October 2008

Optimal Test for Parameter Instability When the Unstable Process and the Error Distribution are Unknown

Dong Jin Lee
University of Connecticut

University of Connecticut

Department of Economics Working Paper Series

Optimal Test for Parameter Instability When the Unstable Process and the Error Distribution are Unknown

Dong Jin Lee
University of Connecticut

Working Paper 2008-40

October 2008

341 Mansfield Road, Unit 1063
Storrs, CT 06269–1063
Phone: (860) 486–3022
Fax: (860) 486–4463
http://www.econ.uconn.edu/

Abstract

This paper proposes asymptotically optimal tests for unstable parameter process under the feasible circumstance that the researcher has little information about the unstable parameter process and the error distribution, and suggests conditions under which the knowledge of those processes does not provide asymptotic power gains. I first derive a test under known error distribution, which is asymptotically equivalent to LR tests for correctly identified unstable parameter processes under suitable conditions. The conditions are weak enough to cover a wide range of unstable processes such as various types of structural breaks and time varying parameter processes. The test is then extended to semiparametric models in which the underlying distribution in unknown but treated as unknown infinite dimensional nuisance parameter. The semiparametric test is adaptive in the sense that its asymptotic power function is equivalent to the power envelope under known error distribution.

Journal of Economic Literature Classification: C12, C14, C22

Keywords: Adaptation, optimal test, parameter instability, semiparametric modl, semiparametric power envelope, structural break, time varying parameter

1. Introduction

The instability of economic relationships is a common problem and is of central importance in econometric modeling. As a result, there has been substantial literature on testing for parameter instability. (See the review paper by Perron (2006).) A distinctive property of the parameter instability test is that there exists a large variety of ways for un unstable parameter to occur, such as a single structural break, multiple structural breaks, and various unstable time varying parameters. Majority of the tests assume that the unstable processes are correctly specified. However, economic theory or the information on the model provides little knowledge about which specific alternative process to use for the test. Attempts to resolve the problem by deriving optimal tests against a wide range of parameter instabilities are done by Nyblom (1989), and Elliott and Müller (2006). However, Nyblom (1989)'s test is locally most powerful only under the counterfactual assumption that the initial point of the parameter is known. Elliott and Müller (2006)'s test is optimal only in linear regression models with Gaussian error distribution.

Another problem of these tests is that their optimalities are maintained only when the underlying distribution is known, despite it is more likely that the error distribution is incorrectly specified in many data set. The optimal tests work through this problem by providing distribution-free size property to the test, but at the expense of losing efficiency. Unfortunately, no work has been devoted to discovering an efficient parameter instability test under unknown error distribution.

The main contribution of this paper is to propose asymptotically optimal tests under the feasible assumption that both the unstable parameter process and the underlying distribution are not identified. I first derive a test under known error distribution, which is asymptotically equivalent to likelihood ratio tests for correctly identified unstable parameter processes under suitable conditions. The conditions are weak enough to cover a broad set of local unstable processes such as various types of structural breaks and time varying parameter processes as long as they are asymptotically described by the Wiener processes. The test is considered as the generalization of Elliott and Müller (2006) into nonlinear nongaussian models and is consequently equivalent to their optimal invariant test in testing linear regression coefficient with Gaussian error distribution. This setup does not only provide a benchmark for the semiparametric analysis, but also can be applied to models with a rather restricted family of error distribution such as exponential family or generalized t-distribution, in which a finite-dimensional parameter determines the shape of the distribution.

The test is then extended to semiparametric models in which the underlying distribution is unknown but treated as an unknown infinite dimensional nuisance parameter. The suggested test is derived based on the kernel estimate of the score function. As long as the unstable parameter locally follows a mean zero Brownian motion, the semiparametric test is adaptive under mild conditions in that the power is asymptotically equivalent to the parametric power envelope. Consequently, there is no loss of asymptotic power by not knowing the true underlying distribution. Since the seminal work by Bickel (1982), numerous authors have employed adaptation in testing problems. Choi, Hall, and Schick (1996) show that the test based on adaptive estimation is also efficient. Banerjee (2005), and Murphy and der Vaart (1997) examine the property of likelihood ratio tests in semiparametric models. Benghabrit and Hallin (1998), and Hallin and Jurečková (1999) use adaptivity to derive asymptotically efficient tests in AR model. Shin and So (1999), and Ling (2003) use it for unit root tests.

Most research has focused on standard testing problems in which the *locally asymptotic normal (LAN)* property of the class of likelihood is involved. However, the parameter instability test is nonstandard in the sense that the parameter of interest is nonstationary random. Hence, the inference based on LAN is not applicable straightforward to this set-up. Recent research extends the adaptation to such nonstandard settings as *locally asymptotic quadratic (LAQ)* likelihood ratio, in which the quadratic term of the local approximation stays random even in the limit. (See Jeganathan (1995), and Ling and McAleer (2003) for examples.) Jansson (2008) extends the LAQ to a unit root testing problem.

The testing problem in this paper is different from LAQ because the asymptotic randomness does not come from the Fisher information matrix but from the unstable parameter process. The model can be regarded as a weighted average of LAN where the weight function is determined by the measure of the unstable parameter process. One of the main finding in this paper is that this non-standard testing problem is still amenable to adaptation by using extant semiparametric methods developed for standard problems. In this sense, this paper provides an example of the extent to which one can get adaptive tests in models far from LAN.

This paper is organized as follows: Section 2 introduces the model and the hypothesis to be tested. Section 3 studies efficient tests under the assumption that the underlying distribution is known. Section 4 extends the result of section to semiparametric models. Section 5 performs Monte Carlo studies. And Section 6 concludes.

2. The Model and the Breaking Processes

This section defines the model and the test hypothesis. Consider a stochastic process $(y, X) \equiv Z \equiv \{Z_t : \Omega \to \mathbb{R}^{r+1},\ r \in \mathbb{N},\ t = 1, ..., T\}$ defined on a complete probability space $(\Omega, \mathfrak{F}, P)$ where $\mathfrak{F} = \{\mathfrak{F}_t,\ t = 1, ..., T\}$ and $\mathfrak{F}_t$ denotes the smallest σ-algebra that Z_t is adapted to, i.e. $\mathfrak{F}_t \equiv\ \sigma(Z_1, ..., Z_t)$. y_t is an endogenous variable with conditional distribution function $F_t(y) = Pr(Y_t \leq y_t|\mathfrak{F}_{t-1}, X_t)$ and the corresponding conditional density function $f(y_t)$, which is measurable both under the null and the alternative hypotheses. X_t is a vector of explanatory variables with the conditional density $f_X(x_t|\mathfrak{F}_{t-1})$. Consider the model

$$\epsilon_t = \frac{1}{\sigma(X_t, \theta_t, \gamma)} \left(y_t - m(X_t, \theta_t, \gamma)\right) \tag{2.1}$$

where $m(\cdot)$ is a measurable function which is continuous and differentiable with respect to (θ_t, γ). $m(\cdot)$ contains various types of linear and nonlinear times series models but does not consider the nonparametric or partially nonparamtric model because the parameters (θ_t, γ) are finite dimensional. $\theta_t \in \Theta \subseteq \mathbb{R}^p$ is the vector of the parameter of interest, and $\gamma \in \Gamma \subseteq \mathbb{R}^s$ is the vector of nuisance parameters. I split γ into two components, $\theta_0 \in \Gamma_\theta \subseteq \mathbb{R}^p$, and $\gamma_0 \in \Gamma_\gamma \subseteq \mathbb{R}^q, q = s - p$. θ_0 coincides with θ_t so that the parameter vector can be rearranged as $(\theta_0 + \theta_t, \gamma_0)$. and γ_0 are held constant over times. The model corresponds to pure structural breaks or pure time varying parameters if no γ_0 appears on the model in which the whole parameters are subject to be unstable. And the appearance of γ_0 would lead to partial structural break/time varying parameter models. ϵ_t is an error term with a moment restriction and a continuous density $g(\cdot)$. Consequently, the conditional distribution of y_t can be represented as $f_t(y) = (1/\sigma(X_t, \theta_t, \gamma))g((y_t - m(X_t, \theta_t, \gamma))/\sigma(X_t, \theta_t, \gamma))$.

The objective of this paper is to test whether the unstable parameter process $\{\theta_t\}$ presents in the model. Under the null hypothesis of stability, $\{\theta_t\}$ are zeros for all $t = 1, \ldots, T$, such that the parameter vector would be (θ_0, η_0). To examine asymptotic local powers, the alternative hypothesis is considered to be local to the null by assuming that $\{\theta_t\}$ take the form $\theta_t = \frac{1}{T}\delta_t \;\; \forall t = 1, \ldots, T$. Unlike the standard testing problem, the appropriate neighborhood in order for the test to have nontrivial asymptotic power is where θ_t is of order T^{-1} in probability. The reason for this is that the test focuses on alternatives with a persistently varying $\{\theta_t\}$, in that permanent change of the parameter has more implications in both economic and statistic concepts. It is implicit in the formulation that (y_t, X_t), δ_t, and their distributions may depend on T, but I suppress the dependency for the purpose of notational convenience.

The alternative hypothesis is not defined in a single form because there exist a large variety of ways in which θ_t is not stable. Any specific assumption on unstable θ_t would lead to a different alternative hypothesis resulting in a different testing problem. Existing instability tests can be categorized into two big streams based on types of unstable processes: One is the test of structural breaks, and the other is the test of time varying parameters. Structural break tests consider a model in which θ_t permanently shift N times in a sample period. For a single break example, the parameter vector equals (θ_0, η_0) for $t = 1, \ldots, \tau$ and $(\theta_0 + \bar{\theta}, \eta_0)$ for $t = \tau + 1, \ldots, T$. Time varying parameter tests posit random process of θ_t. Even within time-varying parameter approaches there are many possible alternatives based on the distributional properties of θ_t. However, economic theory generally does not provide enough information to pick a specific alternative process. Consequently, an alternative process are often arbitrarily chosen depending on what the researcher has in mind.

Elliott and Müller (2006), and Nyblom (1989) get around the problem by providing only minimal identifying conditions on the unstable process. Nyblom (1989) assumes that the unstable processes is martingale. Elliott and Müller (2006) consider any processes which are asymptotically described by the Wiener processes. Their idea is that the seemingly different approaches of structural breaks and time varying parameters are in fact not distinctive. Both are considered as specific forms of a unified framework. For example, if we let $\Delta\theta_t$ have a continuous distribution with probability p and equal zero with probability $(1-p)$, then this time varying parameter process is reduced to a multiple structural break with $(T \cdot p)$ expected breaks. On the other hand, it becomes a random walk if $\Delta\theta_t$ is *iid* normal. However, Nyblom (1989)'s test is locally most powerful only under the counterfactual assumption that θ_0 is known, and Elliott and Müller (2006)'s optimality is restricted only to linear regression models with Gaussian error distribution. One of the main finding in this paper is that Elliott and Müller (2006)'s optimality can be extended to more general circumstance with nonlinear function and non Gaussian error distribution. Specifically, I consider unstable processes that satisfy the following condition.

Condition 1. *i)* $\{T\Delta\theta_t\}$ *is uniform mixing with mixing coefficient of size -r/(2r-2) or strong mixing of size -r/(r-2), r>2*

ii) $E[\Delta\theta_t] = 0$ *and* $E[|T\Delta\theta_{t,i}|^r] < K < \infty$ *for all t=1,...,T, and i=1,...,k.*

iii) $\{T\Delta\theta_t\}$ *is globally covariance stationary with nonsingular long-run covariance matrix,* Ω

iv) The initial value of $\{\theta_t\}$ *satisfies* $\frac{1}{T}\sum_{t=1}^{T}\theta_t = 0$

Condition 1 i) to iii) are conditions for the heterogeneous mixing FCLT in White (2001). Admitting both heteroscedasticity and dependency makes Condition 1) capture many possible persistent breaking processes such as multiple breaks, clustered breaks, regularly occurring breaks, or smooth changing parameters.(See Elliott and Müller (2006) for details.) Any other conditions for FCLT could replace them.

Part (iv) of Condition 1) is necessary to identify the process $\{\theta_t\}$. It implies that the average value of the unstable parameter path is always the same as that under the stable model. Consequently, the test in this set-up detects permanent variation in the parameter, rather than differences between the average value of the parameters. Another benefit of this condition is that it provides the best power under unknown θ_0 in the sense of least favorable parametric submodels. θ_0 is unknown in practice and should be replaced by an estimator, which generally causes the loss of power. Condition iv) plays the role of the least favorable direction of the hypothesis, in which the test has the minimal loss by unknown θ_0. In order to construct the likelihood ratio, we need additional assumptions on $\{\epsilon_t\}$ and $\{X_t\}$ as follows.

Condition 2. *i)* ϵ_t *is iid and conditionally independent of* X_t *given* $\mathfrak{F}_{t-1}$. *The error distribution* $\{g(\epsilon_t)\}$ *does not depend on* θ_t *in the null hypothesis.*

ii) γ *is an interior point of* Γ.

iii) X_t *has conditional distribution* $f_X(X_t|\Im_{t-1})$ *with respect to some* σ*-finite measures,* $\{f_X(X_t|\Im_{t-1})\}$ *does not depend on parameters* γ *and* θ_t *for all* $t = 1, \ldots, T$.

iv) Under H_0, $\{X_t\}$ *are mixing with either* ϕ *of size -r/2(r-1),* $r \geqq 2$ *or* α *of size -r/(r-2), r> 2.*

v) Under H_0*,* $E\left[|X_{t,i}|^r\right] < \Delta < \infty$ *for all* $t = 1, \ldots, T$ *and* $i = 1, \ldots, k$*.* $T^{-1}\sum_{t=1}^{[sT]} \dot{m}(X_t)\dot{m}(X_t)' \rightarrow sJ_m$ *uniformly in* s *where* $\dot{m}(\cdot)$ *is the 1st derivative of* $m(\cdot)$ *with respect to* θ_t*, and* $J_m = E[\dot{m}(X_t)\dot{m}(X_t)']$*.* $T^{-1}\sum_{t=1}^{T} \dot{m}(X_t)\dot{m}(X_t)'$ *is uniformly positive definite.*

The *iid* condition in Condition 2 i) is crucial in order to obtain optimality. However, it can be extended to the non *iid* case in which some finitely parameterized transformation of the data leads back to the *iid* model such as (non)stationary ARMA (Akharif and Hallin (2003)), GARCH (Drost and Klaassen (1997), Ling and McAleer (2003)), and quantile ARCH (Koenker and Zhao (1996)) Models. Moreover, the suggested test would still be correct asymptotic size even though ϵ_t is not iid, if the partial sum of the score function satisfies some asymptotic properties. Condition 2 ii) is the standard maximum likelihood condition. It is also required for the null distribution to be contiguous to the alternative distribution. Condition 2 iii) implies that $\{X_t\}$ is weakly exogenous in the sense of Engle, Hendry, and Richard (1983). In such circumstance, $f_X(\cdot)$ need not be known in order for one to construct a likelihood ratio function. Condition 2 iv) and v) are the condition for the heterogeneous CLT for $\{X_t\}$.

3. **Asymptotically Optimal Tests in Parametric Models**

3.1. **The Optimal Test Function.** This section derives an asymptotically efficient test for condition 1) unstable processes under the assumption that the underlying distribution is known. This paper defines the optimal test based on Neyman-Pearson lemma so that the likelihood ratio function or its equivalents are defined to be optimal. The likelihood ratio in this set-up depends on two unknown density functions; the densities of $\{\theta_t\}$, and $\{\epsilon_t\}$. This section focuses on the first part by assuming that the latter density is known, and the latter part will be considered in the next section. The test function in this section would not only provide a benchmark for tests under more realistic distributional assumptions by providing the upper bound of their asymptotic power envelopes. It is also worthwhile itself in the sense that many researches are likely to use parametric model by choosing a specific family of error distributions for various reasons. For simplicity purpose, I assume that γ_0 does not appear on the model so that all the parameters are subject to be unstable. The models with γ_0

are to be considered in the next section. Under Condition 1) and 2), the conditional density of the data under H_0 is

$$f_0(y, X|\theta_0) = \prod_{t=1}^{T} f(y_t|\theta_0, X_t, \Im_{t-1}) f_X(X_t|\Im_{t-1}) \tag{3.1}$$

The conditional density under the alternative hypothesis is

$$f_1(y, X|\theta_0, \theta_t) = \int \prod_{t=1}^{T} f(y_t|\theta_0, \theta_t, X_t, \Im_{t-1}) f_X(X_t|\Im_{t-1}) d\nu_\theta \tag{3.2}$$

where ν_θ is the measure of $\theta = (\theta_1', \ldots, \theta_T')'$. If ν_θ is known, the Neymann-Pearson Lemma implies that rejecting H_0 for a large value of the likelihood ratio statistic, defined as

$$LR_T = \int \prod_{t=1}^{T} \frac{f(y_t|\theta_0, \theta_t, X_t, \Im_{t-1})}{f(y_t|\theta_0, X_t, \Im_{t-1})} d\nu_\theta \tag{3.3}$$

has the best power against the alternative distribution (3.2). Most optimal tests for parameter instability are manipulations of (3.3) that make the test feasible and easy to compute. Since LR_T depends on $f(y|\cdot)$ and ν_θ, the different types of optimal test statistics might come from the choice of ν_θ and $f(y|\cdot)$.

The likelihood ratio (3.3) is infeasible to use in practice because ν_θ is unknown. Even though ν_θ is specified, it has an integral in its form which makes the computation too complicated to be used in practice. The method proposed in this section resolves the problem by suggesting another easy-to-compute test function, $\hat{B}(\Omega)$, which is asymptotically equivalent to LR_T, but does not depend on ν_θ other than Ω.

In order to define the feasible optimal test function, I introduce some notations and definitions. Let $\dot{\ell} = (\dot{\ell}_1', \ldots, \dot{\ell}_T')'$ be the first derivative of the log likelihood function, and $J_\theta = E[\dot{\ell}_t \dot{\ell}_t']$. And let $\Omega^* = J_\theta^{\frac{1}{2}} \Omega J_\theta^{\frac{1}{2}}$. I decompose Ω^* into the orthonormal matrix of its eigenvectors, P, and the diagonal matrix of the eigenvalues, $\Lambda = diag(a_1^2, \ldots, a_k^2)$, such that $P\Lambda P' = \Omega^*$ and $a_i > 0, \forall i$. Let I_T be the $T \times T$ identity matrix, and e be the $T \times 1$ vector of ones. The first derivative normalized to have unit variance and zero covariance can be written as $\dot{\ell}^*(\hat{\theta}) = (I_T \otimes P' J_\theta^{-1/2}) \dot{\ell}(\hat{\theta})$ or $\dot{\ell}_t^*(\hat{\theta}) = P' J_\theta^{-1/2} \dot{\ell}_t(\hat{\theta})$. Furthermore, define $\dot{\ell}_{i,t}^*$ be the i^{th} element of $\dot{\ell}_t^*(\hat{\theta})$ and ζ_i be the vector of the partial sum of $\dot{\ell}^*_{i,t}$, i.e. j^{th} element of ζ_i is $\sum_{t=1}^{j} \dot{\ell}^*_{i,t}$. The test statistic I suggest is

$$\hat{B}(\Omega) = \sum_{i=1}^{k} \zeta_i'(\widehat{\theta}, \widehat{J}_\theta)' \left[\frac{a_i^2}{T^2} I_T - F M_e F' \right]^{-1} \zeta_i(\widehat{\theta}, \widehat{J}_\theta) \tag{3.4}$$

where $M_e = I_T - \frac{1}{T} ee'$, $F = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \dots & \dots & 1 \end{pmatrix}$, and $\widehat{\theta}$ and $\widehat{J}_\theta$ are the maximum likelihood estimators under H_0. $\hat{B}(\Omega)$ does not have the integral so that the computation is tractable. Note that $\hat{B}(\Omega)$ depends on the distribution of $\{\theta_t\}$ only through the eigenvalues of Ω. Consequently, proving the optimality of $\hat{B}(\Omega)$ implies that the the knowledge of the unstable process other than Ω is asymptotically irrelevant under the suggested conditions.

I now present the outline of the proof of the optimality of $\hat{B}(\Omega)$. The proof requires several steps. First, I focus on the integrand of LR_T, denoted by L_T, to suggest an asymptotically equivalent formula. At this time I assume that θ_0 is known. Second, I give an alternative formula which weakly converges to L_T. A test function $B(\Omega)$ is then provided based on the alternative of L_T. Third, I show that the weak convergency in the second step is sufficient for the asymptotic equivalence of LR_T and $B(\Omega)$ both under H_0 and H_1. Finally, I replace θ_0 by its maximum likelihood estimator to make $\hat{B}(\Omega)$ and show that $\hat{B}(\Omega)$ converges in probability to $B(\Omega)$.

3.2. **Asymptotic Optimality of the Test Statistic.** First, I simplify the integrand of LR_T. Since the integrand can be regarded as the likelihood ratio for specific values of alternative parameters, θ, a simple and powerful method for simplification is to use the Taylor expansion of the logarithm of the likelihood. However, it can be made rigorous under moment or continuity conditions on the 2nd derivative of the log likelihood that many distributions do not satisfy. Consequently, I impose an alternative single condition that only involves a first derivative, i.e. the square roots of $f(\cdot)$ correspond to unit vectors in space of square integrable functions, as follows.

Condition 3. *Let $\xi_t(\cdot|\theta_0,\theta_t)$ be the square root of the density, $f(\cdot)$. Under H_0,*
1) There exists a $k\times 1$ random vector $\dot{\xi}^\theta_t(\cdot|\theta_0,\theta_t)$ such that $\mathbb{E}\|\dot{\xi}^\theta_t(\cdot|\theta_0,\theta_t)\|^2<\infty$, and

$$\mathbb{E}\left(\left[\left(\frac{\xi_t(\cdot|\theta_0,h)}{\xi_t(\cdot|\theta_0,0)}-1\right)-h'\frac{\dot{\xi}^\theta_t(\cdot|\theta_0,0)}{\xi_t(\cdot|\theta_0,0)}\right]^2\right)\to 0 \text{ as } \|h\|\longrightarrow 0,\quad \forall t\leq T \tag{3.5}$$

2) $J_\theta(s)=\frac{1}{T}\sum_{t=1}^{[sT]}4\frac{\dot{\xi}^\theta_t(\cdot|\theta_0,0)\dot{\xi}^\theta_t(\cdot||\theta_0,0)'}{\xi_t(\cdot|\theta_0,0)^2}\longrightarrow sJ_\theta$
for any $s\in[0,1]$ and $J_\theta(1)$ is positive definite for all t

The derivative $\dot{\xi}_t(\cdot,\theta_t)$ is called Hellinger derivative, and the score function $\dot{\ell}^\theta=(\dot{\ell}^\theta_1(\theta_0),\ldots,\dot{\ell}^\theta_T(\theta_0))$, is then defined as $\dot{\ell}^\theta_t(\theta_0)=2\frac{\dot{\xi}^\theta_t(\cdot|\theta_0,0)}{\xi_t(\cdot|\theta_0,0)}$. Part (1) of Condition 3, called *differentiability in quadratic mean* (DQM), is weak enough to be satisfied by a wide variety of densities and strong enough to deliver the approximation similar to the Taylor expansion. Local asymptotic approximation of a likelihood ratio statistic under Condition 1) is widely developed in standard testing problems (LeCam (1970)) and nonstandard problems (Jeganathan (1995), Ling and McAleer (2003), and Jansson (2008)). The set up in this paper is different in that the square of $\{\theta_t\}$ stays random even in the large sample. The following lemma shows that the similar quadratic approximation is possible in this set up.

Lemma 1. *Let $\dot{\ell}^\theta=(\dot{\ell}^\theta_1(\theta_0)',\ldots,\dot{\ell}^\theta_T(\theta_0)')'$ be the score function based on the Hellinger derivative. Under Condition 1) to 3),*

$$L_T=(1+o_p(1))exp\left[\dot{\ell}^{\theta\prime}(M_e\otimes I_k)\theta-\frac{1}{2}\theta'(M_e\otimes J_\theta)\theta\right] \tag{3.6}$$

This approximation can be considered as a *locally asymptotic quadratic (LAQ)* approximation defined by Jeganathan (1995) in the sense that the quadratic term is random because of the random $\{\theta_t\}$, and the null and the alternative distribution is contiguous, which is shown in Theorem 1). But it is different from standard concept of LAQ because the information function J_θ is nonrandom. Accordingly, I denote (3.6) by LAQ^*.

As a next step, I deal with the main problem that ν_θ is unknown. I replace $\{\theta_t\}$ by another random sequence $\{\widetilde{\theta}_t\}$ and show that L_T based on $\{\widetilde{\theta}_t\}$ weakly converges to the same limit as that of L_T of Condition 1) $\{\theta_t\}$. The random vector $\widetilde{\theta}=(\widetilde{\theta}'_1,\ldots,\widetilde{\theta}'_T)'$ is defined as a multivariate random walk process, i.e. $T\Delta\widetilde{\theta}_t\sim iid\ N(0,\Omega)$. (See

Lemma 6 in the appendix.) Using Lemma 1, it can be shown that, if $\theta = \widetilde{\theta}$, LR_T is asymptotically equivalent to

$$\widetilde{LR}_T = \int exp\left[\dot{\ell}'(M_e \otimes I_k)\widetilde{\theta} - \frac{1}{2}\widetilde{\theta}'(M_e \otimes J_\theta)\widetilde{\theta}\right] d\nu_{\widetilde{\theta}} \tag{3.7}$$

The advantage of replacing $\{\theta_t\}$ by $\{\widetilde{\theta}_t\}$ is that the integral is easily calculated because both the integrand $\widetilde{L}_T$ and $d\nu_{\widetilde{\theta}}$ are of exponential quadratic form. Through some matrix manipulations, we get the following lemma.

Lemma 2. *Let $B(\Omega)$ be defined as (3.9) with replacing ζ_i with $\zeta_i^* = \sum_{t=1}^{j}[\dot{\ell}^*_{i,t}(\theta_0) - \frac{1}{T}\sum_{\tau=1}^{T}\dot{\ell}^*_{i,\tau}(\theta_0)]$*

$$B(\Omega) = \frac{1}{2}ln\widetilde{LR}_T + c$$

where $c = -\sum_{i=1}^{k} log\left(\frac{2a_i \exp[-a_i]}{1-\exp[-2a_i]}\right)$ is constant.

Lemma 2 implies that $B(\Omega)$ is asymptotically optimal if $|\widetilde{LR}_T - LR_T|$ converges to zero in probability both under H_0 and H_1. Note that the integrands of LR_T and $\widetilde{LR}_T$ weakly converge to the same limit. Theorem 1 shows that the weak convergency is enough for the L_2 convergence in this set up, using the fact that a weakly converging uniformly bounded sequence L_1 converges.

Let $\phi_T(Z|\Omega,\theta_0)$ be a critical function for any of Condition 1) processes. That is, $\phi_T(Z|\Omega,\theta_0)$ is a $[0,1]$ valued function determined by Z. I consider an asymptotically α-significant test, i.e. $\lim_{T\to\infty}\int \phi_T(Z|\Omega,\theta_0)f_0(Z|\theta_0)\,dZ = \alpha$. The power function of the test is defined as $\int \phi_T(Z|\Omega,\theta_0)f_1(Z|\theta_0)dZ$. The following theorem gives the optimality of $B(\Omega)$.

Theorem 1. *Let $\psi_T(Z|\Omega)$ be a critical function for $B(\Omega)$ and $\Psi(\Omega)$ be the asymptotic power function of $\psi_T(Z|\Omega)$, i.e. $\Psi(\Omega) = \underline{\lim}_{T\to\infty}\int \psi_T(Z|\Omega)f_1(Z|\theta_0)dZ$. Under Conditions 1) to 3),*

$$\overline{\lim}_{T\to\infty}\int \phi_T(Z|\Omega)f_1(Z|\theta_0)dZ \leq \Psi(\Omega)$$

Theorem 1 implies that the powers of optimal tests do not depend on the particular distributional form of θ_t other than its global covariance matrix, Ω. It leads to the following implication: The knowledge of the exact distribution of the unstable process is asymptotically inappropriate for conducting an optimal test, as long as the process satisfies Condition 1). As T increases, there is little loss of power by using $B(\Omega)$ rather than tailored LR_T.

As a next step, I replace θ_0 and J_θ with $\hat{\theta}$ and $\hat{J}_\theta$ in order to make $B(\Omega)$ feasible. It is known that a test generally loses its power if the true nuisance parameters are replaced by their estimators. An interesting finding, however, in this paper is that $B(\Omega)$ does not lose any asymptotic power even though the estimators $\hat{\theta}$, and $\hat{J}$ are plugged into $B(\Omega)$ as long as the estimator satisfies some regularity conditions given below. The reason is that Condition 1) iv), $\frac{1}{T}\sum_{t=1}^{T}\theta_t = 0$, makes a role to provide the *least favorable hypothesis* in which the perturbation of unknown θ_0 is defined around the alternative value, and the best power function, say $\bar{\phi}(\theta_t, \theta_0^1)$, is defined as $\bar{\phi}(\theta_t, \theta_0^1) = \inf_{\theta_0} \phi(\theta_t, \theta_0^1; \theta_0)$, where $\phi(\theta_t, \theta_0^1; \theta_0)$ represents the power envelope of any asymptotically size α tests for given θ_0^1 and θ_0, where the superscript of θ_0 means the value of θ_0 under H_0 and H_1.

It is well known that the infimum can be achieve geometrically in the standard testing by projecting the score function of θ_t, $\dot{\ell}$ onto the linear subspace V generated by all possible score function for the nuisance parameter,$\dot{\ell}_0$, which implies that the perturbation of θ_0 lies on the orthogonal complement of V, i.e. $\theta_0^1 = \theta_0^0 - E[\dot{\ell}_0\dot{\ell}_0']^{-1}E[\dot{\ell}_0\dot{\ell}']\theta$ in the local area. It can be shown that Condition 1) iv) can be reinterpreted as the local restriction for θ_0, and implies the above orthogonality. Theorem 2 shows that the orthogonality still provides the infimum in such nonstandard testing problem as in this set-up, if $\hat{J}_\theta$ and $\hat{\theta}$ satisfy the following condition.

Condition 4. *Under the null hypothesis,*

(1) $T^{-1/2}\sum_{t=1}^{[sT]}\dot{\ell}_t(\theta_0 + T^{-1/2}\delta_0) = T^{-1/2}\sum_{t=1}^{[sT]}\dot{\ell}_t(\theta_0) - sK(\theta_0)\delta_0 + o_p(1)$
(2) $\sqrt{T}(\hat{\theta} - \theta_0) = O_p(1)$ *and* $\hat{J}_\theta = J_\theta + o_p(1)$

uniformly for $s \in (0,1)$, *and any nonrandom* $K(\theta_0)$, *where* $\|\delta_0\| < M < \infty$.

Consider a class of tests that have limiting size α for all values at $\sqrt{T}-$neighborhood of θ_0, i.e. $\lim_{T\to\infty}\int \hat{\phi}_T(Z|\Omega, \theta_0) f_0(Z|\theta_0 + \frac{1}{\sqrt{T}}\delta_0)dZ = \alpha$ for every δ_0 where $\|\delta_0\| < M < \infty$, and $\hat{\phi}_T(Z|\Omega, \theta_0)$ is the critical function of the test. The following theorem shows that $\hat{B}(\Omega)$ has the best asymptotic power among asymptotically similar size tests.

Theorem 2. *(1) Under Condition 1) to 4),*

$$\hat{B}(\Omega) = B(\Omega) + o_p(1) \qquad \textit{under } H_0 \textit{ and } H_1$$

(2) Let $\hat{\psi}_T(Z|\Omega)$ be a critical function for $\hat{B}(\Omega)$. Under Conditions 1) i), ii), iii) 2), 3), 4),

$$\overline{\lim_{T\to\infty}} \int \hat{\phi}_T(Z|\Omega) f_1(Z|\theta_0, \delta_0) dZ \leq \underline{\lim_{T\to\infty}}\, t \int \hat{\psi}_T(Z|\Omega) f_1(Z|\theta_0, \delta_0) dZ$$

The first part of Theorem 2 implies that there is no loss of asymptotic power of $B(\Omega)$ even though the true θ_0 is replaced by the estimator. The second part comes straight from the first part in the sense that $\Psi(\Omega)$ can be no worse than the asymptotic power envelope of the tests under unknown θ_0. It implies that the test has asymptotically the best power among all tests that are asymptotically size-α with unknown θ_0. The asymptotic null distribution of $\hat{B}(\Omega)$ is given in the following lemma.

Lemma 3. *Under Condition 1) to 4), the asymptotic null distribution of $\hat{B}(\Omega)$ is*

$$\begin{aligned} \hat{B}(\Omega) \quad \longrightarrow \quad & \sum_{i=1}^{k} [a_i J_i(1)^2 + a_i^2 \int_0^1 J_i(s)^2 ds + \\ & \frac{2a_i}{1-e^{-2a_i}} \{e^{-a_i} J_i(1) + a_i \int_0^1 e^{-a_i s} J_i(s) ds\}^2 - \{J_i(1) + a_i \int J^i(s) ds\}^2] \end{aligned} \tag{3.8}$$

where $J_i(s) = W_i(s) - sW_i(1) - \int_0^s e^{\lambda - s}[W_i(\lambda) - \lambda W_i(1)] d\lambda$, and W_i is the i_{th} element of the independent $k \times 1$ standard Wiener process W.

Note that $\hat{B}(\Omega)$ is derived based on the assumption that Ω is known, which is unobservable in practice. Consequently, there is no uniformly most powerful test in this framework. Instead, if we focus on one point in the alternative parameter space, we can find a most powerful test in the neighborhood of the predetermined point. Such a test is called a *point optimal test*. (see King (1988), and Nyblom (1986) for details.) Following this idea, I choose $\hat{\Omega}$ such that $J_\theta^{1/2}{}'\hat{\Omega} J_\theta^{1/2} = c^2 I_k$ where $c = 10$ as Elliott and Müller (2006). Replacing with $\hat{\Omega}$, the point optimal test statistic, $\hat{B}(\hat{\Omega})$, is given by

$$\hat{B}(\hat{\Omega}) = \sum_{i=1}^{k} \zeta_i'(\widehat{\theta}, \widehat{J}_1)' \{\frac{c^2}{T^2} I_T - F M_e F'\}^{-1} \zeta_i(\widehat{\theta}, \widehat{J}_1) \tag{3.9}$$

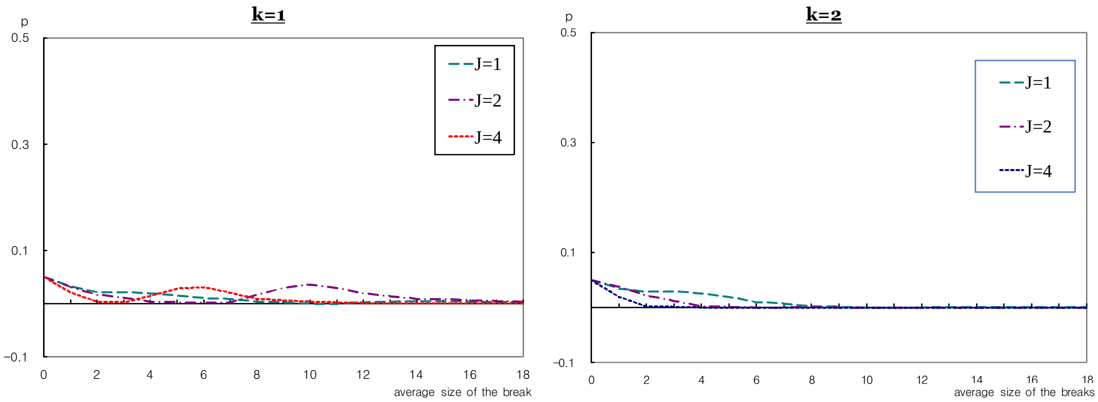


FIGURE 1. **Asymptotic power loss of the Point Optimal Test $\hat{B}(\hat{\Omega})$**
note) Powers are plotted from 20,000 draws using 2,000 standard normal steps to approximate Wiener Processes. Power losses are calculated as the difference between $\Psi(\hat{\Omega})$ and $\Psi(\Omega)$. The average size of the breaks are measured by the eigenvalue of Ω.

Selected asymptotic upper tail percentiles of $\hat{B}(\hat{\Omega})$ are calculated by Elliott and Müller (2006). In addition to the simplicity, using C also has merit because it enables $\hat{B}(\hat{\Omega})$ to be invariant with respect to re-parameterizations. Since $\dot{\ell}_i^*(\hat{\theta})$ does not change to any parameterization and $\{I_T - \frac{100}{T^2}FM_eF'\}$ is constant, we immediately observe that $\hat{B}(\hat{\Omega})$ is invariant to reparameterization. The invariance may be reinterpreted as meaning that the direction of breaks under the alternative should not affect the outcome of the test. Figure 1 shows the loss of asymptotic power by using $\hat{\Omega}$ rather than the true Ω. For both k=1, and k=2, the power loss does not exceed 5 % p in any levels of J_θ which would imply that the unknown Ω is a minor problem.

4. **Asymptotically Optimal Tests in Semiparametric Models**

The optimal test $B(\Omega)$ is based on the counterfactual assumption that $f(y_t|\cdot)$, or equivalently $g(\epsilon_t)$, is correctly specified. This section extends the previous result by investigating asymptotically efficient tests under unknown $g(\cdot)$ in which $g(\cdot)$ is treated as an unknown infinite-dimensional nuisance parameter. This relaxation modifies the model in the previous section into the semiparametric one with a real valued parametric component $\theta = (\theta_0', \theta_1', \ldots, \theta_T')' \in R^{k(T+1)}$, and a single nonparametric component $g \in \mathcal{G}$ which denotes the unknown distribution of the error term, where $\mathcal{G}$ is the collection of all probability measures on the sample space. I first consider the case where g is known to belong to a specific parametric family of distribution indexed by a finite dimensional parameter $\eta \in \Upsilon$, and suggest conditions under which the asymptotic power envelope is equivalent to $\Psi(\Omega)$. The model is then extended to semiparametric ones in which η is infinite dimensional.

4.1. **The Optimal Test with A Finite Dimensional Nuisance Parameter.** The true set of conditional densities of y_t is characterized as a parametric family $\mathcal{P}_\eta = \{F_t(y|\theta,\eta) : \theta \in R^{k(T+1)}, \eta \in R^q\}$ with dominating measure μ and corresponding densities $f_t(y|\eta) = dF_t(y|\eta)/dy$ such that $g(\epsilon_t) = \sigma_t f(y_t|\eta)$. The model with this parametrization is called a parametric submodel.

The parametric submodel has its own relevancy to empirical analysis as well as provides a inference on semiparametric analysis. A familiar case is testing partial structural breaks in which θ are suspected to have structural breaks while η remain constant($\eta = \gamma_0$). Another case occurs when testing stability of the coefficient of a linear regression model, in which ϵ is from a generalized family of distribution, such as an asymmetric exponential family, where unknown η determines the shape of the distribution.

In this section, I confine my attention to contiguous alternatives for η. Define a $\sqrt{T}$ neighborhood of the true nuisance parameter η_0 as $\eta = \eta_0 + \frac{1}{\sqrt{T}}h$ for bounded $h \in \mathcal{H}_\theta$ where $\mathcal{H}_\theta$ is a *Hilbert space*. In order to ensure that the asymptotic power envelope covers the unknown perturbation of the nuisance parameter h, we need an additional restriction to the test. One widespread way is to confine *asymptotically similar tests* $\phi_T(Z)$ which have the invariant asymptotic size regardless of h, i.e. for a fixed $\alpha > 0$, $\lim_{T\to\infty} \int \phi_T(Z) f_0(Z|h)dZ \leq \alpha$ for every h. This requirement is crucial and plays the role of restriction to regular estimates in estimation theory. (see Hall and Mathiason (1990) for details.) Following the way I analyzed the previous section, my investigation is based on the LAQ^* of the integrand. The likelihood ratio function associated with $\mathcal{P}_\eta$ is

$$LR_T^S = \int \prod_{t=1}^{T} \frac{f(y_t|\theta_0 + \theta_t, \eta_0 + \frac{1}{\sqrt{T}}h)}{f(y_t|\theta_0, \eta_0)} d\nu_\theta \tag{4.1}$$

Analogous to the parametric model case, we need a differentiability condition for $f(\cdot|\eta)$ as follows.

Condition 3') *Let* $\xi_t^S(\cdot|\theta_t,\eta)$ *be the square root of the error density,* $f(y_t|\theta_t,\eta)$ *and* b *be the* $(k+1)\times 1$ *vector. Define* $\theta_t^\eta = (\theta_t', \eta')'$, *and* $\theta_0^\eta = (0', \eta_0')'$. *Under* H_0,

1) There exists a $(k+1)\times 1$ *random vector* $\dot{\xi}_t^S(\cdot,\theta_t^\eta) = \left(\dot{\xi}_t^{\theta\prime}, \dot{\xi}_t^{\eta\prime},\right)'$ *such that* $\mathbb{E}_\theta \|\dot{\xi}_t^S(\cdot,\theta_t^\eta)\|^2 < \infty$ *and*

$$\mathbb{E}\left(\left[\left(\frac{\xi_t^S(\cdot,\theta_0^\eta + b)}{\xi^S(\cdot,\theta_0^\eta)} - 1\right) - b'\frac{\dot{\xi}_t^S(\cdot,\theta_0^\eta)}{\xi_t^S(\cdot,\theta_0^\eta)}\right]^2\right) \to 0 \text{ as } \|b\| \longrightarrow 0$$

2) $$J^S(s) = \tfrac{1}{T}\textstyle\sum_{t=1}^{[sT]} 4\frac{\dot{\xi}_t^S(\cdot,\theta_t^\eta)\dot{\xi}_t^S(\cdot,\theta_t)'}{\xi_t^S(\cdot,\theta_t^\eta)^2} \longrightarrow sJ^S$$

for some positive definite nonrandom $(k+q)\times(k+q)$ *matrix function* J^S *and for any* $s\in[0,1]$ *and* $J^S(1)$ *is positive definite for all* t

$\{\dot{\xi}_t^S(\cdot,\theta_t^\eta)\}$ is still a function of θ_0 but I suppress the dependency for the purpose of convenience. Lemma 4 gives LAQ^* of the integrand in LR_T^S. For the simplicity, I assume that h is scalar.

Lemma 4. *Let's define* $\dot{\ell}_t^\eta = 2\frac{\dot{\xi}_t^\eta(\theta_0^\eta)}{\xi_t^S(\theta_0^\eta)}$, *and* J_η *is the lower right* $q\times q$ *part of* J. *Under Condition 1), 2) and 3)', the integrand of* (4.1), *denoted by* L_T^S, *is equivalent to*
(4.2)
$$L_T^S = (1+o_p(1))\exp\left[\dot{\ell}^{\theta\prime}(M_e\otimes I_k)\theta - \frac{1}{2}\theta'(M_e\otimes J_\theta)\theta\right]\cdot\exp\left[\frac{h}{\sqrt{T}}\sum_{t=1}^T \dot{\ell}_t^\eta - \frac{h^2}{2}J_\eta\right]$$

Using (4.2), it can be shown that LR_T^S is asymptotically equivalent to

$$\widetilde{LR}_T^S = \int exp\left[\dot{\ell}^{\theta\prime}(M_e\otimes I_k)\theta - \frac{1}{2}\theta'(M_e\otimes J_\theta)\theta\right]d\nu_\theta\cdot\exp\left[\frac{h}{\sqrt{T}}\sum_{t=1}^T \dot{\ell}_t^\eta - \frac{h^2}{2}J_\eta\right] \tag{4.3}$$

Note that the integral part in (4.3) is the same as LR_T except $\dot{\ell}^\theta$ depends on η_0. Throughout deriving the power envelope, I act as if η_0 is known, and then show that the asymptotic power envelope is attainable by replacing η_0 by its consistent estimator.

The method used to derive the asymptotic power envelope in this setup exploits the concept of the limits of experiments. An implication in the limits of experiments is that if a sequence of experiments converges to a limit experiment, the best asymptotic power function is the best power function in the limit experiment. In such cases as the existence of the nuisance parameter, finding the power envelope of the limit experiment is much easier than using a classical method. The asymptotic null distribution of $log(LR_T^S)$ is

$$log(LR_T^S) \to^d \Lambda^S(\Omega, h) = c + \Lambda(\Omega) + hW^\eta(1) - \frac{h^2}{2} J_\eta \tag{4.4}$$

where $c = -\sum_{i=1}^k log\left(\frac{2a_i \exp[-a_i]}{1-\exp[-2a_i]}\right)$, $\Lambda(\Omega)$ is the limiting counterpart of $B(\Omega)$ in the parametric model, and W^η is a multivariate brownian motion with variance J_η. Since the convergence holds for all subset I where $\theta \in I \subset \Theta \times \Upsilon$, the sequence of the models converges to a limit experiment so that we can focus on the power envelope of the limit experiment.

The power envelope under non zero h, $\Psi^S(\Omega) = \sup E\left[\phi(Z)exp\left(\Lambda^S(\Omega, h)\right)\right]$ is generally less than the parametric case $\Psi(\Omega)$, because the the former does not achieve $g(\cdot)$ as long as h is nonzero. However, Theorem 3 below shows the interesting result that $\Psi^S(\Omega) = \Psi(\Omega)$ in this set-up. The intuition is as follows; $\Lambda^S(\Omega, h)$, or equivalently $\widetilde{LR}_T^S$, is factored into two parts, of which the second part, $hW^\eta(1) - h^2/2J_\eta$, does not depend on θ. The power function is determined only by the first part, and the asymptotic size restriction is imposed only to the second part. Consequently, the test based on the first part, $c + \Lambda(\Omega)$, is expected to provide the power envelope, while it avoids the size dependency of unknown h. Since $\Lambda(\Omega)$ is equivalent to the limit experiment of $B(\Omega)$, it is possible to construct a test based on $\Lambda(\Omega)$, that hits the parametric power envelope $\Psi(\Omega)$. Let's define the limit power function Ψ^S as

$$\Psi^S(\Omega) = E\left[1_{\{\Lambda(\Omega) > k_h^\alpha\}} exp\left(\Lambda^S(\Omega, h)\right)\right] \tag{4.5}$$

where k_h^α is the continuous function that ensures $E\left[1_{\{\Lambda(\Omega) > k_h^\alpha\}}\right] = \alpha$. The following theorem proves the argument.

Theorem 3. *Under Conditions 1), 2) and 3'), any asymptotic similar test function $\phi_T(Z|\Omega, \eta)$ associated with $\mathcal{P}_\eta$ satisfies*

$$\overline{\lim_{T\to\infty}} \int \phi_T(Z|\Omega, \eta) f_1(Z|\theta, \eta) dZ \leq \Psi^S(\Omega) = \Psi(\Omega) \tag{4.6}$$

Theorem 3 implies that it is possible not to lose any power even though we do not know the true value of η_0, as T gets large. The intuition is because θ_t is invariant to the parametric transformation in a locally linearized neighborhood. In general, the invariance property implies that the likelihood function is represented as a function of $\dot{\ell}_t^{\theta}$ only through its effective score function, which is defined as $\dot{\ell}_t^{\theta e} = \dot{\ell}_t^{\theta} - J_{\theta\eta}J_{\eta}^{-1}\sum_{i=1}^{T}\dot{\ell}_i^{\eta}$, where $J_{\theta\eta} = E[\dot{\ell}_t^{\theta}\dot{\ell}_t^{\eta}]$. $\dot{\ell}_t^{\theta e}$ lies on the orthonormal complement of the space spanned by $\dot{\ell}_t^{\eta}$ so that $\sum \dot{\ell}_t^{\theta e}$ and $\sum \dot{\ell}_t^{\eta}$ are asymptotically independent. Since $\widetilde{LR}_T^S$ is a function of $\dot{\ell}_t^{\theta}$ through $\dot{\ell}_t^{\theta} - \sum_{i=1}^{T}\dot{\ell}_i^{\theta}$, subtracting $J_{\theta\eta}J_{\eta}^{-1}\sum_{i=1}^{T}\dot{\ell}_i^{\eta}$ from the first term and adding it to the second term gives $\dot{\ell}_t^{\theta e} - \sum_{i=1}^{T}\dot{\ell}_i^{\theta e} = \dot{\ell}_t^{\theta} - \sum_{i=1}^{T}\dot{\ell}_i^{\theta}$, which implies that the test is locally invariant to η.

The intuition is similar to Stein's necessary condition for adaptation which is that $J_{\theta\eta}$ is zero. Under Stein's condition, $\dot{\ell}_t^{\theta e}$ is always equivalent to the $\dot{\ell}_t^{\theta}$ so that the invariance property always holds. The set-up in this section does not necessarily satisfy Stein's condition while it obtains the same inference. The orthogonality in this set-up does not come from the property of the error distribution, but from the property of the alternative process, θ_t.

The power envelope $\Psi^S(\Omega)$ is sharp if we have a $\sqrt{T}$-consistent estimator of η_0 that satisfies Condition 4. Let $B^S(\Omega)$ be the small sample counterpart of $\Lambda(\Omega)$, i.e. $B^S(\Omega)$ is the same as $B(\Omega)$ in (3.9) except $\dot{\ell}_t^{\theta}$ depends also on η_0, and let $\widehat{B}^S(\Omega)$ be the plug-in version of $B^S(\Omega)$. Since $B^S(\Omega)$ achieves $\Psi(\Omega)$, it suffices to show that $\widehat{B}^S(\Omega)$ converges in probability to $B^S(\Omega)$ under both H_0 and H_1. Lemma 5 below proves the argument.

Lemma 5. *Suppose there exist $\sqrt{T}$-consistent estimators $\hat{\eta}$ and $\hat{\theta}$, and a consistent estimator $\hat{J}_{\theta}$. Assume that $\dot{\ell}_t^{\theta}(\eta)$ satisfies condition 4) for both η_0 and θ_0. Under Condition 1),2), and 3)'*

$$|\widehat{B}^S(\Omega) - B^S(\Omega)| \longrightarrow 0 \quad \textit{in probability under } H_0 \textit{ and } H_1 \tag{4.7}$$

Lemma 5 also provides the motivation to use an error distribution which is more general than normal. Note that $\Psi^S(\Omega)$ is an increasing function of $\Omega^* = J_{\theta}^{\frac{1}{2}}\Omega J_{\theta}^{\frac{1}{2}}$ which is proportional to the Fisher information of $g(\epsilon_t)$. Accordingly, $\Psi^S(\Omega)$ is strictly increasing in the Fisher information. Consider, for example, Fernandez and Steel (1998)'s generalized exponential family $g(\epsilon_t) = A(\eta)\exp[B(\eta)|\epsilon_t|^{\eta}]$. The fisher information of this type of density ranges $[1, \infty]$ where it is one when $g(\epsilon_t)$ is normal and increases if $g(\epsilon_t)$ is away from normal. Consequently, any non-Gaussian density in

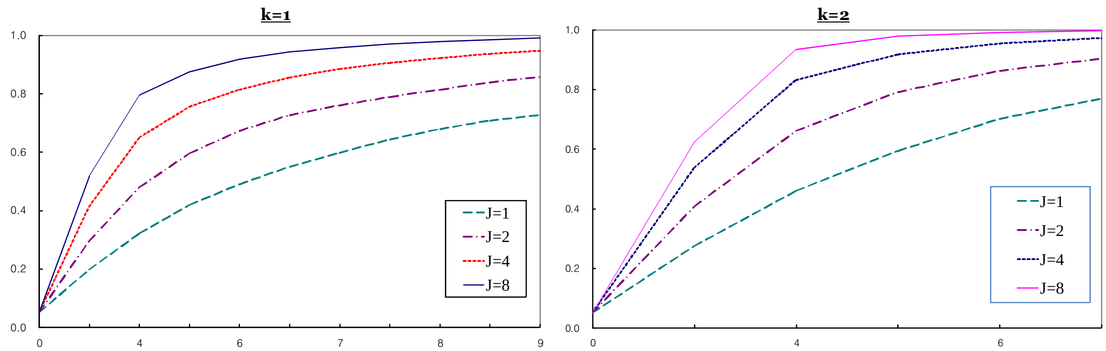


FIGURE 2. **Asymptotic Power Envelopes for Various Fisher Information**
note) Powers are plotted from 10,000 draws using 1,000 standard normal steps to approximate Wiener Processes.

the family would have higher $\Psi^S(\Omega)$. Since $\Psi^S(\Omega)$ is sharp, we may get a significant power gains by using a generalized one whenever $\eta \neq 2$. Figure 2 presents asymptotic power envelopes for various values of the Fisher information in the generalized exponential family, where the bottom line represents Gaussian case. It shows a large increase in power, which justifies the use of the test with non Gaussian error density.

4.2. **Asymptotically Optimal Tests in Semiparametric Models.** The previous section investigates an optimal test under which finite dimensional η in $g(\cdot)$ are unknown, while it is known that g is in a specific set $\mathcal{G}$. This section extends the idea to a model in which g is entirely unknown. Rather than allowing for g to be fully nonparametric, I give a mild restriction that g is parameterized by an infinite dimensional unknown nuisance parameter η. Consequently, the true density $f(\cdot)$ is only known to belong to a class $\mathcal{S}$ which contains all parametric families.

The set $\mathcal{S}$ can be considered as the union of all parametric submodels $\mathcal{P}_\eta$ in which the semiparametric power envelope, say $\Psi^e(\Omega)$, can be defined to be $\inf_{\mathcal{P}_\eta \in \mathcal{S}} \Psi^S(\Omega, h)$. The previous section shows that $\Psi^S(\Omega, h)$ is equivalent to $\Psi(\Omega)$ regardless of h, which implies that $\Psi^e(\Omega) = \Psi(\Omega)$. Unlike the previous section, however, the plug-in version of the efficient test, say $B^*(\Omega)$, is inappropriate because $\sqrt{T}$-consistent estimator of the infinite dimensional η is not available.

This problem is known to be the existence of an adaptive test. It is well known that, in standard LAN set up, an adaptive test is possible if an adaptive estimator exist. (See Choi, Hall, and Schick (1996).) Jansson (2008) extends this finding to nonstandard unit root test. An important finding in this section is that, our nonstandard testing problem is still amenable to adaptation by using extant semiparametric methods developed for standard problems. The purpose is to find a feasible test statistic $B^*(\Omega)$ which converges in probability to $B(\Omega)$ both under H_0 and H_1. Based on (3.6), it implies that there exist estimators $\{\hat{\dot{\ell}}_t^\theta\}$ and $\hat{J}_\theta$ which satisfy

$$\begin{aligned}\sum_{t=1}^{T}(\theta_t - \frac{1}{T}\sum_{i=1}^{T}\theta_i)\hat{\dot{\ell}}_t^\theta &= \sum_{t=1}^{T}(\theta_t - \frac{1}{T}\sum_{i=1}^{T}\theta_i)\dot{\ell}_t^\theta + o_p(1)\\ \hat{J}_\theta &= J_\theta + o_p(1)\end{aligned} \tag{4.8}$$

The objective of this section is to show the existence of the estimators that satisfy (4.8), and to demonstrate that it provides the existence of an adaptive test function. A possible construction of the efficient estimator is to use a kernel estimation method. Using data and the consistent estimator of θ_0, compute the residuals $\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_T$ with $\tilde{\epsilon}_t = \epsilon(y_1, \ldots, y_t, X_1, \ldots, X_T, \widehat{\theta})$ for $t = 1, \ldots, T$. A kernel density estimator is defined as for all e in a small neighborhood of each value of $\tilde{\epsilon}_t$

$$\hat{f}_T(e; \tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_T) = \frac{1}{(T-1)a_T}\sum_{i\neq t} k\left(\frac{e-\tilde{\epsilon}_i}{a_T}\right) \tag{4.9}$$

$$\hat{f}'_T(e; \tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_T) = \frac{1}{(T-1)a_T^2}\sum_{i\neq t} k'\left(\frac{e-\tilde{\epsilon}_i}{a_T}\right) \tag{4.10}$$

where a_T is a bandwidth and the kernel $k(\cdot)$ is three times continuously differentiable with derivative $k^{(i)}$ satisfying $\|k^{(i)}(z)\| < ck(z)$ with $i = 1, 2, 3$ for some positive c, and $\int z^2 k(z)dz < \infty$. $\{\hat{\dot{\ell}}_t^\theta\}$ and $\hat{J}_\theta$ are defined as

$$\hat{\dot{\ell}}_t^\theta(\tilde{\epsilon}_t; \tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_T) = \frac{\hat{f}'_T(\tilde{\epsilon}_t; \tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_T)}{b_T + \hat{f}_T(\tilde{\epsilon}_t; \tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_T)} \tag{4.11}$$

$$\hat{J}_\theta = \frac{1}{T}\sum_{t=1}^{T}\hat{\dot{\ell}}_t^\theta(\tilde{\epsilon}_t)\hat{\dot{\ell}}_t^\theta(\tilde{\epsilon}_t)' \tag{4.12}$$

where $\{b_T\}$ is a sequence of constants such that $(Ta_T^3 b_T)^{-1} \to 0$. Note that $\{\hat{\dot{\ell}}_t^\theta\}$ uses the entire sample data. Most existing research splits the sample period and uses only the observations in one sample period to estimate $\{\hat{\dot{\ell}}_t^\theta\}$ of the other split sample. It splits the sample not because of the elegancy, but because it yields a relatively easy way to obtain the asymptotic result under minimized conditions. From a practical point of view, however, it is desirable to use all sample data in moderate sample sizes in order to avoid the size distortion problem, and to produce a better power. Schick

(1987), and Koul and Schick (1997) suggest a general condition to use the whole data under additional conditions on the boundness of $\dot{m}(\cdot)$ and the memory property of $\{X_T\}$. The method in this section is generally similar to them, and Condition 1) and 2) are shown to be enough to satisfy their conditions, so that no additional condition is required in order to use the whole sample data for adaptation. Let's define the critical function $\psi_T(Z|\Omega) = 1_{[B^*(\Omega)>k_\alpha]}$ where k_α is the continuous function satisfying $E_0[\psi_T(Z|\Omega)] = \alpha$ and $B^*(\Omega)$ as

$$B^*(\Omega) = \sum_{i=1}^{k} \hat{\zeta}_i' \left[\frac{a_i^2}{T^2} I_T - F M_e F' \right]^{-1} \hat{\zeta}_i \tag{4.13}$$

where $\hat{\zeta}_i = (\hat{\zeta}_{i,1}, \ldots, \hat{\zeta}_{i,T})'$, $\hat{\zeta}_{i,j} = \sum_{t=1}^{j} \hat{\dot{\ell}}_{t,i}^{\theta*}$, and $\hat{\dot{\ell}}_{t,i}^{\theta*}$ is the i_{th} element of $\hat{\dot{\ell}}_t^{*\theta}$. Let $\Psi^*(\omega) = \underline{\lim}_{T\to\infty} \int\int \psi_T(Z|\Omega) f_1(Z|\eta) dZ d\nu_\delta$. The following theorem shows that we can construct an adaptive test based on (4.11) and (4.12), without further strict conditions.

Theorem 4. *Under Condition 1) to 4), any asymptotically similar test $\phi(Z|\Omega)_T^*$ associated with $\mathcal{S}$ satisfies*

$$\overline{\lim_{T\to\infty}} \int \phi_T^*(Z|\Omega) f_1(Z|\eta) dZ \leq \Psi^*(\Omega) = \Psi(\Omega)$$

Theorem 4 indicates that $\Psi^*(\Omega)$ provides the asymptotic power envelope in a semiparametric model, and $B^*(\Omega)$ is adaptive in the sense that $\Psi^*(\Omega)$ attains the parametric power envelop $\Psi(\Omega)$. Accordingly, the knowledge of the error distribution is asymptotically irrelevant for conducting an optimal test under mild conditions suggested in this paper.

5. **Comparative Simulation Study**

This section examines the performance of the asymptotically efficient tests in finite samples through Monte Carlo experiments. Parametric and semiparametric set up are separately examined. In parametric model, I consider the linear quantile model with asymmetric Laplace distribution in which the performance of parametric test function $\hat{B}(\hat{\Omega})$ is evaluated. In semiparametric model, various types of the error distributions are considered in linear equation where I examine the performance of semiparametric test $\hat{B}^*(\hat{\Omega})$.

Table 1. Monte carlo estimates of the empirical sizes in parametric models

	Empirical Size(%)					
	$q = 0.3$			$q = 0.5$		
	1 %	5 %	10 %	1 %	5 %	10 %
T= 50	1.72	5.38	9.48	1.11	4.38	8.44
100	1.10	4.58	8.64	1.06	4.96	9.56
200	1.28	5.00	9.48	1.14	5.36	9.70

5.1. **Monte Carlo Simulation in Parametric Models.** Consider the model

$$y_t = X_t'(\beta_0^q + \beta_t^q) + \epsilon_t^q \quad t = 1, \ldots, T \tag{5.1}$$

where y_t is a scalar, and X_t is $k \times 1$ vectors and is assumed to satisfy Condition 2. ϵ_t^q is iid from the asymmetric Laplace distribution which is defined as $\varphi^q(\epsilon) = exp\left[-\frac{1}{q}\epsilon \cdot 1_{\{\epsilon<0\}} + \frac{1}{1-q}\epsilon \cdot 1_{\{\epsilon>0\}}\right]$, where $1_{\{\ \}}$ is an indicator function. In this circumstance, $X_t'(\beta_0^q + \beta_t^q)$ represents q_{th} conditional quantile of y_t, that is, $Pr[y_t > X_t'(\beta_0^q + \beta_t^q)|X_1, \ldots, X_t] = q$. Consequently, ϵ_t^q is not a zero mean disturbance, but has the property that $Pr[\epsilon_t^q < 0] = q$. The score and its covariance with maximum likelihood estimators are defined as $\dot{\ell}_t(\widehat{\beta}_q) = \frac{1}{1-q}X_t - \frac{1}{q(1-q)}X_t 1_{\{y_t<X_t\hat{\beta}_q\}}$ and $\widehat{J}_1 = \frac{1}{Tq(1-q)}\sum_{t=1}^T X_t X_t'$, respectively. The asymmetric laplace distribution is known to be differentiable in quadratic mean. It is easy to show that quantile regression estimator satisfy Condition 4). Consequently, $\hat{B}(\hat{\Omega})$ is asymptotically point optimal in this setup.

I simulate the empirical sizes and the powers of the test. I consider $\{X_t\} = \{(1,\ Z_t)\}$, where $\{Z_t\}$ are generated from AR(1) model with *iid* Gaussian error. I examine 18 combinations of 3 different critical levels (1%, 5%, and 10%), 3 sample sizes (50, 100, and 200), and 2 quantile levels (0.3, 0.5). 5,000 replications are generated for each of 18 combinations. Table 1 shows the experimental result of the empirical sizes. The test performs fairly well for all significant levels. The differences between empirical sizes and actual sizes do not exceed one percent, even when the sample size is as small as 50.

As a next step, I calculate small sample powers of the test and compare them with those of other existing tests. Various types of alternative processes are examined: single break, multiple breaks (2 and 4 breaks), and random walk breaks. The powers are compared with those of *SupF* test, Andrews and Ploberger (1994)'s test (*ExpLM*), and Nyblom (1989)'s test (*Nyb*).

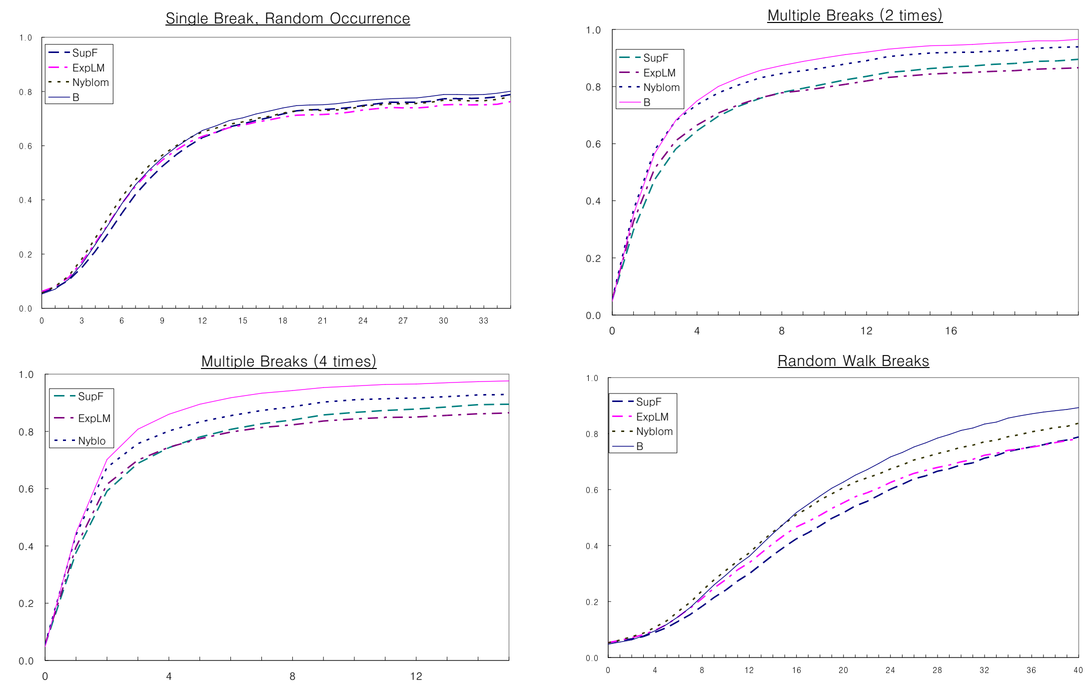


FIGURE 3. **Small Sample Powers, Quantile Models, T=100, q=0.3**

The size adjusted small sample powers are shown at figure 3. The figures show that $\hat{B}(\hat{\Omega})$ performs the best among 4 test statistics. $\hat{B}(\hat{\Omega})$ has the best power against the random walk process and the multiple breaks. The gaps become larger as the number of breaks increases. The powers of $\hat{B}(\hat{\Omega})$ for the single break alternatives are pretty close to $ExpLM$ and $SupF$ even though both $ExpLM$ and $SupF$ explicitly consider single break alternatives.

The differences of the powers, however, are mild for all unstable processes. Even though $SupF$ and $ExpLM$ are not designed for time varying parameter processes, the two tests show pretty reasonable power properties against the random walk case. Note that the breaking processes considered in $SupF$, $ExpLM$, and Nyb do not satisfy Condition 1. This gives an important empirical implication: The asymptotic equivalence of the optimal tests shown in the previous section can be more or less applied even in small samples and in the breaking process which are a little apart from Condition 1. The loss of power by misspecifying the true unstable parameter process is allowable. I also perform the simulation for different sample sizes and quantile levels. I don't present the simulation results for them because they are similar to what I present here.

Table 2. Monte carlo estimates of the empirical sizes in the semiparametric model

	$T = 100$			$T = 200$		
Model	1%	5%	10%	1%	5%	10%
A1) Standard Normal	1.46	4.16	8.86	0.82	5.12	10.58
A2) Symmetric Laplace	0.78	4.70	9.68	1.26	5.42	10.26
A3) Asymmetric Laplace	1.48	6.42	12.30	1.46	6.12	10.80
A4) Student t(4)	1.20	5.54	9.86	1.00	4.64	9.86
A5) Bimodal	1.24	5.78	10.71	1.39	5.78	10.74

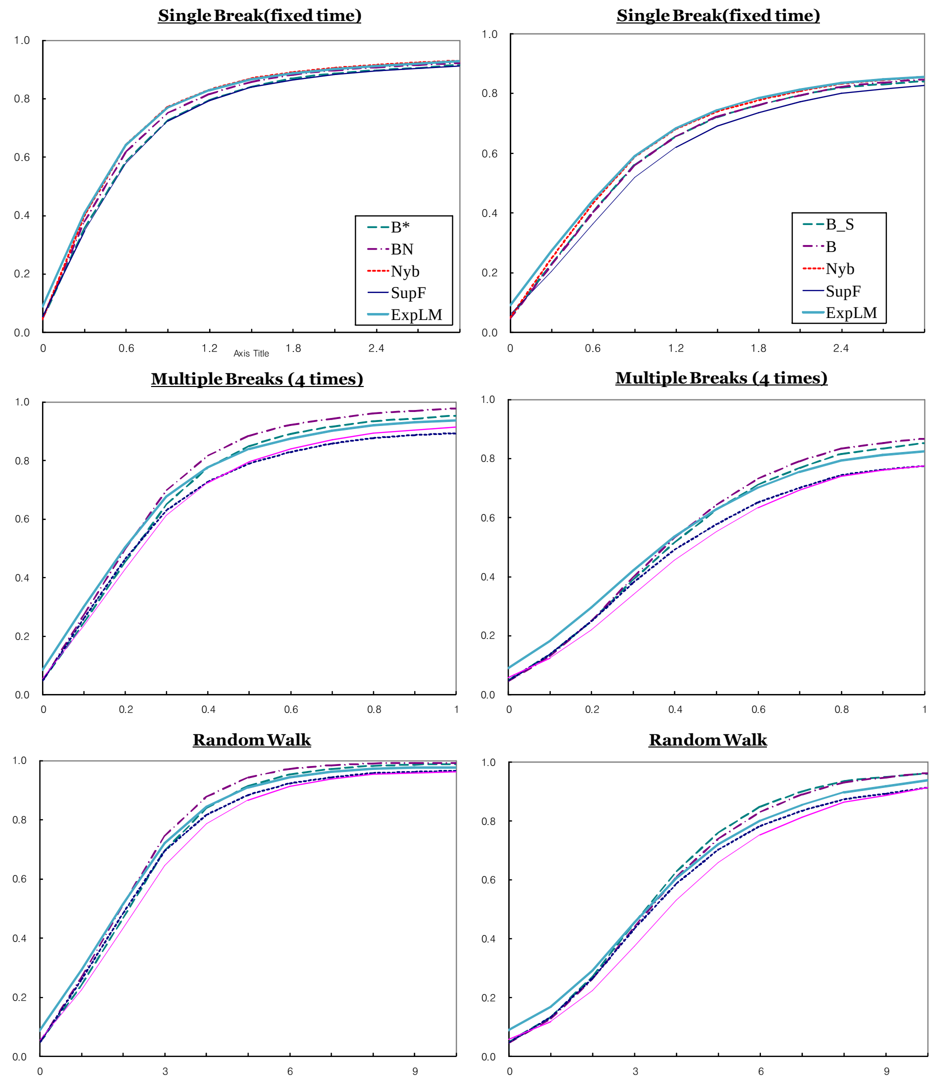
5.2. **Monte Carlo Simulation in Semiparametric Models.** Consider the simple linear regression model.

$$y_t = X_t'(\beta_0 + \beta_t) + \epsilon_t \quad t = 1, \ldots, T \tag{5.2}$$

The setup is the same as in the previous simulation except ϵ_t is now assumed to mean zero and has different error distributions. For the estimate of the density, I use standard Gaussian kernel estimation where the bandwidth is chosen by an optimal window width method based on Gaussian distribution. Reasonable changes of kernel, such as logistic and Epanechnikov do not significantly alter the result. b_T is chosen to be $0.001 \times a^{1/3}$. Five different error distributions are designed, which are listed below.

A1) Standard Normal Distribution
A2) Asymmetric Laplace Distribution(location shifted to $E[\epsilon_t] = 0$).
A3) Student t-distribution with $\nu = 4$ degree of freedom
A4) Mixture of two standard Normal distributions with mean 2, and -2, respectively

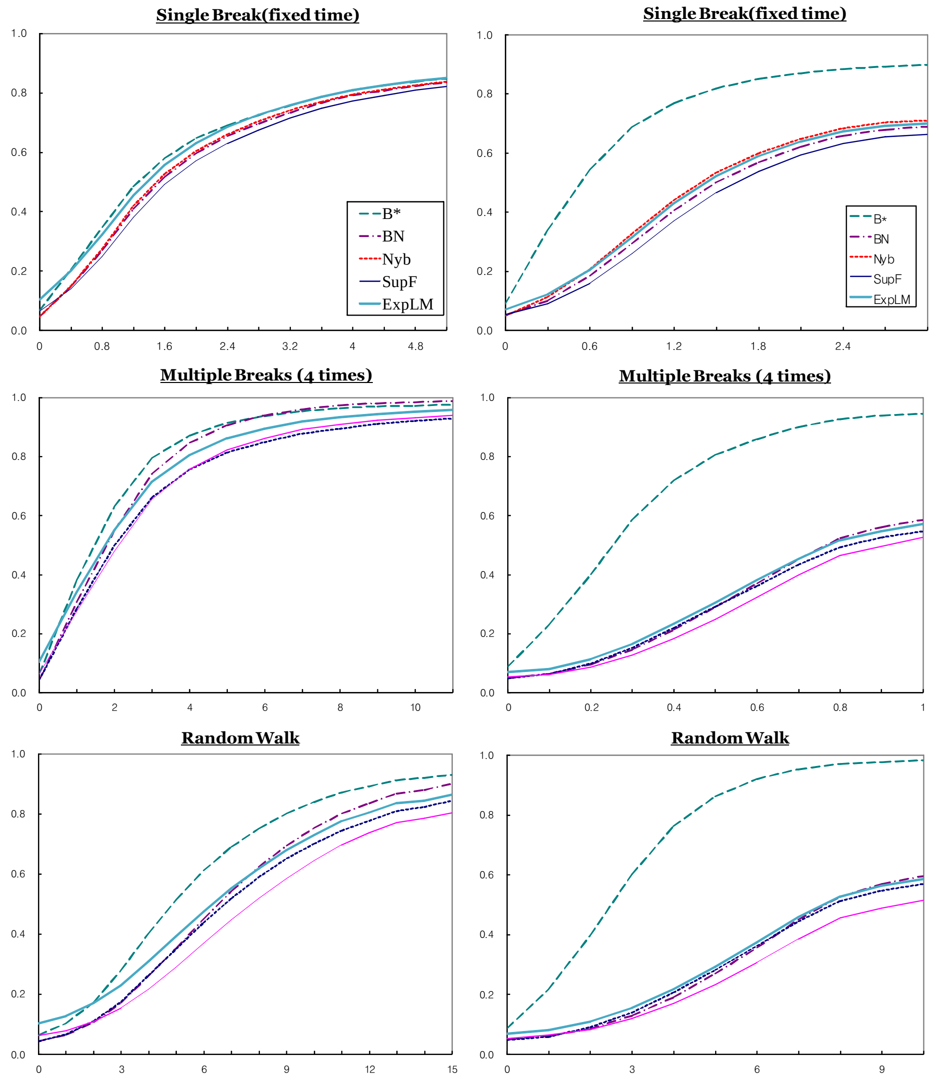
Table 2 shows the experimental result of the empirical sizes in which the small sample sizes performance of $B^*(\hat{\Omega})$ is shown to be fairly good in all distributions. The selected results of the simulated small sample powers are shown in figures 4 and 5, where $B^*(\hat{\Omega})$ is compared with $B(\hat{\Omega})$, $SupF$, $ExpLM$, and Nyb, which are set up based on the Gaussian error distribution. Therefore, these tests might have the best powers in A1 but lose some powers in the other distributions. The powers of all six tests are close to each other when the error distribution is unimodal and symmetric. The left hand side of figure 4 shows that $B^*(\hat{\Omega})$ has similar powers to the others even when the error distribution is normal. It implies that $B^*(\hat{\Omega})$ is little outperformed by the existing tests based on Gaussian distribution, even in the worst



(Standard Normal) (Student t(4))

FIGURE 4. **Small Sample Powers in the semiparametric model, T=100**

case. The right hand side of figure 4 shows that in t-distribution, $B^*(\Omega)$ performs the best against multiple breaks and random walk parameter. However, the power gaps between $B^*(\hat{\Omega})$ and others are small. Unlike the large sample case (figure 2), substantial power gains by using non-Gaussian error distribution are not clear in this small sample instance. The result in the Laplace distributional case is similar to the t-distribution case, and I do not present the results in this paper. Since the distinctive feature of Gaussian, Laplace and student-t distributions is thickness of tail, these results imply that the relative finite sample powers are not very sensitive to tail behavior of error distribution. Figure 5 shows that $B^*(\hat{\Omega})$ performs the best when



(Asymmetric Laplace) (Bimodal)

FIGURE 5. **Small Sample Powers in the semiparametric model, T=100 (continued)**

the error distribution is skewed and the gaps become larger as the number of breaks increase. The gaps are relatively bigger than previous distributions. The power gaps become fairly consequential in bimodal error distribution, as shown in figure 5. $B^*(\hat{\Omega})$ has the powers 62%p greater than the best of the others, at its greatest extent. In summary, there is considerable power improvement of the adaptive test $B^*(\hat{\Omega})$. The degree of the improvement depends on the modality and the skewness, rather than the tail behavior.

6. Conclusion

Parameter instability is of central importance in time series models. This paper gives three implications for testing parameter stability. First, asymptotically optimal tests for parameter instability do not require information about the exact form of the unstable parameter processes. Many tests are designed to have good powers against specific alternative processes. The result in this paper implies that a tailored test for specific instability does not have any power gain in the asymptotic sense, which means that attempts to derive tailor-made tests are asymptotically irrelevant. Monte carlo simulation results show that misspecifying the unstable process results in only a mild loss of powers even in small samples.

Second, Adaptation has shown to be possible in such nonstandard testing problem as unstable parameter process. It implies that an attempt to find a well-fitted error distribution is asymptotically inappropriate under mild conditions because one may not gain any asymptotic power. This asymptotic irrelevancy is consequential because widely assumed normal density is generally far from macroeconomics and financial data, and choosing another specific density often might be too discretionary.

Finally, I suggest two easy-to-compute asymptotically optimal test statistics. $\hat{B}(\hat{\Omega})$ is used when the error distribution restricted to a certain parametric family, while $B^*(\hat{\Omega})$ can be applied if any restriction of the error distribution is irrelevant. By avoiding the sample-split method, the test $B^*(\Omega)$ also shows good size and power performance even in small samples. Small sample simulations show that the test statistics have correct sizes and improved powers against the existing tests for almost all unstable processes.

Appendix A. Proofs

A.1. **proof of Lemma 1.** Let $\xi_t^0 = \xi_t(\cdot|\theta_0, 0)$, $\xi_t^1 = \xi_t(\cdot|\theta_0, \theta_t)$. Condition 3 implies

$$\xi_t^1 = \xi_t^0 + \theta_t' \dot{\xi}_t^0 + r_t \tag{A.1}$$

where $E[(\frac{r_t}{\xi_t^0})^2] = o_p(\|(\theta_t)\|^2)$. By using (A.1), the square root of the integrand of the LR statistics in (3.3) can be written as,

$$\sqrt{L_T} = \prod_{t=1}^{T} \left(\frac{\xi_t^1}{\xi_t^0} \right) = \prod_{t=1}^{T} \left(\frac{\xi_t^1 - \xi_t^0}{\xi_t^0} + 1 \right) = \prod_{t=1}^{T} \left(\theta_t' \frac{\dot{\xi}_t^0}{\xi_t^0} + \frac{r_t}{\xi_t^0} + 1 \right) = \prod_{t=1}^{T} (1 + \eta_t) \tag{A.2}$$

where $\eta_t = \theta_t' \frac{\dot{\xi}_t^0}{\xi_t^0} + R_t$ and $R_t = \frac{r_t}{\xi_t^0}$. Therefore L_T can be rewritten as,

$$L_t = exp\left[\sum_{t=1}^{T} log(1+\eta_t)\right]$$

Note that $\sum_{t=1}^{T} log(1+\eta_t) = \sum_{t=1}^{T} \eta_t - \frac{1}{2}\sum_{t=1}^{T} \eta_t^2 + o_p(1)$ if $\max_t |\eta_t| = o_p(1)$, and $\sum_{t=1}^{T} \eta_t^2 = O_p(1)$. Since Condition 1 implies that $\sum \dot{\ell}_t \sum_i \theta_i = 0$ and $J_\theta \sum \theta_t = 0$, Lemma 1 is proved by showing

(1) $\sum_{t=1}^{T} \eta_t = \frac{1}{2}\sum_{t=1}^{T} \theta_t' \dot{l}_t - \frac{1}{8}\sum_{t=1}^{T} \theta_t' J_\theta \theta_t + o_p(1)$
(2) $\sum_{t=1}^{T} \eta_t^2 = \frac{1}{4}\sum_{t=1}^{T} \theta_t' J_\theta \theta_t + o_p(1)$
(3) $\max_t |\eta_t| = o_p(1)$

Proof of (1) : Let $\delta_t = T\theta_t$. To prove (1), we have only to show that $\sum_{t=1}^{T} R_t = -\frac{1}{8T^2}\sum_{t=1}^{T} \delta_t' J_\theta \delta_t + o_p(1)$. Squaring both sides of (A.1) gives

$$\begin{aligned}(\xi_t^1)^2 &= (\xi_t^0)^2 + r_t^2 + \frac{2}{T}\xi_t^0 \delta_t' \dot{\xi}_t^0 + 2\xi_t^0 r_t + \frac{2}{T}\delta_t' \dot{\xi}_t^0 r_t + \frac{1}{T^2}\delta_t' \dot{\xi}_t^0 \dot{\xi}_t^{0\prime} \delta_t \\ \Rightarrow 2R_t &= 2\frac{r_t}{\xi_t^0} = \left(\frac{(\xi_t^1)^2}{(\xi_t^0)^2} - 1\right) - R_t^2 - \frac{1}{T}\delta_t' \dot{\ell}_t - \frac{1}{T}\delta_t' \dot{\ell}_t R_t - \frac{1}{4T^2}\delta_t' \dot{\ell}_t \dot{\ell}_t' \delta_t\end{aligned}$$

By taking conditional expectation with respect to δ_t, we get

$$\begin{aligned}2E[R_t|\delta_t] &= \left(E\left[\frac{(\xi_t)^2}{(\xi_t^0)^2}|\delta_t\right] - 1\right) - E\left[R_t^2|\delta_t\right] - \frac{1}{T}\delta_t' E[\dot{\ell}_t|\delta_t] - \frac{1}{T}\delta_t' E[\dot{\ell}_t R_t|\delta_t] - \\ &\quad \frac{1}{4T^2}\delta_t' E[\dot{\ell}_t \dot{\ell}_t'|\delta_t]\delta_t\end{aligned}$$

Let $\widetilde{R}_t = \mathbf{1}\{\|\delta_t/\sqrt{T}\| < M_T\}R_t$ denote a truncated version of R_t where $\frac{M_T}{\sqrt{T}} \to 0$ and $M_T \to \infty$. The sequences $\widetilde{R}_t$ and R_t are asymptotically equivalent in the sense that $\sum_{t=1}^{T} R_t = \sum_{t=1}^{T} \widetilde{R}_t + o_p(1)$. Note that $\max_{\{\|\delta_t/\sqrt{T}\|<M_T\}} (\frac{1}{T^2}d_s' d_s)^{-1} E[R_s^2|d_s] = o_p(1)$ from (A.1) and $\frac{1}{T^2}\sum \delta_t \delta_t' = O_p(1)$ from Condition 1. Consequently,

$$\begin{aligned}\sum_{t=1}^{T} E[\widetilde{R}_t^2|\delta_t] = \sum_{t=1}^{T} \mathbf{1}_{\{\|\delta_t/\sqrt{T}\|<M_T\}} E[R_t^2|\delta_t] \le \sum_{t=1}^{T} \max_{\{\|\delta_t/\sqrt{T}\|<M_T\}} \left(\left(\frac{1}{T^2}d_s' d_s\right)^{-1} E[R_s^2|d_s]\right)\frac{1}{T^2}\delta_t'\delta_t \\ = \max_{\{\|\delta_t/\sqrt{T}\|<M_T\}} \left((\frac{1}{T^2}d_s' d_s)^{-1} E[R_s^2|d_s]\right)\frac{1}{T^2}\sum_{t=1}^{T}\delta_t'\delta_t = o_p(1) \times O_p(1) = o_p(1)\end{aligned}$$

Also, using Chebychev inequality,

$$\begin{aligned}\frac{1}{T}d_{ti}E[\dot{\ell}_{ti}R_t|\delta_t] &\le \frac{1}{T}d_{ti}E[\dot{\ell}_{t,i}^2|\delta_t]^{1/2}E[R_t^2|\delta_t]^{1/2} \\ &= O_p(T^{-1/2}) \times (O_p(1))^{1/2} \times (o_p(T^{-1/2}))^{1/2} = o_p(T^{-1}) \qquad \text{(A.3)}\end{aligned}$$

Note that $E[\dot{\ell}_t|\delta_t] = 0$, and $E[\dot{\ell}_t\dot{\ell}_t'|\delta_t] = J_\theta$,(see Vaart (1998)). Using (A.1) and (A.3), (1) is proved because

$$\sum_{t=1}^{T} R_t = \sum_{t=1}^{T} E[R_t|\delta_t] + o_p(1) = \frac{1}{8T^2}\sum_{t=1}^{T}\delta_t' J_\theta \delta_t + o_p(1) \tag{A.4}$$

Proof of (2):

$$\begin{aligned}\sum_{t=1}^{T}\eta^2 &= \frac{1}{4T^2}\sum_{t=1}^{T}\delta_t'\dot{\ell}_t\dot{\ell}_t'\delta_t + \frac{1}{T}\sum_{t=1}^{T}\delta_t'\dot{\ell}_t R_t + \sum_{t=1}^{T} R_t^2 \\ &= \left(\frac{1}{4}\sum_{t=1}^{T}\delta_t' J_\theta\delta_t + o_p(1)\right) + o_p(1) + o_p(1)\end{aligned}$$

where the last two terms of the last equality comes from (A.3) and (A.4).

Proof of (3):

$$\begin{aligned}\max_t \eta_t &\le \frac{1}{2}\max_{t,\|\frac{1}{\sqrt{T}}\delta_t\|\le M_T} |\frac{1}{\sqrt{T}}\delta_t|' \cdot |\frac{\dot{\ell}_t}{\sqrt{T}}| + \max_t R_t + o_p(1) \\ &\le \frac{1}{2}\max_{\|\frac{1}{\sqrt{T}}\delta_t\|\le M_T} \|\frac{1}{\sqrt{T}}\delta_t\| \cdot \|\frac{\dot{\ell}_t}{\sqrt{T}}\| + \max_t R_t + o_p(1) \\ &\le \frac{M_T}{\sqrt{T}}\|\dot{\ell}_t\| + \max_t R_t + o_p(1) = o_p(1) + o_p(1) + o_p(1) = o_p(1)\end{aligned}$$

The first term of the 2nd inequality comes from Cauchy-Schwarz inequality, the first term of the last equality comes from $E[\dot{\ell}_t^2] \le \infty$ and the second term comes from

$$max_t|R_t|^2 \le \sum_{t=1}^{T} R_t^2 = o_p(1)$$

which completes the proof. $\diamond$

A.2. **Proof of Lemma 2.** In order to Prove Lemma 2, we need the following lemma.

Lemma 6. *Let's define* $\widetilde{L}_T$ *as*

$$\widetilde{L}_T = exp\left[\dot{\ell}'(M_e \otimes I_k)\widetilde{\theta} - \frac{1}{2}\widetilde{\theta}'(M_e \otimes J_\theta)\widetilde{\theta}\right] \tag{A.5}$$

Under Condition 1) to 3), $|\widetilde{L}_T - L_T| \Longrightarrow 0$ *under* H_0*, where* $\Rightarrow$ *represents weak convergency.*

Proof) Lemma 6 can be proved by showing that for any $\delta = T\theta$ that satisfies Condition 1, $\frac{1}{T}\delta'(M_e \otimes I_k)\dot{\ell}$ and $\frac{1}{2T^2}\delta'(M_e \otimes J_1)\delta$ converge to well defined limiting variables. The first term can be rewritten as

$$\begin{aligned}\frac{1}{T}\delta'(M_e \otimes I_k)\dot{\ell} &= \frac{1}{T}\delta'\dot{\ell} - \frac{1}{T^2}\delta'(ee' \otimes I_k)\dot{\ell} = \frac{1}{T}\delta'\dot{\ell} - \frac{1}{T^2}[(e' \otimes I_k)\delta]'[(e' \otimes I_k)\dot{\ell}] \\ &= \frac{1}{T}\sum_{t=1}^{T}\delta_t'\dot{\ell}_t - \frac{1}{T^2}(\sum_{t=1}^{T}\delta_t)'(\sum_{t=1}^{T}\dot{\ell}_t)\end{aligned}$$

Consequently, I prove that each term of the last equation converges to a well defined limiting distributions.

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\delta_t'\dot{\ell}_t &= tr[\Omega^{*\frac{1}{2}}\frac{1}{T}\sum_{t=1}^{T}\Omega^{-\frac{1}{2}}\delta_t'\dot{\ell}_t J_1^{-\frac{1}{2}}] \\ &\Rightarrow tr[\Omega^{*\frac{1}{2}}\int W_\delta dW_\ell'] = \int W_\delta'\Omega^{*\frac{1}{2}}dW_\ell\end{aligned}$$

where W_δ and W_ℓ are multivariate standard Wiener processes.

$$\begin{aligned}\frac{1}{T^2}(\sum_{t=1}^{T}\delta_t)'(\sum_{t=1}^{T}\dot{\ell}_t) &= tr\left[\Omega^{*\frac{1}{2}}(T^{\frac{3}{2}}\sum_{t=1}^{T}\Omega^{-\frac{1}{2}}\delta_t)(T^{\frac{1}{2}}\sum_{t=1}^{T}J^{-\frac{1}{2}}\dot{\ell}_t)'\right] \\ &\Rightarrow tr\left[\Omega^{*\frac{1}{2}}\int W_\delta(r)drW_\ell(1)'\right] = \int W_\delta(r)dr\Omega^{*\frac{1}{2}}W_\ell(1)\end{aligned}$$

The convergence of the second term of L_T can be proved as

$$\begin{aligned}\frac{1}{T^2}\delta'(M_e \otimes J_\theta)\delta &= \frac{1}{T^2}\delta'(I_T \otimes J_\theta)\delta - \frac{1}{T^3}\delta'(ee' \otimes J_\theta)\delta \\ &= \frac{1}{T^2}\delta'(I_T \otimes J_\theta)\delta - \frac{1}{T^3}[(e' \otimes J_\theta^{1/2})\delta]'[(e' \otimes J_\theta^{1/2})\delta] \\ &= \frac{1}{T}\sum_{t=1}^{T}\delta_t'J_\theta\delta - \frac{1}{T^3}(\sum_{t=1}^{T}\delta_t)'J_\theta(\sum_{t=1}^{T}\delta_t)\end{aligned}$$

$$\begin{aligned}\frac{1}{T^2}\sum_{t=1}^{T}\delta_t'J_\theta\delta_t &= tr\left[\Omega^*\frac{1}{T^2}\sum_{t=1}^{T}\left(\Omega^{-\frac{1}{2}}\delta_t\right)\left(\Omega^{-\frac{1}{2}}\delta_t\right)'\right] \\ &\Rightarrow tr\left[\Omega^*\int W_\delta(r)W_\delta(r)'dr\right] = \int W_\delta(r)'\Omega^*W_\delta(r)dr\end{aligned}$$

$$\begin{aligned}\frac{1}{T^3}\left(\sum_{t=1}^{T}\delta_t\right)'J_\theta\left(\sum_{t=1}^{T}\delta_t\right) &= tr\left[\Omega^*\left(T^{\frac{3}{2}}\sum_{t=1}^{T}\Omega^{-\frac{1}{2}}\delta_t\right)\left(T^{\frac{3}{2}}\sum_{t=1}^{T}\Omega^{-\frac{1}{2}}\delta_t\right)'\right] \\ &\Rightarrow tr\left[\Omega^*\int W_\delta(r)dr\int W_\delta(r)'dr\right] = \int W_\delta(r)'dr\Omega^*\int W_\delta(r)dr\end{aligned}$$

which completes the proof. $\diamond$

Proof of Lemma 2): Let's denote the variance of $\widetilde{\theta}$, $FF'/T^2 \otimes \Omega$ as K. $\widetilde{LR}_T$ can be written as

$$\begin{aligned}
\widetilde{LR}_T &= \int (2\pi)^{-\frac{k(T-1)}{2}} |K|^{-\frac{1}{2}} \exp\left[\dot{\ell}'(M_e \otimes I_k)\widetilde{\theta} - \frac{1}{2}\widetilde{\theta}'(M_e \otimes J_1)\widetilde{\theta} - \frac{1}{2}\widetilde{\theta}' K^{-1}\widetilde{\theta}\right] d\widetilde{\theta} \\
&= |K(M_e \otimes J_1) + I_{Tk}|^{1/2} \exp\left[\frac{1}{2}\dot{\ell}'(M_e \otimes I_k)\{(M_e \otimes J_1) + K^{-1}\}^{-1}(M_e \otimes I_k)\dot{\ell}\right] \\
&\quad \times \int (2\pi)^{-\frac{k(T-1)}{2}} \left|(M_e \otimes J_1) + K^{-1}\right|^{\frac{1}{2}} \exp[-\frac{1}{2}\left(\widetilde{\theta} - \{(M_e \otimes J_1) + K^{-1}\}(M_e \otimes I_k)\dot{\ell}\right)' \\
&\quad \times \{(M_e \otimes J_1) + K^{-1}\}\left(\widetilde{\theta} - \{(M_e \otimes J_1) + K^{-1}\}(M_e \otimes I_k)\dot{\ell}\right)] d\widetilde{\theta} \\
&= |K(M_e \otimes J_1) + I_{Tk}| \exp\left[\frac{1}{2}\dot{\ell}'(M_e \otimes I_k)\{(M_e \otimes J_1) + K^{-1}\}^{-1}(M_e \otimes I_k)\dot{\ell}\right] \\
&= c \cdot \exp\left[\frac{1}{2}\dot{\ell}'(M_e \otimes I_k)\{M_e \otimes J_\theta + T^2(FF')^{-1} \otimes \Omega^{-1}\}^{-1}(M_e \otimes I_k)\dot{\ell}\right] \\
&= c \cdot \exp[\frac{1}{2}\dot{\ell}'(M_e \otimes I_k)(I \otimes J^{-1/2}P)\{M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda^{-1}\}^{-1} \\
&\qquad \times (I_T \otimes J^{-1/2}P)'(M_e \otimes I_k)\dot{\ell}]
\end{aligned}$$

$$= c \cdot \exp\left[\frac{1}{2}\bar{\dot{\ell}}^{*\prime}\{M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda^{-1}\}^{-1}\bar{\dot{\ell}}^{*\prime}\right] \tag{A.6}$$

where $c = |K(M_e \otimes J_1) + I_{Tk}|$, $\bar{\dot{\ell}}^* = (\bar{\dot{\ell}}_1^{*\prime}, \ldots, \bar{\dot{\ell}}_T^{*\prime})'$, and $\bar{\dot{\ell}}_j^* = \dot{\ell}_j^* - \frac{1}{T}\sum_{t=1}^T \dot{\ell}_t^*$. I then change the expression of the test statistic. Let's define $(a_i^2, \ldots, a_k^2)$ be the vector of the diagonal elements of Λ, and ι_i be the $k \times 1$ vector which is one at i^{th} element and zeros otherwise.

$$M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda^{-1} = M_e \otimes I_k + \sum_{i=1}^k a_i^{-2}(\frac{FF'}{T^2})^{-1} \otimes \iota_i\iota_i' = \sum_{i=1}^k (M_e + K_{ai}^{-1}) \otimes \iota_i\iota_i' \tag{A.7}$$

where $K_{ai} = a_i^2\left(\frac{FF'}{T^2}\right)$. Note that $\iota_i\iota_i' \cdot \iota_j\iota_j'$ is $k \times k$ zero matrix if $i \neq j$ and $\iota_i\iota_i'$ if $i = j$. It makes the inverse of $M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda$ easy as below.

$$(M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda)^{-1} = \sum_{i=1}^k (M_e + K_{ai}^{-1})^{-1} \otimes \iota_i\iota_i' \tag{A.8}$$

because $\left[\sum_{i=1}^k \left(M_e + K_{ai}^{-1}\right) \otimes \iota_i\iota_i'\right]\left[\sum_{i=1}^k (M_e + K_{ai}^{-1})^{-1} \otimes \iota_i\iota_i'\right] = \sum_{i=1}^k I_T \otimes \iota_i\iota_i' + \sum_{i=1}^k \sum_{j\neq i}\left(M_e + K_{ai}^{-1}\right)\left(M_e + K_{aj}^{-1}\right)^{-1} \otimes (\iota_i\iota_i')(\iota_j\iota_j') = I_T \otimes I_k$. Consequently,

$$\begin{aligned}
\bar{\dot{\ell}}^{*\prime}(M_e \otimes I_k \quad &+ \quad (\frac{FF'}{T^2})^{-1} \otimes \Omega^{*-1})^{-1}\bar{\dot{\ell}}^{*\prime} = \bar{\dot{\ell}}^{*\prime}\left[\sum_{i=1}^k (M_e + K_{ai}^{-1})^{-1} \otimes \iota_i\iota_i'\right]\bar{\dot{\ell}}^{*\prime} \\
&= \sum_{i=1}^k \bar{\dot{\ell}}^{*\prime} F\left(FM_eF + (\frac{T^2}{a_i^2})I_T\right)^{-1} F'\bar{\dot{\ell}}^{*\prime} \qquad \text{(A.9)}
\end{aligned}$$

Taking log of (A.6) and applying (A.9) completes the proof. $\diamond$

A.3. **proof of Theorem 1.** Theorem 1 can be proven by showing that $P[|LR_T - \widetilde{LR}_T| > \epsilon] \to 0$ under both the null and the alternative hypothesis.

(1) Proof of the convergence under the null hypothesis: For $0 < M < \infty$, define

$$\begin{aligned} LR_T(M) &= \int \Pi_{t=1}^{T} \frac{f(\epsilon_t^1|\theta_0)}{f(\epsilon_t^0|\theta_0,\theta_t)} \mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}} d\nu_\theta \\ \widetilde{LR}_T(M) &= \int exp\left[\dot{\ell}'(M_e \otimes I_k)\widetilde{\theta} - \frac{1}{2}\widetilde{\theta}'(M_e \otimes J_\theta)\widetilde{\theta}\right] \mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}} d\nu_{\widetilde{\theta}} \end{aligned}$$

Note that for any $\epsilon > 0$, the following is satisfied

$$\begin{aligned} P[|LR_T - \widetilde{LR}_T| > 3\epsilon] &\leq P[|LR_T - LR_T(M)| > \epsilon] && (i) \\ &+ P[|\widetilde{LR}_T - \widetilde{LR}_T(M)| > \epsilon] && (ii) \\ &+ P[|LR_T(M) - \widetilde{LR}_T(M)| > \epsilon] && (iii) \end{aligned} \tag{A.10}$$

Consequently, it suffices to show that each term of (A.10) converges to zero, respectively.

(i) $P[|LR_T - LR_T(M)| > \epsilon] \leq \epsilon^{-1}E[|LR_T - LR_T(M)|]$

$$= \epsilon^{-1}E\left[\int_{\|\sqrt{T}\theta\|>M} \Pi_{t=1}^{T} \frac{f(\epsilon_t|\theta_0,\theta_t)}{f(\epsilon_t|\theta_0,0)} d\nu_\theta\right] = \epsilon^{-1}\int_{\|\sqrt{T}\theta\|>M} d\nu_\theta \;=\; \epsilon^{-1}P[\|\sqrt{T}\theta\| > M]$$

The first inequality comes from Chebychev inequality. The second equality uses *Fubini Theorem*. The right hand side of the last equality can be made arbitrarily small for all T by taking M large enough by the property of θ defined in Condition 1.

(ii) $\left|\widetilde{LR}_T - \widetilde{LR}_T(M)\right| = \int \widetilde{L}_T d\nu_{\widetilde{\theta}} - \int_{\|\sqrt{T}\theta\|<M} \widetilde{L}_T d\nu_{\widetilde{\theta}}$

$$\begin{aligned} &= c\exp\left[\frac{1}{2}\bar{\ell}^{*\prime}\{M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda^{-1}\}^{-1}\bar{\ell}^{*\prime}\right] \int (2\pi)^{-\frac{k(T-1)}{2}} \left|(M_e \otimes J_1) + K^{-1}\right|^{\frac{1}{2}} \\ &\quad \times \exp[-\frac{1}{2}\left(\widetilde{\theta} - \{(M_e \otimes J_1) + K^{-1}\}(M_e \otimes I_k)\dot{\ell}\right)' \{(M_e \otimes J_1) + K^{-1}\} \\ &\quad \times \left(\widetilde{\theta} - \{(M_e \otimes J_1) + K^{-1}\}(M_e \otimes I_k)\dot{\ell}\right)] d\nu\widetilde{\theta} \\ &= c\exp\left[\frac{1}{2}B^*(\theta_0, J_1, \Omega)\right] \int_{\|\sqrt{T}\theta\|>M} d\nu\widetilde{\theta} \end{aligned} \tag{A.11}$$

The first term on the last equation is $O_p(1)$ by Lemma 3, and the second term can be made arbitrarily small by taking M large by Condition 1. In consequence, $P[|\widetilde{LR}_T - \widetilde{LR}_T(M)| > \epsilon]$ can be made arbitrarily small for all T large by taking M sufficiently large.

Proof of (iii): Let's define

$$
\begin{aligned}
L_T(M) &= \prod_{t=1}^{T} \frac{f(\epsilon_t^1|\theta_t)}{f(\epsilon_t^0)} \cdot \mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}} = L_T \cdot \mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}} \\
\widetilde{L}_T(M) &= \exp\left[\dot{\ell}'(M_e \otimes I_k)\widetilde{\theta} - \frac{1}{2}\widetilde{\theta}'(M_e \otimes J_\theta)\widetilde{\theta}\right] \cdot \mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}} = \widetilde{L}_T(\widetilde{\theta}) \cdot \mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}} \\
L_T^*(M) &= \exp\left[\dot{\ell}'(M_e \otimes I_k)\theta - \frac{1}{2}\theta'(M_e \otimes J_\theta)\theta\right] \cdot \mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}} = L_T^*(\theta) \cdot \mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}}
\end{aligned}
$$

The test statistics are defined as $LR_T(M) = \int L_T(M)d\nu_\theta$, $\widetilde{LR}_T(M) = \int \widetilde{L}_T(M)d\nu_{\widetilde{\theta}}$. We define additional test statistic, $LR_T^*(M) = \int L_T^*(M)d\nu_\theta$. I prove (iii) by showing that $LR_T(M) - LR_T^*(M) \to^p 0$ and $\widetilde{LR}_T(M) - \widetilde{LR}_T^*(M) \to^p 0$. The first convergence is proved as

$$
LR_T(M) = \int L_T(M)d\nu\delta = \int (1 + o_p(1))L_T^*(M)d\nu_\theta = LR_T^*(M) + o_p(1)
$$

The second equality follows from Lemma 1. The third equality uses $LR_T^*(M)$ is bounded in probability which is shown as

$$
\begin{aligned}
P[LR_T^*(M) > K] &\le K^{-1}E\left[LR_T^*(M)\right] = K^{-1}\int E\left[LR_T^*(M)\right]d\nu_\theta \\
&= K^{-1}\int E[L_T|\theta]\mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}}d\nu_\theta
\end{aligned}
$$

which can be made arbitrarily small by choosing K sufficiently large. To prove the second convergence, we use an additional indicator function $\mathbf{1}_{\{B(\Omega)>K\}}$ and define new test statistics $LR_T(M,K)$, $\widetilde{LR}_T(M,K)$, and $\widetilde{LR}_T^*(M,K)$ as $LR_T(M)$, $\widetilde{LR}_T(M)$, and $\widetilde{LR}_T^*(M)$ multiplied by $\mathbf{1}_{\{B(\cdot)>K\}}$, respectively. Note that for any $\epsilon > 0$,

$$
\begin{aligned}
P[|\widetilde{LR}_T(M) - \widetilde{LR}_T*(M)| > 3\epsilon] &\le P[|\widetilde{LR}_T(M) - \widetilde{LR}_T(M,K)| > \epsilon] \\
&+ P[|\widetilde{LR}_T^*(M) - \widetilde{LR}_T^*(M,K)| > \epsilon] + P[|\widetilde{LR}_T(M,K) - \widetilde{LR}_T^*(M,K)| > \epsilon]
\end{aligned}
$$

The convergence of the first term can be easy to show by using the similar method of (A.11). The convergence of the second term can be shown as

$$
\begin{aligned}
P[|\widetilde{LR}_T^*(M) &- \widetilde{LR}_T^*(M,K)| > \epsilon] \le \frac{1}{\epsilon}E\left[\left|\widetilde{LR}_T^*(M) - \widetilde{LR}_T^*(M,K)\right|\right] \\
&= \frac{1}{\epsilon}\int\int L_T^*\left(1 - \mathbf{1}_{\{B^*(\cdot)>K\}}\right)\mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}}d\nu_\theta d\nu_z \\
&\le \frac{1}{\epsilon}\int \mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}}d\nu_\theta = P\left[\|\sqrt{T}\theta\| < M\right] \qquad \text{(A.12)}
\end{aligned}
$$

where the third inequality comes from $\int L_T^* d\nu_z = 1$. (A.12) can be made arbitrarily small for all T by taking M sufficiently large. In order to prove the convergence of the third term, we define additional random elements γ and $\widetilde{\gamma}$, which have the same distribution as θ and $\widetilde{\theta}$, respectively and are independent of θ and $\widetilde{\theta}$ and of each other. We prove $LR_T^*(M) - \widetilde{LR}_T(M)$ convergence in mean square. Note that $LR_T^*(M)$ and $\widetilde{LR}_T(M)$ can be alternatively written as integrals with respect to the measure of γ and $\widetilde{\gamma}$, respectively. Let $LR_T^*(M,K,\theta)$ and $\widetilde{LR}_T(M,K,\theta)$ be $LR_T^*(M,K)$ and $\widetilde{LR}_T(M,K)$ integrated with respect to the measure of θ.

$$
\begin{aligned}
&E[(LR_T^*(M,K) - \widetilde{LR}_T(M,K))^2] \\
&\quad = E\left[(LR_T^*(M,K,\theta) - \widetilde{LR}_T(M,K,\widetilde{\theta}))(LR_T^*(M,K,\gamma) - \widetilde{LR}_T(M,K,\widetilde{\gamma}))\right] \\
&\quad = E\left[LR_T^*(M,K,\theta)LR_T^*(M,K,\gamma)\right] - E\left[LR_T^*(M,K,\theta)\widetilde{LR}_T(M,K,\widetilde{\gamma})\right] - \\
&\qquad\qquad E\left[\widetilde{LR}_T(M,K,\widetilde{\theta})LR_T^*(M,K,\gamma)\right] + E\left[\widetilde{LR}_T(M,K,\widetilde{\theta})\widetilde{LR}_T(M,K,\widetilde{\gamma})\right] \\
&\qquad = \int\int\int \widetilde{L}_T(\theta)\mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}}\widetilde{L}_T(\gamma)\mathbf{1}_{\{\|\sqrt{T}\gamma\|<M\}}\mathbf{1}_{\{B^*(\cdot)>K\}}d\nu_\theta d\nu_\gamma d\nu_z - \\
&\qquad \int\int\int \widetilde{L}_T(\theta)\mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}}\widetilde{L}_T(\widetilde{\gamma})\mathbf{1}_{\{\|\sqrt{T}\widetilde{\gamma}\|<M\}}\mathbf{1}_{\{B^*(\cdot)>K\}}d\nu_\theta d\nu_{\widetilde{\gamma}} d\nu_z - \\
&\qquad \int\int\int \widetilde{L}_T(\widetilde{\theta})\mathbf{1}_{\{\|\sqrt{T}\widetilde{\theta}\|<M\}}\widetilde{L}_T(\gamma)\mathbf{1}_{\{\|\sqrt{T}\gamma\|<M\}}\mathbf{1}_{\{B^*(\cdot)>K\}}d\nu_{\widetilde{\theta}} d\nu_\gamma d\nu_z + \\
&\qquad \int\int\int \widetilde{L}_T(\widetilde{\theta})\mathbf{1}_{\{\|\sqrt{T}\widetilde{\theta}\|<M\}}\widetilde{L}_T(\widetilde{\gamma})\mathbf{1}_{\{\|\sqrt{T}\widetilde{\gamma}\|<M\}}\mathbf{1}_{\{B^*(\cdot)>K\}}d\nu_{\widetilde{\theta}} d\nu_{\widetilde{\gamma}} d\nu_z
\end{aligned}
$$

Lemma 6 implies that the integrands of all four terms weakly converge to the same limiting distribution. Thus, Crystal Ball condition give us that it is enough to show that $SupE[\widetilde{L}_T(M,K)^{2+\delta}]$ is finite. It can be proved by computations close to those in the proof of Lemma 2.

$$
\begin{aligned}
E[\widetilde{L}_T(M,K)^a] &= \int\int (2\pi)^{-\frac{k(T-1)}{2}}|K|^{-\frac{1}{2}}\exp[a\dot{\ell}'(M_e\otimes I_k)\widetilde{\theta} - a\widetilde{\theta}'(M_e\otimes J_1)\widetilde{\theta} - \frac{1}{2}\widetilde{\theta}'K^{-1}\widetilde{\theta}] \\
&\qquad \times\mathbf{1}_{[\|\sqrt{T}\theta\|<M]}\mathbf{1}_{[B^*(\cdot)<K]}d\widetilde{\theta}d\nu_z \\
&= c_1\cdot\int\exp[\frac{a^2}{4}\dot{\ell}'(M_e\otimes I_k)(a(M_e\otimes J_1)+K^{-1})^{-1}(M_e\otimes I_k)\dot{\ell}]\int(2\pi)^{\frac{-k(T-1)}{2}} \\
&\quad \times|a(M_e\otimes J_1)+K^{-1}|^{\frac{1}{2}}\ \exp[-\frac{1}{2}(\widetilde{\theta} - a(aM_e\otimes J_1 + K^{-1})(M_e\otimes I_k)2\dot{\ell})' \\
&\quad \times(a(M_e\otimes J_1)+K^{-1})(\widetilde{\theta} - a(aM_e\otimes J_1+K^{-1})(M_e\otimes I_k)2\dot{\ell})]\mathbf{1}_{[\|\sqrt{T}\theta\|<M]}\mathbf{1}_{[B^*(\cdot)<K]}d\widetilde{\theta}d\nu_z \\
&= c_1\int\exp\left[\frac{a^2}{4}\bar{\dot{\ell}}^{*\prime}(M_e\otimes I_k + (\frac{FF'}{T^2})^{-1}\otimes\frac{1}{a}\Omega^{*-1})^{-1}\bar{\dot{\ell}}^{*\prime}\right]\mathbf{1}_{[B^*(\cdot)<K]d\nu_z P\left[\|\sqrt{T}\theta\|<M\right]} \\
&c_1 P\left[\|\sqrt{T}\theta\|<M\right]\int\exp\left[\frac{a}{4}B(\Omega,\sqrt{a}J_1,\theta_0)\right]\mathbf{1}_{[B^*(\cdot)<K]}d\nu_z
\end{aligned}
$$

$$=\leq c_1 P\left[\|\sqrt{T}\theta\| < M\right]\exp\left[\frac{a}{4}K\right]$$

so that, for sufficiently large K there exits S such that $SupE[\widetilde{L}_T(M,K)^{2+\delta}] < S$.

(2) Proof of convergence under the alternative hypothesis: The proof can be done by showing that the distribution under the alternative hypothesis, $f_1(y|\theta_0,\theta_t)$ is contiguous to that under the null hypothesis, $f_0(y|\theta_0)$. The contiguity of the distribution in which the likelihood ratio has the asymptotic distribution as (3.8) has already been shown by Elliott and Müller (2006).

A.4. **Proof of Theorem 2.** (1) Let's rewrite $\hat{\theta} = \theta_0 + T^{-\frac{1}{2}}W_T$ where W_T is a $k \times 1$ random variable with $P[|W_T| > M] \to 0$ for arbitrarily large M. By using Condition 4 and continuous mapping theorem, we could get

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{[sT]}\widehat{J}_1^{-1/2}\dot{\ell}_t(\hat{\theta}) &= T^{-1/2}\sum_{t=1}^{[sT]}\widehat{J}_1^{-1/2}\dot{\ell}_t(\theta_0 + T^{-1/2}W_T) + o_p(1) \\ &= T^{-1/2}\sum_{t=1}^{[sT]}\widehat{J}_1^{-1/2}\dot{\ell}_t(\theta_0) - sK(\theta_0)W_T + o_p(1) \end{aligned}$$

Since $W_T(1)$ is constant for all $t \leq T$, The theorem can be easily proved by showing that $B(\Omega)$ doesn't change for the transformation from $\{\dot{\ell}_i(\theta_0)\}$ to $\{\dot{\ell}_i(\theta_0) + c\}$ where c is the $T \times 1$ vector of constants. Note that $M_e\dot{\ell}_i^*(\theta_0) = M_e\left[\dot{\ell}_i^*(\theta_0) + c\right]$. We could get

$$\begin{aligned} B(\Omega) &= \sum_{i=1}^{k}\dot{\ell}_i^{*\prime}(\theta_0)\{M_e - G_{a_i}\}\dot{\ell}_i^*(\theta_0) = \sum_{i=1}^{k}\dot{\ell}_i^{*\prime}(\theta_0)M_e[M_e + K_{ai}^{-1}]^{-1}M_e\dot{\ell}_i^*(\theta_0) \\ &= \sum_{i=1}^{k}\left[\dot{\ell}_i^{*\prime}(\theta_0) + c\right]M_e[M_e + K_{ai}^{-1}]^{-1}M_e\left[\dot{\ell}_i^*(\theta_0) + c\right] \\ &= \sum_{i=1}^{k}\dot{\ell}_i^{*\prime}(\widehat{\theta})\{M_e - G_{a_i}\}\dot{\ell}_i^*(\widehat{\theta}) + o_p(1) = \hat{B}(\Omega) + o_p(1) \end{aligned} \tag{A.13}$$

which shows the asymptotic equivalency under the null hypothesis. The equivalency under the alternative comes from the contiguity which is proven in Theorem 1.

(2) Note that the test $B(\Omega)$ can be interpreted as asymptotically most powerful for testing that $f(Z|\theta_0)$ versus $f(Z|\frac{1}{T}\delta_t, \theta_0 + \frac{1}{T}\bar{\delta}_0(\delta_t))$ when $\bar{\delta}_0(\delta_t)$ is equivalent to Condition 1) iv). Since $\hat{\phi}_T(Z|\Omega)$ has asymptotic α-size for $\bar{\delta}_0(\delta_t)$, Neyman-Pearson Lemma gives

$$h(\Omega) \leq \int\int \hat{\phi}_T(Z|\Omega) f\left(Z|\frac{1}{T}\delta, \theta_0 + \frac{1}{T}\bar{\delta}_0(\delta)\right) dZ d\nu_\delta + o_p(1) \tag{A.14}$$

Consequently (1) implies (2) which completes the proof. $\diamond$

A.5. **proof of Lemma 3.** Let's define $A_i = I_T + K_{ai}^{-1}$. The inverse of A_i can be expressed as,

$$A_i^{-1} = (I + K_{ai}^{-1})^{-1} = K_{ai}(I + K_{ai})^{-1} = I - (I + K_{ai})^{-1}$$

By Sherman-Morrison Lemma,

$$\begin{aligned}
&[M_e + K_{ai}^{-1}]^{-1} = A_i^{-1} - (A_i^{-1}e)(1 + e'A_i^{-1}e/T)^{-1}(e'A_i^{-1}) \\
&= I - (I + K_{ai})^{-1} + (1 + e'A_i^{-1}e/T)^{-1}\left[ee' - 2(I + K_{ai})^{-1}ee' + (I + K_{ai})^{-1}ee'(I + K_{ai})^{-1}\right]
\end{aligned}$$

Define $T \times (T-1)$ vector B_e as $B_eB_e' = M_e$.

$$\begin{aligned}
&M_e[M_e + K_{ai}^{-1}]^{-1}M_e \\
&= M_e - M_e(I + K_{ai})^{-1}M_e + (1 + e'A_i^{-1}e/T)^{-1}M_e(I + K_{ai})^{-1}ee'(I + K_{ai})^{-1}M_e \\
&= M_e - B_e(B_e(I + K_{ai})B_e')^{-1}B_e' = M_e - G_{ai}
\end{aligned}$$

where $G_a = H_a^{-1} - H_a^{-1}e(e'H_a^{-1}e)^{-1}e'H_a^{-1}$, $H_a = r_a^{-1}FA_aA_a'F'$, $A_a = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -r_a & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & \cdots & -r_a & 1 \end{pmatrix}$ and $r_a = 1 - aT^{-1}$. The third equality uses the fact that $M_ee = 0$. The last equality is by Lemma 4 of Elliott and Müller (2006). Consequently the test statistic can be written as

$$B(\Omega) = \sum_{i=1}^{k} \dot{\ell}_i^{*\prime}(\theta_0)\{M_e - G_{a_i}\}\dot{\ell}_i^*(\theta_0) \tag{A.15}$$

Lemma 6 of Elliott and Müller (2006) gives us the distribution of the test statistic which is the same as (3.8). $\diamondsuit$

A.6. **Proof of Lemma 4 and 5.** The proofs are similar to the proofs of Lemma 1, and 3, respectively, and are therefore omitted in the interest of brevity. $\diamondsuit$

A.7. **Proof of Theorem 3.** Since the test function ϕ_T is bounded in probability, Prohorov's Theorem implies that for every subsequence $\phi_{T'}$, there exists a further subsequence with $\phi_{T''} \Rightarrow \phi$ as $T'' \to \infty$ under H_0. Theorem 6.6 of Vaart (1998) gives the asymptotic distribution of $\phi_{T''}$ under H_1 as $L = I_{\{\phi\}}exp[\Lambda^S]$. Accordingly the following convergence holds.

$$\lim_{T'' \to \infty} E\left[\phi_{T''}(Z_T)\right] \longrightarrow^d E\left[\phi(S_\theta, W_\eta)exp[\Lambda^S(\Omega, h)]\right] \tag{A.16}$$

where $S_\theta = (\int W_\theta' dW_\epsilon - \int W_\theta' W_\epsilon(1), \int W_\theta' W_\theta - (\int W_\theta)'(\int W_\theta))$, W_θ is a Brownian motion independent of W_η and W_ϵ, and W_ϵ is a Brownian motion of which the covariance with W_η is $J_{\theta\eta}$. (A.16) enables us to use the limits of experiments to obtain the asymptotic power envelope for the testing problem. Let's define the two power functions in the limit experiments as follows.

$$\begin{aligned}
\Psi(\Omega) &= E\left[1_{\{\Lambda(\Omega) > k_\alpha\}}\exp[\Lambda(\Omega)]\right] \\
\Psi^S(\Omega, h) &= E\left[1_{\{\Lambda(\Omega) > k_\alpha\}}\exp[\Lambda^S(\Omega, h)]\right]
\end{aligned} \tag{A.17}$$

$\Psi(\Omega)$ gives the asymptotic power envelope in parametric models by theorem 1. By construction $\Psi^S(\Omega,h) \leq \Psi(\Omega)$. Therefore it is enough to show that

$$\Psi^S(\Omega,h) = \Psi(\Omega) \qquad \text{for all } \Omega, h$$

$$\begin{aligned}\Psi^S(\Omega,h) &= E\left[1_{\{\Lambda>k_\alpha\}}\exp[\Lambda^S]\right] = E\left[1_{\{\Lambda>k_\alpha\}}\exp[\Lambda]\exp\left[hW_\eta - \frac{h^2}{2}J_\eta\right]\right] \\ &= E\left[1_{\{\Lambda>k_\alpha\}}\exp[\Lambda]E\left[\exp\left[hW_\eta - \frac{h^2}{2}J_\eta\right]|S_\theta\right]\right]\end{aligned}$$

Note that W_η has zero covariance with $\int W'_\theta dW_\epsilon - \int W'_\theta W_\epsilon(1)$ so that W_η is independent of S_θ and normal with zero mean and variance J_η. Therefore,

$$\begin{aligned}E\left[\exp\left[hW_\eta - \frac{h^2}{2}J_\eta\right]|S_\theta\right] &= \int \exp\left[hW_\eta - \frac{h^2}{2}J_\eta\right]\exp\left[-\frac{1}{2}W'_\eta J_\eta^{-1}W_\eta\right]dW_\eta \\ &= \int \exp\left[-\frac{1}{2}(W_\eta - hJ_\eta)'J_\eta^{-1}(W_\eta - hJ_\eta)\right]dW_\eta = 1 \end{aligned} \tag{A.18}$$

Consequently, we get

$$\Psi^S(\Omega,h) = E\left[1_{\{\Lambda>k_\alpha\}}\exp[\Lambda]\right] = \Psi(\Omega) \tag{A.19}$$

which completes the proof. $\diamond$

A.8. **proof of Theorem 4.** Let's define $\widehat{LR}_T$ as

$$\widehat{LR}_T = \int exp\left[\hat{\dot{\ell}}^{\theta'}(M_e\otimes I_k)\theta - \frac{1}{2}\theta'(M_e\otimes J_\theta)\theta\right]d\nu_\theta = \int \hat{L}_T d\nu_\theta$$

Theorem 4 is proven by showing that $P[|\widetilde{LR}_T - LR^*_T| > \epsilon] \to 0$ under H_0 and H_1 where $LR^*_T = \int L^*_T d\nu_\theta$ and L^*_T is as defined in Theorem 1. Since LR^*_T is contiguous as shown in the proof of Theorem 1, it suffices to show the convergence only under the null hypothesis. Throughout the proof, I assume that θ_0 is known. The asymptotic invariancy of replacing θ_0 by $\hat{\theta}$ is straightforward from Theorem 2. be the same as $\widetilde{LR}_T$ except $\dot{\ell}^{\theta'}$ and J_θ) are replaced by $\hat{\dot{\ell}}^{\theta'}$ and $\hat{J}_\theta$), respectively. both the null and the alternative hypothesis. For $0 < M < \infty$, define $\widehat{LR}_T = \int \hat{L}_T \mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}}d\nu_\theta$, and $LR^*_T = \int L^*_T \mathbf{1}_{\{\|\sqrt{T}\theta\|<M\}}d\nu_\theta$. Note that for any $\epsilon > 0$, the following is satisfied

$$\begin{aligned}P[|LR^*_T - \widehat{LR}_T| > 3\epsilon] &\leq P[|LR^*_T - LR^*_T(M)| > \epsilon] & (i) \\ &+ P[|\widehat{LR}_T - \widehat{LR}_T(M)| > \epsilon] & (ii) \\ &+ P[|LR^*_T(M) - \widehat{LR}_T(M)| > \epsilon] & (iii)\end{aligned} \tag{A.20}$$

Therefore, it suffices to show that each term of (A.20) converges to zero, respectively. (i), and (ii) can be proved in the similar way as Theorem 1. Hence I will prove (iii) only by showing that

$$\ln(L_T^*) = \ln(\widehat{L}_T) + o_p(1) \tag{A.21}$$

so that

$$LR_T^*(M) = \int L_T^*(M) d\nu\theta = \int (1 + o_p(1)) \widehat{L}_T(M) d\nu_\theta = \widehat{LR}_T(M) + o_p(1) \tag{A.22}$$

For the notational convenience, The proof is done based on univariate θ_t and constant σ_t, The extension is straigtforward. Let $\theta_t^* = \theta_t \mathbf{1}\{|\sqrt{T}\theta_t| \leq M\}$. Then (A.21) is proved by showing that

$$\begin{aligned} \sum_{t=1}^{T} (\theta_t^* - \frac{1}{T}\sum_{i=1}^{T} \theta_i^*)' \dot{m}(X_t) \hat{\dot{\ell}}^g(\epsilon_t) &= \sum_{t=1}^{T} (\theta_t^* - \frac{1}{T}\sum_{i=1}^{T} \theta_i^*)' \dot{m}(X_t) \dot{\ell}^g(\epsilon_t) + o_p(1) && \text{(A.23)} \\ \hat{J}_\theta &= J_\theta + o_p(1) && \text{(A.24)} \end{aligned}$$

where $\dot{\ell}^g(\epsilon_t)$ is the 1st derivative of $\ln g(\epsilon_t)$. To simplify the proof, I replace $\dot{m}(X_t)$ by $\dot{m}(X_t)^* = \dot{m}(X_t)\mathbf{1}\{|\dot{m}(X_t)| \leq M_m\}$. It can be easily shown that the replacement does not affect the result by using exactly the same way as the proof of (i) and (ii) in Theorem 1. The proof of Lemma 4.3 of Schick (1987) implies that if $\sqrt{T}\int \hat{\dot{\ell}}^g(\epsilon) d\epsilon \nrightarrow 0$, then $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{\dot{\ell}}^g(\epsilon_t) - \dot{\ell}^g(\epsilon_t)\right) = \sqrt{T}\int \hat{\dot{\ell}}^g(\epsilon) g(\epsilon) d\epsilon + o_p(1)$. It implies that (A.23) can be obtained if we have the following

$$\sum_{t=1}^{T} \theta_t^* \dot{m}^*(X_t) \left(\hat{\dot{\ell}}^g(\epsilon_t) - \bar{\dot{\ell}}^g(\epsilon_t)\right) = T\bar{\theta}^* \bar{\dot{m}}^* \int \left(\hat{\dot{\ell}}^g(\epsilon) - \bar{\dot{\ell}}^g(\epsilon)\right) g(\epsilon) d\epsilon + o_p(M) \tag{A.25}$$

where $\bar{\theta}^*$, $\bar{\dot{\ell}}^g(\epsilon_t)$ and $\bar{\dot{m}}^*$ are their sample means. I first show that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \dot{m}^*(X_t)\left(\hat{\dot{\ell}}^g(\epsilon_t) - \dot{\ell}^g(\epsilon_t)\right) = \sqrt{T}\bar{\dot{m}}^*(X_t)\int \hat{\dot{\ell}}^g(\epsilon) g(\epsilon) d\epsilon + o_p(1) \tag{A.26}$$

Theorem 6.2 of Koul and Schick (1997) implies that (A.26) holds if for some sequence $< \tau_T >$ of positive integers tending to infinity, the following is satisfied (See pp.269-271 of Koul and Schick (1997)))

$$\frac{1}{T}\sum_{1 \leq l,t \leq T}\sum_{|t-l| > \tau_t} E\left(|\dot{m}^*(X_t) - E[\dot{m}^*(X_t)|\epsilon_1, \ldots, \epsilon_{l-1}, \epsilon_{l+1}, \ldots, \epsilon_T]|^2\right) = o_p(1) \tag{A.27}$$

Note that $E\left[\dot{m}^*(X_t)|\epsilon_1, \ldots, \epsilon_{l-1}, \epsilon_{l+1}, \ldots, \epsilon_T\right] = E[\dot{m}^*(X_t)|\epsilon_1, \ldots, \epsilon_{l-1}]$ if $l > t$ because of Condition 2). Consequently we have only to show that

$$\frac{1}{T}\sum_{1 \leq t,l \leq T}\sum_{|t-l| > \tau_t} E\left(|\dot{m}^*(X_t) - E[\dot{m}^*(X_t)|\epsilon_1, \ldots, \epsilon_l]|^2\right) = o_p(1) \tag{A.28}$$

for all $l < t$. Let's set $\tau_t = T^{1/2-\alpha}$ where $0 < \alpha < 1/2$. Then,

$$\begin{aligned}
& \frac{1}{T} \sum_{1 \le t,l \le T} \sum_{|t-l| > \tau_t} E\left(|\dot{m}^*(X_t) - E[\dot{m}^*(X_t)|\epsilon_1, \ldots, \epsilon_l]|^2 \right) \\
= & \frac{1}{T} \sum_{1 \le t,l \le T} \sum_{|t-l| > \tau_t} E\left(|(\dot{m}^*(X_t) - E[\dot{m}^*(X_t)]) + (E[\dot{m}^*(X_t)] - E[\dot{m}^*(X_t)|\epsilon_1, \ldots, \epsilon_l])|^2 \right) \\
\le & \frac{1}{T} \sum_{1 \le t,l \le T} \sum_{|t-l| > \tau_t} E\left(|\dot{m}^*(X_t) - E[\dot{m}^*(X_t)]|^2 \right) + E\left(|E[\dot{m}^*(X_t)] - E[\dot{m}^*(X_t)|\epsilon_1, \ldots, \epsilon_l]|^2 \right)
\end{aligned}$$

The first term is $O_p(T^{-2\alpha})$ because $\frac{1}{T}\sum_{1 \le t,l \le T}\sum_{|t-l|>\tau_t} E\left(|\dot{m}^*(X_t) - E[\dot{m}^*(X_t)]|^2\right) < T^{-2\alpha}M_x^2 = O_p(T^{-2\alpha})$. Theorem 4.2 of Davidson (1994) implies that the second therm is also is $O_p(T^{-2\alpha})$ because,

$$\begin{aligned}
& \frac{1}{T} \sum_{1 \le t,l \le T} \sum_{|t-l| > \tau_t} E\left(|E[\dot{m}^*(X_t)] - E[\dot{m}^*(X_t)|\epsilon_1, \ldots, \epsilon_l|^2 \right) \\
\le & \frac{1}{T} \sum_{t=[T^{1/2-\alpha}]+1}^{T} (t - [T^{\frac{1}{2}-\alpha}]) E[36 \cdot |\dot{m}^*(X_t)|^2] \le T^{-2\alpha} M_x = O_p(T^{-2\alpha})
\end{aligned}$$

where $[x]$ is the the largest integer less than x. It satisfies (A.28). Proving (A.25) based on (A.26) is equivalent to proving (A.26) based on (A.8). Therefore we have only to show that $\sqrt{T}\theta_t^*$ satisfies (A.27). Note that θ_t^* is independent of $\{\epsilon_t\}$ and by Condition 1) $E\left[\theta_t^*|\epsilon_1, \ldots, \epsilon_{l-1}, \epsilon_{l+1}, \ldots, \epsilon_T\right]$ for all l. Consequently,

$$\begin{aligned}
& \frac{1}{T} \sum_{1 \le l,t \le T} \sum_{|t-l| > \tau_t} E\left(\left\| \sqrt{T}\theta_t^* - E[\sqrt{T}\theta_t^*|\epsilon_1, \ldots, \epsilon_{l-1}, \epsilon_{l+1}, \ldots, \epsilon_T] \right\|^2 \right) \\
= & \frac{1}{T} \sum_{1 \le l,t \le T} \sum_{|t-l| > \tau_t} E\left(\left\| \sqrt{T}\theta_t^* \right\|^2 \right) \le \frac{1}{T} \sum_{1 \le t,l \le T} \sum_{|t-l| > \tau_t} M = O_p(T^{-2\alpha})
\end{aligned}$$

which satisfies (A.28). Convergence of $\hat{J}_\theta$ is proved by Schick (1987) which completes the proof. $\diamond$

References

Akharif, A., and M. Hallin (2003): "Efficient Detection of Random Coefficients in Autoregressive Models," *The Annals of Statistics*, 31(2).

Andrews, D., and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica*, 62(3).

Banerjee, M. (2005): "Likelihood Ratio Tests under Local Alternatives in Regular Semiparametric Models," *Statistica Sinica*, 15.

Benghabrit, Y., and M. Hallin (1998): "Locally Asymptotically Optimal Tests for AR(p) against diagonal bilinear dependence," *Journal of Statistical Planning and Inference*, 68.

Bickel, P. (1982): "The 1980 Wald Memorial Lectures: On Adaptive Estimation," *Annals of Statistics*, 10(3).

Choi, S., W. Hall, and A. Schick (1996): "Asymptotically Uniformly Most Powerful Tests in Parametric and Semiparametric Models," *The Annals of Statistics*, 24(2).

Davidson, J. (1994): *Stochastic Limit Theory: Advanced Texts in Econometrics*. Oxford, London, 1st edition edn.

Drost, F. C., and C. A. J. Klaassen (1997): "Efficient Estimation in Semiparametric GARCH Models," *Journal of Econometrics*, 81.

Elliott, G., and U. Müller (2006): "Efficients Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economics Studies*, 73, 907–940.

Engle, R., D. Hendry, and J. Richard (1983): "Exogeneity," *Econometrica*, 51, 277–304.

Fernandez, C., and M. Steel (1998): "On Bayesian Modeling of Fat Tails and Skewness," *Journal of the American Statistical Association*, 93(441), 359–371.

Hall, W., and D. Mathiason (1990): "On Large-Sample Estimation and Testing in Parametric Models," *International Statistical Review*, 58(1), 77–97.

Hallin, M., and J. Jurečková (1999): "Optimal Tests for Autoregressive Models Based on Autoregression Rank Scores," *The Annals of Statistics*, 27(4).

Jansson, M. (2008): "Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis," *Econometrica*, 76.

Jeganathan, P. (1995): "Some Aspects of Asymptotic Theory with Applications to Time Series Models," *Econometric Theory*, 11.

King, M. (1988): "Towards a Theory of Point Optimal Testing," *Econometric Reviews*, 6, 169–218.

Koenker, R., and Q. Zhao (1996): "Conditional Quantile Estimation and Inference for ARCH Models," *Econometric Theory*, 12, 793–813.

Koul, H. L., and A. Schick (1997): "Efficient Estimation in Nonlinear Autoregressive Time-Series Models," *Bernouilli*, 3(3), 247–277.

LeCam, L. (1970): "On the Assumptions Used to Prove Asymptotic Normality of Maximum Likelihood Estimators," *Annals of Mathematical Statistics*, 41(1), 802–828.

Ling, S. (2003): "Adaptive Estimators and Tests of Stationary and Nonstationary Short- and Long-Memory ARFIMA-GARCH Models," *Journal of The American Statistical Association*, 98(464), 955–967.

Ling, S., and M. McAleer (2003): "On Adaptive Estimation in Nonstationary ARMA Models with GARCH Errors," *The Annals of Statistics*, 31(2), 642–674.

Murphy, S., and A. V. der Vaart (1997): "Semiparametric Likelihood Ratio Inference," *The Annals of Statistics*, 25(4), 1471–1509.

Nyblom, J. (1986): "Testing for the Deterministic Linear-Trend in Time Series," *Journal of the American Statistical Association*, 81, 545–549.

——— (1989): "Testing for the Constancy of Parameters over Time," *Journal of the American Statistical Association*, 84(405), 348–368.

Perron, P. (2006): "Dealing with Structural Breaks," *Palgrave Handbook of Econometrics, Vol. 1 : Econometric Theory, K. Patterson and T.C. Mills (eds.)*, pp. 278–352.

Schick, A. (1987): "A Note on the Construction of Asymptotically Linear Estimators," *Journal of Statistial Planning and Inference*, 16, 89–105.

Shin, D. W., and B. S. So (1999): "Unit Root Tests Based on Adaptive Maximum Likelihood Estimation," *Econometric Theory*, 15, 1–23.

Vaart, A. v. d. (1998): *Asymptotic Statistics*. Cambridge University Press, New York, 1st edn.

White, H. (2001): *Asymptotic Theory for Econometrics*. Academic Press, San Diego, revised edn.