, notice that

$$PO\left(W_n,\pi\right)=\frac{\int_{H_{1n}}\prod_{i=1}^{n}\frac{f_h(w_i)}{f_0(w_i)}\pi(h)dh}{\int_{H_{0n}}\prod_{i=1}^{n}\frac{f_h(w_i)}{f_0(w_i)}\pi(h)dh}=\frac{\int_{H_{1n}}Z_n(h)\pi(h)dh}{\int_{H_{0n}}Z_n(h)\pi(h)dh}$$

where $Z_n(\cdot)$ is the likelihood ratio process. If $Z_n(\cdot)$ weakly converges to $Z_\infty(\cdot)$, the limit likelihood ratio process, then by the continuous mapping theorem, $PO\left(W_n,\pi\right)$ weakly converges to

$$\frac{\int_{H_{1\infty}}Z_\infty(h)\pi(h)dh}{\int_{H_{0\infty}}Z_\infty(h)\pi(h)dh}\equiv PO_\infty\left(\pi\right),$$

where $H_{0\infty}=\lim\limits_{n\to\infty}H_{0n}$ and $H_{1\infty}=\lim\limits_{n\to\infty}H_{1n}$ are usually subspaces, half subspaces or unions of them.[7] Now, the critical value b is determined by

$$P\left(PO_\infty\left(\pi\right)>b\right)=\alpha.$$

To calculate power under the local parameter h_0, we need the asymptotic distribution of $Z_n(h)$ under h_0 instead of θ_0. Rewrite $PO\left(W_n,\pi\right)$ as

$$PO\left(W_n,\pi,h_0\right)=\frac{\int_{H_{1n}}\prod_{i=1}^{n}\frac{f_h(w_i)}{f_{h_0}(w_i)}\pi(h)dh}{\int_{H_{0n}}\prod_{i=1}^{n}\frac{f_h(w_i)}{f_{h_0}(w_i)}\pi(h)dh}\equiv\frac{\int_{H_{1n}}Z_n(h,h_0)\pi(h)dh}{\int_{H_{0n}}Z_n(h,h_0)\pi(h)dh},$$

which converges under h_0 to

$$PO_\infty\left(\pi,h_0\right)=\frac{\int_{H_{1\infty}}Z_\infty(h,h_0)\pi(h)dh}{\int_{H_{0\infty}}Z_\infty(h,h_0)\pi(h)dh}\tag{5}$$

by the continuous mapping theorem, where $Z_\infty(\cdot,h_0)$ is the limit likelihood ratio process under h_0.[8] Note that $PO_\infty\left(\pi\right)=PO_\infty\left(\pi,0\right)$. Now, the local power at $h_0\in H_{1\infty}$ is

$$P_{h_0}^\infty\left(PO_\infty\left(\pi,h_0\right)>b\right),$$

where $P_{h_0}^\infty\left(\cdot\right)$ is the asymptotic distribution under h_0.

Come back to the discussion in the last subsection. If the distribution of $PO_\infty\left(\pi\right)$ is pivotal to $h\in H_{0\infty}$, which is equivalent to $PO_\infty\left(\pi,h_0\right)$ not depending on h_0 when $h_0\in H_{0\infty}$, then b can be found to maximize

[7] The notion of convergence of sets is used here, which is defined, for example, on page 101 of Van der Vaart (1998).

[8] From this form of the posterior odds, we need only know the conditional distribution of the data given weakly exogenous variables.

the WAP over H_{1n}. In this case, the asymptotic WAP is

$$\int_{H_{1\infty}} P_{h_0}^{\infty}\left(PO_{\infty}\left(\pi, h_0\right) > b\right) \pi\left(h_0\right) dh_0.$$

Usually, $Z_{\infty}(h, h_0)$ is exponentially decaying at the tail of h, so $PO_{\infty}(\pi, h_0)$ is well defined as long as $\pi(h)$ has a polynomial majorant. But the power at h_0 is usually greater than α, and converges to 1 when $|h_0|$ goes to infinity. To make the WAP a finite number, $\int_{H_{1\infty}} \pi(h_0)\, dh_0$ must be finite. In other words, $\pi(h_0)$ can be normalized as a density on $H_{1\infty}$. In the following discussion, we will focus on the case that $\int_{H_{0\infty}} \pi(h)\, dh = \int_{H_{1\infty}} \pi(h)\, dh = 1$.[9]

The Bayesian decision rule in Andrews (1994) is that the posterior odds is greater than 1. When we assume $\int_{H_{0\infty}} \pi(h)\, dh = \int_{H_{1\infty}} \pi(h)\, dh = 1$, the posterior odds is equivalent to the Bayes factor statistic in the Bayesian literature.[10] Jeffreys (1961) provides critical values for the Bayes factor which can also apply to the posterior odds. In the examples in Section 3 and 4, the performance of Jeffreys (1961)'s decision rule is evaluated from a frequentist perspective.

2.3 Special Cases

There is a special case where b can be found even if n is finite. In particular, the selection of b is possible when the null hypothesis H_0 is simple; that is, $H_0 : \theta = \theta_0$. In this case, $H_{0n} = \{0\}$, a singleton, and H_{1n} converges to $H_{1\infty} = \mathbb{R}^k \backslash \{0\}$. If $\pi(h)$ is a density, then $PO(W_n, \pi)$ always equals ∞, and H_0 is always rejected. To make the problem nontrivial, assume there is a unit point mass at $h = 0$ and a density on $h \in H_{1n}$.[11] In other words, π is a mixture of a discrete component and a continuous component. This makes testing different from estimation where a point mass in the prior is seldom assumed.[12] Under this assumption,

$$\begin{aligned} PO(W_n, \pi) &= \int_{H_{1n}} \prod_{i=1}^{n} \frac{f_h(w_i)}{f_0(w_i)} \pi(h) dh \\ &= \int_{H_{1n}} Z_n(h) \pi(h) dh. \end{aligned}$$

In finite samples, b is selected such that $P_0^n\left(\int_{H_{1n}} Z_n(h)\pi(h)dh > b\right) = \alpha$. In large samples, b is selected such that $P_0^{\infty}\left(\int_{\mathbb{R}^k\backslash\{0\}} Z_{\infty}(h)\pi(h)dh > b\right) = \alpha$. The corresponding test statistic in practice is the integrated likelihood ratio test statistic:

$$\int_{H_{1n}} \frac{L_n(\theta_0 + \varphi_n h)}{L_n(\theta_0)} \pi(h) dh.$$

[9] This is equivalent to $\pi = \frac{1}{2}$ in Andrews (1994).

[10] The key difference between the Bayes factor statistic and posterior odds is that the posterior odds takes into account the prior odds of H_0 and H_1. Since the Bayes factor $BF = \frac{\int_{\Theta_1} L_n(\theta)\pi(\theta|H_1)d\theta}{\int_{\Theta_0} L_n(\theta)\pi(\theta|H_0)d\theta} = \frac{P(\text{Data}|H_1)}{P(\text{Data}|H_0)} = \frac{\text{Average Likelihood Conditional on } H_1}{\text{Average Likelihood Conditional on } H_0} = \frac{\int_{\Theta_1} L_n(\theta)\pi(\theta|H_1)\pi(H_1)d\theta}{\int_{\Theta_0} L_n(\theta)\pi(\theta|H_0)\pi(H_0)d\theta} / \frac{\pi(H_1)}{\pi(H_0)} = \frac{PO(W_n,\pi)}{\pi(H_1)/\pi(H_0)} = \frac{\text{Posterior Odds}}{\text{Prior Odds}}$, only $\pi(\theta|H_1)$ and $\pi(\theta|H_0)$ appear in BF, which does not depend on the prior beliefs concerning H_0 and H_1. So the extra information in the posterior odds beyond the Bayes factor is the relative prior belief on H_0 and H_1. See Section 4.3.3 of Berger (1985) for an introduction to the Bayes Factor.

[11] If Θ_1 also includes only finitely many points, then π is assumed to be discrete on H_{1n}, but this is not the emphasis of this paper.

[12] In the estimation problem, θ_0 is unknown, so a prior can not be imposed on the localized parameter space. Also, θ_0 has the same importance as any other point in Θ since any point in Θ can be the true value. But in the testing problem, the "point" θ_0 has the same importance as the "set" $\Theta\backslash\{\theta_0\}$ since testing is a dichotomatic decision problem, so putting a point mass on θ_0 is not unreasonable. In an AR(1) estimation problem, one might have a point and slab prior with mass point at unity and a density over $(-1, 1)$. A Bayesian could then update this prior based on sample data. But such a prior specification is obviously stimulated by unit root tests which make unity a special point.

The power function becomes

$$P_{h_0}^{\infty}\left(\int_{H_{1\infty}} \frac{Z_\infty(h,h_0)}{Z_\infty(0,h_0)}\pi(h)dh > b\right).$$

The cornerstone of testing in the frequentist literature is the Neyman-Pearson lemma, which is a special case of the above argument. For a fixed h_0, the prior $\pi(h)$ is assumed to put unit point masses on 0 and h_0 and have no density elsewhere, then $PO\left(W_n,\pi\right)$ reduces to the likelihood ratio $Z_n(h_0)$ and $B_n\left(d_n,\pi\right)$ reduces to $b\ \alpha + 1 - \text{Power}_n(h_0)$, where b is selected to guarantee the type I error to be α. So the likelihood ratio test is equivalent to maximize $\text{Power}_n(h_0)$ with the significance level α. The asymptotic critical value is determined by

$$P_0^{\infty}\left(Z_\infty(h_0) > b\right) = \alpha,$$

and the asymptotic power is

$$P_{h_0}^{\infty}\left(\frac{1}{Z_\infty(0,h_0)} > b\right).$$

In the literature (e.g., Andrews and Ploberger, 1994), H_1 is formulated as a simple hypothesis when the WAP criterion is used:

$$H_1 : W_n \sim \int_{H_{1n}} L_n\left(\theta_0 + \varphi_n h\right)\pi(h)dh.$$

When $PO_\infty\left(\pi,h_0\right)$ is pivotal to $h_0 \in H_{0\infty}$, the Neyman-Pearson lemma can be used to find the optimal test in the WAP sense, which is just $PO\left(W_n,\pi\right)$. When $PO_\infty\left(\pi,h_0\right)$ is not pivotal to $h_0 \in H_{0\infty}$, the WAP maximizing test is typically found by the Neyman-Pearson test of

$$H_0 : W_n \sim \int_{H_{0n}} L_n\left(\theta_0 + \varphi_n h\right)\pi'(h)dh$$

versus H_1 above, where π' is the least favorable distribution for h and is hard to identify in many problems; see Section 3.8 of Lehmann and Romano (2005) for more discussion. Fortunately, the quantity $PO_\infty\left(\pi,h_0\right)$ in the examples of Section 3 and 4 is pivotal to $h_0 \in H_{0\infty}$. As mentioned in the introduction, Theorem 2 in Andrews and Ploberger (1994) actually claims that their test has the smallest weighted average type I and type II errors, not the maximum WAP. This is because the null hypothesis there is composite not simple as they assumed when using the Neyman-Pearson Lemma. Fortunately, their test indeed maximizes the WAP from the discussion in Section 3 below since it is pivotal to $h_0 \in H_{0\infty}$.

2.4 Likelihood Ratio Tests

For completeness and comparison, we also report the asymptotic results for the general likelihood ratio test. Usually, the test statistic

$$\Lambda_n = 2\log\frac{\sup_{\theta\in\Theta} L_n\left(\theta\right)}{\sup_{\theta\in\Theta_0} L_n\left(\theta\right)} \tag{6}$$

is used instead of

$$\widetilde{\Lambda}_n = \frac{\sup_{\theta\in\Theta_1} L_n\left(\theta\right)}{\sup_{\theta\in\Theta_0} L_n\left(\theta\right)}.$$ [13]

[13] $\Lambda_n = 2\log\left(\widetilde{\Lambda}_n \vee 1\right)$.

Define

$$\widehat{\theta}_{n,0} = \arg\max_{\theta\in\Theta_0} L_n(\theta),$$
$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} L_n(\theta).$$

The asymptotic distribution of Λ_n can be derived by a similar procedure as above. Note that

$$\Lambda_n = 2\log\frac{L_n\left(\widehat{\theta}_n\right)}{L_n\left(\widehat{\theta}_{n,0}\right)} = 2\log\frac{\sup_{h\in H_n}\frac{L_n(h)}{L_n(0)}}{\sup_{h\in H_{0n}}\frac{L_n(h)}{L_n(0)}} = 2\log\frac{\sup_{h\in H_n} Z_n(h)}{\sup_{h\in H_{0n}} Z_n(h)} \overset{\theta_0}{\rightsquigarrow} 2\log\frac{\sup_{h\in H_\infty} Z_\infty(h)}{\sup_{h\in H_{0\infty}} Z_\infty(h)} \equiv \Lambda_\infty$$

where $L_n(h) = \prod_{i=1}^{n} f_h(w_i)$ is the local likelihood function. When H_0 is simple,

$$\Lambda_n = 2\log\sup_{h\in H_n} Z_n(h) \overset{\theta_0}{\rightsquigarrow} 2\log\sup_{h\in H_\infty} Z_\infty(h).$$

An asymptotic critical value b is selected by

$$\sup_{h\in H_{0\infty}} P_h^\infty(\Lambda_\infty > b) \le \alpha.$$

When H_0 is simple, b is just selected by

$$P_0^\infty(\Lambda_\infty > b) = \alpha.$$

The power calculation is also similar as above. For the alternative h_0, rewrite Λ_n as

$$\Lambda_n(h_0) = 2\log\frac{\sup_{h\in H_n}\frac{L_n(h)}{L_n(h_0)}}{\sup_{h\in H_{0n}}\frac{L_n(h)}{L_n(h_0)}} = 2\log\frac{\sup_{h\in H_n} Z_n(h,h_0)}{\sup_{h\in H_{0n}} Z_n(h,h_0)} \overset{h_0}{\rightsquigarrow} 2\log\frac{\sup_{h\in H_\infty} Z_\infty(h,h_0)}{\sup_{h\in H_{0\infty}} Z_\infty(h,h_0)} \equiv \Lambda_\infty(h_0),$$

so the power is

$$P_{h_0}^\infty(\Lambda_\infty(h_0) > b).$$

When H_0 is simple,

$$\Lambda_\infty(h_0) = 2\log\sup_{h\in H_\infty}\frac{Z_\infty(h,h_0)}{Z_\infty(0,h_0)}.$$

Note that $\Lambda_\infty = \Lambda_\infty(0)$. Roughly speaking, the difference between the likelihood ratio test statistic and posterior odds is the substitution of an integral by a sup operator. The likelihood ratio test is Bahadur efficient as illustrated in Bahadur (1967).

In the discussion above, it is assumed that there is always a b such that the significance level α can be attained. This can happen if $PO_\infty(\pi)$ and Λ_∞ are continuously distributed. Usually, the distribution of $PO_\infty(\pi)$ is continuous since $PO(W_n,\pi)$ is an average, but this is not always true for Λ_∞. For example, Chernoff (1954) showed that "if one tests whether θ is on one side or the other of a smooth $(k-1)$-dimensional surface in k-dimensional space and θ lies on the surface, the asymptotic distribution of λ (Λ_n in this paper) is that of a chance variable which is zero half the time and which behaves like χ^2 with one degree of freedom the other half of the time". As we will show in Section 4, this also happens in the test on the location of threshold point, although the reason for discreteness is different.

When there is a discrete component in the distribution of T_∞, where T_∞ is the asymptotic limit of some test statistic T_n, then there may not exist a b such that $P_0^\infty(T_\infty > b) = \alpha$ even in the simple H_0 case.[14] For example, for a fixed α, it may happen that

$$P_0^\infty(T_\infty > b) < \alpha \text{ and } P_0^\infty(T_\infty \geq b) > \alpha.$$

In this case, randomized tests are used, and the results above follow with minor modification. A randomized test is defined as

$$d_n = \begin{cases} 1, & \text{if } T_n > b, \\ c, & \text{if } T_n = b, \\ 0, & \text{if } T_n < b, \end{cases} \tag{7}$$

such that

$$E_0^\infty[d_\infty] = P_0^\infty(T_\infty > b) + cP_0^\infty(T_\infty = b) = \alpha.$$

So

$$c = \frac{\alpha - P_0^\infty(T_\infty > b)}{P_0^\infty(T_\infty = b)} \in (0,1).$$

But $E_0^n[d_n]$ does not necessarily converge to α. Furthermore, the power under h_0 does not necessarily converge to

$$E_{h_0}^\infty[d_\infty] = P_{h_0}^\infty(T_\infty(h_0) > b) + cP_{h_0}^\infty(T_\infty(h_0) = b),$$

where $T_\infty(h_0)$ follows the asymptotic distribution of T_n under h_0 such as $\Lambda_\infty(h_0)$ above.

2.5 Summary

The most important element in the above discussion is the likelihood ratio process $Z_n(h, h_0)$. This is not a coincidence, since $Z_n(h, h_0)$ includes all relevant information about the parameters. The asymptotic representation theorem in Van der Vaart (1991) makes this more clear.[15] The theory of limits of experiments introduced by Le Cam (1972) and detailed in Van der Vaart (1996) underlies most of the optimality discussion, and this paper is not an exception. See the examples in Section 3 and 4 to appreciate more about the power of Le Cam's theory. In both examples, the basic idea is the same. First, find the optimal test in the limit experiment. Second, match the optimal test in finite samples.

The above arguments for optimality are summarized in the following Theorem 1. To make the statement of the theorem easier, we will simplify the testing structure discussed above although it can be easily extended to a more general framework. This simplified structure will be used in the examples below. Such a simplication also appears in Andrews (1994) and Choi, Hall and Schick (1996). Suppose a reparametrization of $\theta = (\vartheta', \zeta')'$ makes $H_0 : \vartheta = \vartheta_0$ and $H_1 : \vartheta \neq \vartheta_0$, where $\vartheta \in \mathbb{R}^r$. Assume further that $H_\infty \equiv \lim_{n\to\infty} \varphi_n^{-1}\left(\Theta - (\vartheta_0', \zeta')'\right) = \mathbb{R}^k$ does not depend on ζ, then $H_{0\infty} = 0 \times \mathbb{R}^{\underline{k}}$ is a $\underline{k}$ dimensional space, and $H_{1\infty} = \mathbb{R}^r \backslash \{0\} \times \mathbb{R}^{\underline{k}}$, where $\underline{k} = k - r$. Denote the local parameters at (ϑ, ζ) as (h_ϑ, h_ζ) and $h = (h_{\vartheta_0}, h_\zeta)$.

Definition 1 *A test d_n is* of asymptotic level $\alpha \in (0,1)$ at ζ *if*

$$\overline{\lim_{n\to\infty}} E_{(0,h_\zeta)}[d_n] \leq \alpha \textit{ for every } h_\zeta \in \mathbb{R}^{\underline{k}},$$

where $E_{(0,h_\zeta)}[\cdot]$ is the expectation under the local parameter $h = (0, h_\zeta)$.

[14] In the composite null hypothesis case, there may not exist b such that $\sup_{h \in H_{0\infty}} P_h^\infty(T_\infty > b) = \alpha$.

[15] In an estimation problem, only $Z_n(h)$ is relevant, but the theory of limits of experiments requires the convergence of $Z_n(h, h_0)$. The power calculation above makes up this gap.

As mentioned in Choi, Hall and Schick (1996), this definition plays the role of a restriction to regular estimates in estimation theory.

Definition 2 *A test d_n is* asymptotically π-averaged most powerful of level α at ζ, *written in short form as AAMP(α,π,ζ), if*

$$\overline{\lim_{n\to\infty}}\int_{H_{1n}} E_{h_0}\left[\xi_n\right]\pi\left(h_0\right)dh_0 \leq \underline{\lim_{n\to\infty}}\int_{H_{1n}} E_{h_0}\left[d_n\right]\pi\left(h_0\right)dh_0 \tag{8}$$

for every other asymptotic level α test ξ_n at ζ. A test is AAMP(α,π) if it is AAMP(α,π,ζ) for every nuisance parameter ζ.

Definition 3 *Two tests are* asymptotically equivalent, *or simply* equivalent, *if their difference converges to 0 in probability under both H_0 and H_1.*

Now, we state the theorem with high-level assumptions. The checking of these assumptions will be illustrated in the examples of Section 3 and 4.

Theorem 1 *If there is a test statistic T_n such that*

(i) $T_n - PO\left(W_n,\pi,h_0\right) \xrightarrow{p} 0$ *under any $h_0 \in H_\infty$, or they are equivalent;*

(ii) $PO\left(W_n,\pi,h_0\right) \overset{h_0}{\rightsquigarrow} \frac{\int_{H_{1\infty}} Z_\infty(h,h_0)\pi(h)dh}{\int_{H_{0\infty}} Z_\infty(h,h_0)\pi(h)dh} \equiv PO_\infty\left(\pi,h_0\right)$, *which has a continuous distribution, for any $h_0 \in H_\infty$, where $Z_\infty(\cdot,h_0)$ is the limit likelihood ratio process under h_0;*

(iii) $PO_\infty\left(\pi,h_0\right)$ *is pivotal to $h_0 \in H_{0\infty}$;*

Then d_n defined by $\mathbf{1}\left(T_n > b\right)$ is AAMP(α,π,ζ), where b satisfies

$$P_0^\infty\left(PO_\infty\left(\pi\right) > b\right) = \alpha$$

with $PO_\infty\left(\pi\right)$ the common distribution of $PO_\infty\left(\pi,h_0\right)$ under $h_0 \in H_{0\infty}$. If $PO_\infty\left(\pi,h_0\right)$ is pivotal to both $h_0 \in H_{0\infty}$ and ζ in (iii), then d_n is AAMP(α,π). Furthermore,

$$E_{h_0}^n\left[d_n\right] \to P_{h_0}^\infty\left(PO_\infty\left(\pi,h_0\right) > b\right),$$

for any $h_0 \in H_{1\infty}$.

From Theorem 1, three questions should be answered to find the optimal test in the sense of (8).

(1) What is $PO_\infty\left(\pi,h_0\right)$? Does $PO_\infty\left(\pi,h_0\right)$ satisfy (iii) in Theorem 1? If so, what is the power envelope?

(2) What is the feasible test statistic T_n such that (i) in Theorem 1 is satisfied?

(3) How to find the critical value b in practice?

The asymptotic distribution of the likelihood ratio statistic is also stated in the following Theorem under high-level conditions.

Theorem 2 *Suppose the following four assumptions are satisfied:*

(i) $\widehat{\theta}_n$ *and* $\widehat{\theta}_{n0}$ *are both φ_n-consistent under $h_0 \in H_\infty$.*

(ii) *There is an approximation* $\log Z_n^a(h,h_0)$ *of* $\log Z_n(h,h_0)$ *such that for any* $M<\infty$,

$$\sup_{\|h\|\le M}\left|\log Z_n(h,h_0)-\log Z_n^a(h,h_0)\right|\xrightarrow{p}0$$

under h_0.

(iii) *There is a metric on the space of the sample path of* $Z_n^a(h,h_0)$ *such that* $\sup_{\|h\|\le M}$ *is a continuous operator on that space and*

$$Z_n^a(h,h_0)\overset{h_0}{\rightsquigarrow}Z_\infty(h,h_0)$$

for h *on any compact set.*

(iv) $\lim_{\|h\|\to\infty}Z_\infty(h,h_0)=0$ *almost surly.*

Then

$$\Lambda_n(h_0)\overset{h_0}{\rightsquigarrow}\Lambda_\infty(h_0)$$

for every $h_0\in H_\infty$.

For comparative purposes, the following Table 1 provides some analogs between testing and estimation problems in the decision framework.

Testing	Estimation
Prior on the Local Parameter Space	Loss Function on the Local Parameter Space
Asymptotic Average Power	Asymptotic Average Risk
Likelihood Ratio Test	Maximum Likelihood Estimator[16]
Posterior Odds	Bayes Estimator

Table 1: Analogs Between Testing and Estimation

3 Tests in LAN Models

Suppose the sequence of models P_θ^n is LAN: in other words, there exist matrices φ_n and I_θ and random vectors Δ_θ^n such that $\Delta_\theta^n\overset{\theta}{\rightsquigarrow}N(0,I_\theta)$ and for every converging sequence $h_n\to h$,

$$\ln\frac{dP^n_{\theta+\varphi_n h_n}}{dP_\theta^n}=h'\Delta_\theta^n-\frac{1}{2}h'I_\theta h+o_{P_\theta^n}(1).$$

Andrews (1994) proves the asymptotic equivalence between the posterior odds tests and the three asymptotically equivalent tests especially the likelihood ratio test under classical second-order smoothness assumptions which imply the LAN models. The following discussion essentially follows Andrews (1994) and Andrews and Ploberger (1994), but relaxes the required assumptions. The second purpose of the discussion is to provide an intuitive explanation for why the equivalence holds in LAN models, which contrasts with the results in the next section where such equivalence breaks down. The last purpose of this section is to fill a logical gap in Andrews and Ploberger (1994), as mentioned in the introduction.

[16] The maximum likelihood estimator (MLE) can be treated as a special Bayes estimator, but the likelihood ratio test statistic is hard to be treated as a special posterior odds.

From chapter 7 of Van der Vaart (1998), a decision in a LAN model is equivalent to the decision based on a single observation z_h which follows $N(h, I^{-1})$ in the limit experiment, where $I = I(\theta_0)$ is the information matrix when the true value is θ_0. Now,

$$Z_n(h, h_0) \overset{h_0}{\rightsquigarrow} Z_\infty(h, h_0) = \frac{dN\left(h, I^{-1}\right)}{dN\left(h_0, I^{-1}\right)}\left(z_{h_0}\right) = \exp\left\{-\frac{1}{2}\left(h - h_0\right)' I \left(h - h_0\right) + \left(h - h_0\right)' I z\right\}, \tag{9}$$

for any $h_0 \in H_\infty = \mathbb{R}^k$, where z follows $N(0, I^{-1})$.

Suppose there are r constraints in H_0:

$$\begin{aligned} H_0 &: \ a(\theta_0) = 0, \\ H_1 &: \ a(\theta_0) \neq 0, \end{aligned}$$

where $a(\cdot) : \mathbb{R}^k \to \mathbb{R}^r$ is a smooth function and $r \leq k$. Since $a(\theta_0) = 0$ is locally equivalent to $Ah = 0$, where $A = \frac{\partial a(\theta_0)}{\partial \theta'}$ is of full row rank, the problem reduces to the test

$$\begin{aligned} H_0 &: \ Ah = 0, \\ H_1 &: \ Ah \neq 0. \end{aligned}$$

Andrews (1994) considers the special case $A = \begin{pmatrix} \mathbf{I}_r & \mathbf{0}_{r\times \underline{k}} \end{pmatrix}$ with $\mathbf{I}_r$ a $r \times r$ identity matrix. In this case, ϑ is the first r coordinates of θ as in the simplified framework at the end of Section 2, but the following discussion applies to a general A.

We answer the three questions in Section 2 in an intuitive way and summarize the results rigorously in the following Theorem 3. First, we find the WAP envelope. In this case,

$$\begin{aligned} \Lambda_\infty\left(h_0\right) &= 2\log\frac{\sup_{h\in H_\infty} Z_\infty(h, h_0)}{\sup_{h\in H_{0\infty}} Z_n(h, h_0)} \\ &= 2\log\frac{\sup_{h\in\mathbb{R}^k}\exp\left\{-\frac{1}{2}\left(h - h_0\right)' I\left(h - h_0\right) + \left(h - h_0\right)' I z\right\}}{\sup_{Ah=0}\exp\left\{-\frac{1}{2}\left(h - h_0\right)' I\left(h - h_0\right) + \left(h - h_0\right)' I z\right\}} \\ &= z'Iz - \left(z'P'IPz - h_0 P^{\perp\prime} I P^{\perp} h_0 - 2h_0' P^{\perp\prime} I P^{\perp} z\right) \\ &= z' P^{\perp\prime} I P^{\perp} z + h_0 P^{\perp\prime} I P^{\perp} h_0 + 2h_0' P^{\perp\prime} I P^{\perp} z \\ &= z_{h_0}' P^{\perp\prime} I P^{\perp} z_{h_0}, \end{aligned}$$

where $\arg\sup_{Ah=0}\exp\left\{-\frac{1}{2}\left(h - h_0\right)' I\left(h - h_0\right) + \left(h - h_0\right)' I z\right\} = \left[\mathbf{I} - I^{-1}A'\left(AI^{-1}A'\right)^{-1}A\right]\left(z + h_0\right) \equiv Pz_{h_0}$ with $\mathbf{I}$ a $k \times k$ identity matrix, and $P^\perp = I^{-1}A'\left(AI^{-1}A'\right)^{-1}A$ satisfies $P^{\perp\prime}IP = P'IP^\perp = 0$, $P^{\perp\prime}IP^\perp = P^{\perp\prime}I = IP^\perp$, $P'IP = P'I = IP$. Actually, P is the projection matrix on the subspace $\{Ah = 0\}$ with respect to the inner product $\langle h_1, h_2\rangle = h_1' I h_2$, and $P^\perp$ is the projection matrix on the orthogonal space of $\{Ah = 0\}$.[17] For any h such that $Ah = 0$, $z_h' P^{\perp\prime} I P^\perp z_h$ follows the same χ^2_r distribution, so the asymptotic distribution of the likelihood ratio test is pivotal to $h_0 \in H_{0\infty}$.[18] Furthermore, the asymptotic distribution does not depend on θ_0, so it is also pivotal to any true parameter $\theta_0 \in \Theta_0$.

[17] In the special case mentioned above, (4.3) of Andrews (1994) gives the basis for this orthogonal space.

[18] If z_h follows $N\left(h, I^{-1}\right)$ such that $Ah = 0$, then $P^\perp z_h$ follows the same distribution as $P^\perp z$. The χ^2_r distribution follows from, for example, Lemma 16.6 in Van der Vaart (1998).

In this case, if $\pi(h)$ is selected such that

$$PO_{\infty}(\pi, h_0) = \frac{\int_{Ah\neq 0} Z_{\infty}(h,h_0)\pi(h)dh}{\int_{Ah=0} Z_{\infty}(h,h_0)\pi(h)dh} = G\left(z_{h_0}' P^{\perp\prime} I P^{\perp} z_{h_0}\right),$$

where $G(\cdot)$ is an increasing function, then $z_{h_0}' P^{\perp\prime} I P^{\perp} z_{h_0}$ is asymptotically sufficient for $PO_{\infty}(\pi, h_0)$. The decision based on $PO_n(W_n, \pi)$ is asymptotically equivalent to that based on Λ_n. Such a $\pi(h)$ indeed exists, for example,

$$\pi(h)dh = \mathbf{1}(Ah=0)\cdot \pi_1(Ph)\, d(Ph) + \mathbf{1}(Ah\neq 0)\cdot \pi_2\left(P^{\perp}h\right) d\left(P^{\perp}h\right)\cdot \pi_1(Ph)\, d(Ph),$$

with $\pi_1(\cdot)$ an arbitrary density function, and

$$\pi_2\left(P^{\perp}h\right) = \sqrt{\frac{c^r |I|}{(2\widetilde{\pi})^r}} \exp\left\{-\frac{c}{2} h' P^{\perp\prime} I P^{\perp} h\right\}.$$

The prior on $H_{0\infty}^{\perp}$, the orthogonal space of $H_{0\infty}$, is $N\left(0, (cI)^{-1}\right)$. The larger is c, the more weight is given to alternatives for which $P^{\perp}h$ is large. It is noninformative when $c=0$, and is a point mass at 0 when $c=\infty$. Andrews (1994) takes $\pi_1(\cdot)$ as a point mass at $\mathbf{0}$, which makes sense only because the asymptotic distribution of the test statistic is pivotal to $\theta_0 \in \Theta_0$, but this is not generally true as shown in next section.[19] The motivation for this specification is that the power on $H_{0\infty}^{\perp}$ is of main interest, so the prior is specified in a separable way with respect to $H_{0\infty}$ and $H_{0\infty}^{\perp}$. We will see that this prior is conjugate to $Z_{\infty}(h,h_0)$ such that the common terms related to $H_{0\infty}$ in the numerator and denominator of $PO_{\infty}(\pi, h_0)$ are canceled.

$$\begin{aligned}
& PO_{\infty}(\pi, h_0) \\
= & \frac{\int_{Ah\neq 0} \exp\left\{-\frac{1}{2}(h-h_0)' I (h-h_0) + (h-h_0)' I z\right\} \sqrt{\frac{c^r|I|}{(2\widetilde{\pi})^r}} \exp\left\{-\frac{c}{2} h' P^{\perp\prime} I P^{\perp} h\right\} d\left(P^{\perp}h\right)\cdot \pi_1(Ph)\, d(Ph)}{\int_{Ah=0} \exp\left\{-\frac{1}{2}(h-h_0)' I (h-h_0) + (h-h_0)' I z\right\} \mathbf{1}(Ah=0)\cdot \pi_1(Ph)\, d(Ph)} \\
= & \frac{\begin{array}{c}\int_{Ah\neq 0} \exp\left\{-\frac{1}{2}(h-h_0)'\left(P+P^{\perp}\right)' I \left(P+P^{\perp}\right)(h-h_0) + (h-h_0)'\left(P+P^{\perp}\right)' I \left(P+P^{\perp}\right) z\right\} \\ \sqrt{\frac{c^r|I|}{(2\widetilde{\pi})^r}} \exp\left\{-\frac{c}{2} h' P^{\perp\prime} I P^{\perp} h\right\} d\left(P^{\perp}h\right)\cdot \pi_1(Ph)\, d(Ph)\end{array}}{\begin{array}{c}\int_{Ah=0} \exp\{-\frac{1}{2}(h-h_0)'\left(P+P^{\perp}\right)' I \left(P+P^{\perp}\right)(h-h_0) \\ + (h-h_0)'\left(P+P^{\perp}\right)' I \left(P+P^{\perp}\right) z\} \mathbf{1}(Ah=0)\cdot \pi_1(Ph)\, d(Ph)\end{array}} \\
= & \frac{\int_{\mathbb{R}^r} \sqrt{\frac{c^r|I|}{(2\widetilde{\pi})^r}} \exp\left\{-\frac{1}{2}(h-h_0)' P^{\perp\prime} I P^{\perp}(h-h_0) + (h-h_0)' P^{\perp\prime} I P^{\perp} z - \frac{c}{2} h' P^{\perp\prime} I P^{\perp} h\right\} d\left(P^{\perp}h\right)}{\exp\left\{-\frac{1}{2} h_0' P^{\perp\prime} I P^{\perp} h_0 - h_0' P^{\perp\prime} I P^{\perp} z\right\}} \\
= & \int_{\mathbb{R}^r} \sqrt{\frac{c^r|I|}{(2\widetilde{\pi})^r}} \exp\left\{-\frac{1}{2}\left[(h-h_0)' P^{\perp\prime} I P^{\perp}(h-h_0) - h_0' P^{\perp\prime} I P^{\perp} h_0 - 2h' P^{\perp\prime} I P^{\perp} z + c h' P^{\perp\prime} I P^{\perp} h\right]\right\} d\left(P^{\perp}h\right) \\
= & \int_{\mathbb{R}^r} \sqrt{\frac{c^r|I|}{(2\widetilde{\pi})^r}} \exp\left\{-\frac{1}{2}\left[(1+c)\left(s - \frac{P^{\perp}(z+h_0)}{1+c}\right)' I \left(s - \frac{P^{\perp}(z+h_0)}{1+c}\right) - \frac{(z+h_0)' P^{\perp\prime} I P^{\perp}(z+h_0)}{(1+c)}\right]\right\} ds \\
= & \left(\frac{c}{1+c}\right)^{r/2} \exp\left\{\frac{(z+h_0)' P^{\perp\prime} I P^{\perp}(z+h_0)}{2(1+c)}\right\} \\
\equiv & G\left(z_{h_0}' P^{\perp\prime} I P^{\perp} z_{h_0} | c, r\right)
\end{aligned}$$

[19]See Assumption 5 in Andrews (1994) for a more general setup of the priors on $H_{0\infty}^{\perp}$.

does not depend on Ph_0 and shares the same power as Λ_n.[20] Now, b is set to $G\left(\chi^2_{r,1-\alpha}|c,r\right)$, where $\chi^2_{r,1-\alpha}$ is the $1-\alpha$ quantile of χ^2_r distribution, so b depends on α, and also on the prior through c. The larger the significance level, the smaller the weight on the type I error. But b need not depend on c in a monotonic way, which depends on the value of $r-\chi^2_{r,1-\alpha}$.[21] Push to extreme cases: $c=0$, $b=0$; $c=\infty$, $b=1$; $\alpha=0$, $b=\infty$; $\alpha=1$, $b=\left(\frac{c}{1+c}\right)^{r/2}>0$.

The local power based on $PO_n\left(W_n,\pi\right)$ is

$$\begin{aligned} P(PO_\infty\left(\pi,h_0\right) &> b) = P\left(z'_{h_0}P^{\perp\prime}IP^{\perp}z_{h_0}>\chi^2_{r,1-\alpha}\right) \\ &= P\left(\chi^2_r\left(\delta\right)>\chi^2_{r,1-\alpha}\right) \\ &= \sum_{j=0}^{\infty}e^{-\delta/2}\frac{(\delta/2)^j}{j!}\frac{\Gamma\left(j+r/2,\chi^2_{r,1-\alpha}/2\right)}{\Gamma\left(j+r/2\right)} \end{aligned}$$

where $\chi^2_r\left(\delta\right)$ is the noncentral chi-square distribution with the noncentrality parameter δ, $\Gamma\left(a,x\right)=\int_x^\infty t^{a-1}e^{-t}dt$ is the upper incomplete gamma function, and $\Gamma\left(\cdot\right)$ is the Gamma function.[22] $\delta=h_0'P^{\perp\prime}IP^{\perp}h_0$ only depends on $P^\perp h_0$ not on Ph_0, so the power is the same on the ellipsoid $h_0'P^{\perp\prime}IP^{\perp}h_0=c$ for a fix c. This phenomenon is first observed in Wald (1943). For a fixed r, the larger the significance level, the more powerful the test, so there is a trade-off between the type I error and the type II error as shown in Section 2. When δ goes to infinity or α goes to 1, the power goes to 1. This power does not depend on the prior since the effect of the prior is offset by the choice of b.[23] As discussed in Section 2, such a test maximizes the asymptotic WAP:

$$\begin{aligned} &\int_{Ah_0\neq0}P(PO_\infty\left(\pi,h_0\right)>b)\pi(h_0)dh \\ =&\int_{Ah_0\neq0}P\left(\chi^2_r\left(\delta\right)>\chi^2_{r,1-\alpha}\right)\pi(h_0)dh_0 \\ =&\int_{Ah_0\neq0}P\left(\chi^2_r\left(\delta\right)>\chi^2_{r,1-\alpha}\right)\sqrt{\tfrac{c^r|I|}{(2\pi)^r}}\exp\left\{-\tfrac{c}{2}h_0'P^{\perp\prime}IP^{\perp}h_0\right\}d\left(P^\perp h_0\right)\cdot\pi_1\left(Ph_0\right)d\left(Ph_0\right) \\ =&\int_{\mathbb{R}^{\underline{k}}}\pi_1\left(Ph_0\right)d\left(Ph_0\right)\cdot\int_{\mathbb{R}^r\setminus\{0\}}P\left(\chi^2_r\left(\delta\right)>\chi^2_{r,1-\alpha}\right)\sqrt{\tfrac{c^r|I|}{(2\pi)^r}}\exp\left\{-\tfrac{c}{2}h_0'P^{\perp\prime}IP^{\perp}h_0\right\}d\left(P^\perp h_0\right) \\ =&\int_{\mathbb{R}^{\underline{k}}}\pi_1\left(Ph_0\right)d\left(Ph_0\right)\cdot\int_{\mathbb{R}^r\setminus\{0\}}P\left(\chi^2_r\left(s'Is\right)>\chi^2_{r,1-\alpha}\right)dN\left(0,\left(cI\right)^{-1}\right)(s), \end{aligned}$$

depends on the priors on Ph and $P^\perp h$ in a separable way. Figure 1 shows the power of this testing procedure when $\alpha=5\%$. When r gets larger, or there are more constraints under H_0, the power envelope is lower. This makes sense because there are more directions to violate the null hypothesis when there are more constraints, which makes rejecting H_0 more difficult.

---

[20] This form of PO_∞ is a little different from the form in Andrews (1994) because c plays the same role as $\frac{1}{c}$ there.

[21] If $r-\chi^2_{r,1-\alpha}\geq0$, then b increases from 0 to 1 as c increases from 0 to ∞. When $r-\chi^2_{r,1-\alpha}<0$, then b increases for $c\in\left(0,\frac{r}{\chi^2_{r,1-\alpha}-r}\right]$, and decreases for $c\in\left(\frac{r}{\chi^2_{r,1-\alpha}-r},\infty\right)$. The maximum of b is greater than 1 and is attained at $c=\frac{r}{\chi^2_{r,1-\alpha}-r}$. $r-\chi^2_{r,1-\alpha}<0$ is equivalent to $P\left(\chi^2_r\leq r\right)<1-\alpha$. But $P\left(\chi^2_r\leq r\right)=\frac{\gamma(r/2,r/2)}{\Gamma(r/2)}$ is a decreasing function with a limit 0.5 as r goes to ∞, where $\gamma\left(a,x\right)=\int_0^x t^{a-1}e^{-t}dt$ is the lower incomplete gamma function, so as long as $\alpha<1-P\left(\chi^2_1\leq1\right)=0.3173$, the second case above happens for all r. In this case, there are two c's such that $PO_\infty\left(\pi\right)=1$ instead of one solution in Table 1 of Andrews (1994).

[22] The asymptotic distribution under the alternative is derived by Le Cam's third lemma in the classical literature, but it is straightforward using the limits of experiments framework here.

[23] If $\pi_2(h)$ is only bowl-shaped, the analysis above does not follow. In this sense, the testing problem is a little different from the estimation problem, and $\pi(h)$ is not completely equivalent to the loss function. For example, let $\pi_2(P^\perp h)=\sqrt{\frac{c^r}{(2\pi)^r}}\exp\left\{-\frac{c}{2}h'P^{\perp\prime}P^\perp h\right\}$; that is, the variance matrix is $(c\mathbf{I})^{-1}$ instead of $(cI)^{-1}$, where $\mathbf{I}$ is a $r\times r$ identity matrix, then it can be shown that $PO_\infty\left(\pi,h_0\right)=\sqrt{\frac{c^r}{|I+c\mathbf{I}|}}\exp\left\{\frac{1}{2}z'_{h_0}P^{\perp\prime}I\left(I+c\mathbf{I}\right)^{-1}IP^\perp z_{h_0}\right\}$. The decision based on $PO_n\left(W_n,\pi\right)$ is different from that based on the likelihood ratio statistic, since c can not be absorbed by b unless $r=1$ or $I=\mathbf{I}$. So for a general prior, the likelihood ratio test need not be optimal in the weighted average power sense. This phenomenon is only loosely parallel to the estimation theory in LAN models, where the MLE and the Bayes estimator are asymptotically equivalent for a large class of loss functions, but not equivalent for every loss function.

In Figure 1, the point-wise power envelope is also drawn. This power envelope is computed against $P^{\perp}h_0 = c$ for each c by the Neyman-Pearson lemma. Notice that

$$\begin{aligned}
&\frac{\sup_{P^{\perp}h=c} Z_\infty(h,h_0)}{\sup_{Ah=0} Z_\infty(h,h_0)} \\
= \quad & \frac{\sup_{P^{\perp}h=c} \exp\left\{-\frac{1}{2}(h-h_0)' P'IP(h-h_0) + (h-h_0)' P'IPz\right\}}{\sup_{Ah=0} \exp\left\{-\frac{1}{2}(h-h_0)' P'IP(h-h_0) + (h-h_0)' P'IPz\right\} \exp\left\{-\frac{1}{2}c'Ic - c'IP^{\perp}z\right\}} \\
= \quad & \frac{\sup_{Ax=0} \exp\left\{-\frac{1}{2}(x-Ph_0)' I(x-Ph_0) + (x-Ph_0)' IPz\right\}}{\sup_{Ax=0} \exp\left\{-\frac{1}{2}(x-Ph_0)' I(x-Ph_0) + (x-Ph_0)' IPz\right\} \exp\left\{-\frac{1}{2}c'Ic - c'IP^{\perp}z\right\}} \\
= \quad & \exp\left\{\frac{1}{2}c'Ic + c'IP^{\perp}z\right\}
\end{aligned}$$

The critical value is determined by

$$P\left(\exp\left\{c'IP^{\perp}z - \frac{1}{2}c'Ic\right\} > b\right) = \alpha,$$

so

$$b = \exp\left\{\sqrt{c'Ic}\, z_{1-\alpha} - \frac{1}{2}c'Ic\right\}.$$

The power envelope at $P^{\perp}h_0 = c$ is

$$P\left(\exp\left\{c'I\left(P^{\perp}z + c\right) - \frac{1}{2}c'Ic\right\} > \exp\left\{\sqrt{c'Ic}\, z_{1-\alpha} - \frac{1}{2}c'Ic\right\}\right) = 1 - \Phi\left(z_{1-\alpha} - \sqrt{c'Ic}\right) = \Phi\left(\sqrt{c'Ic} - z_{1-\alpha}\right)$$

which only depends on $c'Ic$.

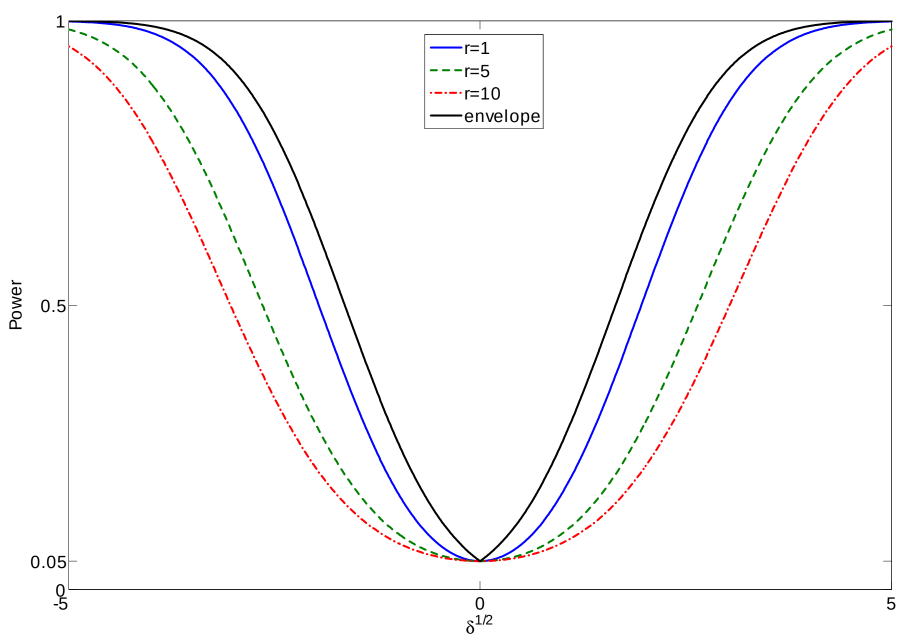

Figure 1: Local Power in LAN Models

To compare with the decision procedure of the Bayes factor proposed in Jeffreys (1961) and detailed in Kass and Raftery (1995), we also copy the critical values of the Bayes factor in Table 2 below. In Table 2, we also report the p values for the critical values. Note that the p values depend on both c and r; that is, depends on the prior and the number of restrictions in H_0. we only report the cases for $c = 0.1, 1, 10$ and $r = 1, \cdots, 9$. From Table 1, the Bayes factor seems conservative in all cases compared to the frequentists with the significance level 5%.

The questions (2) and (3) in Section 2 are answered simultaneously in the above discussion, and no randomization is required. Commonly, the likelihood ratio test is used instead of the posterior odds due to the complexity of the Bayes integral evaluation in the posterior odds. Since the score test and the Wald test are asymptotically equivalent to the likelihood ratio test, they are also optimal in the WAP sense with the above weighting scheme.

The discussion above is summarized in the following Theorem 3.

Theorem 3 *Under Assumptions C0-C4, $PO(W_n, \pi)$ and $\left(\frac{c}{1+c}\right)^{r/2} \exp\left\{\frac{\Lambda_n}{2(1+c)}\right\}$ are equivalent, so Λ_n is AAMP(α, π).*

Because Λ_n is AAMP(α, π), Andrews and Ploberger (1994)'s test indeed maximizes the WAP.

p values (%) ↘ BF ↓ r →	1	2	3	4	5	6	7	8	9	Strength of Evidence against H_0
$c = 0.1$										
<1:1	10.44	7.15	4.78	3.21	2.17	1.47	1.00	0.69	0.47	Negative (supports H_0)
1:1 to 3:1	2.46	2.14	1.60	1.14	0.81	0.57	0.40	0.28	0.19	Barely worth mentioning
3:1 to 10:1	0.55	0.57	0.47	0.36	0.26	0.19	0.14	0.10	0.07	Substantial
10:1 to 30:1	0.15	0.17	0.15	0.12	0.09	0.07	0.05	0.04	0.03	Strong
30:1 to 100:1	0.04	0.05	0.04	0.04	0.03	0.02	0.02	0.01	0.01	Very strong
>100:1	-	-	-	-	-	-	-	-	-	Decisive
$c = 1$										
<1:1	23.90	25.00	24.48	23.58	22.58	21.57	20.60	19.66	18.78	Negative (supports H_0)
1:1 to 3:1	1.62	2.78	3.59	4.15	4.53	4.78	4.95	5.04	5.08	Barely worth mentioning
3:1 to 10:1	0.11	0.25	0.39	0.52	0.65	0.75	0.85	0.93	0.99	Substantial
10:1 to 30:1	0.01	0.03	0.05	0.07	0.10	0.13	0.15	0.18	0.20	Strong
30:1 to 100:1	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.03	0.03	Very strong
>100:1	-	-	-	-	-	-	-	-	-	Decisive
$c = 10$										
<1:1	30.59	35.05	36.98	38.04	38.71	39.15	39.45	39.66	39.81	Negative (supports H_0)
1:1 to 3:1	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	Barely worth mentioning
3:1 to 10:1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Substantial
10:1 to 30:1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Strong
30:1 to 100:1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Very strong
>100:1	-	-	-	-	-	-	-	-	-	Decisive

Table 2: Critical Values and p values of the Bayes Factor in Regular Cases
(p values are for the right end point of the ranges)

4 Tests on the Location of a Threshold Point

In (1), if the hypotheses are (2) in the introduction, then $\vartheta = \gamma$, and $\zeta = \underline{\theta}$ in the general framework of Section 2. $r = 1$, $\underline{k} = 2\ell + 2 + d_\eta$, and $\varphi_n = \begin{pmatrix} \frac{1}{n} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{n}}\mathbf{I}_{\underline{k}} \end{pmatrix}$.

In this model,

$$Z_n(h, h_0) = \prod_{i=1}^{n} \frac{f_h(w_i)}{f_{h_0}(w_i)} = \exp\left(\sum_{i=1}^{n} \ln \frac{f_{e|x,q}(w_i|\theta_0 + \varphi_n h)}{f_{e|x,q}(w_i|\theta_0 + \varphi_n h_0)} \right),$$

where

$$f_{e|x,q}(w|\theta) = \frac{1}{\sigma_1} f_{e|x,q}\left(\frac{y - x'\beta_1}{\sigma_1} |x, q; \eta \right) \mathbf{1}(q \le \gamma) + \frac{1}{\sigma_2} f_{e|x,q}\left(\frac{y - x'\beta_2}{\sigma_2} |x, q; \eta \right) \mathbf{1}(q > \gamma),$$

and $h = (v, u')'$, $h_0 = (v_0, u_0')'$ with v the local parameter for γ and u the local parameters for $\underline{\theta}$.

4.1 Weighted Average Power Envelope

From Yu (2007),

$$Z_\infty(h, h_0) = \exp\left\{ -\frac{1}{2}(u - u_0)' \mathcal{J}(u - u_0) + (u - u_0)' \mathcal{J} W + D_{v_0}(v) \right\}, \tag{10}$$

where $\mathcal{J} = \mathcal{J}(\theta_0)$ is the information matrix for regular parameters at θ_0, $W \sim N\left(0, \mathcal{J}^{-1}\right)$, and

$$D_{v_0}(v) = \begin{cases} \sum_{i=1}^{N_1(|v - v_0|)} z_{1i}, \text{ if } v \le v_0; \\ \sum_{i=1}^{N_2(v - v_0)} z_{2i}, \text{ if } v > v_0; \end{cases}$$

is a compound Poisson process. In $D_{v_0}(v)$,

$$\left\{ \{z_{1i}\}_{i\ge 1}, \{z_{2i}\}_{i\ge 1}, N_1(\cdot), N_2(\cdot) \right\}$$

are independent from each other, $N_i(\cdot)$, $i = 1, 2$, is a Poisson process with intensity $f_q(\gamma_0)$, z_{1i} is the limiting conditional distribution of $\ln \frac{\frac{\sigma_{10}}{\sigma_{20}} f_{e|x,q}\left(\frac{\sigma_{10} e_i + x_i'(\beta_{10} - \beta_{20})}{\sigma_{20}} |x_i, q_i; \eta_0 \right)}{f_{e|x,q}(e_i|x_i, q_i; \eta_0)}$ given $\gamma_0 + \Delta < q_i \le \gamma_0$, $\Delta < 0$ with $\Delta \uparrow 0$, and z_{2i} is the limiting conditional distribution of $\ln \frac{\frac{\sigma_{20}}{\sigma_{10}} f_{e|x,q}\left(\frac{\sigma_{20} e_i - x_i'(\beta_{10} - \beta_{20})}{\sigma_{10}} |x_i, q_i; \eta_0 \right)}{f_{e|x,q}(e_i|x_i, q_i; \eta_0)}$ given $\gamma_0 < q_i \le \gamma_0 + \Delta$, $\Delta > 0$ with $\Delta \downarrow 0$. Furthermore, W and $D_{v_0}(\cdot)$ are independent from each other and the process $D_{v_0}(v)$ is a cadlag process with $D_{v_0}(v_0) = 0$.

Suppose $\pi(h) = \pi_1(u)\pi_2(v)$, where $\pi_1(u)$ is an arbitrary density, and $\pi_2(v)$ puts a point mass at 0 and a continuous density on $\mathbb{R}\backslash\{0\}$. In other words, the priors on the regular and nonregular parameter spaces are independent. Then

$$\begin{aligned} PO_\infty(\pi, h_0) &= \frac{\int_{\mathbb{R}^{\underline{k}}} \exp\left\{ -\frac{1}{2}(u - u_0)' \mathcal{J}(u - u_0) + (u - u_0)' \mathcal{J} W \right\} \pi_1(u) du \cdot \int_{\mathbb{R}\backslash\{0\}} \exp\left\{ D_{v_0}(v) \right\} \pi_2(v) dv}{\int_{\mathbb{R}^{\underline{k}}} \exp\left\{ -\frac{1}{2}(u - u_0)' \mathcal{J}(u - u_0) + (u - u_0)' \mathcal{J} W \right\} \pi_1(u) du \cdot \int_{\{0\}} \exp\left\{ D_{v_0}(v) \right\} \pi_2(v) dv} \\ &= \int_{\mathbb{R}\backslash\{0\}} \exp\left\{ D_{v_0}(v) - D_{v_0}(0) \right\} \pi_2(v) dv \equiv PO_\infty(\pi, v_0) \end{aligned} \tag{11}$$

is pivotal to $u_0 \in \mathbb{R}^{\underline{k}}$ which is the nuisance local parameter under H_0, so condition (iii) of Theorem 1 is satisfied. Notice that

$$PO_\infty(\pi) = \int_{\mathbb{R}\backslash\{0\}} \exp\{D(v)\}\pi_2(v)dv,$$

where $D(v) = D_0(v)$. If b is set such that $P\left(\int_{\mathbb{R}\backslash\{0\}} \exp\{D(v)\}\pi_2(v)dv > b\right) = \alpha$, then any test statistic that is equivalent to $PO(W_n, \pi)$ achieves the best asymptotic WAP with the asymptotic significance level α. Interestingly, $PO_\infty(\pi)$ does not depend on $\pi_1(u)$ asymptotically if $\pi(\cdot)$ is separable as in the regular case. This result is also parallel to the efficient estimation result in Yu (2007), where the asymptotic distribution of the Bayes estimator of γ does not depend on the loss function on the regular parameters as long as the loss function is separable. But $PO_\infty(\pi)$ depends on the whole $\pi_2(\cdot)$. This is different from the regular case in Section 2, where $PO_\infty(\pi)$ depends on $\pi_2(\cdot)$ only through a constant c such that its effect can be absorbed by the selection of b. The best asymptotic WAP

$$\begin{aligned}
& \int_{H_{1\infty}} P_{h_0}^{\infty}\left(PO_\infty(\pi, h_0) > b\right)\pi_1(u_0)\pi_2(v_0)du_0dv_0 \\
= & \int_{H_{1\infty}} P_{v_0}^{\infty}\left(PO_\infty(\pi, v_0) > b\right)\pi_1(u_0)\pi_2(v_0)du_0dv_0 \\
= & \int_{\mathbb{R}^{\underline{k}}} \pi_1(u_0)du_0 \cdot \int_{\mathbb{R}\backslash\{0\}} P_{v_0}^{\infty}\left(PO_\infty(\pi, v_0) > b\right)\pi_2(v_0)dv_0
\end{aligned}$$

depends on $\pi_1(\cdot)$ and $\pi_2(\cdot)$ in a separable way as in the regular case, where $P_{h_0}^{\infty}$ reduces to $P_{v_0}^{\infty}$ in the first equality because $PO_\infty(\pi, h_0)$ is pivotal to u_0. But the value of the WAP depends on the value of $\underline{\theta}_0$, since $D_{v_0}(v)$ depends on $\underline{\theta}_0$. This result is parallel to the estimation result in Yu (2007), where the asymptotic risk of the Bayes estimator depends on the loss function in a separable way when the loss function is separable, and the asymptotic risk of the Bayes estimator of γ depends on $\underline{\theta}_0$.

A natural class of prior functions $\pi_2(\cdot)$ is $\pi_2(v) = \sqrt{\frac{c}{2\pi}}\exp\left\{-\frac{c}{2}v^2\right\}$ on $v \neq 0$; that is, a normal density with mean zero and variance $1/c$. Note that the optimal test is different for different c's. The calculation of asymptotic critical values and the local power function is nontrivial. An algorithm is developed in Appendix D to carry out this kind of calculation.

To aid intuition, power functions for a simple threshold regression model are shown in Figure 2 and 3. The model is specified as

$$y = \begin{cases} \beta_1 + \sigma_1 e, & q \le \gamma; \\ \sigma_2 e, & q > \gamma. \end{cases}, \ q \sim U[0,1], e \sim N(0,1), \text{ and } q \text{ is independent of } e,$$

where β_{10} is set to be 1, and the only unknown parameter is γ. In this simple model,

$$\begin{aligned}
z_{1i} &= \ln\left(\frac{\sigma_{10}}{\sigma_{20}}\right) + \frac{1}{2}\left(e_i^{-2} - \frac{\left(\sigma_{10}e_i^- + \beta_{10}\right)^2}{\sigma_{20}^2}\right), \\
z_{2i} &= \ln\left(\frac{\sigma_{20}}{\sigma_{10}}\right) + \frac{1}{2}\left(e_i^{+2} - \frac{\left(\sigma_{20}e_i^+ - \beta_{10}\right)^2}{\sigma_{10}^2}\right),
\end{aligned}$$

where e_i^- and e_i^+ have the same distribution as e, and $N_i(|v|) \sim Poisson(|v|)$, $i = 1, 2$. Figure 2 illustrates the power functions for $\sigma_{10} = 0.2$, $\sigma_{20} = 0.4$, and $\pi_2(v) = \sqrt{\frac{c}{2\pi}}\exp\left\{-\frac{c}{2}v^2\right\}$ with $c = 0.1, 1$ and 10. For comparison, we also report the case with $\pi_2(v) = 1$. Figure 3 considers the power function for $\sigma_{10} = \sigma_{20} = 0.3$ with the same priors as in Figure 2.

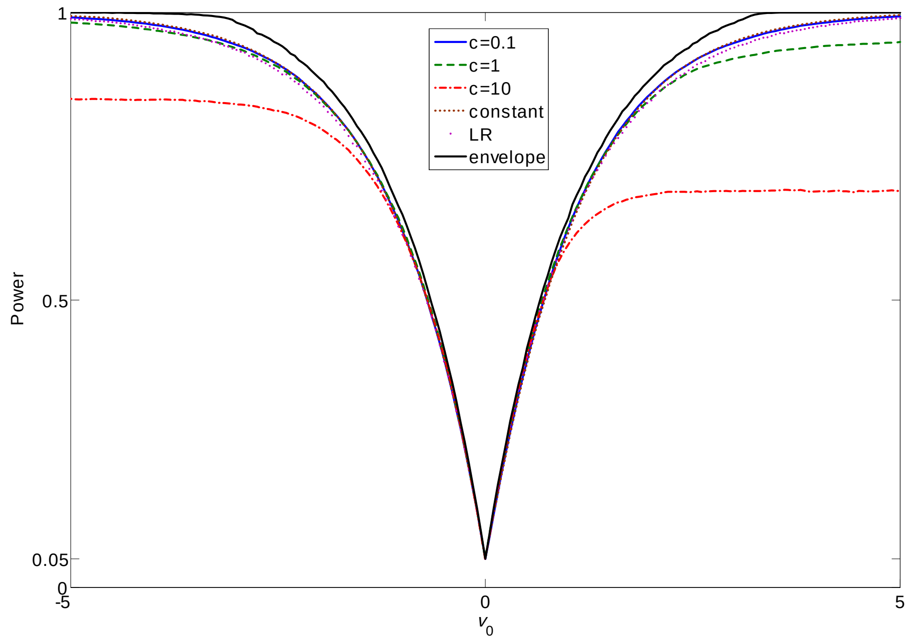

Figure 2: Local Power with $\sigma_{10} = 0.2$, $\sigma_{20} = 0.4$ and Normal Priors

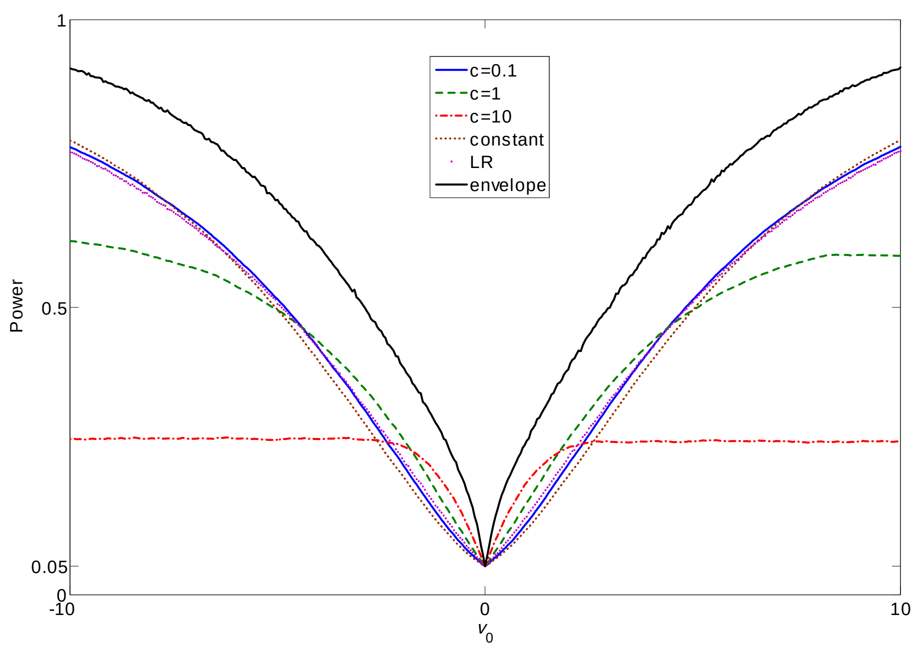

Figure 3: Local Power with $\sigma_{10} = \sigma_{20} = 0.3$ and Normal Priors

From Figure 2, the power functions are different for different c's although the priors are all normally distributed. This is very different from the regular case in Section 3. When c gets larger, more power is gained for v_0 in the neighborhood of 0, since a larger weight is put on the neighborhood of 0. But the power gain in the neighborhood of 0 is not very significant, and the power loss off the neighborhood of 0 is huge, so it seems a smaller c is preferable. Another observation is that the power function is not symmetric. This is mainly because z_{1i} and z_{2i} have different distributions in this setup. When $\sigma_{10} = \sigma_{20}$ as in Figure 3, the power functions are symmetric. Comparing Figure 2 and Figure 3, the threshold effect in variance provides substantial information on γ_0 as shown in the estimation environment of Yu (2007), since the power functions in Figure 2 are much higher than those in Figure 3.[24]

In the calculation of the WAP envelope above, no randomized tests are used. This is true even in extreme cases. Suppose $\sigma_{10} = \sigma_{20} = 0$, then $\exp\{D(v)\}$ is a two-sided Bernoulli process as shown in Yu (2007). In this case,

$$\int_{\mathbb{R}\setminus\{0\}} \exp\{D(v)\} \pi_2(v)dv = \int_{-T_1}^{0} \pi_2(v)dv + \int_{0}^{T_2} \pi_2(v)dv,$$

where T_1 and T_2 are independent standard exponential random variables. Suppose $\pi_2(v) = 1$, then

$$\int_{\mathbb{R}\setminus\{0\}} \exp\{D(v)\} \pi_2(v)dv = T_1 + T_2$$

follows Gamma(2,1) distribution which is continuous, so no randomization is needed.

Figures 2 and 3 also graph the point-wise power envelope as a benchmark. Since we are only interested in the power in the direction of v, fix a v_0, then the limit likelihood ratio under H_0 is

$$\frac{Z_\infty((u, v_0), (u, 0))}{Z_\infty((u, 0), (u, 0))} = \exp\{D(v_0)\},$$

independent of u, and under v_0 is

$$\frac{Z_\infty((u, v_0), (u, v_0))}{Z_\infty((u, 0), (u, v_0))} = \exp\{-D_{v_0}(0)\}.$$

Notice that there is a point mass $e^{-f_q(\gamma_0)|v_0|}$ at 1 in the distribution of $\exp\{D(v_0)\}$, so if

$$P(\exp\{D(v_0)\} > 1) < \alpha \text{ and } P(\exp\{D(v_0)\} \geq 1) > \alpha,$$

then there is no b such that

$$P(\exp\{D(v_0)\} > b) = \alpha.$$

In this case, the randomized test is used:

$$d_\infty = \begin{cases} 1, & \text{if } D(v_0) > 0, \\ \frac{\alpha - P(D(v_0) > 0)}{e^{-f_q(\gamma_0)|v_0|}}, & \text{if } D(v_0) = 0, \\ 0, & \text{if } D(v_0) < 0, \end{cases}$$

[24] The standard deviation 0.3 in Figure 3 is the average of σ_{10} and σ_{20} in Figure 2 to make the two figures comparable, although such a comparison is not rigorous.

and the power envelope at v_0 is

$$\begin{aligned} & P\left(D_{v_0}(0)<0\right)+\frac{\alpha-P\left(D\left(v_0\right)>0\right)}{e^{-f_q(\gamma_0)|v_0|}} P\left(D_{v_0}(0)=0\right) \\ = \; & \alpha+P\left(D_{v_0}(0)<0\right)-P\left(D\left(v_0\right)>0\right). \end{aligned}$$

Otherwise, the critical value is determined by

$$P\left(\exp\left\{D\left(v_0\right)\right\}>b\right)=\alpha,$$

and the power envelope is

$$P\left(\exp\left\{-D_{v_0}(0)\right\}>b\right).$$

When v_0 goes to infinity, $\exp\left\{D\left(v_0\right)\right\}=0$ almost surely, so a purely randomized test is used.

For comparison, the power for the likelihood ratio test is also shown in Figure 2 and 3. In this nonregular case, the Wald and LM tests are not well defined, so the usual trinity that applies in regular cases breaks down. From Section 2,

$$\begin{aligned} \Lambda_\infty\left(h_0\right) &= 2\log\frac{\sup_{h\in H_\infty} Z_\infty(h,h_0)}{\sup_{h\in H_{0\infty}} Z_\infty(h,h_0)} \qquad (12)\\ &= 2\log\frac{\sup_{h\in\mathbb{R}^k}\exp\left\{-\frac{1}{2}\left(u-u_0\right)'\mathcal{J}\left(u-u_0\right)+\left(u-u_0\right)'\mathcal{J}^{-1}W+D_{v_0}(v)\right\}}{\sup_{h\in H_{0\infty}}\exp\left\{-\frac{1}{2}\left(u-u_0\right)'\mathcal{J}\left(u-u_0\right)+\left(u-u_0\right)'\mathcal{J}^{-1}W+D_{v_0}(v)\right\}} \\ &= 2\log\sup_{v\in\mathbb{R}}\exp\left\{D_{v_0}(v)-D_{v_0}(0)\right\} \\ &= 2\sup_{v\in\mathbb{R}}\left\{D_{v_0}(v)-D_{v_0}(0)\right\}, \end{aligned}$$

and

$$\Lambda_\infty=\Lambda_\infty(0)=2\sup_{v\in\mathbb{R}}\left\{D(v)\right\}.$$

Notice that there is a point mass p_{10} at zero in the distribution of Λ_∞, where $p_{10}=P\left(\max D(v)=0\right)$. Such a point mass appears because the value $D(v)$ is fixed as 0 for v in a neighborhood of 0; see Yu (2007) for more discussion about the process $D(v)$. So if $\alpha>1-p_{10}$, then there is no b such that

$$P\left(\sup_{v\in\mathbb{R}}\left\{D(v)\right\}>b\right)=\alpha.$$

In this case, the randomized test is used again:

$$d_\infty=\begin{cases} 1, & \text{if } \Lambda_\infty>0, \\ \frac{\alpha-(1-p_{10})}{p_{10}}, & \text{if } \Lambda_\infty=0. \end{cases}$$

From Appendix D, the power function under h_0 is

$$\begin{aligned} & P\left(\sup_{v\in\mathbb{R}}\left\{D_{v_0}(v)-D_{v_0}(0)\right\}>0\right)+\frac{\alpha-(1-p_{10})}{p_{10}}P\left(\sup_{v\in\mathbb{R}}\left\{D_{v_0}(v)-D_{v_0}(0)\right\}=0\right) \\ = & \begin{cases} 1-(1-\alpha)\left[e^{-f_q(\gamma_0)|v_0|}+\sum_{k=1}^{\infty}\frac{e^{-f_q(\gamma_0)|v_0|}\left(f_q(\gamma_0)|v_0|\right)^k}{k!}\frac{p_{2k}}{p_{10}}\right], & \text{if } v_0<0, \\ 1-(1-\alpha)\left[e^{-f_q(\gamma_0)v_0}+\sum_{k=1}^{\infty}\frac{e^{-f_q(\gamma_0)|v_0|}\left(f_q(\gamma_0)v_0\right)^k}{k!}\frac{p_{1k}}{p_{10}}\right], & \text{if } v_0>0, \end{cases} \end{aligned}$$

where $\{p_{1k}; p_{2k}\}_{k=1}^{\infty}$ are defined in Appendix D. If $\alpha \leq 1 - p_{10}$, then the critical value is determined by

$$P\left(\sup_{v\in\mathbb{R}} \{D(v)\} > b\right) = \alpha,$$

and the power function under h_0 is

$$P\left(\sup_{v\in\mathbb{R}} \{D_{v_0}(v) - D_{v_0}(0)\} > b\right).$$

To further appreciate why randomization is necessary when the likelihood ratio test is used, the extreme case with $\sigma_{10} = \sigma_{20} = 0$ is considered. In this case, $\Lambda_\infty = 0$ almost surely, so randomization must be used to let the type I error be α.

From Figure 2 and 3, the likelihood ratio test is not dominated by the posterior odds with any prior. The powers are closer to the power envelope when $\sigma_{10} = 0.2$, $\sigma_{20} = 0.4$ than the case $\sigma_{10} = \sigma_{20} = 0.3$.

Table 3 reports the p values of the Bayes factor for the case when $\pi_2(\cdot)$ is a density. The Bayes factor is very conservative in the case $\sigma_{10} = 0.2$, $\sigma_{20} = 0.4$, but liberal in the case $\sigma_{10} = \sigma_{20} = 0.3$. Combining with the results in Section 3, it seems that the performance of Bayes factor depends on the prior and nuisance parameters of the model.

In the regular model, when there is no UMP test, some restrictions such as unbiasedness or similarity are put on a test procedure to identify a unique power envelope. In this nonregular case, such restrictions will not identify the unique power envelope. For example, in the simple setups of Figure 2 and 3, all $PO(W_n, \pi)$ with different π's are similar with respect to the nuisance parameter u and unbiased.[25] This is very different from regular cases, where the WAP criterion, unbiasedness and similarity all identify the same power envelope. Furthermore, in regular cases, efficient estimation and optimal testing are essentially the same problem, since the test based on the efficient estimation is optimal in some sense as shown in, for example, chapter 13 of Lehmann and Romano (2005). But the duality between point estimation and hypothesis testing breaks down in this nonregular case. In summary, the WAP criterion is a more natural criterion than the classical ones when no UMP tests exist.

p values (%) $\searrow$	$\sigma_{10} = 0.2$, $\sigma_{20} = 0.4$			$\sigma_{10} = \sigma_{20} = 0.3$			Strength of Evidence
$BF \downarrow$ $c \rightarrow$	0.1	1	10	0.1	1	10	against H_0
<1:1	2.94	4.38	4.61	62.53	57.80	44.05	Negative (supports H_0)
1:1 to 3:1	1.22	1.51	0.92	57.98	50.77	30.90	Barely worth mentioning
3:1 to 10:1	0.38	0.49	0.31	53.67	45.68	26.66	Substantial
10:1 to 30:1	0.10	0.13	0.07	49.92	41.68	23.39	Strong
30:1 to 100:1	0.03	0.02	0.01	45.79	37.35	20.20	Very strong
>100:1	-	-	-	-	-	-	Decisive

Table 3: Critical Values and p values of the Bayes Factor in Tests on the Location of a Threshold Point in Threshold Regression (p values are for the right end point of the ranges)

[25] Another popular restriction is invariance, which is not suitable in this case.

4.2 Feasible Test Statistic and Critical Value

The question left is to find a test statistic equivalent to $PO(W_n, \pi)$. The following test statistic is suggested:

$$T_n\left(\widetilde{\underline{\theta}}\right) = \int_{n(\Gamma-\gamma_0)\backslash\{0\}} \frac{L_n\left(\widetilde{\underline{\theta}}, \gamma_0 + n^{-1}v\right)}{L_n\left(\widetilde{\underline{\theta}}, \gamma_0\right)} \cdot \pi_2(v)dv,$$

where $L_n(\cdot,\cdot)$ is the likelihood function, $\widetilde{\underline{\theta}}$ is a consistent estimator of $\underline{\theta}$ under both H_0 and H_1, e.g., $\widetilde{\underline{\theta}}$ could be the MLE or Bayes estimator (BE) in Yu (2007), and Γ is the parameter space of γ. For simplicity, use T_n for $T_n\left(\widetilde{\underline{\theta}}\right)$. The following theorem shows that T_n is asymptotically equivalent to $PO(W_n, \pi)$.

Theorem 4 *Suppose Assumptions L0-L7 and L9 holds, then T_n and $PO(W_n, \pi)$ are equivalent, so T_n is AAMP($\alpha, \pi, \underline{\theta}$), but not AAMP($\alpha, \pi$).*

For completeness, the asymptotic distribution of Λ_n is stated in the following theorem.

Theorem 5 *Suppose Assumptions L0-L8 holds, then*

$$\Lambda_n(h_0) \overset{h_0}{\rightsquigarrow} \Lambda_\infty(h_0)$$

for any $h_0 \in H_\infty$, where $\Lambda_\infty(h_0)$ is defined in (12).

Compared to $PO_\infty(\pi)$ in regular cases, $PO_\infty(\pi)$ in this section is only pivotal to $h \in H_{0\infty}$, but not to $\theta_0 \in \Theta_0$, which makes the critical values of $PO_\infty(\pi)$ hard to obtain. $PO_\infty(\pi) = \int_{\mathbb{R}\backslash\{0\}} \exp\{D(v)\}\pi_2(v)dv$ depends on the nuisance parameter $\underline{\theta}$ and also on conditional random variables $\{z_{1i}, z_{2i}\}_{i=1}^\infty$ which we don't know how to simulate. The arguments in Section 3.3 of Yu (2008a) can still apply here.

From Yu (2008b), the parametric bootstrap works although the nonparametric bootstrap fails in the estimation problem, so the parametric bootstrap is suggested for finding the asymptotic critical values. Following Hansen (1996), we use the p-value transformation to make the acceptance-rejection decision. Let $F^0(\cdot)$ denote the distribution of $PO_\infty(\pi)$, and define $p_n = 1 - F^0(T_n)$. Tests based on T_n and p_n are equivalent since F^0 is monotonic and continuous. p_n converges weakly to $p^{v_0} = 1 - F^0(PO_\infty(\pi, v_0))$ from Theorem 4. In particular, it converges to a uniform distribution U on $[0,1]$ under H_0. Our test is to reject H_0 if $p_n \leq \alpha$. The asymptotic power function associated with this test is

$$\lim_{n\to\infty} P_{h_0}^n(p_n \leq \alpha) = P_{v_0}^\infty\left(F^0(PO_\infty(\pi, v_0)) \geq 1-\alpha\right).$$

The task remaining is to estimate p_n, and the following Algorithm B is used for this purpose.

Algorithm B:

Step B1: Find a consistent estimator $\widetilde{\underline{\theta}}$ of $\underline{\theta}_0$ under both H_0 and H_1. For example, the joint estimate of $\underline{\theta}$ and γ by maximum likelihood or Bayes estimation can serve this purpose.

Step B2: Simulate $\{w_i\}_{i=1}^n$ from the joint distribution $f\left(y, x, q|\widetilde{\underline{\theta}}, \gamma_0\right)$, where γ_0 is the γ value in H_0.

Step B3: Calculate T_n using the data $\{w_i\}_{i=1}^n$ in Step B2, where $\widetilde{\underline{\theta}}$ in T_n is the consistent estimator in Step B1.

Step B4: Repeat Step B2 and B3 J times to get $\left\{T_n^1, \cdots, T_n^J\right\}$, and estimate p_n by $\widehat{p}_n \equiv \frac{1}{J}\sum_{j=1}^{J} \mathbf{1}\left(T_n^j \geq T_n\right)$.

The following theorem claims that $\widehat{p}_n$ is a valid estimation of p_n asymptotically.

Theorem 6 *Suppose Assumptions L0-L7 and L9 holds, then*

$$\widehat{p}_n = p_n + o_p(1).$$

Hence,

$$\widehat{p}_n \overset{h_0}{\rightsquigarrow} p^{v_0},$$

and the asymptotic distribution of $\widehat{p}_n$ is U under H_0.

In the example of the last subsection, suppose β_{10} is unknown and is the only nuisance parameter. The true γ_0 is equal to 0.5. Table 4 reports the finite-sample size using Algorithm B. When calculating the size,

$$H_0 : \gamma = 0.5.$$

$\underline{\theta}$ is estimated by maximum likelihood method in Step B1 to save simulation time, although Yu (2007) shows the Bayes method has some efficiency benefit. Suppose the empirical distribution of $\left\{T_n^1, \cdots, T_n^J\right\}$ in Step B4 is $\widehat{F}_n$, then the size of the test is estimated as $P\left(1 - \widehat{F}_n\left(T_n\right) \leq \alpha\right)$, where the γ_0 in T_n and Step B2 is 0.5. Note that the asymptotic size is always α and is not reported in Table 4. In all simulations, 1000 simulated samples were drawn, and $J = 1000$. From Table 4, the parametric bootstrap works very well even when the prior is improper.

$n \rightarrow$	100				400			
α (%)$\rightarrow$	10	5	2.5	1	10	5	2.5	1
Size ↘	$\sigma_{10} = 0.2$, $\sigma_{20} = 0.4$							
$c = 0.1$	10.8	5.9	3.7	1.1	10.7	5.1	2.2	1.1
$c = 1$	10.7	5.3	3.1	1.1	9.9	5.3	2.8	1.2
$c = 10$	10.7	4.9	2.4	1.4	9.6	5.5	3.1	1.7
Constant	11.7	6.0	3.5	1.1	11.4	5.2	2.2	1.0
Size ↘	$\sigma_{10} = \sigma_{20} = 0.3$							
$c = 0.1$	9.1	4.4	2.2	1.2	9.6	5.5	3.0	1.4
$c = 1$	9.8	5.3	2.6	1.4	9.2	4.5	2.8	1.7
$c = 10$	9.5	4.3	2.8	1.2	9.5	5.3	3.1	1.3
Constant	10.2	4.8	2.4	1.2	10.2	5.6	3.4	1.4

Table 4: Size Using the Parametric Bootstrap (Based on 1000 Repetitions)

The power using the parametric bootstrap above is reported in Table 5. In this case,

$$H_0 : \gamma = 0.49.$$

When calculating power, the same formula $P\left(1 - \widehat{F}_n\left(T_n\right) \leq \alpha\right)$ is used but the γ_0 in T_n and Step B2 is replaced by 0.49. The local parameter v_0 is $n\left(0.5 - 0.49\right) = 1$ when $n = 100$, and is 4 when $n = 400$, so the asymptotic power is also reported for comparison. As expected, the power increases as the type I error increases and the sample size increases. When there is a threshold effect in variance, the finite-sample power is close to the asymptotic power. When there is no threshold effect in variance, the finite-sample power is comparable with the case with threshold effect in variance, and much higher than the asymptotic power. This suggests that there is significant finite-sample refinement using parametric bootstrap when the identification of the threshold point is weak.

Based on these simulations, the parametric bootstrap is recommended for practical computation of the critical values.

$n \to$	100				400			
α (%)$\to$	10	5	2.5	1	10	5	2.5	1
Power (%) $\searrow$	$\sigma_{10} = 0.2$, $\sigma_{20} = 0.4$							
$c = 0.1$	70.1	64.1	58.1	48.9	99.1	97.6	96.6	94.1
$c = 1$	71.3	65.3	60.7	52.5	94.1	93.0	88.1	83.0
$c = 10$	67.4	61.4	52.3	41.5	70.7	67.4	55.1	42.1
Constant	69.5	63.2	57.4	47.3	99.5	98.1	96.7	95.5
Asymptotic	$v_0 = 1$				$v_0 = 4$			
$c = 0.1$	67.4	62.5	58.2	49.5	99.1	98.1	97.1	95.1
$c = 1$	68.0	63.0	59.4	52.1	95.3	94.0	90.2	84.9
$c = 10$	64.9	59.6	53.1	41.7	71.6	69.0	55.3	43.6
Constant	66.7	61.9	57.6	48.3	99.5	98.4	97.4	96.0
Power (%) $\searrow$	$\sigma_{10} = \sigma_{20} = 0.3$							
$c = 0.1$	67.0	62.9	58.8	55.3	98.7	97.6	97.0	96.0
$c = 1$	68.1	63.1	59.9	55.9	95.8	94.6	93.4	91.8
$c = 10$	66.0	61.0	56.9	50.9	75.3	72.9	63.8	56.8
Constant	66.6	61.6	58.4	54.8	98.9	98.0	97.2	96.4
Asymptotic	$v_0 = 1$				$v_0 = 4$			
$c = 0.1$	21.7	12.8	7.3	3.4	55.2	42.8	32.3	21.6
$c = 1$	25.2	16.0	10.0	5.5	55.1	44.1	34.8	25.4
$c = 10$	28.2	19.2	13.2	8.0	36.4	26.6	19.7	13.1
Constant	20.5	11.6	6.3	2.7	53.8	40.6	29.6	18.2

Table 5: Power Using the Parametric Bootstrap (Based on 1000 Repetitions)

5 Conclusion

This paper proposes a unified framework for asymptotically optimal tests. The general framework is applied to two specific tests. The first is a classical test in LAN models, where the posterior odds and the likelihood ratio are asymptotically equivalent and both are optimal in the WAP sense. The second is a test about the location of a threshold point. In this test, the likelihood ratio test is not optimal in the WAP sense, and the optimal test is based on the posterior odds which depends on the prior on the local parameter space and is not unique.

References

Andrews, D.W.K., 1994, The Large Sample Correspondence Between Classical Hypothesis Tests and Bayesian Posterior Odds Tests, *Econometrica*, 62, 1207-1232.

Andrews, D.W.K., 2001, Testing When a Parameter is on the Boundary of the Maintained Hypothesis, *Econometrica*, 69, 683-734.

Andrews, D.W.K. and W. Ploberger, 1994, Optimal Tests when a Nuisance Parameter is Present Only Under an Alternative, *Econometrica*, 62, 1383-1414.

Bahadur, R.R., 1967, An Optimal Property of the Likelihood Ratio Statistic, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (1965/66) I, 13-26, University of California Press, Berkeley.

Beran, R., 1997, Diagnosing Bootstrap Success, *Annals of the Institute of Statistical Mathematics*, 49, 1-24.

Berger, J.O., 1985, *Statistical Decision Theory and Bayesian Analysis*, 2nd edition, New York: Springer-Verlag.

Chamberlain, G., 2007, Decision Theory Applied to an Instrumental Variables Model, *Econometrica*, 75, 609-652.

Chernoff, H., 1954, On the Distribution of the Likelihood Ratio Statistic, *Annals of Mathematical Statistics*, 25, 573-578.

Choi, S., Hall, W. and Schick, A., 1996, Asymptotically Uniformly Most Powerful Tests in Parametric and Semiparametric Models, *Annals of Statistics*, 24, 841-861.

Hansen, B.E., 1996, Inference when a Nuisance Parameter is not Identified under the Null Hypothesis, *Econometrica*, 413-430.

Hansen, B.E., 2000, Sample Splitting and Threshold Estimation, *Econometrica*, 575-603.

Ibragimov, I. and R. Has'minskii, 1981, *Statistical Estimation: Asymptotic Theory*, New York : Springer-Verlag.

Jeffreys, H., 1961, *Theory of Probability*, 3rd edition, Oxford: Oxforx University Press.

Kass, R.E. and A.E. Raftery, 1995, Bayes Factors, *Journal of the American Statistical Association*, 90, 773-795.

Le Cam, L., 1972, Limits of Experiments, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 245-261.

Lee, S. and M.H. Seo, 2008, Semiparametric Estimation of a Binary Response Model with a Change-point Due to a Covariate Threshold, *Journal of Econometrics*, 144, 492-499.

Lehmann, E.L. and J.P. Romano, 2005, *Testing Statistical Hypotheses*, 3rd edition, New York: Springer Verlag.

Newey, W.K. and D.L. McFadden, 1994, Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Vol. 4, R.F. Eagle and D.L. McFadden, eds., Elsevier Science B.V., Ch. 36, 2113-2245.

Oosterhoff, J. and van Zwet, W., 1979, A Note on Contiguity and Hellinger Distance, *Contributions to Statistics*, Reidel, Dordrecht-Boston, Mass. London, 157-166.

Van der Vaart, A.W., 1991, An Asymptotic Representation Theorem, *International Statistical Review*, 59, 97-121.

Van der Vaart, A.W., 1996, *Limits of Experiments*, unpublished manuscript.

Van der Vaart, A.W., 1998, *Asymptotic Statistics*, New York : Cambridge University Press.

Wald, A., 1943, Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large, *Transactions of the American Mathematical Society*, 54, 426-482.

Yu, P., 2007, Likelihood-based Estimation and Inference in Threshold Regression, unpublished manuscript, Department of Economics, University of Wisconsin at Madison.

Yu, P., 2008a, Semiparametric Efficient Estimation in Threshold Regression, unpublished manuscript, Department of Economics, University of Wisconsin at Madison.

Yu, P., 2008b, Bootstrap in Threshold Regression, unpublished manuscript, Department of Economics, University of Wisconsin at Madison.

Appendix A: Regularity Conditions

Tests in LAN Models

First, some notations are collected for reference in all assumptions, lemmas and proofs under the title "Tests in LAN Models".

$$\begin{aligned}
S(w|\theta) &= \frac{\partial \ln f(w|\theta)}{\partial\theta} \text{ is the score function,} \\
I(\theta) &= E\left[S(w|\theta)\, S(w|\theta)'\right] \text{ is the information matrix, } I = I(\theta_0), \\
\overline{z} &= I^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S(w_i|\theta_0), \\
LR_n(S,h,h_0) &= -\frac{1}{2}(h-h_0)'\, I\,(h-h_0) + (h-h_0)'\, I\overline{z}, \\
LR_n(S,h) &= LR_n(S,h,0)
\end{aligned}$$

Assumption C0: θ_0 is an interior point of Θ.

Assumption C1: w has a density $f(\cdot|\theta)$ which is continuously differentiable in a neighborhood of θ_0 for almost every w.

Assumption C2: $I(\theta)$ is continuous and nonsingular at θ_0.

Assumption C3: For every θ_1 and θ_2 in a neighborhood of θ_0, there exists a measurable function $m(w)$ such that $E_{\theta_0}\left[m(w)^2\right] < \infty$,

$$|\ln f(w|\theta_1) - \ln f(w|\theta_2)| \le m(w)\, \|\theta_1 - \theta_2\|.$$

Assumption C4: $\widehat{\theta}_{n,0}$ and $\widehat{\theta}_n$ are consistent under θ_0.

Remark 1 *All assumptions are standard but weaker than those in Andrews (1994) since no twice continuous differentiability is assumed. For example, the Laplace density is covered in these assumptions but excluded by Andrews (1994). The m function in Assumption C3 is called the* slope function *for the Lipschitz function* $\ln f(w|\cdot)$*. Assumption C4 can be verified by, for example, Theorem 2.5 of Newey and McFadden (1994). Only local Lipschitz condition is required in Assumption C3, which is because C4 has already constrained θ in a neighborhood of θ_0. Without C4, a global Lipschitz condition is usually required to prove the consistency.*

Tests on the Location of a Threshold Point

First, some notations are collected for reference in all assumptions, lemmas and proofs under the title "Tests on the Location of a Threshold Point".

$$H_{vn} = n(\Gamma - \gamma_0) \setminus \{0\},$$

$$\begin{aligned}
Z_n^d(v, v_0|\theta) = \exp&\left\{\sum_{i=1}^{n} \overline{z}_{1i}(\underline{\theta})\, \mathbf{1}\left(\gamma + \frac{v}{n} < q_i \le \gamma + \frac{v_0}{n}\right)\right\} \\
&+ \exp\left\{\sum_{i=1}^{n} \overline{z}_{2i}(\underline{\theta})\, \mathbf{1}\left(\gamma + \frac{v_0}{n} < q_i \le \gamma + \frac{v}{n}\right)\right\}
\end{aligned}$$

with

$$\overline{z}_{1i}(\underline{\theta}) = \overline{z}_1(w_i|\underline{\theta}),\ \overline{z}_1(w|\underline{\theta}) = \ln \frac{\frac{\sigma_1}{\sigma_2} f_{e|x,q}\left(\frac{\sigma_1 e + x'(\beta_1-\beta_2)}{\sigma_2}|x,q;\eta\right)}{f_{e|x,q}(e|x,q;\eta)},\ \overline{z}_{1i} = \overline{z}_1(w_i|\underline{\theta}_0),$$

$$\overline{z}_{2i}(\underline{\theta}) = \overline{z}_2(w_i|\underline{\theta}),\ \overline{z}_2(w|\underline{\theta}) = \ln \frac{\frac{\sigma_2}{\sigma_1} f_{e|x,q}\left(\frac{\sigma_2 e - x'(\beta_1-\beta_2)}{\sigma_1}|x,q;\eta\right)}{f_{e|x,q}(e|x,q;\eta)},\ \overline{z}_{2i} = \overline{z}_2(w_i|\underline{\theta}_0),$$

$$Z_n^d(v|\cdot) = Z_n^d(v,0|\cdot),\ Z_n^d(v) = Z_n^d(v,0|\theta_0)$$

$$S(w|\theta) = \begin{pmatrix} -\frac{\partial \ln f_{e|x,q}}{\partial e}(e|x,q;\eta)\frac{x'}{\sigma_1}\mathbf{1}(q\le\gamma) \\ -\frac{\partial \ln f_{e|x,q}}{\partial e}(e|x,q;\eta)\frac{x'}{\sigma_2}\mathbf{1}(q>\gamma) \\ -\frac{1}{\sigma_1}\left(1+\frac{\partial \ln f_{e|x,q}}{\partial e}(e|x,q;\eta)e\right)\mathbf{1}(q\le\gamma) \\ -\frac{1}{\sigma_2}\left(1+\frac{\partial \ln f_{e|x,q}}{\partial e}(e|x,q;\eta)e\right)\mathbf{1}(q>\gamma) \\ \frac{\partial \ln f_{e|x,q}}{\partial \eta}(e|x,q;\eta) \end{pmatrix} \equiv \begin{pmatrix} S_{\beta_1}(\theta) \\ S_{\beta_2}(\theta) \\ S_{\sigma_1}(\theta) \\ S_{\sigma_2}(\theta) \\ S_{\eta}(\theta) \end{pmatrix}$$

is the score function for $\underline{\theta}$

$$\mathcal{J}(\underline{\theta},\gamma_0) \equiv E\left[S(w|\underline{\theta},\gamma_0)\,S'(w|\underline{\theta},\gamma_0)\right]$$

is the information matrix for regular parameters evaluated at $\underline{\theta}$, and

$$\mathcal{J} = \mathcal{J}(\theta_0)$$

$$\overline{z} = \frac{\mathcal{J}^{-1}}{\sqrt{n}}\sum_{i=1}^{n} S(w_i|\theta_0)$$

$$LR_n\left(S, Z_n^d, h, h_0\right) = -\frac{1}{2}(u-u_0)'\mathcal{J}(u-u_0) + (u-u_0)'\mathcal{J}\overline{z} + \ln Z_n^d(v,v_0|\theta_0)$$

and

$$LR_n\left(S, Z_n^d, h\right) = LR_n\left(S, Z_n^d, h, 0\right)$$

Assumption L0: θ_0 is an interior point of Θ which is bounded. $\sigma_{10} \neq \sigma_{20}$, $\beta_{10} \neq \beta_{20}$, σ_{10} and σ_{20} are bounded away from zero.

Assumption L1: (x,q) has a marginal density $f_{x,q}$, and e has a conditional density $f_{e|x,q}(e|x,q;\eta)$ which is continuously differentiable in both e and η for η in a neighborhood of η_0 and for almost every (e,x,q).

Assumption L2: $\mathcal{J}(\underline{\theta},\gamma_0)$ is continuous, nonsingular and finite for $\underline{\theta}$ in an open neighborhood of $\underline{\theta}_0$.

Assumption L3: For every (μ_1,σ_1,η_1) and (μ_2,σ_2,η_2) with μ_1,σ_1, μ_2,σ_2 in a bounded set and η_1 and η_2 in a neighborhood of η_0, there exists a slope function $m(w)$ such that $E_{\theta_0}\left[m(w)^4\right] < \infty$,

$$\left|\ln f_{e|x,q}(\mu_1+\sigma_1 e|x,q;\eta_1) - \ln f_{e|x,q}(\mu_2+\sigma_2 e|x,q;\eta_2)\right| \le m(w)\left(|\mu_1-\mu_2| + |\sigma_1-\sigma_2| + \|\eta_1-\eta_2\|\right).$$

Assumption L4: $E[\|x\|^4] < \infty$.

Assumption L5: $f_q(\cdot)$ is continuous, and $0 < \underline{f}_q \le f_q(q) \le \overline{f}_q < \infty$ for q in a neighborhood of γ_0.

Assumption L6:

$$
\begin{aligned}
E\left[\sup_{\underline{\theta}\in\mathcal{N}}\left|\overline{z}_1\left(w|\underline{\theta}\right)\right|\right] &< \infty, \\
E\left[\sup_{\underline{\theta}\in\mathcal{N}}\left|\overline{z}_2\left(w|\underline{\theta}\right)\right|\right] &< \infty.
\end{aligned}
$$

where $\mathcal{N}$ an open neighborhood of $\underline{\theta}_0$.

Assumption L7: Both z_{1i} and z_{2i} have absolutely continuous distributions.

Assumption L8: $\widehat{\theta}_n$ is consistent under θ_0.

Assumption L9: $\widetilde{\underline{\theta}}$ is consistent under both H_0 and H_1.

Remark 2 *These assumptions are weaker than those in Yu (2007). The assumptions on the regular parameters are basically the same as those in LAN models. $LR_n\left(S, Z_n^d, h\right)$ is an approximation of the log likelihood ratio statistic. S is the asymptotically sufficient statistic for regular parameters, and Z_n^d is the asymptotically sufficient statistic for the nonregular parameter. Checking of Assumption L8 and L9 can be found in Yu (2007).*

In all proofs and lemmas related to a general h_0, only the case for $h_0 = 0$ is proved. The general case only complicates notations without changing the essential idea.

Appendix B: Proofs

Proof of Theorem 1:. From the discussion in the main context,

$$
\begin{aligned}
& \overline{\lim_{n\to\infty}}\, b\int_{H_{0n}} P_h^n\left(d_n = 1\right)\pi(h)dh + \int_{H_{1n}}\left(1 - P_h^n\left(d_n = 1\right)\right)\pi(h)dh \qquad (13) \\
\leq\ & \underline{\lim}_{n\to\infty}\, b\int_{H_{0n}} P_h^n\left(\xi_n = 1\right)\pi(h)dh + \int_{H_{1n}}\left(1 - P_h^n\left(\xi_n = 1\right)\right)\pi(h)dh
\end{aligned}
$$

for any test ξ_n. From Assumptions (i) and (ii), $T_n \overset{h_0}{\rightsquigarrow} PO_\infty\left(\pi, h_0\right)$ for any $h_0 \in H_\infty$, so the critical value b and the the asymptotic power are valid. The left hand side (lhs) of (13) is

$$
\begin{aligned}
& \overline{\lim_{n\to\infty}}\, b\int_{H_{0n}} P_h^n\left(d_n = 1\right)\pi(h)dh + \int_{H_{1n}}\left(1 - P_h^n\left(d_n = 1\right)\right)\pi(h)dh \\
=\ & \overline{\lim_{n\to\infty}}\, b\int_{H_{0n}} P_h^n\left(T_n > b\right)\pi(h)dh + \int_{H_{1n}}\left(1 - P_h^n\left(T_n > b\right)\right)\pi(h)dh \\
=\ & b\int_{H_{0\infty}} \alpha\pi(h)dh + \overline{\lim_{n\to\infty}}\int_{H_{1n}}\left(1 - P_h^n\left(T_n > b\right)\right)\pi(h)dh
\end{aligned}
$$

where the first equality is from the definition of d_n, and the second equality is from Fubini's theorem. The right hand side (rhs) is

$$
\begin{aligned}
& \underline{\lim}_{n\to\infty}\, b\int_{H_{0n}} P_h^n\left(\xi_n = 1\right)\pi(h)dh + \int_{H_{1n}}\left(1 - P_h^n\left(\xi_n = 1\right)\right)\pi(h)dh \\
\leq\ & b\int_{H_{0\infty}} \alpha\pi(h)dh + \underline{\lim}_{n\to\infty}\int_{H_{1n}}\left(1 - P_h^n\left(\xi_n = 1\right)\right)\pi(h)dh,
\end{aligned}
$$

where the inequality is from Fubini's theorem and the assumption that ξ_n is of asymptotic level α. Combining the above two formula with (13), we have

$$\overline{\lim_{n\to\infty}} \int_{H_{1n}} \left(1 - P_h^n\left(T_n > b\right)\right) \pi(h) dh \le \underline{\lim_{n\to\infty}} \int_{H_{1n}} \left(1 - P_h^n\left(\xi_n = 1\right)\right) \pi(h) dh, \tag{14}$$

which reduces to the result we want. Actually, $\overline{\lim}_{n\to\infty}$ on the lhs of (14) is $\lim_{n\to\infty}$ by Fubini's theorem and the continuity of the asymptotic distribution of T_n. ■

Remark 3 *The proof critically depends on the continuity of the distribution of $PO_\infty(\pi)$. If $PO_\infty(\pi)$ is not continuously distributed, $E_h^n[d_n]$ does not necessarily converge to α for $h \in H_{0n}$ if b and c are defined in (7).*

Proof of Theorem 2:. Since both $\widehat{\theta}_{n,0}$ and $\widehat{\theta}_n$ are φ_n-consistent, for any $\varepsilon > 0$, there exists M which may depend on ε such that

$$P\left(\left\|\varphi_n^{-1}\widehat{\theta}_{n,0}\right\| > M\right) < \varepsilon, \text{ and } P\left(\left\|\varphi_n^{-1}\widehat{\theta}_n\right\| > M\right) < \varepsilon.$$

Fix arbitrary nonnegative number c such that $P(\Lambda_\infty = c) = 0$,

$$\begin{aligned}
&P(\Lambda_n \le c) \\
&\overset{(1)}{=} P\left(2\sup\nolimits_{h\in H_n} \log Z_n(h) - 2\sup\nolimits_{h\in H_{0n}} \log Z_n(h) \le c\right) \\
&\overset{(2)}{=} P\left(2\sup\nolimits_{H_n\cap\|h\|\le M} \log Z_n(h) - 2\sup\nolimits_{H_{0n}\cap\|h\|\le M} \log Z_n(h) \le c\right) + \epsilon \\
&\overset{(3)}{=} P\left(2\sup\nolimits_{H_n\cap\|h\|\le M} \log Z_n^a(h) - 2\sup\nolimits_{H_{0n}\cap\|h\|\le M} \log Z_n^a(h) + o_p(1) \le c\right) + \epsilon \\
&\overset{(4)}{\longrightarrow} P\left(2\sup\nolimits_{H_\infty\cap\|h\|\le M} \log Z_\infty(h) - 2\sup\nolimits_{H_{0\infty}\cap\|h\|\le M} \log Z_\infty(h) \le c\right) + \epsilon
\end{aligned}$$

where (1) is from the definition of Λ_n, (2) follows by breaking the whole sample space into $\left\{\left\|\varphi_n^{-1}\widehat{\theta}_{n,0}\right\| \le M, \left\|\varphi_n^{-1}\widehat{\theta}_n\right\| \le M\right\}$ and its complement, ϵ is some number no larger than 2ε, (3) is from Assumption (ii) with $Z_n^a(h) = Z_n^a(h, 0)$, and (4) holds because of the continuous mapping theorem and Assumption (iii). Note also that $H_n \cap \|h\| \le M$ converges to $H_\infty \cap \|h\| \le M$, and $H_{0n} \cap \|h\| \le M$ converges to $H_{0\infty} \cap \|h\| \le M$.

From Assumption (iv),

$$P\left(\left|\sup_{H_\infty\cap\|h\|\le M} \log Z_\infty(h) - \sup_{H_\infty} \log Z_\infty(h)\right| > \varepsilon\right) < \varepsilon,$$

and

$$P\left(\left|\sup_{H_{0\infty}\cap\|h\|\le M} \log Z_\infty(h) - \sup_{H_{0\infty}} \log Z_\infty(h)\right| > \varepsilon\right) < \varepsilon$$

by taking ε small and M large. Since c is a continuous point on the cdf of Λ_∞,

$$P\left(2\sup_{H_\infty\cap\|h\|\le M} \log Z_\infty(h) - 2\sup_{H_{0\infty}\cap\|h\|\le M} \log Z_\infty(h) \le c\right) - P(\Lambda_\infty \le c) < \varepsilon.$$

In the above argument, ε can be arbitrarily small, so

$$P(\Lambda_n \le c) \to P(\Lambda_\infty \le c).$$

■

Proof of Theorem 3:. From Lemma 2,

$$PO\left(W_n,\pi,h_0\right)-\frac{\int_{H_{1\infty}}\exp\left\{LR_n\left(S,h,h_0\right)\right\}\pi(h)dh}{\int_{H_{0\infty}}\exp\left\{LR\left(S,h,h_0\right)\right\}\pi(h)dh}\xrightarrow{p}0.$$

By a similar argument as in the main context,

$$\frac{\int_{H_{1\infty}}\exp\left\{LR_n\left(S,h,h_0\right)\right\}\pi(h)dh}{\int_{H_{0\infty}}\exp\left\{LR\left(S,h,h_0\right)\right\}\pi(h)dh}=\left(\frac{c}{1+c}\right)^{r/2}\exp\left\{\frac{\left(\overline{z}+h_0\right)'P^{\perp\prime}IP^{\perp}\left(\overline{z}+h_0\right)}{2\left(1+c\right)}\right\}.$$

From Lemma 3,

$$\Lambda_n\left(h_0\right)-\left(\overline{z}+h_0\right)'P^{\perp\prime}IP^{\perp}\left(\overline{z}+h_0\right)\xrightarrow{p}0.$$

By the continuous mapping theorem, the result holds. ■

Proof of Theorem 4:. We only prove $PO\left(W_n,\pi\right)-T_n\xrightarrow{p}0$ under θ_0 since all other models indexed by h are contiguous to the model under θ_0 from Lemma 7. From the proof of Lemma 8,

$$PO\left(W_n,\pi\right)-\int_{0<\|v\|\leq M}Z_n^d\left(v\right)\pi_2(v)dh=o_p(1)$$

for M large enough. When n is large enough,

$$\begin{aligned}T_n &= \int_{H_{vn}}\frac{L_n\left(\underline{\widetilde{\theta}},\gamma_0+n^{-1}v\right)}{L_n\left(\underline{\widetilde{\theta}},\gamma_0\right)}\cdot\pi_2(v)dv\\ &= \int_{0<\|v\|\leq M}\frac{L_n\left(\underline{\widetilde{\theta}},\gamma_0+n^{-1}v\right)}{L_n\left(\underline{\widetilde{\theta}},\gamma_0\right)}\cdot\pi_2(v)dv+\int_{H_{vn}\cap\|v\|>M}\frac{L_n\left(\underline{\widetilde{\theta}},\gamma_0+n^{-1}v\right)}{L_n\left(\underline{\widetilde{\theta}},\gamma_0\right)}\cdot\pi_2(v)dv.\end{aligned}$$

So it is sufficient to prove that (i) for any ε, there is $M>0$ which may depend on ε such that $P\left(\int_{H_{vn}\cap\|v\|>M}\frac{L_n\left(\underline{\widetilde{\theta}},\gamma_0+n^{-1}v\right)}{L_n\left(\underline{\widetilde{\theta}},\gamma_0\right)}\cdot\pi_2(v)dv>\varepsilon\right)<\varepsilon$; (ii) $\int_{0<\|v\|\leq M}Z_n^d\left(v\right)\pi_2(v)dh-\int_{0<\|v\|\leq M}\frac{L_n\left(\underline{\widetilde{\theta}},\gamma_0+n^{-1}v\right)}{L_n\left(\underline{\widetilde{\theta}},\gamma_0\right)}\cdot\pi_2(v)dv\xrightarrow{p}0$ for any $M>0$.

We show (i) first.

$$\begin{aligned}&\frac{L_n\left(\underline{\widetilde{\theta}},\gamma_0+n^{-1}v\right)}{L_n\left(\underline{\widetilde{\theta}},\gamma_0\right)}=Z_n^d\left(v|\underline{\widetilde{\theta}},\gamma_0\right)\\ =&\ \exp\left\{\sum_{i=1}^n\overline{z}_{1i}\left(\underline{\widetilde{\theta}}\right)\mathbf{1}\left(\gamma_0+\frac{v}{n}<q_i\leq\gamma_0\right)+\sum_{i=1}^n\overline{z}_{2i}\left(\underline{\widetilde{\theta}}\right)\mathbf{1}\left(\gamma_0<q_i\leq\gamma_0+\frac{v}{n}\right)\right\}\end{aligned}$$

By Assumption L6 and the strict Jensen's inequality,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\overline{z}_{1i}\left(\underline{\widetilde{\theta}}\right)\xrightarrow{p}E\left[\overline{z}_{1i}\right] &< 0,\\ \frac{1}{n}\sum_{i=1}^n\overline{z}_{2i}\left(\underline{\widetilde{\theta}}\right)\xrightarrow{p}E\left[\overline{z}_{2i}\right] &< 0,\end{aligned}$$

so $\sum_{i=1}^{n} \overline{z}_{1i}\left(\widetilde{\underline{\theta}}\right) \mathbf{1}\left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0\right) + \sum_{i=1}^{n} \overline{z}_{2i}\left(\widetilde{\underline{\theta}}\right) \mathbf{1}\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n}\right) = O_p(1)$ even when v is large. For any $\varepsilon > 0$,

$$
\begin{aligned}
& P\left(\int_{H_{vn}\cap\|v\|>M} \frac{L_n\left(\widetilde{\underline{\theta}}, \gamma_0 + n^{-1}v\right)}{L_n\left(\widetilde{\underline{\theta}}, \gamma_0\right)} \cdot \pi_2(v)dv > \varepsilon\right) \\
= \; & P\left(O_p(1) \cdot \int_{H_{vn}\cap\|v\|>M} \pi_2(v)dv > \varepsilon\right) \\
= \; & P\left(O_p(1) > \varepsilon / \int_{H_{vn}\cap\|v\|>M} \pi_2(v)dv\right).
\end{aligned}
$$

The rhs can be made arbitrarily small by taking M large.

We now prove (ii). Note that for v on any compact set,

$$
\begin{aligned}
& \frac{L_n\left(\widetilde{\underline{\theta}}, \gamma_0 + n^{-1}v\right)}{L_n\left(\widetilde{\underline{\theta}}, \gamma_0\right)} \\
= \; & \exp\left\{\sum\nolimits_{i=1}^{n} \overline{z}_{1i}\left(\underline{\theta}_0\right) \mathbf{1}\left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0\right) + \sum\nolimits_{i=1}^{n} \left(\overline{z}_{1i}\left(\widetilde{\underline{\theta}}\right) - \overline{z}_{1i}\left(\underline{\theta}_0\right)\right) \mathbf{1}\left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0\right)\right\} \\
& \exp\left\{\sum\nolimits_{i=1}^{n} \overline{z}_{2i}\left(\underline{\theta}_0\right) \mathbf{1}\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n}\right) + \sum\nolimits_{i=1}^{n} \left(\overline{z}_{2i}\left(\widetilde{\underline{\theta}}\right) - \overline{z}_{2i}\left(\underline{\theta}_0\right)\right) \mathbf{1}\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n}\right)\right\} \\
= \; & Z_n^d(v) \exp\left\{o_p(1)\right\} \\
= \; & Z_n^d(v) + o_p(1),
\end{aligned}
$$

where the first equality is from breaking $\overline{z}_{1i}\left(\widetilde{\underline{\theta}}\right)$ into $\overline{z}_{1i}\left(\underline{\theta}_0\right) + \left(\overline{z}_{1i}\left(\widetilde{\underline{\theta}}\right) - \overline{z}_{1i}\left(\underline{\theta}_0\right)\right)$ and $\overline{z}_{2i}\left(\widetilde{\underline{\theta}}\right)$ into $\overline{z}_{2i}\left(\underline{\theta}_0\right) + \left(\overline{z}_{2i}\left(\widetilde{\underline{\theta}}\right) - \overline{z}_{2i}\left(\underline{\theta}_0\right)\right)$. The second equality deserves some explanation. From Lemma 4, $\overline{z}_{1i}\left(\widetilde{\underline{\theta}}\right) - \overline{z}_{1i}\left(\underline{\theta}_0\right) \leq m(w_i)\left\|\widetilde{\underline{\theta}} - \underline{\theta}_0\right\|$ and $\overline{z}_{2i}\left(\widetilde{\underline{\theta}}\right) - \overline{z}_{2i}\left(\underline{\theta}_0\right) \leq m(w_i)\left\|\widetilde{\underline{\theta}} - \underline{\theta}_0\right\|$, where $m(w)$ is the slope function for $\overline{z}_{1i}(\theta)$ and $\overline{z}_{2i}(\theta)$. So from Assumption L9, $\overline{z}_{1i}\left(\widetilde{\underline{\theta}}\right) - \overline{z}_{1i}\left(\underline{\theta}_0\right)$ and $\overline{z}_{2i}\left(\widetilde{\underline{\theta}}\right) - \overline{z}_{2i}\left(\underline{\theta}_0\right)$ are $o_p(1)$ for any i. Furthermore, $\sum_{i=1}^{n} \mathbf{1}\left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0\right) = O_p(1)$ and $\sum_{i=1}^{n} \mathbf{1}\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n}\right) = O_p(1)$ by the famous Poisson approximation and Assumption L5. So both $\sum_{i=1}^{n}\left(\overline{z}_{1i}\left(\widetilde{\underline{\theta}}\right) - \overline{z}_{1i}\left(\underline{\theta}_0\right)\right) \mathbf{1}\left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0\right)$ and $\sum_{i=1}^{n}\left(\overline{z}_{2i}\left(\widetilde{\underline{\theta}}\right) - \overline{z}_{2i}\left(\underline{\theta}_0\right)\right) \mathbf{1}\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n}\right)$ are $o_p(1)$ uniformly for all v on any compact set. The third equality follows from the Taylor expansion. Now,

$$
\begin{aligned}
& \int_{0<\|v\|\leq M} \frac{L_n\left(\widetilde{\underline{\theta}}, \gamma_0 + n^{-1}v\right)}{L_n\left(\widetilde{\underline{\theta}}, \gamma_0\right)} \cdot \pi_2(v)dv \\
= \; & \int_{0<\|v\|\leq M} \left(Z_n^d(v) + o_p(1)\right) \pi_2(v)dv \\
= \; & \int_{0<\|v\|\leq M} Z_n^d(v)\, \pi_2(v)dv + o_p(1),
\end{aligned}
$$

where the last equality is from the fact that $o_p(1)$ here is uniform for all v such that $0 < \|v\| \leq M$. ■

Proof of Theorem 5:. This involves checking the conditions of Theorem 2.

(i) From Yu (2007), $\widehat{\theta}_n$ are φ_n-consistent under θ_0. $\widehat{\theta}_{n,0}$ is a special $\widehat{\theta}_n$ with γ_0 known, so is also φ_n-consistent under θ_0.[26]

(ii) From Lemma 5,

$$\sup_{\|h\|\leq M}\left|\ln Z_n(h) - LR_n\left(S, Z_n^d, h\right)\right| \xrightarrow{p} 0 \text{ for every } M.$$

(iii) The Skorohod metric is used for this condition to satisfy. Lemma 6 shows the weak convergence on any compact set.

(iv) $\lim_{v\to\infty} D(v) = -\infty$ almost surely since $E[z_{1i}] < 0$, $E[z_{2i}] < 0$ and $f(\gamma_0) > 0$.

$$\lim_{\|u\|\to\infty}\left\{-\frac{1}{2}u'\mathcal{J}u + u'\mathcal{J}W\right\} = -\infty$$

almost surely since W is tight.

■

Proof of Theorem 6:. From Proposition 1.1 of Beran (1997) and Assumption L9, we need only check that for every sequence $\underline{\theta}_n$ converging to $\underline{\theta}_0$,

$$T_n\left(\underline{\theta}_n\right) \overset{(\underline{\theta}_n,\gamma_0)}{\rightsquigarrow} PO_\infty(\pi),$$

where $PO_\infty(\pi)$ is defined in (11).

By checking the proof of Theorem 4, $T_n\left(\underline{\theta}_n\right) - PO\left(W_n, \pi\right) \xrightarrow{p} 0$ under $(\underline{\theta}_n, \gamma_0)$. The only change in the proof is to substitute $o_p(1)$ by $o(1)$. To prove $PO\left(W_n,\pi\right) \overset{(\underline{\theta}_n,\gamma_0)}{\rightsquigarrow} PO_\infty(\pi)$, the proof procedure of Lemma 8 and Corollary 1 can be used. The only change is to prove

$$\ln Z_n(h) = LR_n\left(S, Z_n^d, h\right) + o_p(1),$$

where the $o_p(1)$ is uniform for h on any compact set in $\mathbb{R}^k$ under $(\underline{\theta}_n, \gamma_0)$ instead of θ_0. But in the proof of Lemma 5, Theorem 7.2 and Lemma 19.31 of Van der Vaart (1998) can be strengthened to $(\underline{\theta}_n, \gamma_0)$ instead of a fixed point θ_0 by Theorem II.1.2 in Ibragimov and Has'minskii (1981). ■

Appendix C: Lemmas

In theorem 1, assumption (ii) is a very high-level assumption. The following Lemma 0 provides some primary conditions and will be used in both examples.

Lemma 0 *If*

(i) *there is an approximation* $\log Z_n^a(h, h_0)$ *of* $\log Z_n(h, h_0)$ *such that for any* $M < \infty$,

$$\sup_{\|h\|\leq M}\left|\log Z_n(h,h_0) - \log Z_n^a(h,h_0)\right| \xrightarrow{p} 0$$

under h_0.

(ii) for any $\varepsilon > 0$, there is $M > 0$ which may depend on ε and n^* such that when $n > n^*$, $Z_n(h, h_0)$ *and* $Z_n^a(h, h_0)$ *satisfy the following two conditions:*

[26] $\widehat{\gamma}_{n,0} = \gamma_0$, so is consistent at any rate.

(1)

$$
\begin{aligned}
P\left(\int_{H_{1n}\cap\|h\|>M} Z_n(h,h_0)\pi(h)dh>\varepsilon\right) &< \varepsilon, P\left(\int_{H_{0n}\cap\|h\|>M} Z_n(h)\pi(h,h_0)dh>\varepsilon\right)<\varepsilon,\\
P\left(\int_{H_{0\infty}\cap\|h\|>M} Z_n^a(h,h_0)\pi(h)dh>\varepsilon\right) &< \varepsilon, P\left(\int_{H_{1\infty}\cap\|h\|>M} Z_n^a(h,h_0)\pi(h)dh>\varepsilon\right)<\varepsilon;
\end{aligned}
$$

(2)

$$
\begin{aligned}
P\left(\int_{H_{0\infty}\cap\|h\|\le M} Z_n^a(h,h_0)\pi(h)dh<\varepsilon\right) &< \varepsilon,\\
P\left(\left|\int_{H_{1\infty}\cap\|h\|\le M} Z_n^a(h,h_0)\pi(h)dh\right|>M\right) &< \varepsilon.
\end{aligned}
$$

Then

$$
PO\left(W_n,\pi,h_0\right)-\frac{\int_{H_{1\infty}} Z_n^a(h,h_0)\pi(h)dh}{\int_{H_{0\infty}} Z_n^a(h,h_0)\pi(h)dh}\xrightarrow{p}0.
$$

If furthermore,

(iii) *there is a metric on the space of the sample path of* $Z_n^a(h,h_0)$ *such that* $\int_{\|h\|\le M}$ *is a continuous operator on that space and*

$$
Z_n^a(h,h_0)\overset{h_0}{\rightsquigarrow} Z_\infty(h,h_0)
$$

for h *on any compact set.*

(iv) for any $\varepsilon>0$, there is $M>0$ which may depend on ε such that

$$
P\left(\int_{H_{1\infty}\cap\|h\|>M} Z_\infty(h,h_0)\pi(h)dh>\varepsilon\right)<\varepsilon,\ P\left(\int_{H_{0\infty}\cap\|h\|>M} Z_\infty(h,h_0)\pi(h)dh>\varepsilon\right)<\varepsilon,
$$

Then

$$
PO\left(W_n,\pi,h_0\right)\overset{h_0}{\rightsquigarrow}\frac{\int_{H_{1\infty}} Z_\infty(h,h_0)\pi(h)dh}{\int_{H_{0\infty}} Z_\infty(h,h_0)\pi(h)dh}.
$$

Proof. First, note that

$$
PO\left(W_n,\pi\right)=\frac{\int_{H_{1n}} Z_n(h)\pi(h)dh}{\int_{H_{0n}} Z_n(h)\pi(h)dh}=\frac{\int_{H_{1n}\cap\|h\|\le M} Z_n(h)\pi(h)dh+\int_{H_{1n}\cap\|h\|\le M} Z_n(h)\pi(h)dh}{\int_{H_{0n}\cap\|h\|\le M} Z_n(h)\pi(h)dh+\int_{H_{0n}\cap\|h\|\le M} Z_n(h)\pi(h)dh}
$$

and

$$
\frac{\int_{H_{1\infty}} Z_n^a(h)\pi(h)dh}{\int_{H_{0\infty}} Z_n^a(h)\pi(h)dh}=\frac{\int_{H_{1\infty}\cap\|h\|\le M} Z_n^a(h)\pi(h)dh+\int_{H_{1\infty}\cap\|h\|\le M} Z_n^a(h)\pi(h)dh}{\int_{H_{0\infty}\cap\|h\|\le M} Z_n^a(h)\pi(h)dh+\int_{H_{0\infty}\cap\|h\|\le M} Z_n^a(h)\pi(h)dh}.
$$

For n large enough, $\{H_{1n} \cap \|h\| \leq M\} = \{H_{1\infty} \cap \|h\| \leq M\}$. So

$$\begin{aligned} & \int_{H_{1n}\cap\|h\|\leq M} Z_n(h)\pi(h)dh \\ = & \int_{H_{1\infty}\cap\|h\|\leq M} \exp\left\{\ln Z_n^a(h) + o_p(1)\right\}\pi(h)dh \\ = & \int_{H_{1\infty}\cap\|h\|\leq M} Z_n^a(h)\exp\left\{o_p\left(1\right)\right\}\pi(h)dh \\ = & \int_{H_{1\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh + o_p\left(1\right) \end{aligned}$$

where the first equality is from assumption (i), and the third equality is because $o_p(1)$ here is uniform for M and a Taylor expansion $\exp\{o_p(1)\} = 1 + o_p(1)$ can apply. Similarly, $\int_{H_{0n}\cap\|h\|\leq M} Z_n(h)\pi(h)dh - \int_{H_{0\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh = o_p(1)$.

For two nonnegative random sequences $\{X_n\}$ and $\{Y_n\}$, if $X_n > \varepsilon$ for some $\varepsilon > 0$ with probability approaching 1, $Y_n = O_p(1)$, and there are two other sequences $\{X_n^a\}$ and $\{Y_n^a\}$ such that $X_n - X_n^a = o_p(1)$ and $Y_n - Y_n^a = o_p(1)$, then

$$\begin{aligned} \left|\frac{Y_n + o_p(1)}{X_n + o_p(1)} - \frac{Y_n^a + o_p(1)}{X_n^a + o_p(1)}\right| &= \left|\frac{Y_n X_n^a - Y_n^a X_n + o_p(1)\left(Y_n - Y_n^a + X_n^a - X_n\right) + o_p(1)}{\left(X_n + o_p(1)\right)\left(X_n^a + o_p(1)\right)}\right| \\ &= \left|\frac{X_n\left(Y_n - Y_n^a\right) + Y_n\left(X_n - X_n^a\right) + o_p(1)}{\left(X_n + o_p(1)\right)\left(X_n^a + o_p(1)\right)}\right| \\ &= \left|\frac{Y_n - Y_n^a}{X_n^a + o_p(1)} + \frac{Y_n\left(X_n - X_n^a\right)}{\left(X_n + o_p(1)\right)\left(X_n^a + o_p(1)\right)} + o_p(1)\right| \\ &= o_p(1). \end{aligned}$$

Let $X_n = \int_{H_{0n}\cap\|h\|\leq M} Z_n(h)\pi(h)dh$, $Y_n = \int_{H_{1n}\cap\|h\|\leq M} Z_n(h)\pi(h)dh$, $X_n^a = \int_{H_{0\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh$, and $Y_n^a = \int_{H_{1\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh$, we prove the first result.

To prove the second result, we need only to prove

$$\frac{\int_{H_{1\infty}} Z_n^a(h)\pi(h)dh}{\int_{H_{0\infty}} Z_n^a(h)\pi(h)dh} \overset{\theta_0}{\rightsquigarrow} \frac{\int_{H_{1\infty}} Z_\infty(h)\pi(h)dh}{\int_{H_{0\infty}} Z_\infty(h)\pi(h)dh}.$$

By the continuous mapping theorem and Assumption (iii), $\int_{H_{1\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh \overset{\theta_0}{\rightsquigarrow} \int_{H_{1\infty}\cap\|h\|\leq M} Z_\infty(h)\pi(h)dh$, and $\int_{H_{0\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh \overset{\theta_0}{\rightsquigarrow} \int_{H_{0\infty}\cap\|h\|\leq M} Z_\infty(h)\pi(h)dh$. From Assumption (ii)(2), $P\left(\int_{H_{0\infty}\cap\|h\|\leq M} Z_\infty(h)\pi(h)dh < \varepsilon\right) < \varepsilon$, and $P\left(\left|\int_{H_{1\infty}\cap\|h\|\leq M} Z_\infty(h)\pi(h)dh\right| > M\right) < \varepsilon$. Combining this fact with Assumption (iv), a similar analysis as above reduces the problem to prove

$$\frac{\int_{H_{1\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh}{\int_{H_{0\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh} \overset{\theta_0}{\rightsquigarrow} \frac{\int_{H_{1\infty}\cap\|h\|\leq M} Z_\infty(h)\pi(h)dh}{\int_{H_{0\infty}\cap\|h\|\leq M} Z_\infty(h)\pi(h)dh},$$

but this holds from an application of the continuous mapping theorem. ■

Remark 4 *$P\left(\int_{H_{0\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh < \varepsilon\right) < \varepsilon$ in (ii)(2) is satisfied in two leading cases: (i) $H_{0\infty} = \{0\}$; that is, the null hypothesis is simple, then $\int_{H_{0\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh = 1$; (ii) the asymptotic distribution of $\int_{H_{0\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh$ is continuous, since $\int_{H_{0\infty}\cap\|h\|\leq M} Z_n^a(h)\pi(h)dh \geq 0$ almost surely.*

Tests in LAN Models

Lemma 1 *Suppose Assumptions C0-C3 hold, then*

$$\ln Z_n(h, h_0) = LR_n(S, h, h_0) + o_p(1),$$

where the $o_p(1)$ *is uniform for* h *on any compact set in* $\mathbb{R}^k$ *under* h_0*.*

Proof. From Lemma 7.6 of Van der Vaart 1998), the model is differentiable in quadratic mean (DQM) at θ_0 by Assumptions C0-C2. Theorem 7.2 of Van der Vaart 1998) shows that,

$$\ln Z_n(h) = LR_n(S, h) + o_p(1),$$

where the residual is $o_p(1)$ only under P_{θ_0}. But from Lemma 19.31 of Van der Vaart (1998) and Assumption C3, this $o_p(1)$ can be strengthened to be uniform for h on any compact set. ■

Remark 5 *A corollary of Lemma 1 is that all models indexed by h are contiguous to each other from Example 6.5 of Van der Vaart (1998).*

Lemma 2 *Suppose Assumptions C0-C3 hold, then*

$$PO(W_n, \pi, h_0) - \frac{\int_{H_{1\infty}} \exp\{LR_n(S, h, h_0)\} \pi(h) dh}{\int_{H_{0\infty}} \exp\{LR(S, h, h_0)\} \pi(h) dh} \xrightarrow{p} 0$$

under h_0*.*

Proof. From Lemma 0, we need only check assumptions (i) and (ii). $\log Z_n^a(h) = LR_n(S, h)$ in this case. (i) is satisfied by Lemma 1, so we concentrate on the proof of (ii).

We show (ii)(1) first. For any $\varepsilon > 0$,

$$\begin{aligned} & P\left(\int_{H_{1n} \cap \|h\| > M} Z_n(h)\pi(h)dh > \varepsilon\right) \\ \leq \; & \varepsilon^{-1} E \int_{H_{1n} \cap \|h\| > M} Z_n(h)\pi(h)dh \\ = \; & \varepsilon^{-1} \int_{H_{1n} \cap \|h\| > M} E[Z_n(h)]\pi(h)dh \\ = \; & \varepsilon^{-1} \int_{H_{1n} \cap \|h\| > M} \pi(h)dh, \end{aligned}$$

where the inequality is from Markov's inequality, the first equality holds by Fubini's theorem, and the second equality holds by $E[Z_n(h)] = 1$. Since $\pi(h)$ is a density on H_{1n}, the rhs can be made arbitrarily small for all n by taking M large. Similarly, $P\left(\int_{H_{0n} \cap \|h\| > M} Z_n(h)\pi(h)dh > \varepsilon\right) < \varepsilon$.

Next,

$$\begin{aligned}
&\int_{H_{1\infty}\cap\|h\|>M}\exp\left\{LR_n\left(S,h\right)\right\}\pi(h)dh\\
=\;&\int_{H_{1\infty}\cap\|h\|>M}\exp\left\{-\frac{1}{2}h'Ih+h'I\overline{z}\right\}\pi(h)dh\\
=\;&\int_{H_{1\infty}\cap\|h\|>M}\exp\left\{-\frac{1}{2}\left(\overline{z}-h\right)'I\left(\overline{z}-h\right)\right\}\exp\left\{\frac{1}{2}\overline{z}'I\overline{z}\right\}\pi(h)dh\\
\leq\;&\int_{H_{1\infty}\cap\|h\|>M}\exp\left\{\frac{1}{2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S\left(w_i|\theta_0\right)\right)'I^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S\left(w_i|\theta_0\right)\right)\right\}\pi(h)dh\\
\leq\;&\exp\left\{\frac{1}{2}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S\left(w_i|\theta_0\right)\right\|^2\left\|I^{-1}\right\|\right\}\int_{H_{1\infty}\cap\|h\|>M}\pi(h)dh,
\end{aligned}$$

where the last inequality follows from Assumption C2. Since $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S\left(w_i|\theta_0\right)=O_p\left(1\right)$ and $I>0$, the rhs can be made arbitrarily small by taking M large. Similarly, $P\left(\int_{H_{0\infty}\cap\|h\|>M}\exp\left\{LR_n\left(S,h\right)\right\}\pi(h)dh>\varepsilon\right)<\varepsilon$.

We now establish (ii)(2). First, note that

$$\begin{aligned}
&\exp\left\{LR_n\left(S,h\right)\right\}\\
=\;&\exp\left\{-\frac{1}{2}h'\left(P+P^{\perp}\right)'I\left(P+P^{\perp}\right)h+h'\left(P+P^{\perp}\right)'I\left(P+P^{\perp}\right)\overline{z}\right\}\\
=\;&\exp\left\{-\frac{1}{2}\left(Ph\right)'I\left(Ph\right)+\left(Ph\right)'IP\overline{z}\right\}\exp\left\{-\frac{1}{2}\left(P^{\perp}h\right)'I\left(P^{\perp}h\right)+\left(P^{\perp}h\right)'IP^{\perp}\overline{z}\right\}\\
=\;&\exp\left\{-\frac{1}{2}\left(Ph-P\overline{z}\right)'I\left(Ph-P\overline{z}\right)+\frac{1}{2}\left(P\overline{z}\right)'IP\overline{z}\right\}\exp\left\{-\frac{1}{2}\left(P^{\perp}h-P^{\perp}\overline{z}\right)'I\left(P^{\perp}h-P^{\perp}\overline{z}\right)+\frac{1}{2}\left(P^{\perp}\overline{z}\right)'IP^{\perp}\overline{z}\right\}.
\end{aligned}$$

Since $\{H_{0\infty}\cap\|h\|\leq M\}=\{\|Ph\|\leq c\}$ for some $c>0$,

$$\begin{aligned}
&\int_{H_{0\infty}\cap\|h\|\leq M}\exp\left\{LR_n\left(S,h\right)\right\}\pi(h)dh\\
=\;&\int_{\|Ph\|\leq c}\exp\left\{-\frac{1}{2}\left(Ph-P\overline{z}\right)'I\left(Ph-P\overline{z}\right)+\frac{1}{2}\left(P\overline{z}\right)'IP\overline{z}\right\}\pi_1\left(Ph\right)d\left(Ph\right)\\
=\;&\exp\left\{\frac{1}{2}\left(P\overline{z}\right)'IP\overline{z}\right\}\int_{\|Ph\|\leq c}\exp\left\{-\frac{1}{2}\left(Ph-P\overline{z}\right)'I\left(Ph-P\overline{z}\right)\right\}\pi_1\left(Ph\right)d\left(Ph\right)
\end{aligned}$$

which has a continuous asymptotic distribution since $\overline{z}$ weakly converges to z under θ_0.

$$\begin{aligned}
&\int_{H_{1\infty}\cap\|h\|\leq M}\exp\left\{LR_n\left(S,h\right)\right\}\pi(h)dh\\
\leq\;&\int_{\|Ph\|\leq M}\exp\left\{-\frac{1}{2}\left(Ph-P\overline{z}\right)'I\left(Ph-P\overline{z}\right)+\frac{1}{2}\left(P\overline{z}\right)'IP\overline{z}\right\}\pi_1\left(Ph\right)d\left(Ph\right)
\end{aligned}$$

$$\int_{\|P^{\perp}h\|\leq M}\exp\left\{-\frac{1}{2}\left(P^{\perp}h-P^{\perp}\overline{z}\right)'I\left(P^{\perp}h-P^{\perp}\overline{z}\right)+\frac{1}{2}\left(P^{\perp}\overline{z}\right)'IP^{\perp}\overline{z}\right\}\sqrt{\frac{c^{r}\left|I\right|}{\left(2\widetilde{\pi}\right)^{r}}}\exp\left\{-\frac{c}{2}\left(P^{\perp}h\right)'IP^{\perp}h\right\}d\left(P^{\perp}h\right)$$
$$\leq \exp\left\{\frac{1}{2}\left(P\overline{z}\right)'IP\overline{z}\right\}\exp\left\{\frac{1}{2}\left(P^{\perp}\overline{z}\right)'IP^{\perp}\overline{z}\right\}$$
$$= O_p(1),$$

where the last equality is from the fact that both $P\overline{z}$ and $P^{\perp}\overline{z}$ are $O_p(1)$. ■

Remark 6 *The convergence in probability to zero can be understood under any probability measure indexed by h_0, since they are all contiguous to each other from Remark 5.*

Remark 7 *A corollary of Lemma 2 is that $PO(W_n,\pi,h_0)\overset{h_0}{\rightsquigarrow}\frac{\int_{Ah\neq 0}Z_{\infty}(h,h_0)\pi(h)dh}{\int_{Ah=0}Z_{\infty}(h,h_0)\pi(h)dh}$, where $Z_{\infty}(h,h_0)$ is defined in (9). This is basically an application of the continuous mapping theorem, since $\frac{\int_{H_{1\infty}}LR_n(S,h-h_0)\pi(h)dh}{\int_{H_{0\infty}}LR(S,h-h_0)\pi(h)dh}$ is a continuous function of $\overline{z}$.*

Lemma 3 *Suppose Assumptions C0-C4 hold, then*

$$\Lambda_n(h_0)-(\overline{z}+h_0)'P^{\perp\prime}IP^{\perp}(\overline{z}+h_0)\xrightarrow{p}0$$

under h_0.

Proof. This is essentially Theorem 16.7 of Van der Vaart (1998). ■

Tests on the Location of a Threshold Point

Lemma 4 (Lipschitz Continuity) *Under Assumptions L0-L1 and L3-L4, both $\overline{z}_1(w|\underline{\theta})$ and $\overline{z}_2(w|\underline{\theta})$ are Lipschitz continuous in $\underline{\theta}$ for $\underline{\theta}$ in a neighborhood of $\underline{\theta}_0$ with the slope function in L^2 space.*

Proof. Only the result for $\overline{z}_1(w|\underline{\theta})$ are proved, since the proof for $\overline{z}_2(w|\underline{\theta})$ is similar.

For $\underline{\theta}$ and $\overline{\underline{\theta}}$ in a neighborhood of $\underline{\theta}_0$, by Assumptions L3 and L0,

$$\left|\ln\frac{\frac{\overline{\sigma_1}}{\overline{\sigma_2}}f_{e|x,q}\left(\frac{\overline{\sigma_1}e+x'\left(\overline{\beta_1}-\overline{\beta_2}\right)}{\overline{\sigma_2}}|x,q;\overline{\eta}\right)}{f_{e|x,q}\left(e|x,q;\overline{\eta}\right)}-\ln\frac{\frac{\sigma_1}{\sigma_2}f_{e|x,q}\left(\frac{\sigma_1e+x'\left(\beta_1-\beta_2\right)}{\sigma_2}|x,q;\eta\right)}{f_{e|x,q}\left(e|x,q;\eta\right)}\right|$$
$$\leq m(w)\left(\frac{\sigma_2+\sigma_1}{\sigma_2\overline{\sigma_2}}\left(\left|\overline{\sigma_1}-\sigma_1\right|+\left|\overline{\sigma_2}-\sigma_2\right|\right)+\frac{\|x\|}{\overline{\sigma_2}}\left(\left\|\overline{\beta_1}-\beta_1\right\|+\left\|\overline{\beta_2}-\beta_2\right\|\right)+\frac{\|\beta_1\|+\|\beta_2\|}{\sigma_2\overline{\sigma_2}}\|x\|\left|\overline{\sigma_2}-\sigma_2\right|+\|\overline{\eta}-\eta\|\right)$$
$$+\frac{\left|\overline{\sigma_1}-\sigma_1\right|}{\min\left\{\sigma_1,\overline{\sigma_1}\right\}}+\frac{\left|\overline{\sigma_2}-\sigma_2\right|}{\min\left\{\sigma_2,\overline{\sigma_2}\right\}}$$
$$\leq c\left(1+m(w)+m(w)\|x\|\right)\left|\overline{\underline{\theta}}-\underline{\theta}\right|.$$

From the Cauchy-Schwarz inequality, $1+m(w)+m(w)\|x\|$ is in L^2 from Assumption L3 and L4. ■

Lemma 5 *Suppose Assumptions L0-L5 hold, then*

$$\ln Z_n(h,h_0)=LR_n\left(S,Z_n^d,h,h_0\right)+o_p(1),$$

where the $o_p(1)$ is uniform for h on any compact set in $\mathbb{R}^k$ under h_0.

Proof. The proof follows from Lemma 2 of Yu (2007). ■

Lemma 6 *Suppose Assumptions L0-L7 hold, then*

$$\ln Z_n(h, h_0) \overset{h_0}{\rightsquigarrow} Z_\infty(h, h_0)$$

for h on any compact set in $\mathbb{R}^k$, where $Z_\infty(h, h_0)$ is specified in (10).

Proof. The result follows from Lemma 6 of Yu (2007). ■

Lemma 7 *Suppose Assumption L0-L7 holds, then the densities $\{L_n(\theta_0 + \varphi_n h) : n \geq 1\}$ of W_n at a fixed $h \in H_n$ is contiguous to the null densities $\{L_n(\theta_0) : n \geq 1\}$.*

Proof. The proof uses Lemma 6.4 (iii) of Van der Vaart (1998). We need only prove that $E[Z_\infty(h)] = 1$. Only the case with $v < 0$ is proved, and the case with $v > 0$ is similar.

$$\begin{aligned}
& E\left[\exp\left\{-\frac{1}{2}u'\mathcal{J}u + u'\mathcal{J}W + D(v)\right\}\right] \\
= \ & \exp\left\{-\frac{1}{2}u'\mathcal{J}u\right\} E\left[\exp\left\{u'\mathcal{J}W\right\}\right] E\left[\exp\left\{D(v)\right\}\right] \\
= \ & \exp\left\{-\frac{1}{2}u'\mathcal{J}u\right\} \exp\left\{\frac{1}{2}u'\mathcal{J}\mathcal{J}^{-1}\mathcal{J}u\right\} E\left[\exp\left\{\sum_{i=1}^{N_1(|v|)} \ln \frac{\frac{\sigma_{10}}{\sigma_{20}} f_{e|x,q}\left(\frac{\sigma_{10}e_i + x_i'(\beta_{10}-\beta_{20})}{\sigma_{20}} |x_i, q_i; \eta_0\right)}{f_{e|x,q}(e_i|x_i, q_i; \alpha_0)} |q_i = \gamma_0 -\right\}\right] \\
= \ & E\left[E\left[\prod_{i=1}^{N_1(|v|)} \frac{\frac{\sigma_{10}}{\sigma_{20}} f_{e|x,q}\left(\frac{\sigma_{10}e_i + x_i'(\beta_{10}-\beta_{20})}{\sigma_{20}} |x_i, q_i; \eta_0\right)}{f_{e|x,q}(e_i|x_i, q_i; \alpha_0)} |q_i = \gamma_0 - \middle| N_1(|v|)\right]\right] \\
= \ & E\left[\prod_{i=1}^{N_1(|v|)} E\left[\frac{\frac{\sigma_{10}}{\sigma_{20}} f_{e|x,q}\left(\frac{\sigma_{10}e_i + x_i'(\beta_{10}-\beta_{20})}{\sigma_{20}} |x_i, q_i; \eta_0\right)}{f_{e|x,q}(e_i|x_i, q_i; \eta_0)}\right]\right] \\
= \ & 1
\end{aligned}$$

where the second equality is from the moment generating function of a multivariate normal distribution, the third equality is from the law of iterated expectation (LIE), the fourth equality is from the independence of $\{z_{1i}\}_{i \geq 1}$ and $N_1(\cdot)$ and the LIE, and the last equality is from the fact that the integral of a density function is always 1. ■

Lemma 8 *Suppose Assumption L0-L5 holds, then*

$$PO(W_n, \pi, h_0) - \frac{\int_{(\mathbb{R}\setminus\{0\})\times\mathbb{R}^{\underline{k}}} \exp\left\{LR_n\left(S, Z_n^d, h, h_0\right)\right\}\pi(h)dh}{\int_{\{0\}\times\mathbb{R}^{\underline{k}}} \exp\left\{LR_n\left(S, Z_n^d, h, h_0\right)\right\}\pi(h)dh} \xrightarrow{p} 0$$

under h_0.

Proof. In this proof, we use an equivalent norm with Euclidean norm as $\|x\| = \max_i x_i$ for a vector $x \in \mathbb{R}^k$. Lemma 0 is used again. Now, $\log Z_n^a(h) = LR_n\left(S, Z_n^d, h\right)$. (i) follows from Lemma 5, so we concentrate on the checking of (ii).

(ii)(1). $P\left(\int_{H_{1n}\cap\|h\|>M} Z_n(h)\pi(h)dh > \varepsilon\right) < \varepsilon$ and $P\left(\int_{H_{0n}\cap\|h\|>M} Z_n(h)\pi(h)dh > \varepsilon\right) < \varepsilon$ can be proved similarly as in Lemma 2.

$$\begin{aligned}&\int_{H_{1\infty}\cap\|h\|>M}\exp\left\{LR_n\left(S,Z_n^d,h\right)\right\}\pi(h)dh\\ =&\int_{\|u\|>M}\exp\left\{-\frac{1}{2}u'\mathcal{J}u+u'\mathcal{J}\bar{z}\right\}\pi_1(u)du\cdot\int_{\|v\|>M}Z_n^d(v)\pi_2(v)dv.\end{aligned}$$

The first term on the rhs is $o_p(1)$ by a similar argument as in Lemma 2. As to the second term on the rhs, first note that $E[\bar{z}_{1i}]<0$, $E[\bar{z}_{2i}]<0$ by the strict Jensen's inequality. Second, from Assumption L5, the interarrival time between jumps is $O_p(1)$. So $Z_n^d(v)$ is exponentially decaying and is uniformly $O_p(1)$ for all v. In consequence, $\int_{\|v\|>M}Z_n^d(v)\pi_2(v)dv=O_p(1)\int_{\|v\|>M}\pi_2(v)dv$ can be made arbitrarily small by taking M large. Similarly, $P\left(\int_{H_{0\infty}\cap\|h\|>M}\exp\left\{LR_n\left(S,Z_n^d,h\right)\right\}\pi(h)dh>\varepsilon\right)<\varepsilon$.

(ii)(2). The first part can be proved similarly as the checking of (ii)(2) in Lemma 2 since it is only an integration on the regular parameters. To prove $\int_{H_{1\infty}\cap\|h\|\leq M}\exp\left\{LR_n\left(S,Z_n^d,h\right)\right\}\pi(h)dh=O_p(1)$, we need only to prove $\int_{\|v\|\leq M}Z_n^d(v)\pi_2(v)dh=O_p(1)$ by (ii)(2) of Lemma 2. As argued in Step (i), $Z_n^d(v)=O_p(1)$ uniformly for v, the result follows. ■

Corollary 1 *Suppose Assumption L0-L7 holds, then*

$$PO\left(W_n,\pi,h_0\right)\overset{h_0}{\rightsquigarrow}\int_{\mathbb{R}\setminus\{0\}}\exp\left\{D_{v_0}(v)-D_{v_0}(0)\right\}\pi_2(v)dv.$$

Proof. From Lemma 0, we need only check assumptions (iii) and (iv). From Lemma 8,

$$\frac{\int_{(\mathbb{R}\setminus\{0\})\times\mathbb{R}^{\underline{k}}}\exp\left\{LR_n\left(S,Z_n^d,h\right)\right\}\pi(h)dh}{\int_{\{0\}\times\mathbb{R}^{\underline{k}}}\exp\left\{LR_n\left(S,Z_n^d,h\right)\right\}\pi(h)dh}=\int_{\mathbb{R}\setminus\{0\}}Z_n^d(v)\pi_2(v)dv.$$

(iii) follows from Lemma 6 with the Skorohod metric used. For (iv), we need only check $P\left(\int_{|v|>M}\exp\left\{D(v)\right\}\pi_2(v)dv>\varepsilon\right)$ $<\varepsilon$. A similar argument as in Lemma 8 shows that $\exp\left\{D(v)\right\}$ is uniformly $O_p(1)$ for all v, so $\int_{|v|>M}\exp\left\{D(v)\right\}\pi_2(v)dv$ can be made arbitrarily small by taking M large. ■

Appendix D: $PO_\infty(\pi,h_0)$ and $\Lambda_\infty(h_0)$ in Section 4

In this appendix, the power function is derived by calculating $PO_\infty(\pi,h_0)$ and $\Lambda_\infty(h_0)$ explicitly.

$PO_\infty(\pi,h_0)$

In this case, finding b and the local power is reduced to the derivation of the distribution of $\int_{\mathbb{R}\setminus\{0\}}\exp\left\{D_{v_0}(v)-D_{v_0}(0)\right\}\pi_2(v)dv$ for any $v_0\in\mathbb{R}$.

Suppose the interarrival times of $D_{v_0}(v)$ are i.i.d. exponential random variables $\{T_{1i}\}_{i=0}^{\infty}$ and $\{T_{2i}\}_{i=0}^{\infty}$

with mean $\frac{1}{f_q(\gamma_0)}$ starting from v_0, then

$$\begin{aligned}
& \int_{\mathbb{R}\backslash\{0\}} \exp\left\{D_{v_0}(v) - D_{v_0}(0)\right\}\pi_2(v)dv \\
= & \int_{-\infty}^{v_0} \exp\left\{D_{v_0}(v) - D_{v_0}(0)\right\}\pi_2(v)dv + \int_{v_0}^{\infty} \exp\left\{D_{v_0}(v) - D_{v_0}(0)\right\}\pi_2(v)dv \\
= & \frac{1}{\exp\left\{D_{v_0}(0)\right\}}\left[\int_{v_0-T_{10}}^{v_0} \pi_2(v)dv + \sum_{i=1}^{\infty}\exp\left(\sum_{j=1}^{i} z_{1j}\right)\int_{v_0-\sum_{j=0}^{i} T_{1i}}^{v_0-\sum_{j=0}^{i-1} T_{1i}} \pi_2(v)dv \right.\\
& \left. + \int_{v_0}^{v_0+T_{20}} \pi_2(v)dv + \sum_{i=1}^{\infty}\exp\left(\sum_{j=1}^{i} z_{2j}\right)\int_{v_0+\sum_{j=0}^{i-1} T_{2i}}^{v_0+\sum_{j=0}^{i} T_{2i}} \pi_2(v)dv\right]
\end{aligned}$$

where the point mass in $\pi_2(v)$ is substituted by any denisty value.

When $\pi_2(v) = \sqrt{\frac{c}{2\pi}}\exp\left\{-\frac{c}{2}v^2\right\}$,

$$\begin{aligned}
& \int_{\mathbb{R}\backslash\{0\}} \exp\left\{D_{v_0}(v) - D_{v_0}(0)\right\}\pi_2(v)dv \\
= & \frac{1}{\exp\left\{D_{v_0}(0)\right\}}\left[\Phi\Big|_{\sqrt{c}(v_0-T_{10})}^{\sqrt{c}v_0} + \sum_{i=1}^{\infty}\exp\left(\sum_{j=1}^{i} z_{1j}\right)\Phi\Big|_{\sqrt{c}\left(v_0-\sum_{j=0}^{i} T_{1i}\right)}^{\sqrt{c}\left(v_0-\sum_{j=0}^{i-1} T_{1i}\right)}\right. \\
& \left. +\Phi\Big|_{\sqrt{c}v_0}^{\sqrt{c}(v_0+T_{20})} + \sum_{i=1}^{\infty}\exp\left(\sum_{j=1}^{i} z_{2j}\right)\Phi\Big|_{\sqrt{c}\left(v_0+\sum_{j=0}^{i-1} T_{2i}\right)}^{\sqrt{c}\left(v_0+\sum_{j=0}^{i} T_{2i}\right)}\right],
\end{aligned}$$

where $\Phi\Big|_{v_1}^{v_2} \equiv \Phi(v_2) - \Phi(v_1)$ and $\Phi(\cdot)$ is the cdf of a standard normal distribution.

When $\pi_2(v) = 1$,

$$\begin{aligned}
& \int_{\mathbb{R}\backslash\{0\}} \exp\left\{D_{v_0}(v) - D_{v_0}(0)\right\}\pi_2(v)dv \\
= & \frac{1}{\exp\left\{D_{v_0}(0)\right\}}\left[T_{10} + \sum_{i=1}^{\infty}\exp\left(\sum_{j=1}^{i} z_{1j}\right)T_{1i} + T_{20} + \sum_{i=1}^{\infty}\exp\left(\sum_{j=1}^{i} z_{2j}\right)T_{2i}\right]
\end{aligned}$$

depends on v_0 only through $\exp\{D_{v_0}(0)\}$.

$\Lambda_\infty(h_0)$

First, we will derive the point-wise power envelope. For this purpose, the distributions of $D(v_0)$ and $-D_{v_0}(0)$ are developed. If $v_0 < 0$, then for any $\Delta \in \mathbb{R}$,

$$
\begin{aligned}
& P(D(v_0) \le \Delta) \\
= \; & P\left(\sum_{i=1}^{N_1(|v_0|)} z_{1i} \le \Delta\right) \\
= \; & \sum_{k=0}^{\infty} P\left(\sum_{i=1}^{N_1(|v_0|)} z_{1i} \le \Delta \,\middle|\, N_1(|v_0|) = k\right) P(N_1(|v_0|) = k) \\
= \; & \sum_{k=0}^{\infty} P\left(\sum_{i=1}^{k} z_{1i} \le \Delta\right) P(N_1(|v_0|) = k) \\
= \; & \sum_{k=0}^{\infty} \frac{e^{-f_q(\gamma_0)|v_0|}\left(f_q(\gamma_0)|v_0|\right)^k}{k!} P\left(\sum_{i=1}^{k} z_{1i} \le \Delta\right).
\end{aligned}
$$

If z_{1i} follows a nonstable distribution, then the cdf of $\sum_{i=1}^{k} z_{1i}$ is hard to derive. In practice, the simulation method can be used to derive this cdf. If z_{1i} follows $N\left(-\frac{\beta_1^2}{2s^2}, \frac{\beta_1^2}{s^2}\right)$ as in the example of Section 4 without threshold effect in variance, then $P\left(\sum_{i=1}^{k} z_{1i} \le \Delta\right) = \Phi\left(\frac{\Delta}{\sqrt{k}\beta_1/s} + \sqrt{k}\frac{\beta_1}{2s}\right)$. The case of $v_0 > 0$ can be similarly developed with z_{1i} substituted by z_{2i} in the final equality above. As mentioned in the main test, there is a point mass at zero in the distribution of $D(v_0)$. The probability of this point mass is $P(N_1(|v_0|) = 0) = e^{-f_q(\gamma_0)|v_0|}$.For a fixed v_0, there may be no b such that $P(D(v_0) \ge \ln b) = \alpha$ and the randomized test is used. Usually, such v_0 falls into an interval around 0, but this is not a general result.[27]

As to $-D_{v_0}(0)$, a similar procedure can be used. If $v_0 < 0$, then for any $\Delta \in \mathbb{R}$,

$$
\begin{aligned}
& P(-D_{v_0}(0) \le \Delta) \\
= \; & P\left(\sum_{i=1}^{N_2(|v_0|)} z_{2i} \ge -\Delta\right) \\
= \; & \sum_{k=0}^{\infty} P\left(\sum_{i=1}^{k} z_{2i} \ge -\Delta\right) P(N_2(|v_0|) = k) \\
= \; & \sum_{k=0}^{\infty} \frac{e^{-f_q(\gamma_0)|v_0|}\left(f_q(\gamma_0)|v_0|\right)^k}{k!} P\left(\sum_{i=1}^{k} z_{2i} \ge -\Delta\right).
\end{aligned}
$$

If z_{1i} follows $N\left(-\frac{\beta_1^2}{2s^2}, \frac{\beta_1^2}{s^2}\right)$, then $P\left(\sum_{i=1}^{k} z_{2i} \ge -\Delta\right) = \Phi\left(\frac{\Delta}{\sqrt{k}\beta_1/s} - \sqrt{k}\frac{\beta_1}{2s}\right)$. For the case $v_0 > 0$, z_{2i} is substituted by z_{1i} in the last equality.

For the likelihood ratio test, the distribution of Λ_∞ is derived to find the critical value. Notice that there is a point mass at zero and no density on $\Delta < 0$ in the distribution of $\frac{\Lambda_\infty}{2}$. The probability of the point

[27] Note that $P(D(v_0) > 0) = \sum_{k=1}^{\infty} \frac{e^{-f_q(\gamma_0)|v_0|}\left(f_q(\gamma_0)|v_0|\right)^k}{k!} P\left(\sum_{i=1}^{k} z_{1i} > 0\right)$ is not generally a decreasing function of $|v_0|$.

mass is p_{10} from Yu (2007). When $\Delta > 0$,

$$\begin{aligned} P\left(\frac{\Lambda_\infty}{2} \leq \Delta\right) &= p_{10} + P\left(0 < \max\left\{Z_1, Z_2\right\} \leq \Delta\right) \\ &= F_1(0) F_2(0) + \int_0^\Delta dF_1(x) F_2(x) \\ &= F_1(0) F_2(0) + F_1(\Delta) F_2(\Delta) - F_1(0) F_2(0) \\ &= F_1(\Delta) F_2(\Delta) \end{aligned}$$

where $F_1(\cdot)$ is the cdf of the random variable $Z_1 \equiv \max\left\{\sum_{i=1}^{k} z_{1i},\ k = 1, 2, \cdots\right\}$, and $F_2(\cdot)$ is similarly understood. From Yu (2007), $F_1(\cdot)$ satisfies a homogeneous Wiener-Hopf equation of the second kind with boundary condition $F_1(-\infty) = 0$, $F_1(\infty) = 1$. So the density of $\frac{\Lambda_\infty}{2}$ on the positive axis is $f_1(x) F_2(x) + F_1(x) f_2(x)$. When the distributions of z_{1i} and z_{2i} are the same, then $F_1(\cdot) = F_2(\cdot)$ and $P\left(\frac{\Lambda_\infty}{2} \leq \Delta\right)$ reduces to $F_1(\Delta)^2$, so the density on the positive axis is $2 f_1(x) F_1(x)$. When $\alpha \leq 1 - p_{10}$, then the critical value $b \geq 0$ is determined by

$$F_1(b) F_2(b) = 1 - \alpha.$$

The power function for $v_0 < 0$ is

$$\begin{aligned} & P\left(\sup_{v\in\mathbb{R}} \left\{D_{v_0}(v) - D_{v_0}(0)\right\} > b\right) \\ = & \sum_{j=0}^{\infty} P\left(\sup_{v\in\mathbb{R}} \left\{D_{v_0}(v) - D_{v_0}(0)\right\} > b, N_2(|v_0|) = j\right) \\ = & \sum_{j=0}^{\infty} \frac{e^{-f_q(\gamma_0)|v_0|} \left(f_q(\gamma_0 |v_0|)\right)^j}{j!} P\left(\sup_{v\in\mathbb{R}} D_{v_0}(v) - \sum_{i=1}^{j} z_{2i} > b\right), \end{aligned}$$

where

$$\begin{aligned} & P\left(\sup_{v\in\mathbb{R}} D_{v_0}(v) - \sum_{i=1}^{j} z_{2i} > b\right) \\ = & P\left(\sum_{i=1}^{j} z_{2i} < -b, \sup_{v\in\mathbb{R}} D_{v_0}(v) = 0\right) + \sum_{k=1}^{\infty} P\left(\sum_{i=1}^{k} z_{1i} - \sum_{i=1}^{j} z_{2i} > b, MaxL = k\right) \\ & + \sum_{k=1}^{\infty} P\left(\sum_{i=1}^{k} z_{2i} - \sum_{i=1}^{j} z_{2i} > b, MaxR = k\right), \end{aligned}$$

$MaxL = k$ means the maximum of $D_{v_0}(v)$ is attained at the kth jump on the left of v_0, and $MaxR = k$ can be similarly understood. There is a recursion form for the above probabilities. Such a recursion solution is a natural extension of Appendix D in Yu (2007). The power on $v_0 > 0$ can be similarly derived. But when $\alpha > 1 - p_{10}$, no such a b exists and the randomized test is used. The power in this case requires the calculation of $P\left(\sup_{v\in\mathbb{R}}\left\{D_{v_0}(v)-D_{v_0}(0)\right\}=0\right)$. Note that for $v_0<0$,

$$\begin{aligned}
& P\left(\sup_{v\in\mathbb{R}}\left\{D_{v_0}(v)-D_{v_0}(0)\right\}=0\right) \\
= & \sum_{k=0}^{\infty} P\left(MaxR=k, N_2\left(|v_0|\right)=k\right) \\
= & \sum_{k=0}^{\infty} \frac{e^{-f_q(\gamma_0)|v_0|}\left(f_q(\gamma_0)\left|v_0\right|\right)^k}{k!} P\left(MaxR=k\right) \\
= & e^{-f_q(\gamma_0)|v_0|} p_{10} + \sum_{k=1}^{\infty} \frac{e^{-f_q(\gamma_0)|v_0|}\left(f_q(\gamma_0)\left|v_0\right|\right)^k}{k!} p_{2k}
\end{aligned}$$

where the second equality follows from the independence between the two events $\{Max=k\}$ and $\{N_2\left(|v_0|\right)=k\}$, and $p_{2k}=P\left(MaxR=k\right)$. A similar result applies to $v_0>0$ with p_{2k} substituted by p_{1k}.