Title
Essays on optimal tests for parameter instability

Author
Lee, Dong Jin

Publication Date
2008

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays on Optimal Tests for Parameter Instability

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in
Economics

by

Dong Jin Lee

Committee in charge:

Professor Graham Elliott, Chair
Professor James D. Hamilton
Professor Ivana Komunjer
Professor Patrick J. Fitzsimmons
Professor Dimitris N. Politis

2008

The dissertation of Dong Jin Lee is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2008

To Jeong-Hee, Lauren, and my parents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Foremost, I would like to express my deepest gratitude to my advisor, Graham Elliott, for his encouragement, patience, and constant invaluable advice. Graham's truly economic intuition on all econometric problems has exceptionally inspired me and enriched my work and thought. I have been amazingly fortunate to have him as an advisor and am indebted to him more than he knows.

I am especially grateful to James D. Hamilton, and Ivana Komunjer for their helpful comments and suggestions. Many thanks also to the other member of my committee: Patrick J. Fitzsimmons, and Dimitris N. Politis.

My sincere thanks are due to Jay-Min Lee, and Byung-Sam Yoo, who inspired me to pursue academics while at Yonsei University. I convey warm acknowledgement to many colleagues at UCSD especially to Young Joon Park, Lei Meng, Jeff Ohara, Michael Yang, Karmen Suen, and other Korean graduate students.

At the same time, I must express my grateful thank to my parents for all their unflagging love and support throughout my life.

Finally, words fail to me to express my gratitude to my wife Jeong-Hee and my daughter Lauren. None of this would have been possible without their love and understanding.

VITA

1996	B.A., Economics, Yonsei University, Seoul, Korea
1999	M.A., Economics, Yonsei University, Seoul, Korea
1999-2002	Economist Bank of Korea, Seoul, Korea
2008	Ph.D., Economics University of California, San Diego.

FIELDS OF STUDY

Major Field: Econometrics (Time Series Analysis, Applied Econometrics (Marcroeconomics, Finance))

Studies in Time Series Analysis
Professors Graham Elliott, James Hamilton, and Ivana Komunjer

Studies in Applied Econometrics
Professors Graham Elliott, James Hamilton, and Ivana Komunjer

ABSTRACT OF THE DISSERTATION

Essays on Optimal Tests For Parameter Instability

by

Dong Jin Lee

Doctor of Philosophy in Economics

University of California, San Diego, Graduation Year

Professor Graham Elliott, Chair

There are a large number of tests for parameter instability designed for specific types of unstable parameter processes and error distributions. However, it is difficult to identify those types in practice based on a priori knowledge. My dissertation studies methods and conditions under which asymptotically efficient tests are obtained without the knowledge of the unstable parameter process and the error distribution.

First, I examine asymptotically optimal tests for parameter instability in which the difficulty in identifying the unstable process is explicitly considered. Elliott and Müller (2006) provide conditions under which a large class of breaking processes lead to asymptotically equivalent optimal tests. Their finding, however, is restricted to linear Gaussian models. I improve upon their work in two ways. First, I show that the asymptotic equivalency of the efficient tests for parameter instability holds even in a broader set of parametric models which includes nonlinear models with non-Gaussian error distributions. It implies that the knowledge of the unstable parameter process is asymptotically irrelevant for testing purposes. Second, I suggest a test statistic that is asymptotically optimal for a broad set of unstable parameter processes.

Second, I study asymptotically efficient tests for parameter instability in general semiparametric models in which the error distribution is unknown but treated as an infinite dimensional nuisance parameter. I first derive the asymptotic power envelope with unknown density and suggest conditions under which a semiparametric model would have the same asymptotic power envelope with known error distribution. The conditions are weak enough to cover a wide range of error distributions by relaxing the twice differentiability and allowing for skewness. An efficient test statistic is then suggested, which is adaptive in the sense that allowing unknown error distribution gives no loss of asymptotic power. This implies that the knowledge of the error distribution is asymptotically irrelevant under mild conditions.

Finally, the suggested parameter instability tests are applied to various quantile models for U.S. inflation process such as Phillips curve, P- star model, and AR models. The tests result shows a strong evidence of parameter instability in most quantile levels of all models.

Chapter I

A Review of Optimal Tests for Parameter Instability

This chapter studies the asymptotic optimality in the hypothesis testing of parameter instability. I first review the behavior of asymptotically point optimal tests in the presence of unknown nuisance parameters, which will provide a conceptual background of the current optimal parameter instability tests and my tests in the dissertation. The classic concept of optimal tests based on a sufficient and complete statistic is evaluated in terms of the most powerful tests in the least favorable parametric submodels. I then examine popular optimal tests for parameter instability; Andrews and Ploberger (1994), Elliott and Müller (2006), and Nyblom (1989). This chapter presents that these optimal tests are interpreted as the weighted average of the asymptotically point optimal tests in the least favorable parametric submodels.

I.1 Introduction

The object of this review chapter is to discuss the asymptotic optimality of the hypothesis test that the parameters of interest are unstable in time series. Consider a parametric model indexed by (β_t, η) for $t = 1, \ldots, T$, $\beta \in \mathbb{R}^k$, and $\eta \in \mathbb{R}^q$. Parameter instability indicates that β_t permanently changes across time. Examples of unstable models most widely used in economics are 'Structural Breaks' and 'Time Varying Parameter'.

Parameter instability is a long-standing problem in econometric modeling when we deal with time series data. Much effort has been devoted to obtaining a powerful test to detect the instability. However, the parameter instability testing problem generally violates regularity conditions for the classical likelihood ratio test to pertain optimality, causing us difficulty in obtaining optimal tests. One of the reason is because the testing problem includes nuisance parameters which are present only under the alternative hypothesis. In structural break models, the parameter representing unknown break point appears only under the alternatives. In time varying parameter models, any parameters that determine the shape of the distribution of the unstable parameters is not identifiable under the null of stable parameters.

In this regard, only a few works suggest asymptotically optimal tests. Andrews and Ploberger (1994) suggest a class of optimal tests for structural breaks in the sense that they provide asymptotically best average power results. Elliott and Müller (2006) suggest an asymptotically optimal invariant test in a linear Gaussian model which is most powerful against a broad set of unstable parameter processes, including both structural breaks and time varying parameters. Nyblom (1989) derives a asymptotically locally most powerful test when the parameter follows martingale process. These tests explicitly and implicitly get around the nonstandard problem by constructing the weighted average of the optimal tests across all possible value of the nuisance parameters that is not identified under the null hypothesis.

This chapter evaluates the optimality of the current popular tests. I first review the behavior of standard asymptotically optimal tests for a regular problem of finite dimensional parameters, under local alternative hypothesis. The regular problem indicates that local asymptotic nomality (LAN) type approximation of likelihood ratio function is available and $\sqrt{T}-$ regular estimates of the nuisance parameters exist. The regular estimates of the parameter of interest does not necessarily have to exist because this chapter consider a testing problem that focus on specific point of alternative hypothesis, which most optimal parameter instability tests do.

I study the classical standard optimal test in terms of the optimal test in the least favorable parametric submodel(LFPS). LFPS method has an advantage in that it avoids complicate mathematics, and can be easily generalized to obtain the asymptotic power envelope in the semiparametric problem in which the nuisance parameter is infinite dimensional.

The existing optimal tests are then reviewed associated with the optimality in the least favorable parametric submodel. Even though their testing problems are non-standard, their optimality coincides with that of LFPS in the sense that the test are interpreted as the weighted average of the optimal tests in LFPS.

I.2 Behaviors of Asymptotic Optimal Tests

This section reviews general concepts of asymptotically optimal tests with and without nuisance parameter. An optimal test is defined as a test that has the maximal power against a particular alternative within a class of tests. Different classes of alternative hypotheses and tests give different concepts of optimal tests such as uniformly most powerful (UMP) tests, locally most powerful test, and point optimal or $\beta-$optimal tests (King (1988)).

A UMP test maximizes the testing power for any parameter process in the alternative hypothesis. Although it is most powerful among all other tests by definition, it is well known that there is no UMP test for parameter instability. A locally most powerful test and a point optimal tests are useful concepts of optimality when UMP test does not exist. A locally most powerful test maximizes the slope of the power function when the parameters of interest are at the boundary of the null and the alternative hypothesis. One problem of this test is that the point at which the test has the maximum power is too close to the null space so that it sometimes fail to adequately reflect that the alternative parameter space is apart from the null in a distinguishable distance. A point optimal test maximizes power at a predetermined point under the alternative hypothesis. An adequate choice of the optimal point allows the point optimal test to have the power close to the optimal in the other points of alternative. In parameter instability tests, Nyblom (1989) suggests a locally most powerful tests, while Andrews and Ploberger (1994) and Elliott and Müller (2006) consider point optimal tests.

One problem of the exact optimal test is that these finite sample theories of optimality are applied only to rather special parametric families. On the other hand, asymptotic optimality will apply more generally to parametric families satisfying smoothness conditions. Most of the exact optimal tests in finite sample have their counterparts for large samples. The existing optimal tests for parameter instability consider asymptotic counterparts rather than looking for an exact optimal test, and the optimal tests derived in the dissertation are also focused on the asymptotic optimality .

In this regard, this section provides the conceptual background of the tests suggested in my dissertation by restricting our interest to the asymptotic counterparts of point optimal tests in general. I first deal with the case when there is no nuisance parameters. Then I generalize to a testing problem when unknown nuisance parameters are present. The latter is more realistic and models with unstable parameters generally contain nuisance parameters such as the initial value of the unstable parameters, and stable part of the parameter set.

I.2.A Optimal Tests without Nuisance Parameter

Consider i.i.d. stochastic process, $Z \equiv \{Z_t : \Omega \rightarrow \mathbb{R}^p,\ p \in \mathbb{N},\ t = 1, ..., T\}$, of which the conditional density is characterized as a parametric model $\mathcal{P} = \{F(z|\beta) : \beta \in R^k\}$ with dominating measure μ and corresponding densities $f(z|\beta) = dF(z|\theta)/dz$. Suppose we are interested in testing $H_0 : \beta = \beta_0$ against a simple alternative $H_1 : \beta = \beta_1$. Let's denote a test function as ϕ_T. If we restrict our interest to the size-α tests, i.e. $E_{\beta_0}[\phi_T] \leq \alpha$, Neyman-Pearson lemma implies that the most powerful test rejects when the log likelihood ratio statistic

$$LR_T \equiv log[L_T(\theta_1)/L_T(\theta_0)]$$

is sufficiently large, where

$$L_T(\theta) = \prod_{t=1}^{T} f(Z_t; \theta) \tag{I.2.1}$$

denotes the likelihood function. For simplicity, I assume that β_t is scalar. We are interested in the asymptotic counterpart of the likelihood ratio. For this purpose, we would like to obtain certain expansion of the likelihood ratio. The classical Taylor expansion, however, requires the twice differentiability of the density function with further assumption of the remainder terms, which is often too strict to be satisfied. In order to avoid such strong assumptions, it turns out to be useful to work with square roots of densities. Furthermore, imposing conditions for a mean property of densities provides a way to mitigate conditions for the remainder terms. The following smoothness condition, called *quadratic mean differentiability* is desirable to obtain the local expansion.

Definition 1 ***(Quadratic Mean Differentiability: QMD)*** *Let* $\xi_t(\cdot,\beta) = \sqrt{f(Z_t|\beta)}$*. The density* $f(\cdot|\beta)$ *is quadratic mean differentiable (QMD) at* β_0 *if there exists a vector of real-valued functions* $\dot{\xi}_t(\cdot,\beta_0)$ *such that*

$$\mathbb{E}_{\beta_0}\left(\left[\left(\frac{\xi_t(\cdot,\beta_0+h_\beta)}{\xi(\cdot,\beta_0)}-1\right)-h_\beta\frac{\dot{\xi}_t(\cdot,\beta_0)}{\xi_t(\cdot,\beta_0)}\right]^2\right)\to 0 \text{ as } \|h_\beta\|\to 0, \quad \forall t\leq T \quad \text{(I.2.2)}$$

The vector-valued function $\dot{\xi}(\cdot,\beta)$, called Hellinger derivative, takes over the role of the classical score vector. It will be shown that, under standard circumstances, the Hellinger derivative has all the information about the random property for testing purpose so that the test based on $(\dot{\xi}_t(\cdot,\beta))$ is asymptotically optimal. *QMD* is weak enough to be satisfied by a wide variety of densities and strong enough to deliver the approximation similar to the Taylor expansion. Let's define $\dot{\ell}_t(\beta_0)$ and the Fisher information I_β as

$$\begin{aligned} \dot{\ell}_t(\beta_0) &= 2 \cdot \frac{\dot{\xi}(\cdot, \beta_0)}{\xi(\cdot, \beta_0)} \\ I_\beta &= 4 \cdot E_{\beta_0}\left[\left(\frac{\dot{\xi}(\cdot, \beta_0)}{\xi(\cdot, \beta_0)}\right)^2\right] \end{aligned}$$

The following lemma, due to LeCam, provides the asymptotic expansion of the likelihood ratio and its distributional property. (Theorem 12.2.3 in Lehman and Romano (2005))

Lemma 1 ***(Local Asymptotic Normality:LAN)*** *Suppose $f(\cdot|\beta)$ is quadratic mean differentiable with $\dot{\ell}_t(\beta_0)$ and the positive definite fisher information I_β. Let's define $S_T = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \dot{\ell}_t(\beta_0)$ and $\beta_1 = \beta_0 + \frac{1}{\sqrt{T}} h_\beta$ where $\|h_\beta\| \leq M < \infty$. Then under H_0, $E_{\beta_0}[\dot{\ell}_t(\beta_0)] = 0$ and*

$$LR_T = h_\beta S_T - \tfrac{1}{2} h_\beta^2 I_\beta + o_{p_{\beta_0}}(1) = \widetilde{LR}_T + o_{p_{\beta_0}}(1) \tag{I.2.3}$$

$$S_T \rightsquigarrow N(0, I_\beta) \tag{I.2.4}$$

where $\rightsquigarrow$ denotes convergence in distribution

Note that the asymptotic expansion and its distribution in Lemma 1 are suggested only under the null hypothesis. However, the power of a test is the property under the alternative hypothesis. Accordingly, in order to obtain an asymptotically optimal test, we need to guarantee that the approximation in Lemma 1 holds also under the alternative hypothesis. An useful concept to derive this property is the contiguity.

Contiguity is an asymptotic version of absolute continuity. In order to motivate the concept, let $Q = F(\cdot;\theta_1)$ and $P = F(\cdot;\theta_0)$ be probability distributions under H_1 and H_0, respectively. Then Q is absolutely continuous with respect to P if $P(A) = 0$ implies $Q(A) = 0$ for every measurable set A. If the alternative distribution is absolutely continuous with respect to the null distribution, the alternative distribution of a test statistic, $T = T(Z)$, and thereby the power of a test can be calculated from the null distribution through the following formula.

$$E_Q[f(T)] = E_P\left[f(T)\frac{dQ}{dP}\right] \tag{I.2.5}$$

where f is some measurable function. Contiguity permits an analogous statement in the large samples, which implies that it has two useful properties: First, if a sequence of a statistic T_T converges in probability to T under H_0, then the convergency holds even under H_1. Second, asymptotic counterpart of (I.2.5) provides a way to obtain the distribution of T_T under H_1. By plugging $1(T)$ into (I.2.5) where $1(\cdot)$ is an indicator function, we get the asymptotic distribution of T_T is

$$T_T \rightsquigarrow E_P\left[1(T)\exp(LR)\right] \qquad \text{under } H_1 \tag{I.2.6}$$

where LR is the asymptotic counterpart of the log-likelihood ratio. (Theorem 6.6 in Vaart (1998)) Consider sequence of the null and the alternative distribution, $\{P_T, Q_T\}$. Suppose we have a test statistic $T_T = T_T(Z)$ of which the asymptotic null distribution and other properties are easily obtained. Our purpose is to evaluate the power property of T_T, and thereby to find out an asymptotically optimal test. It turns out that the local alternative, $\theta_T = \theta_0 + \frac{1}{\sqrt{T}}h_\beta$, together with QMD., provides contiguity of Q_T. Then the contiguity implies that the asymptotic alternative properties of T_T can be obtained from the asymptotic null distribution, which is summarized as

1. (LeCam's first lemma) For any statistics $T_T : \Omega \to \mathbb{R}^p$, if $T_T \to 0$ under H_0, then $T_T \to 0$ under H_1.

2. (LeCam's third lemma) The score function S_T is asymptotically as follows

$$S_T \rightsquigarrow N(I_\beta h_\beta, I_\beta) \tag{I.2.7}$$

These properties provide a convenient way to derive an asymptotically (point)-optimal test. A suggested test statistic is based on the LAN of the likelihood ratio,$\widetilde{LR}_T$. Lemma 1 gives that $|\widetilde{LR}_T - LR_T|$ converges to zero under the null hypothesis. LeCam's first lemma indicates that the asymptotic equivalency holds even under H_1, which implies that any increasing transformations of $\widetilde{LR}_T$ are asymptotically most powerful (AMP) against a point alternative hypothesis h_β[1]. Since the score function S_T is the only random factor in the LAN, An asymptotically point optimal test against a point alternative h_β is to reject the null hypothesis if S_T is sufficiently large, i.e. the asymptotically point optimal test ϕ_T is defined as

$$\phi_T = \begin{cases} 1 & \text{if } S_T \geq c_\alpha \\ 0 & \text{otherwise} \end{cases}$$

Note that the test function is equivalent to Rao's score test which is known to maximizes the derivative of the power function at β_0. It is asymptotically locally most powerful in the sense that it maximizes the slope of the power function. Consequently, we would infer that the asymptotically point optimal tests have equivalent power properties to that of asymptotically locally most powerful test in this standard LAN testing problems. By LeCam's third lemma, the asymptotic power (ψ_h) and the critical value ($c_{h,\alpha}$) can be derived as,

[1]The asymptotic optimality holds even under unknown h in one-sided test. See Lehman and Romano (2005).

$$c_\alpha = I_\beta^{1/2} z_{1-\alpha} \tag{I.2.8}$$

$$\psi_{h_\beta} = 1 - \Psi[z_{1-\alpha} - h_\beta I_\theta^{1/2}] \tag{I.2.9}$$

where $z_{1-\alpha} = \Psi^{-1}(1-\alpha)$ is the $1-\alpha$ quantile of $N(0,1)$ (See Lehman and Romano (2005) Lemma 13.3.1 for details.).

I.2.B Optimal Tests in the Presence of Unknown Nuisance Parameter

The previous section assumes that the parameter of interest β is the only parameter in the model. However, it is more plausible to allow that the nuisance parameters are present. A familiar example is when only a pat of the regression coefficients are to be tested in a linear regression model. In structural break tests, this is called a partial structural break test. Another familiar case is to test the regression coefficients where the error term belongs to some parametric family of distributions indexed by finite numbers of parameters such as asymmetric power distribution.

In this section, I generalize the previous model by introducing a finite dimensional nuisance parameter η. The true set of conditional densities of Z_t is now characterized as a parametric family $\mathcal{P}_\eta = \{F(z|\theta) : \theta = (\beta, \eta), \beta \in \mathbb{R}, \eta \in \mathbb{R}^q\}$ with dominating measure μ and corresponding densities $f_t(z|\theta) = dF_t(z|\theta)/dz$.

In this asymptotic set-up, we consider the local parametrization of θ such that the alternative distribution is contiguous to the null distribution.

$$\beta = \beta_0 + \frac{1}{\sqrt{T}} h_\beta \qquad \eta = \eta_0 + \frac{1}{\sqrt{T}} h_\eta$$

where η_0 is the true value of η and $h = (h_\beta, h'_\eta) \in \mathcal{H}_\theta$ is bounded where the local parameter space $\mathcal{H}_\theta$ is a *Hilbert space*. Analogous to the previous model without nuisance parameter, I impose the smoothness condition that the density $f(z|\theta)$ is *QMD* with respect to both β and η. Consequently, LAN can be written as

$$LR_T^N \equiv log\left(\frac{L_T(\theta_1)}{L_T(\theta_0)}\right) = h'S_T - \frac{1}{2}h'Ih + o_{p_{\theta_0}}(1) = \widetilde{LR}_T + o_{p_{\theta_0}}(1) \tag{I.2.10}$$

$$\begin{aligned} S_T &= \begin{pmatrix} S_T^\beta \\ S_T^\eta \end{pmatrix} = \begin{pmatrix} \frac{1}{T}\sum \dot{\ell}_t^\beta \\ \frac{1}{T}\sum \dot{\ell}_t^\eta \end{pmatrix} \rightsquigarrow N(0, I) \quad \text{under } H_0 \\ I &= \begin{pmatrix} I_\beta & I_{\beta\eta} \\ I'_{\beta\eta} & I_\eta \end{pmatrix} \end{aligned}$$

where $\dot{\ell}_t^\beta$ and $\dot{\ell}_t^\eta$ are the first derivative with respect to β and η, respectively, $I_{\beta\eta} = E[\dot{\ell}_t^\beta \dot{\ell}_t^{\eta'}]$, and $I_\eta = E[\dot{\ell}_t^\eta \dot{\ell}_t^{\eta'}]$. The alternative distribution is shown to be contiguous in the local alternatives so that S_T is also asymptotically normal with mean h and variance I under H_1. Let's define $(1+q)\times 1$ vector $\iota_1 = (1, 0, \ldots, 0)')$. The problem is to test

$$H_0 : \iota_1'\theta = \beta_0 \qquad \text{vs} \qquad H_1 : \iota_1'\theta = \beta_0 + \frac{1}{\sqrt{T}} h_\beta$$

Since the model has unknown perturbation h_η, we need to restrict a set of a test under which an optimality is defined to have the best asymptotic power among the set that covers the perturbation. One way is to restrict the set of tests that have correct asymptotic size property regardless of the perturbation h_η, i.e. I consider a set of tests that are, for a fixed $\alpha > 0$

$$\lim_{T\to\infty} SupE_{(0,h_\eta)}\phi_T(Z) \leq \alpha \qquad \text{for every } h_\eta \tag{I.2.11}$$

In order to obtain an asymptotically optimal test and the power envelope, it is useful to use the method of limits of experiments. An experiment can be regarded as a synonym of a probability model, and a sequence of experiments is defined to converge to a limit experiment if the sequence of likelihood ratio processes converges in distribution to the likelihood ratio process of the limit experiment, i.e. for every $h \in \mathcal{H}$, there exists a probability measure $L(\theta)$ such that

$$\frac{L_T(\theta_1)}{L_T(\theta_0)} \rightsquigarrow \frac{L(\theta_1)}{L(\theta_0)} \quad \text{under } \theta_0$$

The reason that a limit experiment method is a useful tool is because a limit experiment is always statistically easier then a given sequence. Suppose a sequence of tests T_T converges under a given parameter h in distribution to a limit L_h, for every parameter h. Then the asymptotic property of the sequence T_T may be judged from the set of limit laws, $\{L_h\}$. Theorem 9.3 of Vaart (1998) implies that every weakly converging sequence of test statistics converges to a test statistic in the limit experiment. A consequence is that asymptotically no sequence of statistical procedures can be better than the best procedure in the limit experiment. In this way the limit experiment obtains the character of an asymptotic power envelope. The following theorem (Theorem 13.4.1 in Lehman and Romano (2005)) summarizes it.

Theorem 2 *Suppose* $\{Q_{T,h}, h \in \mathcal{H}_\theta\}$ *is an asymptotically normal sequence of mod-*

els with covariance matrix I, Let ϕ_T be a test function. Let $\psi_T(h)$ denote the power of ϕ_T against $Q_{T,h}$. Then for every subsequence $\{T_j\}$, there exists a further subsequence $\{T_{j_m}\}$ and a test ϕ in the limiting experiment $N(h, I^{-1})$ such that, for every h,

$$\psi_{n_{jm}}(h) \longrightarrow \psi(h)$$

where $\psi(h)$ is the power of ϕ.

Now according to Theorem 2, we can approximate the power of a test sequence ϕ_T by the power of a test in the limit experiment $\phi(X)$, where $X = I^{-1}S \sim N(h, I^{-1})$ under H_1. Section 3.9 of Lehman and Romano (2005) implies that there exists an asymptotically point optimal test based on X which rejects for large values of $\iota_1'X$. Note that under H_1,

$$\iota_1'X \sim N(h_\beta, \sigma_0^2) \tag{I.2.12}$$

where

$$\sigma_0^2 = \iota_1'I^{-1}\iota_1 = (1,1)^{th} \text{ element of } I^{-1} \tag{I.2.13}$$

Hence, the critical value (c_α^*) and the power (ψ^*) of the test is then

$$c_\alpha^* = z_{1-\alpha}\sigma_0 \tag{I.2.14}$$

$$\psi_h^* = 1 - \Phi(z_{1-\alpha} - \sigma_0^{-1}h_\beta) \tag{I.2.15}$$

Consequently, Theorem 2 implies that, (I.2.15) provides the asymptotic power envelope of any asymptotically size-α test, ϕ_T. The asymptotic power envelope (I.2.15) is called sharp if it is possible to derive a feasible optimal test sequence which asymptotically hits (I.2.15). It can be shown that the feasible sequence can be constructed if a certain regular estimator sequence is available. Note that the inverse of I can be written in the form

$$I^{-1} = \frac{1}{\sigma_0^2} \begin{pmatrix} 1 & -I'_{\beta\gamma} I_\gamma^{-1} \\ -I_\gamma^{-1} I_{\beta\gamma} & I_\eta^{-1} + I_\gamma^{-1} I_{\beta\gamma} I'_{\beta\gamma} I_\gamma^{-1} \end{pmatrix} \tag{I.2.16}$$

where

$$\sigma_0^2 = I_\beta - I_{\beta\gamma} I_\gamma^{-1} I_{\beta\gamma} \tag{I.2.17}$$

The limit test function $\iota_1' X$ is accordingly rewritten as

$$\iota_1' X = \iota_1' I^{-1} S = \frac{1}{\sigma_0^2} \left[S_\beta - I'_{\beta\gamma} I_\gamma^{-1} S_\eta \right] \equiv \frac{1}{\sigma_0^2} S^* \tag{I.2.18}$$

Since $S_T(\theta_0)$ weakly converges to S both under H_0 and H_1, a sequence of tests based on the finite sample counterpart of S^*, denoted as $S_T^*(\theta_0)(= S_{\beta,T}(\theta_0) - I'_{\beta\gamma} I_\gamma^{-1} S_{\eta,T}(\theta_0))$, attains the asymptotic power envelope. The statistic S_T^* is called effective score and the variance of S_T^* is σ_0^2 which is called the effective information.

The regularity condition for the existence of a feasible optimal test is consequently the condition that enables S_T^* to be asymptotically invariant even if η_0 is replaced by a $\sqrt{T}$-consistent estimator, $\hat{\eta}$. I assume that under H_0 the sequence of the score function satisfy the linear stochastic expansions

$$\begin{aligned} S_{\beta,T}(\eta_0 + \frac{1}{\sqrt{T}} h_\eta) &= S_{\beta,T}(\eta_0) - I'_{\beta\eta} h_\eta + o_{p_0}(1) \\ S_{\eta,T}(\eta_0 + \frac{1}{\sqrt{T}} h_\eta) &= S_{\eta,T}(\eta_0) - I_\eta h_\eta + o_{p_0}(1) \end{aligned} \tag{I.2.19}$$

uniformly for bounded h_η. Consequently, for $\sqrt{T}$-consistent estimator $\hat{\eta} = \eta_0 + \frac{1}{\sqrt{T}}\xi$,

$$\begin{aligned} S_{\beta,T}(\hat{\eta}) &= S_{\beta,T}(\eta_0) - I'_{\beta\eta}\xi + o_{p_0}(1) \\ S_{\eta,T}(\hat{\eta}) &= S_{\eta,T}(\eta_0) - I_\eta\xi + o_{p_0}(1) \end{aligned} \tag{I.2.20}$$

Thus, assuming further that I is continuous, plugging (I.2.20) into the effective score function gives

$$S_T^*(\hat{\eta}) = S_T^*(\eta_0) + o_p(1) \tag{I.2.21}$$

under H_0. And the contiguity implies that (I.2.21) holds even under H_1. Consequently, the test based on $S_T^*(\hat{\eta})$ is asymptotically optimal.

It is interesting to study the influence of the unknown nuisance parameter by comparing the asymptotic power (I.2.15) with the situation in which the nuisance parameters are known. If η are fixed and known, the best limiting power of an asymptotically size-α test is obtained in (I.2.9). Comparing this with (I.2.15), we see that

$$I_\beta^{-1} \geq \sigma_{\theta_0}^2 \equiv I_\beta^{-1} - I_{\beta\eta}I_\eta^{-1}I'_{\beta\eta}$$

which implies that the limiting power under known nuisance parameter dominates that under unknown η. Equality holds when $I_{\beta\eta}$ is a zero vector. since $I_{\beta\eta}$ is the covariance between the score of the parameter of interest,$S_{\beta,T}$ and the score of the nuisance parameter, $S_{\eta,T}$, The zero vector condition implies that there is no loss of asymptotic power if $S_{\beta,T}$ and $S_{\eta,T}$ are orthogonal. Note that this condition is equivalent to Stein (1956)'s necessary condition for the adaptive estimation. A test is adaptive if it has the same asymptotic properties as the one obtained under the assumption that the true distribution is known. Under smooth alternatives with QMD likelihood, parametric models with unknown nuisance parameters provide insight on the optimality in semiparametric models.

The asymptotic optimal tests can be derived from a different concept using the *least favorable parametric submodel.* The following is the intuition: First consider an arbitrary alternative of nuisance parameters h_η. The power envelope of the test under the alternative will be greater than that of any asymptotically size-similar test without the restriction because the information for statistical inference decreases if one enlarges the model, i.e. for any feasible test function $\hat{\phi}$ without knowledge of η_0 and h_η^0,

$$E_{\beta_1}[\hat{\phi}] \leq Sup E_{\beta_1}[\phi(\eta_0, h_\eta^0)] \tag{I.2.22}$$

Since this argument holds for all types of h_ηs, the infimum of the power envelopes over the class of all h_ηs provides an upper bound of the power envelope of the test under unknown nuisance parameters. Geometrically, we get the lower bound by projecting the score function of β onto orthogonal complement of the linear subspace generated by all possible score functions for the nuisance parameter.

In order to demonstrate the idea more precisely, I first assume that η_0 is known and consider an asymptotically optimal test. Applying the Neyman-Pearson lemma to (I.2.10), we find an optimal test of asymptotic level α to be of the form $\tilde{\phi}_T = 1$ if

$$\widetilde{LR}_T \equiv log\left(\frac{L_T(\theta_1)}{L_T(\theta_0)}\right) = h' S_T - \frac{1}{2} h' I h > c_T \tag{I.2.23}$$

and $\tilde{\phi}_T = 0$ otherwise. The asymptotic distribution of $\widetilde{LR}_T$ is $N(-\frac{1}{2}h'Ih, h'Ih)$ under H_0 and $N(\frac{1}{2}h'Ih, h'Ih)$ under H_1. Consequently, the power of the test depends on $h'Ih$. The purpose of this section is to derive the asymptotic power envelope under unknown η_0. Let ψ_T is the power function of a test under unknown η_0. Since $\widetilde{\phi}_T$ is optimal for any known η_0, the following inequality holds,

$$\limsup \psi_T \leq \inf_{h_\eta} \left[\liminf E_h \widetilde{\phi}_T\right] \tag{I.2.24}$$

which implies that the infimum provides the asymptotic power envelope. The asymptotic alternative distribution of $\widetilde{LR}_T$ implies that the infimum is obtained when $h'Ih$ is minimized. Simple algebra shows that $h'Ih$ is minimized in h_η when

$$h_\eta = -I_\eta^{-1} I_{\eta\beta} h_\beta \tag{I.2.25}$$

which we call the least favorable direction. Note that the alternative space lies on the orthnormal complement of the null space under the least favorable direction, under which the null is most difficult to distinguish from the alternative. By plugging (I.2.25) into $\widetilde{LR}_T$ yields

$$\widetilde{LR}_T^* = h_\beta S_T^* - \frac{1}{2} h_\beta I^* h_\beta \tag{I.2.26}$$

where S_T^* and I^* are the effective score and the effective information defined in (I.2.18) and (I.2.17). Consequently, the test based on the effective score S_T^* is asymptotically optimal. But it is the same as asymptotically optimal test derived using the classic method, which implies that the method of the least favorable parametric submodel delivers the same optimality result.

This approach has advantages in the sense that it can be generalized to the testing problem in semiparametric model (Choi et al. (1996)), in which the semiparametric power envelope is defined to be the infimum of the power envelope associated with smooth parametric submodels. It also can be extended to non-standard testing problem in which LAN does not hold. In many important situations, the quadratic term (counterpart of the fisher information I) stays random even in the limit. Such a case is called *locally asymptotically quadratic (LAQ)* if the alternative distribution is contiguous. It is well known that the least favorable direction methods holds under some LAQ circumstances if the score is asymptotically normal and independent of the quadratic term (LAMN).

Even in more general LAQ, this method is still useful to obtain asymptotic optimality in the sense that (I.2.26) provides the asymptotic power envelope in locally asymptotically invariant tests. Let $\widetilde{LR}$ and $\widetilde{LR}^*$ be the asymptotic counterpart of $\widetilde{LR}_T$ and $\widetilde{LR}^*_T$, respectively. A sequence of a test ψ_T is said to be locally asymptotically $\alpha-$invariant if a test function $\psi(S, I)$ satisfies

$$\psi(S,I) = E[\psi(S,I)|S^*, I] \tag{I.2.27}$$

If S_η is conditionally independent of I, then it is shown that $E[\exp(\widetilde{LR})|S^*, I] = \exp(\widetilde{LR}^*)$.(See Jansson (2006) for details.) Combining this with (I.2.27), the following equality holds.

$$E\left[\psi(S,I)\exp[\widetilde{LR}]\right] = E\left[\psi(S,I)\exp[\widetilde{LR}^*]\right] \tag{I.2.28}$$

which implies that LR^* would give the asymptotic power envelope among all location invariant test.

Next section shows that the parameter instability tests are also non-standard and LAN does not applied to them because the asymptotic normality of the score function does not hold under the local alternative hypothesis. It invalidates the direct use of the least favorable direction method because the zero correlation of the efficient scores S_T^* with S_T^η, which is equivalent to the orthogonality of the scores, indicates the independency only under the normality condition. Instead, it is shown that they implicitly use the least favorable direction based on the contiguity property only, which corresponds to LeCam's 1st lemma, as follows: First, they impose the direction $I_{\beta\eta}I_\eta^{-1}h_\beta$ to the local variation of the nuisance parameter h_η which is equivalent to the least favorable direction in standard LAN. Then they derive a feasible test function under unknown nuisance parameters. Finally they show that the feasible test function is asymptotically equivalent to the likelihood ratio with the suggested direction. The contiguity justifies the asymptotic equivalency under the alternative hypothesis too, which proves the asymptotic optimality.

The asymptotic optimal test in LFPS coincides with various methods dealing with unknown nuisance parameters. One example is the method of profile likelihood. Profile likelihood methods are frequently used for eliminating nuisance parameters and for making statistical inference on parameters of interest. Murphy and der Vaart (2000) demonstrate that under local perturbation of the unknown nuisance parameters if the least favorable direction exists, the profile likelihood behaves very much like the ordinary likelihood and correctly selects a least favorable direction. The profile likelihood for β reduces the number of independent parameters to the dimension of β by imposing the nuisance parameter η as a function of β, which maximizes the likelihood function, i.e.

$$pl_T(\beta) \equiv \sup_{\eta} f_T(\beta, \eta)$$

Using the profile likelihood function, we can construct a function analogous to the original likelihood ratio, called profile likelihood ratio, which is defined as

$$PLR_T \equiv \frac{pl_T(\beta_1)}{pl_T(\beta_0)} = \frac{\sup_\eta f_T(\beta_1, \eta)}{\sup_\eta f_T(\beta_0, \eta)} \tag{I.2.29}$$

Note that the profile likelihood ratio is not the true likelihood function, and profiling generally breaks down the iid structure which is one of the main conditions for the existence of the local quadratic approximation. However, Murphy and der Vaart (2000) show that, as long as the least favorable maps exist in the model, one can keep the iid structure so that the local asymptotic quadratic approximation is possible. They show that for any sequence $\beta_t \to \beta_0$,

$$\begin{aligned} &\ln pl_T(\beta_T) - \ln pl_T(\beta_0) = \\ &\quad (\beta_T - \beta_0) S_T^* - \frac{1}{2}(\beta_T - \beta_0)' I^* (\beta_T - \beta_0) + o_{p_{\beta_0,\eta_0}}(\sqrt{T}\|\beta_T - \beta_0\| + 1)^2 \end{aligned} \tag{I.2.30}$$

where $\| \, \|$ is the norm. An interesting finding is that if we replace β_T by the local alternative $\beta_0 + \frac{1}{\sqrt{T}} h_\beta$, then (I.2.31) is asymptotically equivalent to the LAN of the original likelihood ratio in the least favorable parametric submodel. Consequently. we could obtain the result that the test based on the profile likelihood ratio is asymptotically equivalent to the optimal test in LFPS.

I.3 Review of The Optimal Tests for Parameter Instability

This section examines the asymptotic optimality of the existing popular tests for parameter instability. Three tests are considered; Andrews and Ploberger (1994), Elliott and Müller (2006), and Nyblom (1989). This section shows that the first two tests coincides with the optimality considered in the previous section, in the sense that they are the weighted average of the asymptotically point optimal tests in the least favorable parametric submodel. Nyblom (1989)'s test is different from the other two because it maximize the slope of the power function at the boundary of the stable and unstable parameter space. But it can be shown that it has a room to coincide with the other two.

I.3.A Andrews and Ploberger (1994)

Andrews and Ploberger (1994) consider a class of optimal tests for structural break in the sense that they provide a greatest asymptotic weighted average power result. Consider a single structural break case in which the parametric model is indexed by (β_t, η, π) for $t = 1, \ldots, T$. $\eta = (\beta_0, \gamma)$ is a $(k+q) \times 1$ vector of nuisance parameters that are constant for all t. π is the nuisance parameter that represents the time of a break as a portion of the sample size, and β_t is a $k \times 1$ vector of parameters that have a break at time $s = \pi T$. The sample observations is given by $\{Z_t\} = \{(y_t, X_t)\}$ where y_t is endogenous variable and X_t are weakly exogenous variables. The hypothesis to be tested is

$$
\begin{aligned}
H_0 &: \beta_t = 0 \qquad \forall t \leq T \\
H_1 &: \beta_t \neq 0 \qquad \pi T < t \leq T
\end{aligned}
\tag{I.3.1}
$$

Their test is based on the likelihood ratio, but is built as a weighted function of the standard LR tests for all permissable fixed break dates. This configuration is driven by defining the alternative likelihood as the weighted average with respect to the parameters that are identified only under the alternative. Two types of weights are involved. The first applies to parameter π, denoted as $dJ(\pi)$, which represents the possible break dates. The other is related to how far the alternative value is from the null hypothesis within an asymptotic framework that treats alternative values as local to the null hypothesis. If we define the local alternative process as $(\beta, \eta) = \theta = \theta_0 + (1/\sqrt{T})h$ where $\theta_0 = (0, \beta_0, \gamma_0)$, then the weight function is assigned with respect to h, denoted as $DQ_\pi(h)$. Consequently, they consider the likelihood function under the null and the alternative hypothesis as

$$
\text{Under } H_0 : L_T^0 = \prod_{t=1}^{T} f(y_t|\theta_0) f_X(X_i) \tag{I.3.2}
$$

$$
\text{Under } H_1 : L_T^1 = \int \prod_{t=1}^{T} f(y_t|\theta_0 + \frac{1}{\sqrt{T}} h) f_X(X_t) dQ_\pi(h) dJ(\pi) \tag{I.3.3}
$$

A possible reinterpretation of the weight function is that they are the probability measure of the random process $\{\beta_t\}$ which leads to a particular random parameter model with probability measure $dQ_\pi(h)dJ(\pi)$. The likelihood ratio is now defined as

$$LR_T^{AP} = \int \prod_{t=1}^{T} \frac{f(y_t|\theta_0 + \frac{1}{\sqrt{T}}h)}{f(y_t|\theta_0)} dQ_\pi(h)dJ(\pi) \tag{I.3.4}$$

Under the assumption that the likelihood function is twice differentiable, they derive a local quadratic approximation of the integrand of (I.3.4) under H_0, so that the likelihood ratio would be asymptotically equivalent to

$$\widetilde{LR}_T^{AP} = \int \exp\left[h'S_T + h'Ih\right] dQ_\pi(h)dJ(\pi) \tag{I.3.5}$$

where S_T is a vector of the score functions, and I is nonrandom positive definite Fisher information matrix. Note that, given h and π, the integrand in (I.3.4) is reduced to the standard likelihood ratio against the local alternative $\theta = \theta_0 + (1/\sqrt{T})h$. and it can be shown without difficulty that the alternative distribution given h and π is contiguous. It implies that the integrand in (I.3.5) is considered as the exponential of LAN given h and π. Consequently, the likelihood function (I.3.5) is interpreted as a weighted average of LAN with respect to h and π. Note, however, that the local approximation is not exactly the same as the standard concept of LAN because the score function is not asymptotically normal under H_1. The asymptotic normality under the alternative comes when both S_T and LR_T are asymptotically normal under H_0. But since LR_T^{AP} is not asymptotically normal, the asymptotic normality of S_T does not hold in this setup. It implies that we cannot use LeCam's third lemma to obtain an asymptotically optimal test. Instead, they implicitly show that the least favorable direction idea still hold by using only the contiguity property.

In order to get around the problem of unknown nuisance parameter η, Andrews and Ploberger (1994) choose a particular weight function $Q_\pi(h)$. Let V denote the linear subspace of $\mathbb{R}^{k+p}$ defined by

$$V = \{\theta \in \mathbb{R}^{k+p} | \theta = (0', \eta')' \text{for some } \eta \in \mathbb{R}^q\}$$

Then, they consider a weight function $Q_\pi(\cdot)$ on $\mathbb{R}^{k+p}$ that concentrates on the orthogonal complement of V with respect to the inner product $h'IS_T$. The orthogonal complement they consider is now defined as,

$$V^\perp = \left\{ h \in \mathbb{R}^{k+p} | h = \begin{pmatrix} \lambda \\ -I_\eta^{-1} I'_{\beta\eta} \lambda \end{pmatrix} \text{ for some } \lambda \in \mathbb{R}^k \right\} \quad \text{(I.3.6)}$$

But plugging h in (I.3.6) into the likelihood ratio (I.3.5) gives the integrand as a function of the efficient score and the efficient information matrix as .

$$\widetilde{LR}_T^{AP*} = \int \exp\left[\lambda' S_T^* + \lambda I^* \lambda\right] dQ_\pi^\lambda(\lambda) dJ(\pi) \quad \text{(I.3.7)}$$

They show that replacing the integrand in $\widetilde{LR}_T^{AP*}$ by a suggested feasible test function gives a asymptotically equivalent test function. Since any feasible test function satisfies inequality (I.2.22), the suggested feasible test function in the integrand is regarded as the asymptotically optimal test in the least favorable parametric submodel. Consequently, under the assumption that the error term is iid, Andrews and Ploberger (1994)'s test is implicitly interpreted as the weighted average of the asymptotically optimal test in the least favorable parametric submodel. It is also possible that the test function coincides with profile likelihood ratio as in the standard testing problem in the sense that a possible suggested test function in the integral is shown by them to be the profile likelihood ratio.

I.3.B Elliott and Müller (2006)

Elliott and Müller (2006) study asymptotically point optimal invariant tests for instability in coefficients in linear regression models. They consider a variety of unstable parameter processes that could possibly occur in the economy, and provide conditions under which the unstable processes lead to asymptotically equivalent optimal tests. The set of unstable processes they consider includes not only traditional structural breaks but also the case when β is an unstable random variable. They also suggest a feasible test function that is asymptotically point optimal. The linear regression model they consider is

$$y_t = X_t'(\beta_0 + \beta_t) + Z_t'\delta + \epsilon_t \quad t = 1, \ldots, T \tag{I.3.8}$$

where ϵ_t is iid normal. The null hypothesis is that β_t is a zero vector for all $t = 1, \ldots, T$. Under the alternative, $\{\beta_t\}$ are any unstable random vectors which are applicable to heterogeneous mixing functional central limit theorem. (Theorem 7.30 in White (2001)) It implies that the unstable process $\{\beta_t\}$ is asymptotically well approximated by a $k \times 1$ Wiener process. They show that this condition is weak enough to cover a wide range of unstable processes including structural breaks and time varying parameters.

In addition to this, they give a normalization condition that the average value of the random parameter path is always the same as that under the stable model. It implies that the average of the unstable parameter is normalized to zero, i.e. $\frac{1}{T}\sum \beta_t = 0$. This normalization ensures that the likelihood ratio statistic efficiently detects the variation in the coefficient, rather than differences between the average value of the parameter. In this set up, the likelihood functions under H_0 and H_1 are defined as

$$
\begin{aligned}
\text{Under } H_0\text{: } L_T^0 &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{1}{2\sigma^2}\sum_{t=1}^{T} e_t^2\right] \\
\text{Under } H_1\text{: } L_T^1 &= \int \frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{1}{2\sigma^2}\sum_{t=1}^{T}(e_t - X_t'\beta_t)^2\right] d\nu_\beta \qquad \text{(I.3.9)}
\end{aligned}
$$

where $e_t = y_t - X_t'\beta_0 - Z_t'\gamma$ and ν_β is the measure of $\{\beta_t\}$. The likelihood ratio is now defined as

$$
\begin{aligned}
LR_T^{EM} &= \int \frac{\exp\left[\frac{1}{2\sigma^2}\sum_{t=1}^{T}(e_t - X_t'\beta_t)^2\right]}{\exp\left[\frac{1}{2\sigma^2}\sum_{t=1}^{T} e_t^2\right]} d\nu_\beta \\
&= \int \exp\left[\frac{1}{\sigma^2}\sum_{t=1}^{T} e_t X_t'\beta_t - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\beta_t'(X_t e_t)(X_t e_t)'\beta_t\right] d\nu_\beta \qquad \text{(I.3.10)}
\end{aligned}
$$

Note that the integrand in (I.3.10) is equivalent to the likelihood ratio function for fixed $\{\beta_t\}$, leading to the interpretation that (I.3.10) is also regarded as the weighted average of the standard LR, as is the case of Andrews and Ploberger (1994). With this insight, one important determinant of optimal tests of parameter instability is the choice of weight function $d\nu_\beta$, or equivalently, the probability distribution it posits of $\{\beta_t\}$. Different weight functions lead to different unstable parameter processes. Therefore, the choice of weight function implies the choice of specific unstable process. An innovative finding of Elliott and Müller (2006) is that the choice of weight function, or equivalently, the unstable process is asymptotically irrelevant if $\{\beta_t\}$ is asymptotically characterized by a vector of Wiener process.

In order to deal with unknown nuisance parameters (β_0, δ), they focus on deriving the optimal test that is invariant to transformation of the form

$$(y, (X,\ Z)) \ \rightarrow \ (y + X\bar{b} + Z\bar{d},\ (X,\ Z))$$

Using the fact that any invariant test can be written as a function of a maximal invariant group of transformation, they suggest a maximal invariant transformation defined as $(M_Q y,\ (X, Z))$ where $M_Q = I_T - Q\left[QQ'\right]^{-1}Q'$, and $Q = (X, Z)$. Note that M_Q is equivalent to the annihilator in OLS estimation, which indicates that $M_Q y$ is identical to the OLS residuals. Let e_t^* be t^{th} component of $M_Q y$. Then under the maximal invariant transformation, Elliott and Müller (2006) show that the likelihood ratio is asymptotically equivalent to

$$\widetilde{LR}_T^{EM} = \int \exp\left[\frac{1}{\sigma^2}\sum_{t=1}^{T}\hat{\beta}_t' X_t e_t^* - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\hat{\beta}_t'\Sigma_X\hat{\beta}_t\right] d\nu_\beta \tag{I.3.11}$$

where $\hat{\beta}_t = \beta_t - \frac{1}{T}\sum\beta_i$, and $\Sigma_X = E[X_t X_t']$. The likelihood ratio $\widetilde{LR}_T^{EM}$ is interpreted as the weighted average of the likelihood ratio in the least favorable parametric submodel where the probability measure ν_β is assigned as the weight function. Let's define the local alternative $\theta_t = \theta_0 + \frac{1}{\sqrt{T}}h_t$ where $\theta_t = (\beta_t',\ \eta_1')'$, $\theta_0 = (0',\ \eta_0')'$, $h_t = (\sqrt{T}\beta_t',\ h_\eta')'$, and $\eta = (\beta_0',\ \delta')$. For fixed $\{\beta_t\}$, chapter 2 demonstrates that LAQ can be applied to the likelihood ratio so that the local approximation is

$$\tilde{L}_T = \exp\left[\frac{1}{\sqrt{T}}\sum\dot{\ell}_t' h_t - \frac{1}{2T}h_t' I h_t\right] \tag{I.3.12}$$

Note that the first derivative and the Fisher information in a linear Gaussian model are

$$\dot{\ell}_t = (\dot{\ell}_t^{\beta'}, \dot{\ell}_t^{\eta'})', \quad \dot{\ell}_t^{\beta} = \frac{1}{\sigma^2} X_t e_t, \quad \dot{\ell}_t^{\eta} = \begin{pmatrix} \dot{\ell}_t^{\beta_0} \\ \dot{\ell}_t^{\delta} \end{pmatrix} = \frac{1}{\sigma^2} \begin{pmatrix} X_t e_t \\ Z_t e_t \end{pmatrix}$$

$$I = \begin{pmatrix} I_\beta & I_{\beta\eta} \\ I'_{\beta\eta} & I_\eta \end{pmatrix} \quad I_{\beta\eta} = \frac{1}{\sigma^2}[\Sigma_{XZ}] \quad I_\eta = \begin{pmatrix} \Sigma_X & \Sigma_{XZ} \\ \Sigma'_{XZ} & \Sigma_Z \end{pmatrix} \tag{I.3.13}$$

where $\Sigma_{XZ} = E[X_t Z_t]$, and $\Sigma_Z = E[Z_t Z_t]$. Simple calculation shows that applying the candidate for the the least favorable direction $h_t = -I_\eta^{-1} I_{\eta\beta} \beta_t$ gives

$$\frac{1}{\sqrt{T}}\theta_t = (\beta'_t, -\frac{1}{T}\sum \beta'_i,\ 0') \tag{I.3.14}$$

Plugging (I.3.14), and (I.3.13) into (I.3.10) gives that the likelihood ratio is equivalent to the integrand in (I.3.11) except e_t^* is replaced by e_t. Since OLS estimator is the regular maximum likelihood estimator under Gaussian error distribution, we can apply the result in (I.2.21), which implies that replacing e_t^* by e_t provides equivalent asymptotic properties. The likelihood ratio with $\{e_t\}$ and $\{e_t^*\}$ are asymptotically equivalent. Consequently, the integrand of Elliott and Müller (2006)'s likelihood, (I.3.12), is asymptotically equivalent to the likelihood ratio in the least favorable parametric submodel under given $\{\beta_t\}$, which leads to the suggested interpretation.

I.3.C Nyblom (1989)

Nyblom (1989) proposes an asymptotically locally most powerful test against the unstable parameter process that is a martingale, i.e. $E[\beta_t|\Im_{t-1}] = \beta_{t-1}$, where $\Im_t$ is the information set up to time t. Like Elliott and Müller (2006), he shows that the martingale processes cover both structural breaks and time varying parameter processes. A central difference of his test from Andrews and Ploberger (1994) and Elliott and Müller (2006) is that it maximizes the slope of the power function at $\beta_t = 0$. His model does not require the contiguity of the alternative process, nor LAQ of the likelihood ratio. Instead, he focuses on the case in which $\{\beta_t\}$ is very close to zero so that the higher order term than the second one in the Taylor expansion of the likelihood ratio is asymptotically negligible, leading to the likelihood ratio equivalent to

$$\tilde{LR}_T^N = \int \exp\left[\sum \dot{\ell}_t' \beta_t - \frac{1}{2}\beta_t' I_t \beta_t\right] d\nu_\beta \tag{I.3.15}$$

where $I_t = \frac{\partial^2 log f_t(\cdot)}{\partial\beta_t \partial\beta_t'}$ and ν_β is the measure of $\{\beta_t\}$. His test is locally most powerful if the error term is iid so that I_t is nonrandom. One problem of this method is that the optimal property does not hold if the model introduces unknown nuisance parameters. But this condition delivers the integrand in (I.3.15) coincided with LAQ as those in (I.3.5) and (I.3.12). Consequently, (I.3.15) has the possibility to be alternatively interpreted to be the weighted average of the optimal test associated with the likelihood ratios of Andrews and Ploberger (1994) and Elliott and Müller (2006) under suitable conditions that bring forth the local quadratic approximation and contiguity.

I.4 Concluding Remark

This chapter reviews the concept of asymptotic optimality in testing parameter instability of time series models. I first study the standard optimal testing method in the presence of unknown nuisance parameters under which the optimal test is interpreted as the most powerful test in the least favorable parametric submodel. The standard method of optimality cannot be directly applied to the parameter instability test because the parameter instability test has nonstandard properties such as the existence of the unknown nuisance parameter that presents only under the alternative hypothesis. This chapter, however, shows that the current optimal tests coincides with the standard optimal test in the sense that they are interpreted as the weighted average of the standard optimal tests in the least favorable parametric submodel.

Chapter II

Optimal Tests for Parameter Instability in General Time Series Models

It is difficult to select the appropriate test for parameter instability in empirical work because there are a large number of tests designed for different possible unstable processes. Elliott and Müller (2006) resolve this problem by providing conditions under which a large class of breaking processes lead to asymptotically equivalent optimal tests. Their finding, however, is restricted to linear conditional mean equations with normal error distributions. I improve upon their work in two ways. First, I show that the asymptotic equivalency of the efficient tests for parameter instabilities holds even in a broader set of parametric models which includes nonlinear models with non-Gaussian error distribution. It implies that the knowledge of the unstable parameter process is asymptotically irrelevant for testing purposes. Second, I suggest a test statistic that is asymptotically optimal for a broad set of unstable parameter processes which allows for both structural breaks and time varying parameters. Monte Carlo studies show that the suggested test has better small sample powers against various breaking processes, compared to the existing optimal tests.

II.1 Introduction

Structural instability is a common problem in macroeconomic and finance models dealing with time series data. A change in economic policy or market conditions may cause the adjustment of the behavior of economic agents, thereby changing the economic relationship. As a result, parameter instability has always been an important concern in econometric modeling and much effort has been devoted to obtaining convenient and powerful tests for parameter instability.(See the review papers by Perron (2006).) Several distinctive features, However, cause difficulty in testing the parameter instability. First and foremost is the problem that there are many possible ways for an instable parameter to occur. Single break, multiple breaks and time varying parameter processes are examples of such ways. The alternative processes are usually not recognized from data, which presents the problem of how to specify the alternative process. Another difficulty is that there exist nuisance parameters that are not identified under the null hypothesis. In structural break models, the parameter representing unknown break point appears only under the alternatives. In time varying parameter models, any parameter that determines the shape of the parameter distribution is not identifiable under the null of stable parameters. This property violates the regularity conditions for the optimality of the classical Likelihood Ratio, Wald, and Lagrangian Multiplier tests. Furthermore, local asymptotic normality of the tests is inapplicable in general. Therefore, a novel test with new asymptotic distribution is required to obtain the optimality. However, the asymptotic null distribution of the test generally depends on the unidentified nuisance parameters, which implies that the tests relies on the specific instable parameter process that the researcher has in mind.

For these reasons, research has restricted its attention on the specific types of breaking processes. Franzini and Harvey (1983), and Shively (1988) consider models where the parameter is subject to Gaussian breaks of constant variance every period. Nyblom and Makeläinen (1983) consider the random walk parameter processes. Andrews (1993), Bai (1996), and Vogelsang (2005) construct tests for one-time structural change under various circumstances. Andrews and Ploberger (1994), Bai and Perron (1998), and Lavielle and Moulines (2000) generalize to multiple structural breaks cases.

Attempts to cover a wide range of parameter instabilities are done by Nyblom (1989) and Elliott and Müller (2006). Nyblom (1989)'s test is locally most powerful only when the initial point of the parameter is known, which is generally infeasible in economic applications. Elliott and Müller (2006) show that, in a linear model with Gaussian error distribution, any optimal test for specific unstable process has the same asymptotic power against any other unstable processes as long as it is in the broad set they define. However, the linear conditional mean model is sometimes too simple for economic applications. Furthermore, larger movements in economic time series seem to occur more often than would be implied by normality. Therefore, inference designed for a larger variety of economic models and distributions should be relevant for applied econometric work.

This chapter makes two contributions. First, I extend Elliott and Müller (2006)'s finding to general parametric models. This implies that, in a broad set of parametric models, any optimal test for parameter instability is asymptotically equivalent, as long as the instability is in a very general set. The set of instable parameter processes allows for multiple structural breaks, regularly occurring breaks, smooth adjustment of the model to an economic shock, and time varying parameters. The set of parametric models are wide enough to include nonlinear models and quantile models. A wide range of densities are applicable by relaxing the twice differentiability of likelihood functions. Hence I show that the precise form of the instable parameter process is unnecessary for testing purpose in the asymptotic sense.

Second, I suggest a convenient test statistic for this set of instable parameter processes. The test statistic is asymptotically equivalent to the likelihood ratio test statistic under both the null and the alternative hypotheses. This implies that the test statistic is asymptotically optimal. The test statistic is easy to compute because it requires only the maximum likelihood estimation under the null hypothesis. I also calculate small sample powers of the test against various types of the instable parameter processes. The test has better small sample powers than the existing optimal tests for almost all alternative processes. The test is used to investigate the quantile parameter stability in the U.S. inflation model.

This chapter is organized as follows: Section 2 points out some distinctive properties of the testing problem. Section 3 derives an asymptotically optimal test statistic and suggests an implication. Section 4 provides some examples for the optimal test statistic in economic applications. And Section 5 concludes

II.2 Distinctive Features Parameter Instability Tests

This section considers the distinctive features of parameter stability tests and reviews the current test methods. I consider a parametric model indexed by

(β_t, γ) for $t = 1, \ldots, T$. γ is a $(k+q) \times 1$ vector of constant parameters, and β_t is a $k \times 1$ vector of parameters varying across time. The hypothesis to be tested is

$$\begin{aligned} H_0 &: \beta_t = 0 \qquad\qquad \forall t \\ H_1 &: \beta_t \neq 0 \quad \text{for some } t > 1 \end{aligned} \tag{II.2.1}$$

In case of tests of pure parameter instabilities, q is zero and γ is the parameter vector of the initial point of β_t. The hypothesis in (III.2.2) seems identical to that of the standard test problem. However, it has some distinctive features which make it difficult to use standard LM, Wald and LR test. One difficulty is that there exist a large variety of ways in which β_t is not stable. Any specific assumption on unstable β_t would lead to a different testing problem, and a test developed for one alternative β_t might not be useful in another specific β_t. Existing tests can be categorized into two big streams based on types of β_t processes: One is the test of structural breaks, and the other is the test of time varying parameters.

Structural break tests regard β_t as fixed and described by a vector of unknown parameters. Most tests of structural breaks focus on single break cases, in which

$$\begin{aligned} \beta_t &= 0 \qquad \text{for } t \leq \tau T \\ \beta_t &= \overline{\beta} \quad \text{for } \tau T < t \leq T \end{aligned} \tag{II.2.2}$$

where τ is the time of structural change as a fraction of the sample size. Traditional approaches such as Chow (1960), Zellner (1962), Goldfeld and Quandt (1978), and Rothenberg (1984) assume that the time of the structural change is known. In general, however, we do not know the true change point. Even if we know the cause of a structural break, we usually do not know the exact time of the occurrence. For example, one might want to test for structural change occurring sometime during the war period 1939-1949, but it is hard to identify in what exact year the break is suspect to occur. Furthermore, although we know the time of the structural breaks, sometimes change occurs only after a lag of unknown length, or before the event due to anticipation of the event.

In these circumstances, if no structural change occurs, the time of structural change is redundant. The nuisance parameter τ appears under the alternative hypothesis, but not under the null. This feature makes it difficult to use traditional tests such as LR, LM and Wald tests. Davies (1977) shows that the test does not fit into the standard regular testing framework. Consequently LM, Wald, and LR-like tests, constructed with τ treated as a parameter, do not possess their standard large sample asymptotic distributions such as χ^2 distribution.

Quandt (1958, 1960), Davies (1977, 1987) and Hawkins (2000) suggest the use of supremum of LR, LM and Wald tests ($\sup F$ tests) over all values of τ. This search over a set of dependent F-statistics affects the asymptotic distribution of the test, which ceases to be χ^2. Andrews (1993) finds the asymptotic distribution of the statistic in a very general setting, and shows that the tests have non-trivial asymptotic power. The $SupF$ tests are not optimal, however, except in a very restrictive sense. Andrews and Ploberger (1994) propose an optimal test by maximizing weighted average power of LM tests at given τ. The weighting function is designed so that each LM has constant weight on the same ellipses in the parameter space.

Tests for multiple structural breaks are, in principle, extensions of single break cases. Andrews and Ploberger (1994)'s test can be applied to multiple break cases with different weight functions with respect to break times. Bai and

Perron (1998) extend the alternative to multiple structural break processes and examine the maximum of the F-statistic over all combinations of (τ_1,...,τ_N). The computation in practice is, however, cumbersome because they have to calculate the statistic for all possible combinations of the break dates. For example, Bai and Perron (1998)'s test needs to compute $\begin{pmatrix} T \\ p \end{pmatrix}$ different F statistics for T observations and p possible breaks.

Another stream is testing time-varying parameters in which β_t is considered as random. Even in the time varying parameter approaches, there are many possible alternatives based on the assumptions of stationarity and distributions of β_t. Any specific assumption leads to a different testing problem. Engle and Watson (1985) consider stationary autoregressive processes in which β_t deviates only temporarily from zero. Nyblom and Makeläinen (1983) consider β_t as a random walk with a constant X_t. A similar difficulty with the case of structural breaks occurs, since parameters describing the distribution of β_t are identified only under the alternatives. This renders the typical intuition of optimality of general LR, LM, and Wald tests, and asymptotic normality not applicable.

The introduced test functions have good power properties under their specific alternative processes. The problem is to decide what testing method we should use for the specific economic model of interests. Data give little information about the choice of the test statistic. Economic theory also provides restricted guide as to what type of alternative process one would expect in practice. There are few ways, other than to depend on what one has in mind.

Methods to overcome this problem are suggested by Nyblom (1989), Eliasz et al. (2004), and Elliott and Müller (2006). Their methods are based on the idea that structural breaks processes and time varying parameter processes are not truly distinctive. For example, consider

$$\beta_t = \beta_{t-1} + \gamma_t \delta_t \quad t = 1, \ldots, T \tag{II.2.3}$$

(II.2.3) is reduced to random walk process if γ_t is constant for all t and δ_t is iid $N(0, I_k)$ where I_k is $k \times k$ identity matrix. (II.2.3), however, can be defined as a single structural break process. Let $(\gamma_1, \ldots, \gamma_T)$ be a multinomial vector with $Pr(\gamma_t = 1) = p_t$, $Pr(\gamma_t = 0) = 1 - p_t$, and $(\delta_1, \ldots, \delta_T)$ are iid and independent of $(\gamma_1, \ldots, \gamma_T)$. Then exactly one of γ_t is one with others are zero and $\{\beta_t\}$ has only one break at a random time in the sample period. This conformability allows them to construct optimal tests against both structural break and time varying parameters.

Nyblom (1989) and Eliasz et al. (2004) derive a locally most powerful test for the martingale alternative processes. These alternatives provide a wide range of breaking processes that include the finite time structural breaks as well as the random-walk processes. Nyblom (1989)'s test, however, retains optimality only when there are no unknown parameters, γ, under the null hypothesis, which rarely occurs in practice. Eliasz et al. (2004)'s optimality applies to linear Gaussian model only.

Elliott and Müller (2006) provide conditions under which optimal tests are asymptotically equivalent. By allowing for the dependency of the change of the breaks and heteroscedasticity, their condition covers many breaking processes, including relatively few breaks, clustered breaks, regularly occurring breaks, and smooth transitions to change cases. In other words, they find that the precise form of the breaking process β_t under the alternative is irrelevant for the asymptotic power of the tests, as long as they satisfy some conditions. This result simplifies the practice of testing against parameter instability because it allows applied researchers to leave the exact form of the alternative unspecified without forgoing asymptotic power.

The result obtained by Elliott and Müller (2006) is innovative, but, like Eliasz et al. (2004), is restricted to the case with linear conditional mean model and Gaussian errors, even though their suggested test statistic is valid under non-Gaussian distribution. Unfortunately, there are many economic circumstances where the linear Gaussian model is not satisfied. As for the mean linearity, modern macroeconomic theory emphasizes the interaction among representative agents who are, in general, assumed to behave according nonlinear decision rules that are obtained as optimal solutions to dynamic optimization problem. (See Barnett et al. (1992) and Barnett et al. (2000) for details.) Also increasing attention has been devoted to the characteristics other than the mean relationship such as conditional quantiles because they can provide a more informative view of the economic relationship.

As for the normal distribution, there are many evidences that distributions which explicitly accommodate skewness and excess kurtosis often have better explanatory power for variables in macroeconomics and finance. Many types of non Gaussian distribution are suggested in economic modeling. Theodossiou (1998) suggests a type of skewed generalized t distribution. Fernandez and Steel (1998), Komunjer (2006), and Theodossiou (2000), address the use of asymmetric exponential family of distribution. Consequently, relaxing the linear Gaussian assumption is crucial for wide applications.

II.3 An Asymptotically Optimal Test for Parameter Instability

In this section, I construct an asymptotically optimal test for a broad set of unstable processes. The parametric model considered in this section is nonlinear and non-Gaussian so that various types of economic models can be applied. The test uses the average magnitude of the break, which is described by the global covariance, as the only information for the breaking process. It suggests that, as long as the average sizes are equal, any optimal tests for particular unstable processes are asymptotically equivalent.

II.3.A The Model and the Test Statistic

The model I consider is parametric time series model that is suitable for maximum likelihood estimation, and is based on non trending observations, given as

$$y_t = g(X_t, \beta_0, \beta_t) + \epsilon_t \tag{II.3.1}$$

where $g(\cdot)$ is continuous and differentiable with respect to β_t. β_t is the $k \times 1$ vector of parameters to be tested and β_0 is the $k \times 1$ vector of nuisance parameters. To examine asymptotic local power, the alternative hypothesis is considered to be local to the null. Specifically, I assume that $\{\beta_t\}$ take the form.

$$\beta_0^1 = \beta_0 + \frac{1}{T}\delta_0, \qquad \beta_t = \frac{1}{T}\delta_t \;\; \forall t = 1, \ldots, T$$

I define θ_t as $2k \times 1$ vector of all parameters in the model at t, i.e. $\theta_t = \theta_0 + \frac{1}{T} d_t$ where $\theta_0 = (\beta_0', 0_k')'$, $d_t = (\delta_0', \delta_t')'$, and 0_k is the $k \times 1$ vector of zeros. Unlike the standard testing problem, the appropriate neighborhood in order for the test to have nontrivial asymptotic power is where β_t is of order T^{-1} in probability. The reason for this is that the test focuses on alternatives with a persistently varying $\{\delta_t\}$, in the sense that permanent change of the parameter has more implication in both economic and statistic concepts. It is implicit in the formulation that the (y_t, X_t), and θ_t may depend on T, but I suppress the dependency for the purpose of notational convenience. Specifically, I consider unstable processes that satisfy the following condition.

Condition 1

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

2. *E[$\Delta\delta_t$]=0 and $E[|\Delta\delta_{t,i}|^r] < K < \infty$ for all T,t,i*

3. *$\{\Delta\delta_t\}$ is globally covariance stationary with nonsingular long-run covariance matrix, Ω*

Condition 1 does not consider time varying parameter processes as the only type of unstable processes. It includes many kinds of structural breaks by choosing suitable distributional forms. For example, If we let $\Delta\delta_t$ have a continuous distribution with probability p and equal zero with probability $(1-p)$, then the process is expected to have $T \cdot p$ breaks.

Admitting both heteroscedasticity and dependency makes Condition 1 capture almost all persistent breaking processes. Heteroscedasticity of $\Delta\delta_t$ allows different types of breaks to occur in the sample period. Breaks caused by different shocks may not be homogeneous, and the size of the shocks may also be unequal.

Thus, it is more plausible to assume that the breaking processes are heteroscedastic. Heteroscedasticity also covers processes that have fewer breaks in certain periods and more breaks in other periods.

By allowing dependency of $\Delta\delta_t$, Condition 1 allows the parameter to smoothly adjust to a new level after a break. This covers the general set of breaking processes that occur frequently. For example, the oil price shock in 1973 might have had a lagged effect on productivity. Also there may be a movement from one regime to another with a transition period in between.

The sample of observations is given by $\{(y_t, X_t)\}$. The following condition specifies the model and the conditional likelihood function for (y_t, X_t) given $(y_1, \ldots, y_{t-1}, X_1, \ldots, X_{t-1})$.

Condition 2

1. *ϵ_t is independent of $(y_1, \ldots, y_{t-1}, X_1, \ldots, X_t)$ with conditional distribution $f(\epsilon_t|\theta_t)$. The distribution does not depend on d_t in the null hypothesis.*

2. *X_t has conditional distribution $f_X(X_t|\Im_{t-1})$ with respect to some sigma-finite measures, where $\Im_{t-1}$ is a σ-field generated by $(y_1, \ldots, y_{t-1}, X_1, \ldots, X_{t-1})$. $\{f_X(X_t|\Im_{t-1})\}$ does not depend on θ_t for all $t = 1, \ldots, T$.*

3. *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. $E\left[|X_{t,i}|^{r/2+\delta}\right] < \Delta\infty$ for some δ and all $t = 1, \ldots, T$ and $i = 1, \ldots, k$.*

Condition (2) implies that the likelihood function for the data is factored into two pieces, one of which captures the contribution to the distribution of y_t, $f(\epsilon_t|\theta_t)$, and depends on θ_t and the other which contains conditional distribution of X_t and does not depend on θ_t, $f_X(X_t|\Im_t)$. In such likelihood functions, $f_X(\cdot)$

need not be known in order for one to construct the test statistics considered here. Under Condition 2, the likelihood function when no break occurs is

$$f_0(y, X|\theta_0) = \prod_{t=1}^{T} f(\epsilon_t|\theta_0) f_X(X_t|\Im_{t-1}) \tag{II.3.2}$$

The likelihood function under the alternative hypothesis is

$$f_1(y, X|\theta_0, \ldots, \theta_T) = \int \prod_{t=1}^{T} f(\epsilon_t|\theta_0 + \frac{1}{T} d_t) f_X(X_t|\Im_{t-1}) d\nu_d \tag{II.3.3}$$

where $d\nu_d$ is the measure of $d = (d_1', \ldots, d_T')'$. If ν_d is known, the Neymann-Pearson Lemma implies that rejecting H_0 for large value of the likelihood ratio statistic, defined as

$$LR_T = \int \prod_{t=1}^{T} \frac{f(\epsilon_t|\theta_0 + \frac{1}{T} d_t)}{f(\epsilon_t|\theta_0)} d\nu_d \tag{II.3.4}$$

has the best power against the alternative distribution (II.3.3). Most optimal tests for parameter instability are manipulations of (III.3.3) that make the test feasible and easy to compute. Since the test statistic may depend on the distribution $f(\cdot)$ and the distributional properties of breaking processes, ν_d, the different types of optimal test statistics come from the choice of ν_d and $f(\cdot)$.

In principle, LR_T can be used if one specifies the error distribution and the parameter process. However, it has an integral in its form which makes the computation too complicated to be used in practice. The method proposed in this section to resolve the problem is to suggest another test statistic, B, that converges in probability to the same limit as that of (II.3.4) and is easy to compute.

In order to define the test statistic, we need some notations and definitions. Let $\dot{\ell} = (\dot{\ell}'_1, \ldots, \dot{\ell}'_T)'$ be the first derivative of the log of the likelihood $f(\epsilon_t|\cdot)$ with respect to δ_t, and $J_1 = E[\dot{\ell}_t \dot{\ell}'_t]$. Let $\Omega^* = J_1^{\frac{1}{2}} \Omega J_1^{\frac{1}{2}}$. 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 $\dot{\ell}^*(\hat{\beta}) = (I_T \otimes P' J_1^{-1/2}) \dot{\ell}(\hat{\beta})$ or $\dot{\ell}_t^*(\hat{\beta}) = P' J_1^{-1/2} \dot{\ell}_t(\hat{\beta})$. Furthermore, define $\dot{\ell}_{i,t}^*$ be the i^{th} element of $\dot{\ell}_t^*(\hat{\beta})$ 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

$$B(\Omega) = \sum_{i=1}^{k} \zeta_i'(\widehat{\beta}, \widehat{J}_1)' \left[\frac{a_i^2}{T^2} I_T - F M_e F' \right]^{-1} \zeta_i(\widehat{\beta}, \widehat{J}_1) \qquad \text{(II.3.5)}$$

where $M_e = I_T - \frac{1}{T} ee'$, $F = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 1 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \ldots & \ldots & 1 \end{pmatrix}$, and $\widehat{\beta}$ and $\widehat{J}_1$ are the maximum likelihood estimators under H_0. The test statistic, $B(\hat{\beta}, \widehat{J}_1, \Omega)$, does not have the integral so that the computation is tractable. Note that the test statistic depends on the distribution of the unstable parameter process only through the eigenvalues of the covariance matrix. Therefore, proving the optimality of $B(\cdot)$ will provide an evidence of my argument: asymptotic irrelevancy of the knowledge of the unstable process.

II.3.B Asymptotic Optimality of the Test Statistic

The purpose of this section is to present how the optimality of $B(\cdot)$ can be obtained. I suggest another test statistic $\widetilde{LR}_T$ which is an increasing transformation of $B(\cdot)$ and show that $\widetilde{LR}_T$ is asymptotically equivalent to LR_T. I first focus on the integrand of LR_T to suggest an asymptotically equivalent formula that makes LR_T equal $\widetilde{LR}_T$. Then I address that, under some conditions, the equivalence of the integrands is sufficient for the asymptotic equivalence of the two test statistics, both under the null and the alternative hypotheses.

I first simplify the integrand of LR_T. The integrand is considered as the likelihood ratio for specific alternative parameters, d. 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. Fortunately, these density functions have an alternative expansion under a single condition that only involves a first derivative, i.e. the square roots of density functions correspond to unit vectors in space of square integrable functions. The following condition gives the differentiability assumption and additional assumptions for the asymptotic properties of the score function.

Condition 3 *Let $\xi_t(\cdot, \theta_t)$ be the square root of the error density, $f(\epsilon_t|\theta_t)$. Under H_0,*

1. *There exists a $2k \times 1$ random vector $\dot{\xi}_t(\cdot, \theta_t) = \left(\dot{\xi}_t^{0\prime}, \dot{\xi}_t^{t\prime}\right)'$ such that $\mathbb{E}_\theta \|\dot{\xi}_t(\cdot, \theta)\|^2 < \infty$*

$$\mathbb{E}_{\theta}\left(\left[\left(\frac{\xi_t(\cdot,\theta_t+h)}{\xi_t(\cdot,\theta)}-1\right)-h'\frac{\dot{\xi}_t(\cdot,\theta_t)}{\xi_t(\cdot,\theta_t)}\right]^2\right)\rightarrow 0 \ \textit{as} \ \|h\|\rightarrow 0, \quad \forall t\leq T \tag{II.3.6}$$

2.

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

for some positive definite nonrandom $2k\times 2k$ matrix function $J(\theta)$ and for any $s\in[0,1]$ and $\hat{J}(1)$ is positive definite for all t

The derivative $\dot{\xi}(\cdot,\theta)$ is called Hellinger derivative. Part (1) of Condition 3, called *quadratic mean differentiability* (QMD), 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 well developed in standard testing problems (LeCam (1970)) and nonstandard problems (Jansson (2006) and Jeganathan (1995)). However, no work has considered the approximation under the nonstationary time varying parameter alternatives. The following lemma is the extension of the local approximation to the time varying parameter processes.

Lemma 2 *Let's define $\dot{\ell}_t^{\theta}=2\frac{\dot{\xi}_t}{\xi_t}$, and $J=4E[\frac{\dot{\xi}_t}{\xi_t}\frac{\dot{\xi}_t}{\xi_t}']$. Under Condition 1 to 3, the integrand of* (II.3.4)*, denoted as L_T, is equivalent to:*

$$L_T=(1+o_p(1))exp\left[\frac{1}{T}\sum_{t=1}^{T}d_t'\dot{\ell}_t^{\theta}-\frac{1}{2T^2}\sum_{t=1}^{T}d_t'Jd_t\right] \tag{II.3.7}$$

As a next step, I deal with the measure of d, ν_d in order to eliminate the integral in LR_T. The measure, ν_d, consists of two parts: One that represents the Condition 1 process of δ_t, and the other that corresponds to the measure of δ_0. The latter can be interpreted as the weighting function for the unknown nuisance parameter, δ_0. It implies that, (II.3.4) is set to have the best weighted average power for the chosen weight function. With regard to the choice for the weight function, I consider δ_0 that concentrates on the orthogonal complement of the linear subspace, V, generated by all possible score functions for the nuisance parameter. This type of selection provides the power envelope of the tests in which the nuisance parameters are unknown and the sizes are asymptotically similar for all δ_0. The following is the intuition: First consider an arbitrary alternative of nuisance parameters. The power envelope of the test under the alternative will be greater than that of any asymptotically size-similar test without the restriction because the information for statistical inference decreases if one enlarges the model. Since this argument holds for all types of alternative δ_0s, the infimum of the power envelopes over the class of all alternative δ_0s gives an upper bound of the power envelope of the test under unknown nuisance parameters. Geometrically, we get the lower bound by projecting the score function of δ_t onto V, which implies that δ_0 lies on the orthogonal complement of V. This logic corresponds to the concept of Pfanzagl and Wefeylmeyer (1978)'s *least favorable hypothesis* in which the perturbation of the initial parameter is defined around the alternative value, i.e. $\beta_0^0 = \beta_0^1 + \frac{1}{T}\delta_0$, and the power function, say $\bar{\phi}(\beta_t, \beta_0)$, is defined as

$$\bar{\phi}(\beta_t, \beta_0^1) = \inf_{\delta_0} \phi(\beta_t, \beta_0^1; \delta_0) \tag{II.3.8}$$

where $\phi(\beta_t, \beta_0^1; \delta_0)$ represents the power of any asymptotically size α tests for given β_0^1 and δ_0. The following condition specifies it.

Condition 4 *Let $\dot{\ell}_0$ and $\dot{\ell}$ be the score functions for δ_0 and δ, respectively.*

$$\delta_0 = -E[\dot{\ell}_0\dot{\ell}_0']^{-1}E[\dot{\ell}_0\dot{\ell}']\delta \tag{II.3.9}$$

or equivalently,

$$d = (I_{T\times k},\ \ -E[\dot{\ell}_0\dot{\ell}]E[\dot{\ell}_0\dot{\ell}_0']^{-1}])'\delta \tag{II.3.10}$$

Condition 4 can be simplified to $\delta_0 = -\frac{1}{T}\sum_{t=1}^{T}\delta_t$. Therefore, it has another interpretation that the average value of the unstable parameter path is always the same as the one under the model of stable parameter.

As for the measure of δ_t, I replace $\{\delta_t\}$ by another random sequence $\{\widetilde{\delta}_t\}$ and show that the integrand, L_T based on $\{\widetilde{\delta}_t\}$, weakly converges to the same limit as that of L_T based on any $\{\delta_t\}$. The random vector $\widetilde{\delta} = (\widetilde{\delta}_1', \ldots, \widetilde{\delta}_T')'$ is defined as

$$\widetilde{\delta} \sim N(0, FF' \otimes \Omega), \tag{II.3.11}$$

where F is as defined in (III.3.7). (II.3.11) means that $\{\widetilde{\delta}_t\}$ follows a multivariate random walk process, i.e. $\Delta\widetilde{\delta}_t \sim iid\ \ N(0,\Omega)$, which satisfies Condition 1. Therefore, in order to present the asymptotic equivalence of the integrands, it suffices to show that each term of the integrand, (II.3.7), converges to a well-defined random variable regardless of ν_d, as long as it satisfies Condition 1. The following lemma demonstrates the asymptotic equivalency.

Lemma 3 *Let's define $\widetilde{L}_T$ as*

$$\widetilde{L}_T = exp\left[\dot{\ell}'(M_e \otimes I_k)\widetilde{\beta} - \frac{1}{2}\widetilde{\beta}'(M_e \otimes J_1)\widetilde{\beta}\right] \tag{II.3.12}$$

where $\widetilde{\beta} = \frac{1}{T}\widetilde{\delta}$*. Under Condition 1 to 4,* $|\widetilde{L}_T - L_T|$ *weakly converges to zero under* H_0.

Now we have a new test function given as

$$\widetilde{LR}_T = \int exp\left[\dot{\ell}'(M_e \otimes I_k)\widetilde{\beta} - \frac{1}{2}\widetilde{\beta}'(M_e \otimes J_1)\widetilde{\beta}\right] d\nu_{\widetilde{\beta}} \qquad \text{(II.3.13)}$$

Note that $\widetilde{L}_T$ represents the standard form of the optimal test in the presence of unknown nuisance parameters if $\widetilde{\beta}$ is replaced by the standard form of alternative hypothesis $\frac{1}{\sqrt{T}}e \otimes \delta$. *Neyman-Pearson Lemma* asserts that a test based on the effective observation - effective score and information - is most powerful. (See theorem 5.3.2 of Lehman and Romano (2005) and Choi et al. (1996).) Therefore, Condition (4) has another interpretation that it gives the extended version of the optimal test under the existence of unknown nuisance parameters in the sense that the test statistic, (II.3.13), is interpreted as the weighted average of the optimal test statistics in the presence of nuisance parameters, where the weighting function is $d\nu_{\widetilde{\delta}}$

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

Lemma 4 *Let* $\bar{B}(\Omega)$ *be defined as* (III.3.7) *except the normalized score function* $\dot{\ell}_t^*(\hat{\beta}, \hat{J}_1)$ *is replaced by its demeaned counterpart with true null parameters i.e.* $\dot{\ell}_t^*(\beta_0, J_1) - \frac{1}{T}\sum_{i=1}^{T}\dot{\ell}_i^*(\beta_0, J_1)$.

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

where c is constant.

Lemma 4 implies that the test statistic $\bar{B}(\beta_0, J_1, \Omega)$ is asymptotically optimal if $|\widetilde{LR}_T - LR_T|$ converges to zero in probability both under the null and the alternative hypotheses. The convergence of $|\widetilde{LR}_T - LR_T|$ under the null hypothesis is basically the convergence in expectation because LR_T and $\widetilde{LR}_T$ are identical to the expectations of their integrands with respect to β and $\widetilde{\beta}$, respectively. Note that the integrands L_T, and $\widetilde{L}_T$ weakly converge to the same limiting distribution. In Theorem 3, I show that the weak convergence is enough for the convergence in quadratic mean of $|L_T - \widetilde{L}_T|$, and thereby, the convergence in probability of $|\widetilde{LR}_T - LR_T|$.

The convergence under the alternative can be presented by showing that the asymptotic null distribution is contiguous. Let $\phi_T(Z|\Omega, \beta_0)$ be a critical function for testing breaking processes where $Z = (y, X)$. That is, $\phi_T(Z|\Omega, \beta_0)$ is a $[0, 1]$ valued function determined by Z. I consider asymptotically α-significant test, i.e. $\lim_{T\to\infty} \int \phi_T(Z|\Omega, \beta_0) f_0(Z|\beta_0)\ dZ = \alpha$. The power function of the test is defined as $\int \phi_T(Z|\Omega, \beta_0) f_1(Z|\beta_0) dZ$. The following theorem gives the optimality result of the test statistics $\bar{B}(\Omega)$.

Theorem 3 *Let $\psi_T(Z|\Omega, \beta_0)$ be a critical function for $\bar{B}(\beta_0, \Omega)$, i.e. $\psi_T(Z|\Omega, \beta_0) = 1_{[\bar{B}(\beta_0,\Omega) > k_\alpha]}$ where k_α is the continuous function satisfying $\int \psi_T(Z|\Omega, \beta_0) f_0(Z|\beta_0) dZ = \alpha$. Under Conditions 1 to 4, the test $\bar{B}(\beta_0, \Omega)$ satisfies*

$$\overline{\lim}_{T\to\infty} \int\int \phi_T(Z|\Omega, \beta_0) f(Z|\underline{\theta}_0 + \tfrac{1}{T}\underline{d}) d\nu_\delta dZ \leq \underline{\lim}_{T\to\infty} \int\int \psi_T(Z|\Omega, \beta_0) f(Z|\underline{\theta}_0 + \tfrac{1}{T}\underline{d}) d\nu_\delta dZ$$

Theorem 3 implies that the power of the optimal test does not depend on the particular distributional form of δ_t other than its global covariance matrix, Ω. The main argument of this chapter follows from this property: The knowledge of the exact distribution of the unstable process is asymptotically useless for conducting an optimal test, as long as the process satisfies Condition 1. As sample size gets larger, there is little loss of power by relying on $\bar{B}(\Omega)$ rather than tailored LR_T. Theorem 1 also implies that any small sample optimal test designed to detect any specific unstable process will have the same asymptotic power against any other Condition 1 processes. It gives an important, practical implication about how to choose among a variety of different test statistics. One suggestion is to use the test statistic that is easy to compute in practice because one can get the similar result with using other hard-to-compute test statistic even though the unstable parameter process is not correctly specified.

Now I replace the unknown β_0 and J_1 by their consistent estimators to make the test statistic feasible. In general, this replacement causes the loss of power. But I show that replacing β_0 and J_1 by their maximum likelihood estimators does not affect the asymptotic properties of $\bar{B}(\cdot)$, which implies that the feasible test would have the best asymptotic power among the test under unknown initial parameters. This is possible because Condition 4 make the test have the power that is asymptotically equivalent to the power envelope of asymptotically α-similar tests under unknown nuisance parameters. Theorem 4 shows that this convergence holds if $\hat{J}_1$ and $\hat{\beta}$ satisfies the following condition.

Condition 5 *Under the null hypothesis, the maximum likelihood estimator,* $\hat{J}_1$ *and* $\hat{\beta}$*, satisfies*

1. $T^{-1/2}\sum_{t=1}^{[sT]}\dot{\ell}_t(\beta_0+T^{-1/2}h)=T^{-1/2}\sum_{t=1}^{[sT]}\dot{\ell}_t(\beta_0)-sK(\beta_0)h+o_p(1)$
2. $\sqrt{T}(\hat{\beta}-\beta_0)=O_p(1)$ *and* $\hat{J}_1=J_1+o_p(1)$

uniformly for $s \in (0,1)$*, and any nonrandom* $K(\beta_0)$*, where* $h < M < \infty$ *.*

Consider a class of asymptotically similar size tests; that is, tests that have limiting size α for all values of d_0. The critical function of the test is denoted as $\hat{\phi}_T(Z|\Omega, \beta_0)$. The following theorem shows that the test $B(\Omega)$ has the best asymptotic power among asymptotically similar size tests.

Theorem 4 *Let* $\hat{\psi}_T(Z|\Omega)$ *be a critical function for* $B(\Omega)$*. Under Conditions 1 to 3, and 5, the test* $B(\Omega)$ *satisfies*

$$\overline{\lim_{T\to\infty}} \int\int \hat{\phi}_T(Z|\Omega) f(Z|\underline{\theta}_0 + \frac{1}{T}\underline{d}) d\nu_d dZ \leq \underline{\lim_{T\to\infty}} \int\int \hat{\psi}_T(Z|\Omega) f(Z|\underline{\theta}_0 + \frac{1}{T}\underline{d}) d\nu_d dZ$$

Note that Theorem 4 does not require Condition 4. It implies that the test has asymptotically the best power among all tests that are asymptotically size-α and the initial parameter is unknown. The asymptotic null distribution of $B(\Omega)$ is given in the following lemma.

Lemma 5 *Under Conditions 1 to 3 and 5, the asymptotic null distribution of* $B(\Omega)$ *is*

$$\begin{aligned} B(\Omega) \quad \longrightarrow \quad & \sum_{i=1}^{k} [a_i J_i(1)^2 + a_i^2 \int_0^1 J_i(s)^2 ds + \\ & \frac{2a_i}{1-e^{-2a_i}} \{e^{-a_i} J_i(1) + a_i \int_0^1 e^{-a_i s} J_i(s) ds\}^2 - \{J_i(1) + a_i \int J^i(s) ds\}^2] \end{aligned} \tag{II.3.14}$$

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*.*

Selected asymptotic upper tail percentiles are calculated by Elliott and Müller (2006). (See table 1 of Elliott and Müller (2006).)

II.3.C An Asymptotically Point Optimal Test Statistic

I have developed the test statistic $B(\Omega)$ based on the assumption that the covariance matrix of δ_t, Ω, is known. However, the covariance matrix is unknown in practice. Therefore, 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 a specific Ω which implies selecting a point of alternative processes in which the test has maximal power. One possibility is to let Ω^* have a constant value, $\hat{\Omega}$. As long as the eigenvalues of Ω^* are of similar magnitudes, the power of the statistic will become close to the optimal power over a wide range of true Ω. Here I choose $\hat{\Omega} = c^2 I_k$ where $c = 10$. Replacing by C, the point optimal test statistic, $B(\hat{\Omega})$, is given by

$$B(\hat{\Omega}) = \sum_{i=1}^{k} \zeta_i'(\widehat{\beta}, \widehat{J}_1)'\{\frac{c^2}{T^2} I_T - FM_eF'\}^{-1}\zeta_i(\widehat{\beta}, \widehat{J}_1) \qquad \text{(II.3.15)}$$

In addition to the simplicity, using $\hat{\Omega}$ also has merit because it enables $B(\hat{\Omega})$ to be invariant with respect to re-parameterizations. Since $\dot{\ell}_i^*(\hat{\beta})$ does not change to any parameterization and $\{I_T - \frac{100}{T^2} FM_eF'\}$ is constant, we immediately observe that $B(\hat{\Omega})$ 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. The point optimal test, $B(\hat{\Omega})$, distorts the true size in two ways; absolute size, and relative size which is the relative magnitude of one parameter's break compared to those of the other parameters in the model.

Figure II.1 compares the asymptotic power of the point optimal test statistic to the power envelope, which is the power under known Ω^*. The first panel is for k=1. In univariate case, the point optimal test misspecifies the absolute size, but it does not distort the relative size. The asymptotic power is very close to the power envelope, which means that misspecifying how far the alternative process is from the null hypothesis gives little loss of power. It implies that the choice of c has little effect on the asymptotic power of the point optimal test.

The second panel examines the k=2 case where only one eigenvalue of Ω^* is set to have positive value while the other is set to be zero. Since eigenvalues describe the average magnitude of the breaking process in the direction of the corresponding eigenvectors, our setup considers the case that only one component of the 2×1 vector β_t breaks. The panel shows that there is some loss of power due to the replacement of Ω. However, the magnitude of the loss is not severely large. The largest loss occurs when the nonzero eigenvalue is 11 but the loss is less than 6%. The third panel describes when both parameters are non-stable and the average magnitude of the breaking processes differ. The average size of one breaking process is set to be four times greater than the other. In this case, the loss of power due to wrong identification is quite large, which implies that using C may break down the optimality of the test statistic. In conclusion, the goodness of B depends on the difference in the relative sizes of parameters breaking processes.

II.4 Examples

II.4.A Linear Model with Asymmetric Laplace Distribution

Consider the model

$$y_t = X_t'(\beta_0 + \beta_t) + \epsilon_t \quad t = 1, \dots, T \tag{II.4.1}$$

where y_t is a scalar, X_t, β_0 and β_t are $k \times 1$ vectors, $\{y_t, X_t\}$ are observed, β_0, β_t are unknown, and $\{X_t\}$ are assumed to be exogenous and satisfy Condition 2 with $E[X_t X_t'] = \Sigma_X$. ϵ_t is iid from asymmetric Laplace distribution which is defined as

$$\varphi^{\alpha}(\epsilon) = exp\left[-\frac{1}{\alpha}\epsilon \cdot 1_{\{\epsilon<0\}} + \frac{1}{1-\alpha}\epsilon \cdot 1_{\{\epsilon>0\}}\right] \tag{II.4.2}$$

where $1_{\{\ \}}$ is an indicator function. $X_t'(\beta_0 + \beta_t)$ represents α_{th} conditional quantile of y_t, that is,

$$Pr[y_t > X_t'(\beta_0 + \beta_t)|X_1, \dots, X_t] = \alpha \tag{II.4.3}$$

Consequently, ϵ_t is not a zero mean disturbance. It has the property that $Pr[\epsilon_t < 0] = \alpha$. The score and its covariance with maximum likelihood estimators are defined as

Table II.1: Monte Carlo Estimates of the Empirical Sizes (Laplace Error)

Sample Size	α	Empirical Size(%)					
		k=1			k=2		
		1	5	10	1	5	10
50	0.3	1.42	5.22	9.94	1.72	5.38	9.48
	0.5	0.8	4.62	8.88	1.11	4.38	8.44
100	0.3	1.00	4.58	9.02	1.10	4.58	8.64
	0.5	0.96	5.24	10.38	1.06	4.96	9.56
200	0.3	1.24	5.64	10.28	1.28	5.00	9.48
	0.5	1.02	4.76	9.98	1.14	5.36	9.70

$$\begin{aligned} \dot{\ell}_t(\widehat{\beta}) &= \frac{1}{1-\alpha}X_t - \frac{1}{\alpha(1-\alpha)}X_t 1_{\{y_t < X_t\hat{\beta}\}} \\ \widehat{J}_1 &= \frac{1}{T\alpha(1-\alpha)}\sum_{t=1}^{T} X_t X_t' \end{aligned} \tag{II.4.4}$$

Let us check Conditions 2 to 5. Conditions 2 and 4 are satisfied by construction. The Laplace distribution is known to be differentiable in quadratic mean. (See Pollard (1998) for details.) Chapter 4 shows the maximum likelihood estimator satisfies Condition 5. Therefore we can use $B(\hat{\Omega})$ as the optimal test statistic. $\hat{\beta}^{\alpha}$ can be estimated simply by using Koenker and Bassett (1978)'s quantile regression method.

I simulate the empirical size and the power of the test to examine how well the test performs in small samples. Two types of X_t are considered: $\{X_t\} = \{Z_t\}$ and $\{X_t\} = \{(1,\ Z_t)\}$, where $\{Z_t\}$ are generated from AR(1) model with iid Gaussian error. For each X_t, I consider 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 II.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

Table II.2: Alternative Processes for Power Simulations

Single Break :	$\beta_t = \frac{1}{\sqrt{T}}\delta \quad \text{for } t > \tau T$
fixed time :	$\tau = 50\%$, $\tau = 80\%$
random time :	$\tau \sim uniform[\pi, 1-\pi]$
smooth adjustment :	$\beta_{t-1} = 0.7\beta_t \quad \text{for } t \leq \tau T$
Multiple Breaks(fixed time) :	2 breaks, 4 breaks
Time varying parameters :	$T\Delta\beta_t \sim N(0, \delta^2)$

do not exceed one percent, even when the sample size is as small as 50.

As a next step, I perform monte carlo experiments to calculate small sample powers of the test and compare them with those of other optimal tests. I consider various types of alternative processes which are described in Table II.2. The powers are compared with those of *SupF* test, Andrews and Ploberger (1994)'s test (*ExpLM*) and Nyblom (1989)'s test (Nyb). *SupF* and *ExpLM* are designed for single break processes. *Nyb* considers martingale processes which include the single break with random occurrence and the random walk process. *B* covers the random walk and also the single break with random occurrence in small samples.

The size adjusted small sample powers are shown at figure II.2 and II.3. *ExpLM* has a severe size distortion problem in the model. *ExpLM* statistic has size of 8 to 14 percent for $T = 100$ and 7 to 9 percent for T=200. It becomes severer as the number of parameter or the degree of asymmetry increase. The figures show that *B* performs best among 4 test statistics. *B* has the best power against the random walk process and the multiple breaks. The gaps become larger as the number of breaks increase. The powers of B 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 test 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 I showed in the previous section can be more or less applied even in small samples and in the breaking process which are a little bit 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.

II.4.B Mean-Variance Instability under Univariate Normal Distribution

Let $x_1, \ldots, x_t$ be independent random variables drawn from normal distribution with mean μ and variance σ^2. Now Conditions 1 to 5 are almost trivially true.[1] I consider the case of a break in mean and that of breaks in both mean and variance. The former case is considered in linear models such as in Elliott and Müller (2006), while the latter case needs to be considered in the nonlinear set-up. The standard estimation of the scores and Fisher information matrix based on OLS can be used here.

I perform the same Monte Carlo simulations as the previous example and examine the performance of B in this set-up. Table II.3 shows the simulation result for empirical sizes. The empirical sizes are pretty close to the actual sizes in mean

[1]The test in the previous section does not explicitly consider variance parameter. But it is trivial to include σ^2 because the likelihood ratio in (II.3.4) still holds and all the lemmas can be applied. However, the nonzero property of the variance makes its replacement by the normal random variable ((II.3.11)) be somewhat irrelevant, and thereby cause some loss of power in the small sample

Table II.3: Monte Carlo Estimates of the Empirical Sizes (Gaussian Error)

Sample Size	Empirical Size(%)					
	break in μ			breaks in μ and σ^2		
	1	5	10	1	5	10
50	1.24	4.46	9.10	0.78	4.02	8.82
100	1.50	5.76	10.28	0.88	4.20	8.86
200	1.12	5.90	10.34	0.84	4.44	9.02

break case. But they are mildly lower than actual sizes in mean-variance breaks case at all significant levels, which implies that the test, B, is conservative. Figure II.4 and II.5 show the size adjusted small sample powers of the tests. ExpLM still has a size distortion problem. But the distortion is less severe than that in linear quantile models: The empirical sizes range from 8 to 10 percents. The size adjusted powers results are similar to those of the previous example. But the power gaps between B and SupF/ExpLM are bigger for random walk break and multiple breaks cases. The simulation gives a similar implication that the difference of the power among different optimal tests are too big except a few cases. But B is still preferred because it performs better than the other tests in all models and breaking processes considered in the section. I also examine the cases that the conditional mean is the linear function of exogenous variable x_t as in the previous example. The results are practically the same.

II.4.C Binary Choice Model with Logistic Error

Consider the model

$$y_t = I\{(\beta_0 + \beta_t)' X_t + \epsilon_t > 0\} \tag{II.4.5}$$

where I is the indicator function, X_t and β_t are as defined in the previous

Table II.4: Monte Carlo Estimates of the Empirical Sizes (Logit model)

Sample size	Empirical Size(%) 1	5	10
50	0.62	3.60	8.28
100	0.72	4.16	8.84
200	1.08	4.16	8.98

example, and ϵ_t is iid with standard logistic distribution. (II.4.5) is the traditional form of the times series binary choice model. The binary choice model is widely used in macroeconomic and finance time series especially when analyzing discrete or qualitative policy change such as central banks' bank rate. (see Eichengreen et al. (1985), and Dueker (2005) for details.) The first derivative of the log likelihood is defined as

$$\dot{\ell}_t = \begin{cases} \dfrac{\exp\left[(\beta_0+\beta_t)'X_t\right]}{1+\exp\left[(\beta_0+\beta_t)'X_t\right]}X_t & \text{if } y=1 \\ -X_t + \dfrac{\exp\left[(\beta_0+\beta_t)'X_t\right]}{1+\exp\left[(\beta_0+\beta_t)'X_t\right]}X_t & \text{if } y=0 \end{cases} \tag{II.4.6}$$

For the simulation, I consider $\{X_t\} = \{(1,\ Z_t)\}$, where $\{Z_t\}$ are generated from AR(1) model with iid Gaussian error. Monte Carlo simulation results of the empirical sizes are shown at table II.4. The test has reasonable empirical sizes for all sample sizes. The largest gap when the sample size is as small as 50 does not exceed $2\%p$. Similar to the mean-varinace test in the previous example, they are consistently lower than actual sizes, which implies that the test, B, is somewhat conservative.

Figure II.6 shows the size adjusted small sample powers of the tests. All four tests have similar small sample powers in single structural break cases. However, there exist non-negligible power gaps among the tests for multiple breaking processes. For example, B has up to 65 percent more powers than the other tests when there are four breaks in the sample period. The small sample power similarity in the previous examples does not seem to always hold in this nonlinear model. The good performance of B is more clear in this model. B has the power similar to those of the other tests in any type of single break cases. It fairly dominates the other tests in all other alternative processes.

II.5 Conclusion

Parameter instability is an important issue in applied economics because disregarding the instability causes a serious distortion in measuring and forecasting economic relationships. This chapter gives three implications for testing parameter stability.

First, asymptotically optimal tests for parameter instability do not require information about the specific form of the nonconstant parameter process. Many tests are designed to have good powers against specific unstable processes. The result in this chapter 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 useless.

Second, A small sample optimal test designed to detect a specific type of unstable parameter process has powers close to the best asymptotic powers even against other types of unstable processes. Monte carlo simulation results show that misspecifying the unstable process results in only a mild loss of power even in small samples. This suggests that one can choose any specific form of unstable parameter process which is easy to compute.

Finally, I suggest an easy-to-compute asymptotically optimal test statistic. The test statistic requires the maximum likelihood estimation under the null hypothesis only. Small sample simulations show that the test statistic has correct sizes and better powers than those of the other optimal tests for almost all unstable processes even in small samples.

II.A Proofs

II.A.A proof of Lemma 2

Let ξ_t^0, ξ_t^1 be the square root of $f(\epsilon_t^0|\theta_0)$, and, $f(\epsilon_t^1|\theta_t)$ respectively. Differentiability in quadratic mean, which is defined in (II.3.6), implies that ξ_t^1 is expanded by

$$\xi_t^1 = \xi_t^0 + \frac{1}{T} d_t' \dot{\xi}_t^0 + r_t \tag{II.A.1}$$

where $E[(\frac{r_t}{\xi_t^0})^2] = o_p(\|(d_t/T)\|^2)$. By using (III.A.1), the square root of the integrand of the LR statistics in (II.3.4) can be written as,

$$\begin{aligned} \sqrt{L_T} &= \prod_{t=1}^{T} \left(\frac{\xi_t^1}{\xi_t^0} \right) \qquad \text{(II.A.2)} \\ &= \prod_{t=1}^{T} \left(\frac{\xi_t^1 - \xi_t^0}{\xi_t^0} + 1 \right) \\ &= \prod_{t=1}^{T} \left(\frac{1}{T} d_t' \frac{\dot{\xi}_t^0}{\xi_t^0} + \frac{r_t}{\xi_t^0} + 1 \right) \\ &= \prod_{t=1}^{T} (1 + \eta_t) \end{aligned}$$

where $\eta_t = \frac{1}{T} d_t' \frac{\dot{\xi}_t^0}{\xi_t^0} + R_t$ and $R_t = \frac{r_t}{\xi_t^0}$. Thereofre 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 $\sum_{t=1}^{T} \eta_t^2 = O_p(1)$. Hence Lemma 2 is proved by showing

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

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

$$\begin{aligned} (\xi_t^1)^2 &= (\xi_t^0)^2 + r_t^2 + \frac{2}{T}\xi_t^0 d_t' \dot{\xi}_t^0 + 2\xi_t^0 r_t + \frac{2}{T} d_t' \dot{\xi}_t^0 r_t + \frac{1}{T^2} d_t' \dot{\xi}_t^0 \dot{\xi}_t^{0\prime} d_t \\ \Rightarrow 2\xi_t^0 r_t &= (\xi_t^1)^2 - (\xi_t^0)^2 - r_t^2 - \frac{2}{T}\xi_t^0 d_t' \dot{\xi}_t^0 - \frac{2}{T} d_t' \dot{\xi}_t^0 r_t - \frac{1}{T^2} d_t' \dot{\xi}_t^0 \dot{\xi}_t^{0\prime} d_t \\ \Rightarrow 2R_t &= \left(\frac{(\xi_t^1)^2}{(\xi_t^0)^2} - 1\right) - R_t^2 - \frac{1}{T} d_t' \dot{\ell}_t - \frac{1}{T} d_t' \dot{\ell}_t R_t - \frac{1}{4T^2} d_t' \dot{\ell}_t \dot{\ell}_t' d_t \quad \text{(II.A.3)} \end{aligned}$$

By taking conditional expectation with respect to d_t, we get

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

Let $\widetilde{R}_t = \mathbf{1}\{\|d_t/\sqrt{T}\| < M_T\} R_t$ 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 $\max_{\{\|d_t/\sqrt{T}\|<M_T\}} (\frac{1}{T^2} d_s' d_s)^{-1} \times E[R_s^2|d_s] = o_p(1)$ from (III.A.1) and $\frac{1}{T^2}\sum d_t d_t' = O_p(1)$ from Condition 1. Therefore

$$\sum_{t=1}^{T} E[\widetilde{R}_t^2|d_t] = \sum_{t=1}^{T} \mathbf{1}\{\|d_t/\sqrt{T}\| < M_T\} E[R_t^2|d_t] \tag{II.A.5}$$
$$\leq \sum_{t=1}^{T} \max_{\{\|d_t/\sqrt{T}\|<M_T\}} \left((\frac{1}{T^2} d_s' d_s)^{-1} E[R_s^2|d_s] \right) \frac{1}{T^2} d_t' d_t$$
$$= \max_{\{\|d_t/\sqrt{T}\|<M_T\}} \left((\frac{1}{T^2} d_s' d_s)^{-1} E[R_s^2|d_s] \right) \frac{1}{T^2} \sum_{t=1}^{T} d_t' d_t$$
$$= o_p(1) \times O_p(1) = o_p(1)$$

Also, using Chebychev inequality

$$\frac{1}{T} d_{ti} E[\dot{\ell}_{ti} R_t|d_t] \leq \frac{1}{T} d_{ti} E[\dot{\ell}_{t,i}^2|d_t]^{1/2} E[R_t^2|d_t]^{1/2} \tag{II.A.6}$$
$$= O_p(T^{-1/2}) \times (O_p(1))^{1/2} \times (o_p(T^{-1/2}))^{1/2} = o_p(T^{-1})$$

Note that $E[\dot{\ell}_t|d_t] = 0$, and $E[\dot{\ell}_t \dot{\ell}_t'|d_t] = J$,(see Vaart (1998)). Using (II.A.6) together with (II.A.7), we prove (1) by showing that

$$\sum_{t=1}^{T} R_t = \sum_{t=1}^{T} E[R_t|d_t] + o_p(1) \tag{II.A.7}$$
$$= \frac{1}{8T^2} \sum_{t=1}^{T} d_t' J d_t + o_p(1) \tag{II.A.8}$$

proof of (2):

$$\sum_{t=1}^{T} \eta^2 = \frac{1}{4T^2} \sum_{t=1}^{T} d_t' \dot{\ell}_t \dot{\ell}_t' d_t + \frac{1}{T} \sum_{t=1}^{T} d_t' \dot{\ell}_t R_t + \sum_{t=1}^{T} R_t^2$$
$$= \left(\frac{1}{4} \sum_{t=1}^{T} d_t' J d_t + o_p(1) \right) + o_p(1) + o_p(1) \tag{II.A.9}$$

where the last two terms of the last equality comes from (II.A.7) and (II.A.8).

Proof of (3):

$$\begin{aligned}
\max_t \eta_t &\leq \frac{1}{2} \max_{t, \|\frac{1}{\sqrt{T}} d_t\| \leq M_T} |\frac{1}{\sqrt{T}} d_t|' \cdot |\frac{\dot{\ell}_t}{\sqrt{T}}| + \max_t R_t + o_p(1) \\
&\leq \frac{1}{2} \max_{\|\frac{1}{\sqrt{T}} d_t\| \leq M_T} \|\frac{1}{\sqrt{T}} d_t\| \cdot \|\frac{\dot{\ell}_t}{\sqrt{T}}\| + \max_t R_t + o_p(1) \\
&\leq \frac{M_T}{\sqrt{T}} \|\dot{\ell}_t\| + \max_t R_t + o_p(1) \\
&= o_p(1) + o_p(1) + o_p(1) = o_p(1) \qquad \text{(II.A.10)}
\end{aligned}$$

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

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

which completes the proof. $\diamond$

II.A.B proof of Lemma 3

Note that $E[\dot{\ell}_0 \dot{\ell}'] = e' \otimes J_1$ and $E[\dot{\ell}_0 \dot{\ell}_0'] = e'e \otimes J_1$ so that $E[\dot{\ell}_0 \dot{\ell}_0']^{-1} E[\dot{\ell}_0 \dot{\ell}'] = (e'e \otimes J_1)^{-1}(e' \otimes J_1) = \frac{1}{T} e' \otimes I_k$. Also note that $\dot{\ell}^{\theta} = (\dot{\ell}_1', \ldots, \dot{\ell}_T', \sum \dot{\ell}_t')' = (I_T \otimes I_k, e \otimes I_k)' \dot{\ell}$. Under Condition 4, (II.3.7) can be rewritten as

$$
\begin{aligned}
L_T &= \exp\left[\frac{1}{T}d'\dot{\ell}^{\theta} - \frac{1}{2T^2}d'[I_T \otimes I_k, e \otimes I_k]'(I_T \otimes J_1)[I_T \otimes I_k, e \otimes I_k]d\right] \\
&= \exp[\frac{1}{T}\delta'[I_T \otimes I_k, -\frac{1}{T}e \otimes I_k][I_T \otimes I_k, e \otimes I_k]'\dot{\ell} - \frac{1}{2T^2}\delta'[I_T \otimes I_k, -\frac{1}{T}e \otimes I_k] \times \\
&\qquad [I_T \otimes I_k, e \otimes I_k]'(I_T \otimes J_1)[I_T \otimes I_k, e \otimes I_k][I_T \otimes I_k, -\frac{1}{T}e \otimes I_k]'\delta] \\
&= \exp\left[\frac{1}{T}\delta'(M_e \otimes I_k)\dot{\ell} - \frac{1}{2T^2}\delta'(M_e \otimes J_1)\delta\right] \qquad \text{(II.A.11)}
\end{aligned}
$$

Note that (II.A.11) is the same as (II.3.12) except $\widetilde{\delta}$ is replaced by δ. Therefore, I need only to show that for any δ 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.

(1)

$$
\begin{aligned}
\frac{1}{T}\delta'(M_e \otimes I_k)\dot{\ell} &= \frac{1}{T}\delta'\dot{\ell} - \frac{1}{T^2}\delta'(ee' \otimes I_k)\dot{\ell} \\
&= \frac{1}{T}\delta'\dot{\ell} - \frac{1}{T^2}[(e' \otimes I_k)\delta]'[(e' \otimes I_k)\dot{\ell}] \\
&= \frac{1}{T}\sum_{t=1}^{T}\delta_t'\dot{\ell}_t - \frac{1}{T^2}(\sum_{t=1}^{T}\delta_t)'(\sum_{t=1}^{T}\dot{\ell}_t) \qquad \text{(II.A.12)}
\end{aligned}
$$

Therefore, I prove that each term of the last equation converge to well defined limiting distributions.

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\delta_t'\dot{\ell}_t &= tr[\Omega^{*\frac{1}{2}}\frac{1}{T}\sum_{t=1}^{T}\Omega^{-\frac{1}{2}}\delta_t'\dot{\ell}_t J_1^{-\frac{1}{2}}] \\
&= tr[\Omega^{*\frac{1}{2}}\int W_\delta dW_\ell'] \\
&= \int W_\delta'\Omega^{*\frac{1}{2}}dW_\ell \qquad \text{(II.A.13)}
\end{aligned}
$$

where W_δ and W_ℓ are multivariate standard Wiener processes. And,

$$\frac{1}{T^2}(\sum_{t=1}^{T}\delta_t)'(\sum_{t=1}^{T}\dot{\ell}_t) = tr\left[\Omega^{*\frac{1}{2}}(T^{\frac{3}{2}}\sum_{t=1}^{T}\Omega^{-\frac{1}{2}}\delta_t)(T^{\frac{1}{2}}\sum_{t=1}^{T}J^{-\frac{1}{2}}\dot{\ell}_t)'\right]$$
$$= tr\left[\Omega^{*\frac{1}{2}}\int W_\delta(r)drW_\ell(1)'\right]$$
$$= \int W_\delta(r)dr\Omega^{*\frac{1}{2}}W_\ell(1) \qquad \text{(II.A.14)}$$

(2)

$$\frac{1}{T^2}\delta'(M_e\otimes J_1)\delta = \frac{1}{T^2}\delta'(I_T\otimes J_1)\delta - \frac{1}{T^3}\delta'(ee'\otimes J_1)\delta$$
$$= \frac{1}{T^2}\delta'(I_T\otimes J_1)\delta - \frac{1}{T^3}[(e'\otimes J^{1/2})\delta]'[(e'\otimes J_1^{1/2})\delta]$$
$$= \frac{1}{T}\sum_{t=1}^{T}\delta_t' J_1\delta - \frac{1}{T^3}(\sum_{t=1}^{T}\delta_t)'J_1(\sum_{t=1}^{T}\delta_t) \qquad \text{(II.A.15)}$$

$$\frac{1}{T^2}\sum_{t=1}^{T}\delta_t' J\delta_t = tr\left[\Omega^*\frac{1}{T^2}\sum_{t=1}^{T}\left(\Omega^{-\frac{1}{2}}\delta_t\right)\left(\Omega^{-\frac{1}{2}}\delta_t\right)'\right]$$
$$= tr\left[\Omega^*\int W_\delta(r)W_\delta(r)'dr\right]$$
$$= \int W_\delta(r)'\Omega^* W_\delta(r)dr \qquad \text{(II.A.16)}$$

$$\frac{1}{T^3}\left(\sum_{t=1}^{T}\delta_t\right)'J_1\left(\sum_{t=1}^{T}\delta_t\right) = tr\left[\Omega^*\left(T^{\frac{3}{2}}\sum_{t=1}^{T}\Omega^{-\frac{1}{2}}\delta_t\right)\left(T^{\frac{3}{2}}\sum_{t=1}^{T}\Omega^{-\frac{1}{2}}\delta_t\right)'\right]$$
$$= tr\left[\Omega^*\int W_\delta(r)dr\int W_\delta(r)'dr\right]$$
$$= \int W_\delta(r)'dr\Omega^*\int W_\delta(r)dr \qquad \text{(II.A.17)}$$

which completes the proof. $\diamond$

II.A.C Proof of Lemma 4

Let's denote the variance of $\widetilde{\beta}$, $FF'/T^2 \otimes \Omega$ as K. The test statistic $\widetilde{LR}_T$ can be written as

$$
\begin{aligned}
\widetilde{LR}_T &= \int (2\pi)^{-\frac{k(T-1)}{2}} |K|^{-\frac{1}{2}} \exp\left[\dot{\ell}'(M_e \otimes I_k)\widetilde{\beta} - \frac{1}{2}\widetilde{\beta}'(M_e \otimes J_1)\widetilde{\beta} - \frac{1}{2}\widetilde{\beta}'K^{-1}\widetilde{\beta}\right] d\widetilde{\beta} \\
&= |K(M_e \otimes J_1) + I_{Tk}|^{1/2} \exp\left[\frac{1}{2}\dot{\ell}'(M_e \otimes I_k)\{(M_e \otimes J_1) + K^{-1}\}^{-1}\right. \\
&\quad \left.\times (M_e \otimes I_k)\dot{\ell}\right] \int (2\pi)^{-\frac{k(T-1)}{2}} \left|(M_e \otimes J_1) + K^{-1}\right|^{\frac{1}{2}} \\
&\quad \times \exp\left[-\frac{1}{2}\left(\widetilde{\beta} - \{(M_e \otimes J_1) + K^{-1}\}(M_e \otimes I_k)\dot{\ell}\right)' \{(M_e \otimes J_1) + K^{-1}\}\right. \\
&\quad \left.\times \left(\widetilde{\beta} - \{(M_e \otimes J_1) + K^{-1}\}(M_e \otimes I_k)\dot{\ell}\right)\right] d\nu_{\widetilde{\beta}} \\
&= |K(M_e \otimes J_1) + I_{Tk}| \exp\left[\frac{1}{2}\dot{\ell}'(M_e \otimes I_k)\{(M_e \otimes J_1) + K^{-1}\}^{-1}(M_e \otimes I_k)\dot{\ell}\right] \\
&= c \cdot \exp\left[\frac{1}{2}\dot{\ell}'(M_e \otimes I_k)\{M_e \otimes J + T^2(FF')^{-1} \otimes \Omega^{-1}\}^{-1}(M_e \otimes I_k)\dot{\ell}\right] \\
&= c \cdot \exp[\frac{1}{2}\dot{\ell}'(M_e \otimes I_k)(I \otimes J^{-1/2}P)\{M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda^{-1}\}^{-1} \\
&\qquad \times (I_T \otimes J^{-1/2}P)'(M_e \otimes I_k)\dot{\ell}] \\
&= c \cdot \exp\left[\frac{1}{2}\bar{\dot{\ell}}^{*\prime}\{M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda^{-1}\}^{-1}\bar{\dot{\ell}}^{*\prime}\right] \qquad \text{(II.A.18)}
\end{aligned}
$$

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.

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

$$= \sum_{i=1}^{k} (M_e + K_{ai}^{-1}) \otimes \iota_i \iota_i' \qquad \text{(II.A.19)}$$

where $K_{ai} = a_i^2 \left(\frac{FF'}{T^2}\right)^{-1}$. 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.

$$(M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda)^{-1} = \sum_{i=1}^{k} (M_e + K_{ai}^{-1})^{-1} \otimes \iota_i \iota_i' \qquad \text{(II.A.20)}$$

because

$$\left[\sum_{i=1}^{k} \left(M_e + K_{ai}^{-1}\right) \otimes \iota_i \iota_i'\right] \left[\sum_{i=1}^{k} (M_e + K_{ai}^{-1})^{-1} \otimes \iota_i \iota_i'\right] \qquad \text{(II.A.21)}$$

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

Therefore,

$$\bar{\dot{\ell}}^{*\prime} (M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Omega^{*-1})^{-1} \bar{\dot{\ell}}^{*\prime}$$

$$= \bar{\dot{\ell}}^{*\prime} \left[\sum_{i=1}^{k} (M_e + K_{ai}^{-1})^{-1} \otimes \iota_i \iota_i'\right] \bar{\dot{\ell}}^{*\prime}$$

$$= \sum_{i=1}^{k} \bar{\dot{\ell}}^{*\prime} F (F M_e F + (\frac{a_i^2}{T^2}) I_T)^{-1} F' \bar{\dot{\ell}}^{*\prime} \qquad \text{(II.A.22)}$$

Taking log of (II.A.18) and applying (II.A.22) completes the proof. ⋄

II.A.D proof of Lemma 5

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,

$$\begin{aligned}
[M_e + K_{ai}^{-1}]^{-1} &= A_i^{-1} - (A_i^{-1}e)(1 + e'A_i^{-1}e/T)^{-1}(e'A_i^{-1}) \qquad \text{(II.A.23)} \\
&= I - (I + K_{ai})^{-1} + (1 + e'A_i^{-1}e/T)^{-1} \\
&\qquad \left[ee' - 2(I + K_{ai})^{-1}ee' + (I + K_{ai})^{-1}ee'(I + K_{ai})^{-1}\right]
\end{aligned}$$

Let's define $T \times (T-1)$ vector B_e as $B_e B_e' = M_e$. Then

$$\begin{aligned}
&M_e[M_e + K_{ai}^{-1}]^{-1}M_e \\
&\quad = M_e - M_e(I + K_{ai})^{-1}M_e \\
&\qquad +(1 + e'A_i^{-1}e/T)^{-1}M_e(I + K_{ai})^{-1}ee'(I + K_{ai})^{-1}M_e \\
&\quad = M_e - M_e(I + K_{ai})^{-1}M_e \\
&\qquad +(e'(I + K_{ai})^{-1}e)^{-1}M_e(I + K_{ai})^{-1}ee'(I + K_{ai})^{-1}M_e \\
&\quad = M_e - B_e(B_e(I + K_{ai})B_e')^{-1}B_e' \\
&\quad = M_e - G_{ai} \qquad \text{(II.A.24)}
\end{aligned}$$

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

$$B(\beta_0, \Omega) = \sum_{i=1}^{k} \dot{\ell}_i^{*\prime}(\beta_0)\{M_e - G_{a_i}\}\dot{\ell}_i^{*}(\beta_0) \tag{II.A.25}$$

Lemma 6 of Elliott and Müller (2006) gives us the distribution of the test statistic which is given as below,

$$\sum_{i=1}^{k} \Bigg[a_i J_i(1)^2 + a_i^2 \int_0^1 J_i(s)^2 ds + \tag{II.A.26}$$
$$\frac{2a_i}{1 - e^{-2a_i}} \left\{ e^{-a_i} J_i(1) + a_i \int_0^1 e^{-a_i s} J_i(s) ds \right\}^2 - \left\{ J_i(1) + a_i \int J^i(s) ds \right\}^2 \Bigg]$$

where $J_i(s) = T^{-\frac{1}{2}} \sum_{t=1}^{[Ts]} \dot{\ell}_t^{*} - \int_0^s e^{-c(s-\lambda)} (T^{-\frac{1}{2}} \sum_{t=1}^{[T\lambda]} \dot{\ell}_t^{*}) d\lambda$

Under Condition 2 and 3, $\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \dot{\ell}_t^{*} = sW(s)$ where $W(s)$ is multivariate standard wiener processes, which completes the proof.

II.A.E proof of Theorem 3

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

$$LR_T(M) = \int \Pi_{t=1}^{T} \frac{f(\epsilon_t^1|\beta_0)}{f(\epsilon_t^0|\beta_0,\beta_t)} \mathbf{1}\{\|\delta\| < \sqrt{T}M\} d\nu_\delta \tag{II.A.27}$$

$$\widetilde{LR}_T(M) = \int exp\left[\dot{\ell}'(M_e \otimes I_k)\widetilde{\beta} - \frac{1}{2}\widetilde{\beta}'(M_e \otimes J)\widetilde{\beta}\right] \mathbf{1}\{\|\delta\| < \sqrt{T}M\} d\nu_{\widetilde{\delta}} \tag{II.A.28}$$

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

$$\begin{aligned} P[|LR_T - \widetilde{LR}_T| > 3\epsilon] &\leq P[|LR_T - LR_T(M)| > \epsilon] && (i) \\ &+ P[|\widetilde{LR}_T - \widetilde{LR}_T(M)| > \epsilon] && (ii) \\ &+ P[|LR_T(M) - \widetilde{LR}_T(M)| > \epsilon] && (iii) \end{aligned} \tag{II.A.29}$$

Therefore, it suffices to show that each term of (III.A.15) converges to zero, respectively.

$$\begin{aligned} (i)\ P[|LR_T - LR_T(M)| > \epsilon] &\leq \epsilon^{-1}E[|LR_T - LR_T(M)|] \\ &= \epsilon^{-1}E[\int_{\|\delta\|>\sqrt{T}M} \Pi_{t=1}^{T} \frac{f(\epsilon_t|\beta_0)}{f(\epsilon_t^1|\beta_0,\beta_t)} d\nu_\delta] \\ &= \epsilon^{-1}\int_{\|\delta\|>\sqrt{T}M} d\nu_\delta \ = \ \epsilon^{-1}P[\|\delta\| > \sqrt{T}M] \end{aligned} \tag{II.A.30}$$

The first inequality comes from Chebychev inequality. The last 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 β defined in Condition 1.

Proof of (ii):

$$
\begin{aligned}
\left|\widetilde{LR}_T - \widetilde{LR}_T(M)\right| &= \int \widetilde{L}_T d\nu_{\widetilde{\delta}} - \int_{\|\delta\|<\sqrt{T}M} \widetilde{L}_T d\nu_{\widetilde{\delta}} \\
&= c \cdot \exp\left[\frac{1}{2}\bar{\dot{\ell}}^{*\prime}\{M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda^{-1}\}^{-1}\bar{\dot{\ell}}^{*\prime}\right] \\
&= \times \int (2\pi)^{-\frac{k(T-1)}{2}} \left|(M_e \otimes J_1) + K^{-1}\right|^{\frac{1}{2}} \\
&= \times \exp[-\frac{1}{2}\left(\widetilde{\beta} - \{(M_e \otimes J_1) + K^{-1}\}(M_e \otimes I_k)\dot{\ell}\right)' \\
&= \times \{(M_e \otimes J_1) + K^{-1}\}\left(\widetilde{\beta} - \{(M_e \otimes J_1) + K^{-1}\}(M_e \otimes I_k)\dot{\ell}\right)]d\nu\widetilde{\beta} \\
&= c \cdot \exp\left[\frac{1}{2}\bar{B}(\beta_0, J_1, \Omega)\right]\int_{\|\delta\|>\sqrt{T}M} d\nu\widetilde{\beta} \qquad \text{(II.A.31)}
\end{aligned}
$$

The first term on the last equation is $O_p(1)$ by Lemma 5, 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

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

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

$$\begin{aligned} LR_T(M) &= \int L_T(M) d\nu\delta \\ &= \int (1 + o_p(1|\beta) \widetilde{L}^*_T(M) d\nu_\beta \\ &= \widetilde{LR}^*_T(M) + o_p(1) \end{aligned} \tag{II.A.32}$$

The second equality follows from Lemma 2 with Condition 4. The third equality uses $\widetilde{LR}^*_T(M)$ is bounded in probability which is shown as

$$\begin{aligned} P[\widetilde{LR}^*_T(M) > K] &\leq K^{-1} E\left[\widetilde{LR}^*_T(M)\right] \\ &= K^{-1} \int E\left[LR^*_T(M)\right] d\nu_\beta \\ &= K^{-1} \int E[L_T|\beta] \mathbf{1}\{\|\sqrt{T}\beta\| < M\} d\nu_\beta \end{aligned} \tag{II.A.33}$$

which can be made arbitrarily small by choosing K sufficiently large. To prove the second convergence, we use an additional indicator function $\mathbf{1}\{\bar{B}(\cdot) > K\}$ 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 $\epsilon > 0$,

$$P[|\widetilde{LR}_T(M) - \widetilde{LR}_T * (M)| > 3\epsilon]$$
$$\leq P[|\widetilde{LR}_T(M) - \widetilde{LR}_T(M,K)| > \epsilon] + P[|\widetilde{LR}^*_T(M) - \widetilde{LR}^*_T(M,K)| > \epsilon]$$
$$+P[|\widetilde{LR}_T(M,K) - \widetilde{LR}^*_T(M,K)| > \epsilon] \qquad \text{(II.A.34)}$$

The convergence of the first term can be easy to show by using the similar method of (III.A.17), i.e.

$$\|\widetilde{LR}_T(M) - \widetilde{LR}_T(M,K)\| = c\exp\left[\frac{1}{2}\bar{B}(\cdot)\right]\mathbf{1}\{\bar{B}(\cdot) > K\}P\left[\|\sqrt{T}\beta\| < M\right] \qquad \text{(II.A.35)}$$

in which $P[\|\widetilde{LR}_T(M) - \widetilde{LR}_T(M,K)\| > \epsilon]$ can be made arbitrarily small by taking K sufficiently large. The convergence of the second term can be shown as

$$\begin{aligned} P[|\widetilde{LR}^*_T(M) &- \widetilde{LR}^*_T(M,K)| > \epsilon] \leq \frac{1}{\epsilon}E\left[\left|\widetilde{LR}^*_T(M) - \widetilde{LR}^*_T(M,K)\right|\right] \\ &= \frac{1}{\epsilon}\int\int L^*_T\left(1 - \mathbf{1}\{\bar{B}(\cdot) > K\}\right)\mathbf{1}\|\sqrt{T}\beta\| < M d\nu_\beta d\nu_z \\ &\leq \frac{1}{\epsilon}\int \mathbf{1}\|\sqrt{T}\beta\| < M d\nu_\beta = P\left[\|\sqrt{T}\beta\| < M\right] \qquad \text{(II.A.36)} \end{aligned}$$

where the second equality uses Fubini theorem and the third inequality comes from $\int L^*_T d\nu_z = 1$. (II.A.36) 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 β and $\widetilde{\beta}$, respectively and are independent of β and $\widetilde{\beta}$ and of each other. We prove $LR^*_T(M) - \widetilde{LR}_T(M)$ convergence in mean square which implies the convergence in probability. 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.

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

Lemma 3 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 4.

$$
\begin{aligned}
&E[\widetilde{L}_T(M,K)^a] \\
&\quad = \int\int (2\pi)^{-\frac{k(T-1)}{2}}|K|^{-\frac{1}{2}}\exp[a\dot{\ell}'(M_e\otimes I_k)\widetilde{\beta} - a\widetilde{\beta}'(M_e\otimes J_1)\widetilde{\beta} - \frac{1}{2}\widetilde{\beta}'K^{-1}\widetilde{\beta}] \\
&\qquad \times \mathbf{1}[\|\sqrt{T}\beta\| < M]\mathbf{1}[\bar{B}(\cdot) < K]d\widetilde{\beta}d\nu_z \\
&\quad = c_1\cdot\int \exp[\frac{a^2}{4}\dot{\ell}'(M_e\otimes I_k)(a(M_e\otimes J_1)+K^{-1})^{-1}(M_e\otimes I_k)\dot{\ell}] \\
&\qquad \times \int (2\pi)^{-\frac{k(T-1)}{2}}|a(M_e\otimes J_1)+K^{-1}|^{\frac{1}{2}} \\
&\qquad \times \exp[-\frac{1}{2}(\widetilde{\beta} - a(aM_e\otimes J_1 + K^{-1})(M_e\otimes I_k)2\dot{\ell})'(a(M_e\otimes J_1)+K^{-1}) \\
&\qquad \times (\widetilde{\beta} - a(aM_e\otimes J_1 + K^{-1})(M_e\otimes I_k)2\dot{\ell})\mathbf{1}[\|\sqrt{T}\beta\| < M]\mathbf{1}[\bar{B}(\cdot) < K]d\widetilde{\beta}d\nu_z \\
&\quad = c_1\int \exp\left[\frac{a^2}{4}\bar{\dot{\ell}}^{*\prime}(M_e\otimes I_k + (\frac{FF'}{T^2})^{-1}\otimes\frac{1}{a}\Omega^{*-1})^{-1}\bar{\dot{\ell}}^{*\prime}\right] \\
&\qquad \mathbf{1}[\bar{B}(\cdot) < K]d\nu_z P\left[\|\sqrt{T}\beta\| < M\right] \\
&\quad = c_1 P\left[\|\sqrt{T}\beta\| < M\right]\int \exp\left[\frac{a}{4}B(\Omega,\sqrt{a}J_1,\beta_0)\right]\mathbf{1}[\bar{B}(\cdot) < K]d\nu_z \\
&\quad \leq c_1 P\left[\|\sqrt{T}\beta\| < M\right]\exp\left[\frac{a}{4}K\right]
\end{aligned}
\tag{II.A.37}
$$

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, $f(\epsilon_t^1|\beta_0,\beta_t)$ is contiguous to that under the null hypothesis, $f(\epsilon_t^0|\beta_0)$. The condiguity of the distribution in which the likelihood ratio has the asymptotic distribution as (III.3.9) has alredy been shown by Elliott and Müller (2006).

II.A.F proof of Theorem 4

Let's define the power function of $\hat{\phi}_T(Z|\Omega)$ as $h(\Omega)$, i.e.

$$h(\Omega) = \int\int \hat{\phi}_T(Z|\Omega) f(Z|\frac{1}{T}\delta_t, \beta_0 + \frac{1}{T}\delta_0) dZ d\nu_\delta \tag{II.A.38}$$

Note that the test does not depend on the measure of δ_0 because of the asymptotic regularity of the test. Since the test $\bar{B}(\beta_0, \Omega)$ is asymptotically most powerful for testing that $f(Z|\beta_0)$ is the true density versus $f(Z|\frac{1}{T}\delta_t, \beta_0 + \frac{1}{T}\bar{\delta}_0(\delta_t))$ is true and $\hat{\phi}_T(Z|\Omega)$ has asymptotic α-size for $\bar{\delta}_0(\delta_t)$, Neyman-Pearson Lemma gives the following inequality.

$$h(\Omega) \leq \int\int \hat{\phi}_T(Z|\Omega) f\left(Z|\frac{1}{T}\delta, \beta_0 + \frac{1}{T}\bar{\delta}_0(\delta)\right) dZ d\nu_\delta + o_p(1) \tag{II.A.39}$$

Therefore, theorem 2 is proved if $B(\Omega)$ is asymptotically equivalent to $\bar{B}(\beta_0, \Omega)$ under both the null and the alternative hypothesis. Let's rewrite $\hat{\beta}$ as

$$\hat{\beta} = \beta_0 + T^{-\frac{1}{2}} W_T$$

where W_T is a $k \times 1$ random variable with $P[|W_T| > M] \to 0$ for arbitrarily large M. By using Condition 5 and continuous mapping theorem, we could get

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{[sT]} \widehat{J}_1^{-1/2}\dot{\ell}_t(\hat{\beta}) &= T^{-1/2}\sum_{t=1}^{[sT]} \widehat{J}_1^{-1/2}\dot{\ell}_t(\beta_0 + T^{-1/2}W_T) + o_p(1) \\ &= T^{-1/2}\sum_{t=1}^{[sT]} \widehat{J}_1^{-1/2}\dot{\ell}_t(\beta_0) - sK(\beta_0)W_T + o_p(1) \end{aligned} \tag{II.A.40}$$

Since $W_T(1)$ is constant for all $t \leq T$, it can be easily proved by showing that the test statistic $\bar{B}(\beta_0, \Omega)$ doesn't change for the transformation from $\{\dot{\ell}_i(\beta_0)\}$ to $\{\dot{\ell}_i(\beta_0) + c\}$ where c is the $T \times 1$ vector of constants. Note that $M_e \dot{\ell}_i^*(\beta_0) = M_e \left[\dot{\ell}_i^*(\beta_0) + c\right]$. By using (II.A.24) and (II.A.25), we could get

$$\begin{aligned}
\bar{B}(\beta_0, \Omega) &= \sum_{i=1}^{k} \dot{\ell}_i^{*\prime}(\beta_0)\{M_e - G_{a_i}\}\dot{\ell}_i^*(\beta_0) \\
&= \sum_{i=1}^{k} \dot{\ell}_i^{*\prime}(\beta_0) M_e [M_e + K_{ai}^{-1}]^{-1} M_e \dot{\ell}_i^*(\beta_0) \\
&= \sum_{i=1}^{k} \left[\dot{\ell}_i^{*\prime}(\beta_0) + c\right] M_e [M_e + K_{ai}^{-1}]^{-1} M_e \left[\dot{\ell}_i^*(\beta_0) + c\right] \\
&= \sum_{i=1}^{k} \dot{\ell}_i^{*\prime}(\widehat{\beta}_0)\{M_e - G_{a_i}\}\dot{\ell}_i^*(\widehat{\beta}_0) + o_p(1) \\
&= B(\Omega) + o_p(1) \qquad \text{(II.A.41)}
\end{aligned}$$

which shows the asymptotic equivalency under the null hypothesis. The asymptotic equivalency under the alternative hypothesis comes from the contiguity of $\bar{B}(\beta_0, \Omega)$, which completes the proof. $\diamond$

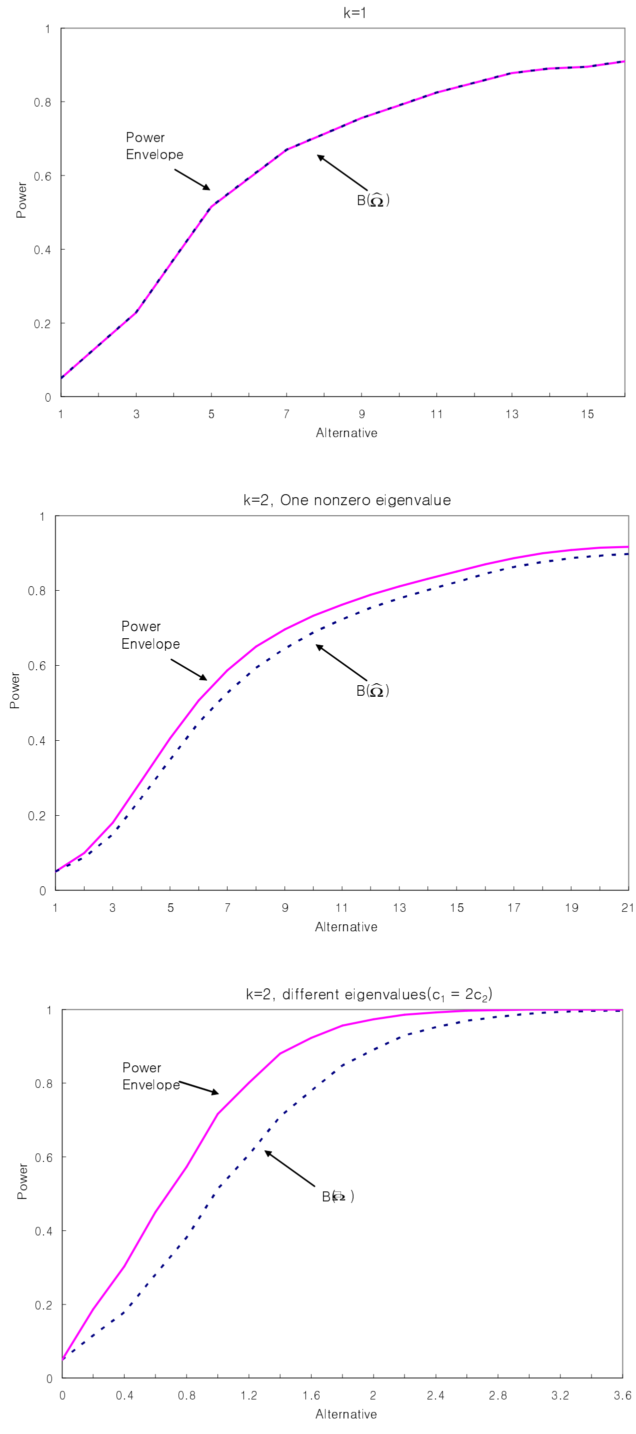

Figure II.1: Asymptotic Local Power

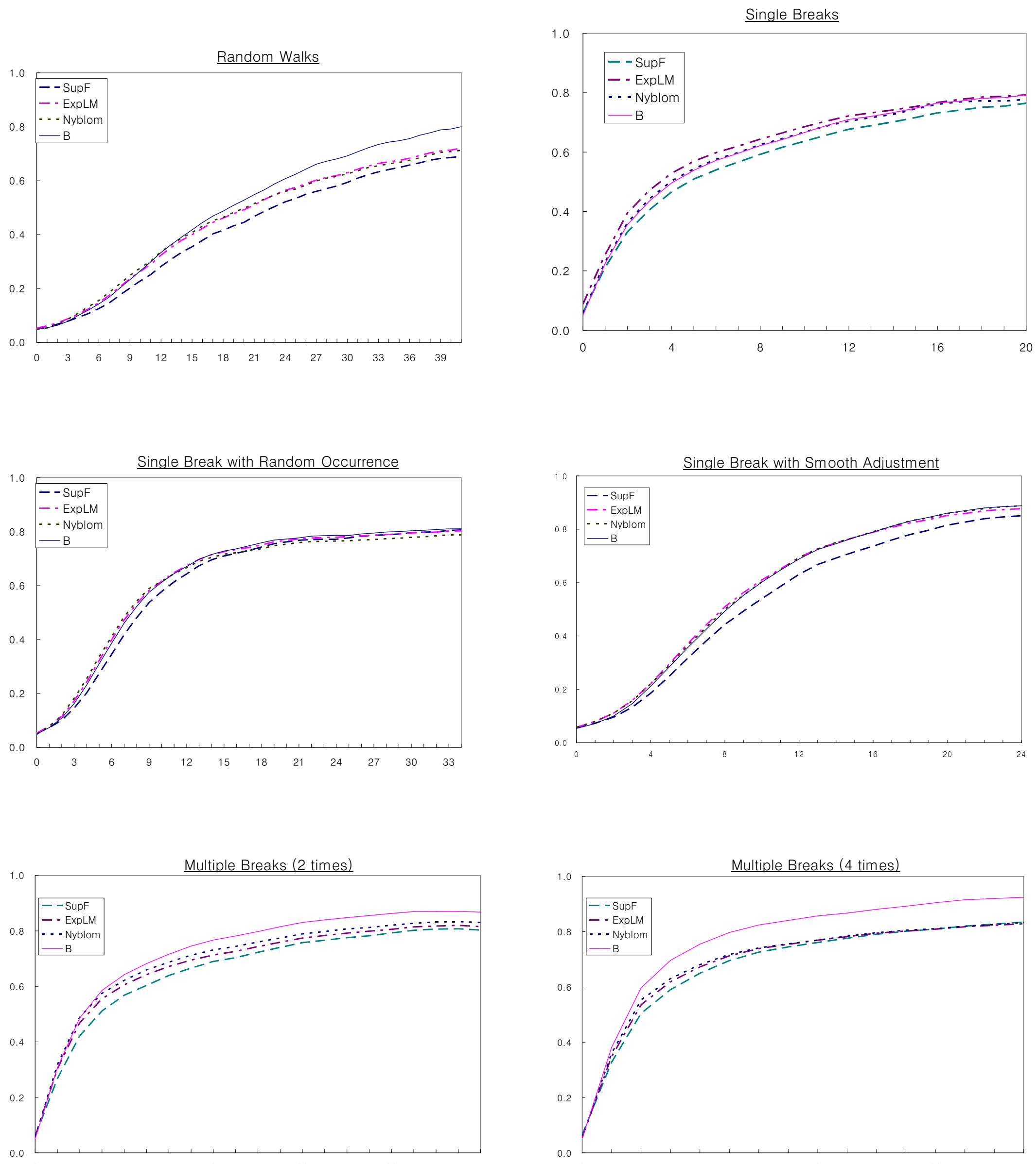

Figure II.2: Small Sample Powers, Quantile Models, k=1, T=100, q=0.3

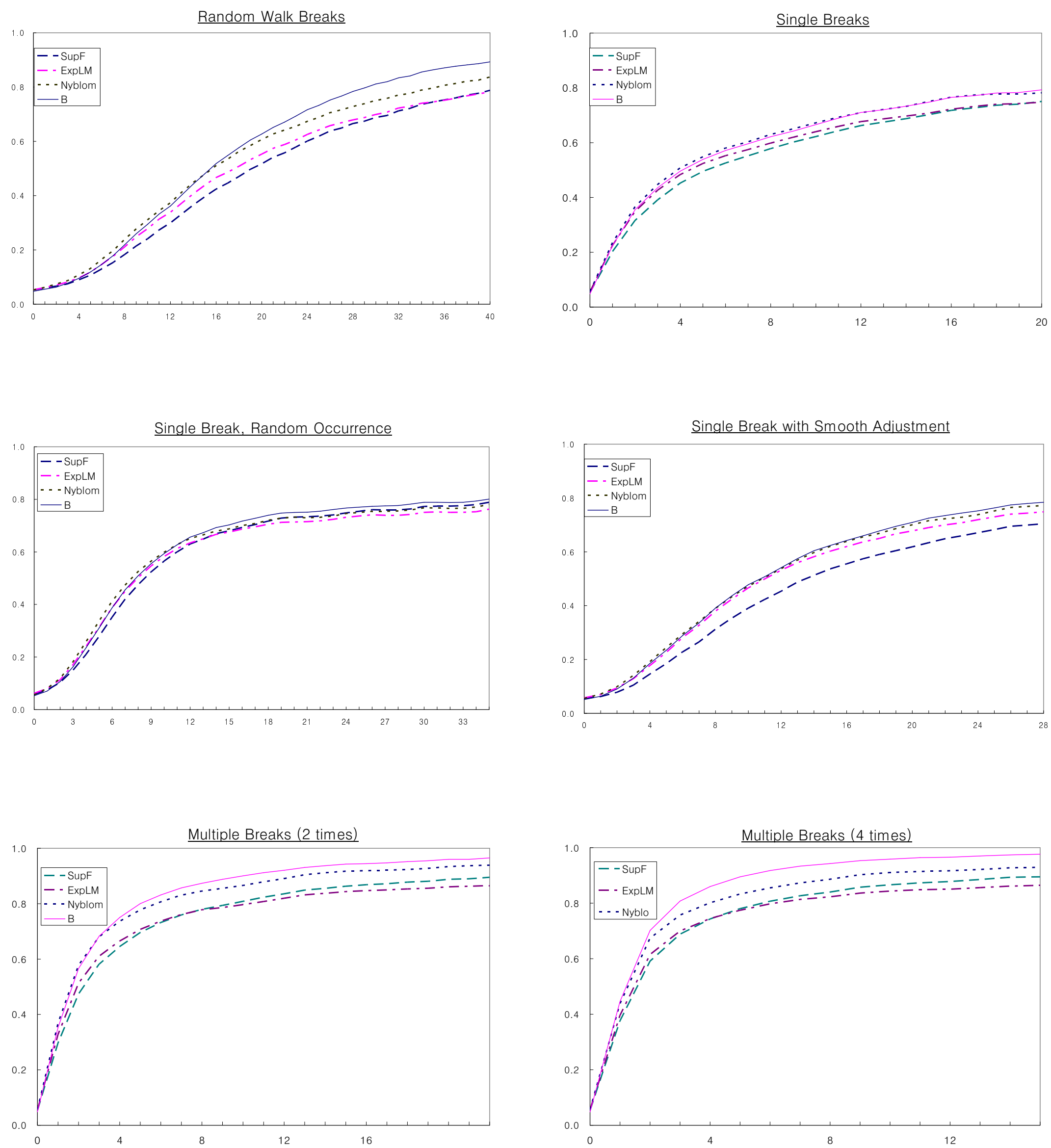

Figure II.3: Small Sample Powers, Quantile Models, k=2, T=100, q=0.3

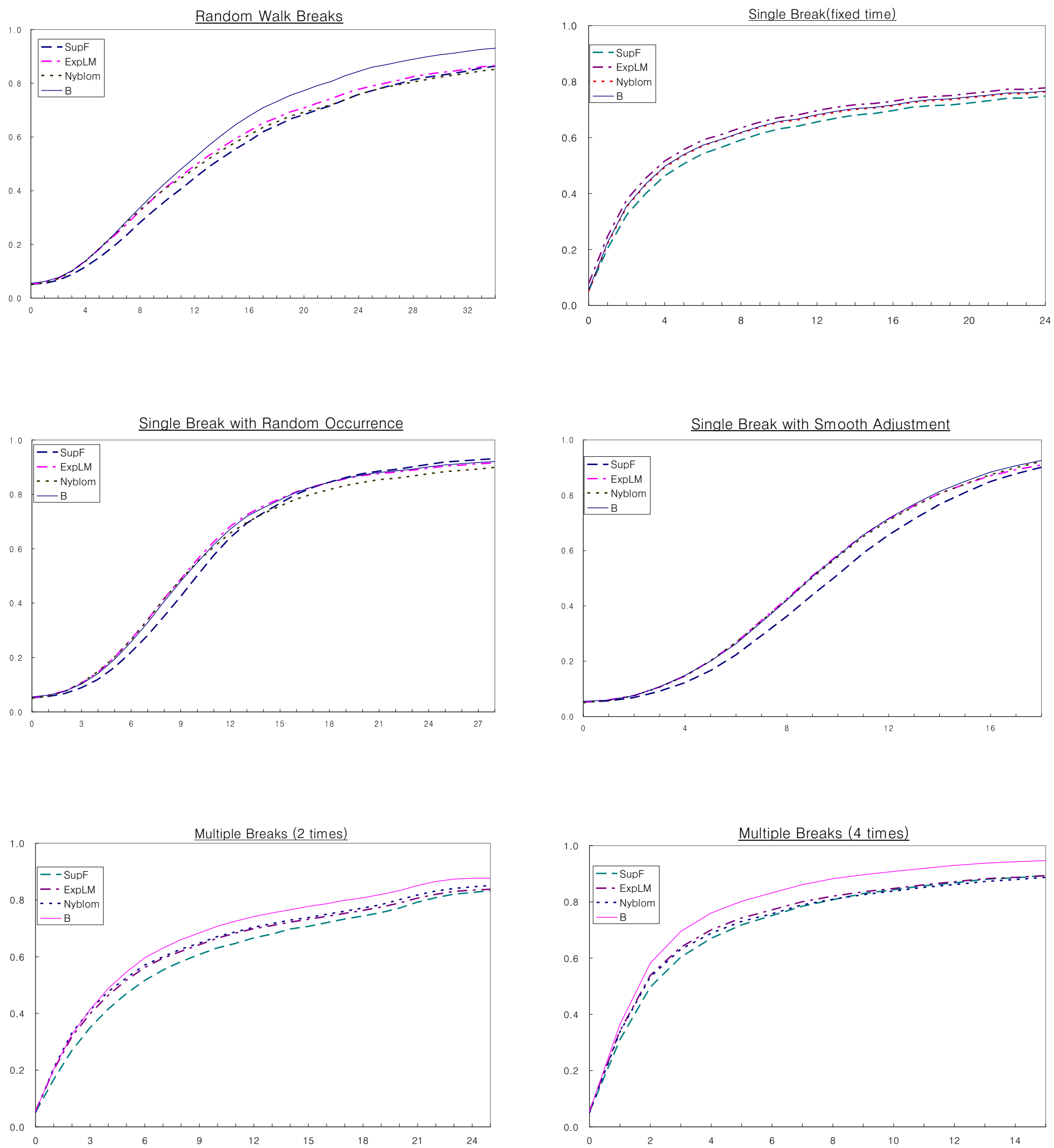

Figure II.4: Small Sample Powers, Breaks in μ in Normal Distribution, T=100

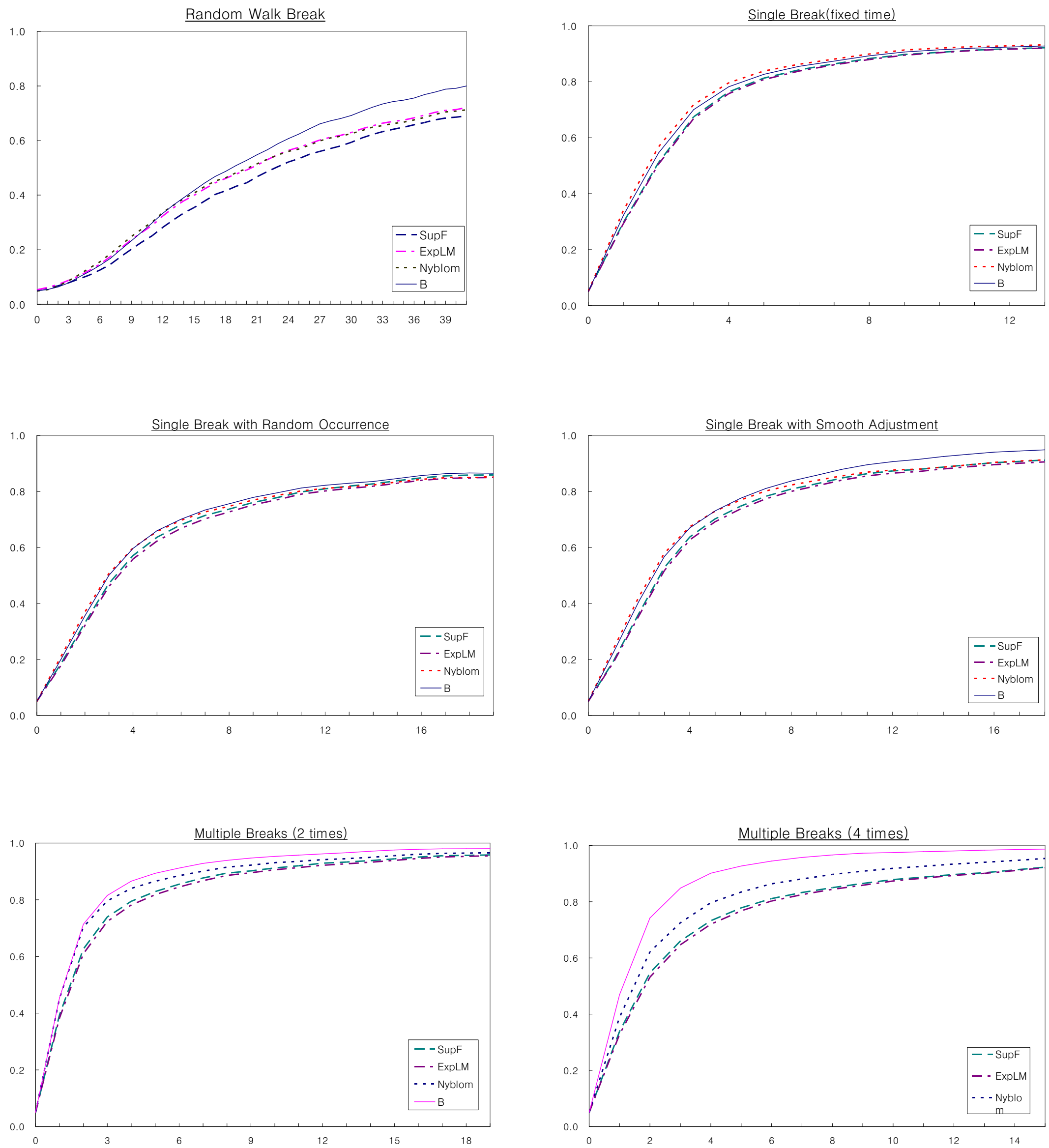

Figure II.5: Small Sample Powers, Breaks in μ and σ^2 in Normal Distribution, T=100

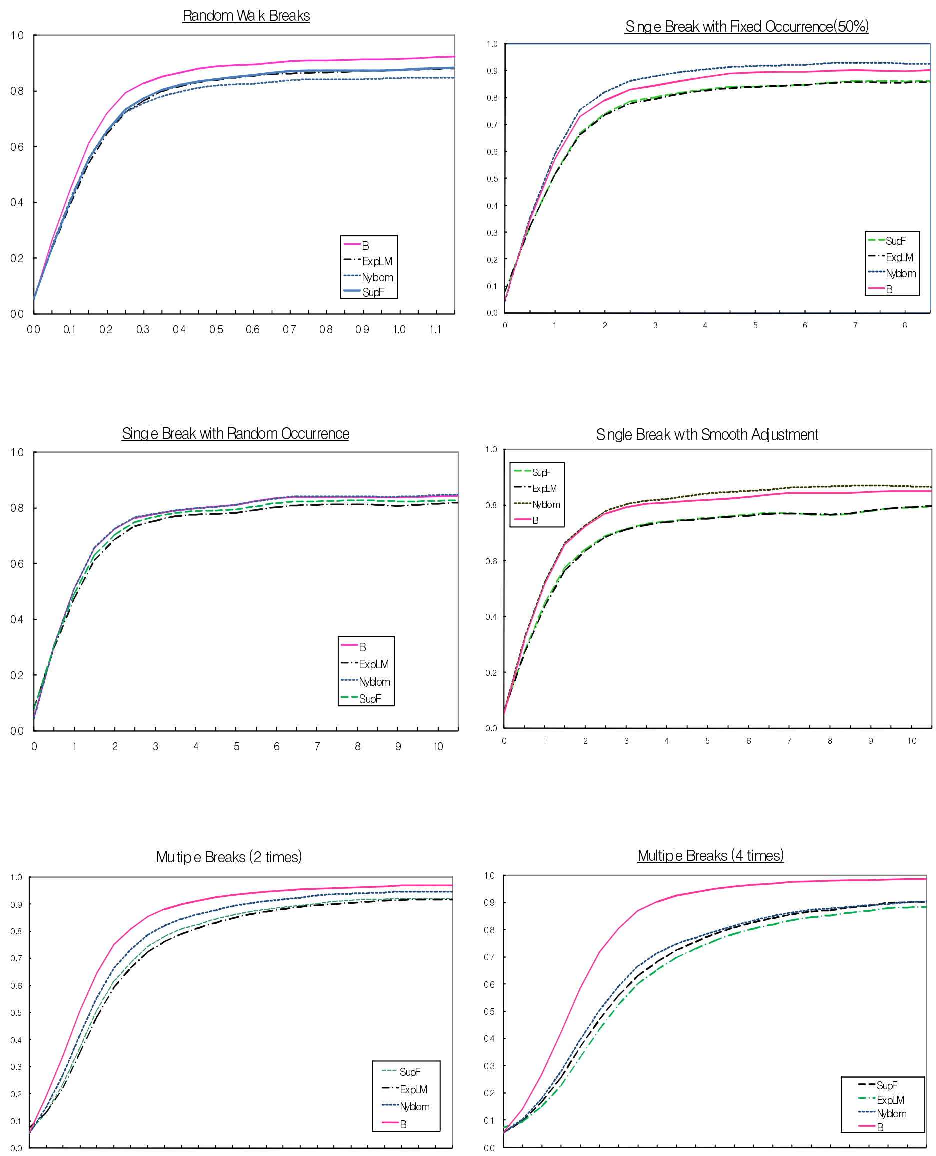

Figure II.6: Small Sample Powers, Logit Model, T=100

Chapter III

Efficient Tests For Parameter Instability in General Models with Unknown Error Distribution

This chapter examines asymptotically efficient tests for parameter instability in general semiparametric models in which the error distribution is unknown but treated as an infinite dimensional nuisance parameter. I first derive the asymptotic power envelope with unknown density and suggest conditions under which a semiparametric model would have the same asymptotic power envelope with known error distribution. The conditions are weak enough to cover a wide range of error distributions by relaxing the twice differentiability and allowing for skewness. An efficient test statistic is then suggested, which is adaptive in the sense that allowing unknown error distribution gives no loss of asymptotic power. This implies that the knowledge of the error distribution is asymptotically irrelevant under mild conditions. Monte Carlo experiments show that the adaptive test has improved small sample powers over the existing tests under various error distributions.

III.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. Recent attention has been paid to obtaining a test that has the best asymptotic power. Andrews and Ploberger (1994) suggest an optimal test for structural breaks which has the asymptotically best average power. Elliott and Müller (2006) provide a test in a linear Gaussian model which is most powerful against a broad set of unstable parameter processes, including both structural breaks and time varying parameters. Chapter 2 generalizes Elliott and Müller (2006)'s test so that it obtains the asymptotic optimality in a wide range of nonlinear non-Gaussian models.

These tests are optimal only when the underlying distribution is known. In many data sets, however, it is more likely that the error distribution is incorrectly specified. In this circumstance, it is to be expected that tests lose validity by mistakenly recognizing outliers. The optimal tests partly 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 test under unknown error distribution.

The purpose of this chapter is to examine the asymptotically efficient tests for parameter instability in a semiparametric set-up in which the fact that the underlying distribution is unknown is explicitly considered. I analyze the test in a single unified framework of the unstable parameter processes which is wide enough to include various nonstationary time varying parameters and permanent structural breaks. Therefore, it will be shown that the test set-up in this chapter is general enough to cover a wide range of unstable parameter processes and error distributions, but requires only modest information about them.

This chapter makes two contributions. First, I derive asymptotic power envelopes for testing general parameter breaking processes in a semiparametric set-up. This asymptotic power bound is sharp in the sense that it is attainable by feasible test statistics. An important finding is that the asymptotic power envelope in a semiparametric model is equivalent to that under known error distribution. The asymptotic equivalency holds even under rather mild conditions that allow for the asymmetry of distributions while most existing work requires symmetry. This equivalency implies that the knowledge of the underlying distribution is asymptotically irrelevant in obtaining an efficient test function. The power envelope does not require the information of the exact parameter breaking process as long as it is in a suggested set. Therefore, this chapter works through the problem of identifying two unobservable random processes in the model, unstable parameters and the error term, by providing conditions under which the attainable power envelope is asymptotically free of their distributional information.

Second, I suggest a test statistic that is asymptotically efficient in the sense that its power converges to the semiparametric power envelope. The test statistic is derived based on the method of adaptation using kernel estimates of the score function. An estimator or a test is adaptive if it has the same asymptotic properties as the one obtained under the assumption that the true distribution is known. Since the seminal work by Bickel (1982), numerous authors have employed adaptation in time series models. Choi et al. (1996) extend this idea to the standard testing problem, to 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 test in semiparametric models. Benghabrit and Hallin (1998), and Hallin and Jurečová (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 testing problem considered in this chapter is nonstandard in the sense that the parameter to be tested 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 a *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 (2006) extends the LAQ to a unit root testing problem. However, the testing problem in this chapter is different from the previous considerations in the sense that the likelihood ratio is not LAQ, but a weighted average of LAQ. This chapter shows that this non-standard testing problem is still amenable to adaptation by using extant semiparametric methods developed for standard problems. In this sense, this chapter provides an example of the extent to which one can obtain adaptive tests in models far from LAN.

This chapter 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 3 to parametric submodels. Section 5 suggests an adaptive test in a semiparametric set-up. Section 6 performs Monte Carlo studies. And Section 7 concludes.

III.2 The Model and the Breaking Processes

This section defines the model and the hypothesis to be tested. 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. $\mathfrak{F}_t \equiv \sigma(Z_1, ..., Z_t)$. Define $F_t(y)$ as the conditional distribution of y_t, and $f_t(y)$ as the corresponding conditional density. Consider the model

$$y_t = m(X_t, \beta_0, \beta_t) + \epsilon_t \tag{III.2.1}$$

where $m(\cdot)$ is continuous and differentiable with respect to β_t. β_t is the $k \times 1$ vector of parameters to be tested and β_0 is the $k \times 1$ vector of nuisance parameters which are constant for all $t = 1, \ldots, T$. ϵ_t is a mean zero error term with a distribution g. The mean zero property is to identify the model and can be replaced by other moment conditions such as quantile restriction, if necessary. The objective of this chapter is to test whether the parameter vector that links the observables X_t to y_t remains stable over time, i.e.

$$\begin{aligned} &H_0 : \beta_t = 0 \qquad && \forall t \\ &H_1 : \beta_t \neq 0 && \text{for some } t > 1 \end{aligned} \tag{III.2.2}$$

so that the parameter vector is β_0 under H_0 and $\beta_0 + \beta_t$ under H_1. To examine asymptotic local powers, the alternative hypothesis is considered to be local to the null by assuming that $\{\beta_t\}$ take the form

$$\beta_t = \frac{1}{T}\delta_t \quad \forall t = 1, \ldots, T$$

Unlike the standard testing problem, the appropriate neighborhood in order for the test to have nontrivial asymptotic power is where β_t is of order T^{-1} in probability. The reason for this is that the test focuses on alternatives with a persistently varying $\{\delta_t\}$, 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.

Note that different specification of the unstable β_t would lead to a different testing problem. For example, the problem is reduced to a structural break test such as Andrews (1993), Andrews and Ploberger (1994), and Bai and Perron (1998) if we regard β_t as fixed and described by a vector of unknown parameters. On the other hand, considering β_t as random variables makes (III.2.2) a test for time varying parameters as in Nyblom and Makeläinen (1983). However, there are few ways to identify in a priori a specific breaking process in one's model. For this reason, an effective test is one that is powerful against a wide range of parameter instabilities. The set-up in this chapter leaves the breaking processes unspecified as long as they are in a set which is broad enough to cover a lot of unstable parameter processes that might happen in the economy. Specifically, I consider unstable processes that satisfy the following condition.

Condition 6 *i) $\{\Delta\delta_t\}$ is uniform mixing with mixing coefficient of size -r/(2r-2) or strong mixing of size -r/(r-2), r>2*

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

iii) The initial value of $\{\delta\}$ satisfies $\delta_0 = -\frac{1}{T}\sum_{t=1}^{T}\delta_t$

iv) $\{\Delta\delta_t\}$ *is globally covariance stationary with nonsingular long-run covariance matrix,* Ω

The basic idea of Condition 6 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 of unstable processes as defined in Condition 6. For example, if we let $\Delta\delta_t$ have a continuous distribution with probability p and equal zero with probability $(1-p)$, then Condition 6 captures a multiple structural break model with $(T \cdot p)$ expected breaks. On the other hand, it is reduced to a random walk parameter model if $\Delta\delta_t$ is *iid* normal.

Admitting both heteroscedasticity and dependency makes Condition 6 capture many possible persistent breaking processes. Heteroscedasticity of $\Delta\delta_t$ allows different types of breaks to occur in a sample period in the sense that breaks caused by different shocks may have different sizes. Heteroscedasticity also covers processes that have fewer breaks in certain periods and more breaks in other periods. Dependency of $\Delta\delta_t$ allows the parameter to smoothly adjust to a new level after a break. This covers the general set of breaking processes that occur frequently. For example, the oil price shock in 1973 did not change the economy at a time, but might have had a lagged effect.

Part (iii) of Condition 6 is necessary to identify the process $\{\delta_t\}$. It implies that the average value of the random 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 the existence of unknown nuisance parameter β_0, in the sense of least favorable parametric submodels. β_0 is generally unknown and should be replaced by an estimator, which causes some loss of power. This condition plays the role of the least favorable direction of the alternative hypothesis, in which the test has the minimal loss by unknown β_0. (See Chapter 2 for details.)

In order to construct the likelihood ratio, we need additional assumptions on the distributions of $\{\epsilon_t\}$ and $\{X_t\}$. The following condition specifies this.

Condition 7 *i)* ϵ_t *is iid with conditional distribution* $g(\epsilon_t|\beta_0, \beta_t)$. ϵ_t *is conditionally independent of* X_t *given* $\mathfrak{F}_{t-1}$. *The error distribution does not depend on* β_t *in the null hypothesis.*

ii) 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* β_0 *and* β_t *for all* $t = 1, \ldots, T$.

iii) 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.*

iv) 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* β_t. $J_m = E[\dot{m}(X_t)\dot{m}(X_t)']$. $T^{-1}\sum_{t=1}^{T} \dot{m}(X_t)\dot{m}(X_t)'$ *is uniformly positive definite.*

Condition 7 implies that the likelihood function for the data is factored into two pieces, one which captures the contribution to the distribution of y_t, $f(y_t|\mathfrak{F}_{t-1}, X_t, \beta_0, \beta_t)$, and depends on (β_0, β_t), and the other which contains conditional distribution of X_t and does not depend on (β_0, β_t), $f_X(X_t|\mathfrak{F}_{t-1})$. In such likelihood functions, $f_X(\cdot)$ need not be known in order for one to construct the test statistics considered here. The *iid* assumption on ϵ_t is crucial in this set-up. 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.

III.3 Asymptotically Optimal Tests in Parametric Models

This section reviews asymptotically efficient tests under the counterfactual assumption that the error distribution, $g(\epsilon|\cdot)$, is known. It would give a benchmark for tests under more realistic distributional assumptions by providing the upper bound of their asymptotic power envelopes. Under Condition 6 and 7, the likelihood function under H_0 is

$$f_0(y, X|\beta_0) = \prod_{t=1}^{T} g(\epsilon_t|\beta_0) f_X(X_t|\Im_{t-1}) \tag{III.3.1}$$

The likelihood function under the alternative hypothesis is

$$f_1(y, X|\beta_0, \beta) = \int \prod_{t=1}^{T} g(\epsilon_t|\beta_0, \beta_t) f_X(X_t|\Im_{t-1}) d\nu_\beta \tag{III.3.2}$$

where $\beta = (\beta_1', \ldots, \beta_T')'$, ν_β is the measure of β. If ν_β is known, the Neymann-Pearson Lemma implies that rejecting H_0 for a large value of the likelihood ratio statistic, defined as

$$LR_T = \int \prod_{t=1}^{T} \frac{g(\epsilon_t|\beta_0, \frac{1}{T}\delta_t)}{g(\epsilon_t|\beta_0)} d\nu_\beta \tag{III.3.3}$$

has the best power against the alternative distribution (III.3.2). The asymptotically efficient test considered in this chapter is based on the local approximation of

(III.3.3). To do this, we need a condition for the differentiability of the likelihood function of the error term. The following condition gives the differentiability assumption and an additional assumption for the asymptotic properties of the score function.

Condition 8 *Let $\xi_t(\cdot|\beta_0, \beta_t)$ be the square root of the error density, $g(\cdot)$. Under H_0,*

i) There exists a $k \times 1$ random vector $\dot{\xi}_t^{\beta}(\cdot|\beta_0, \beta_t)$ such that $\mathbb{E}\|\dot{\xi}_t^{\beta}(\cdot|\beta_0, \beta_t)\|^2 < \infty$, and

$$\mathbb{E}\left(\left[\left(\frac{\xi_t(\cdot|\beta_0, h)}{\xi_t(\cdot|\beta_0, 0)} - 1\right) - h'\frac{\dot{\xi}_t^{\beta}(\cdot|\beta_0, 0)}{\xi_t(\cdot|\beta_0, 0)}\right]^2\right) \to 0 \text{ as } \|h\| \longrightarrow 0, \quad \forall t \leq T \tag{III.3.4}$$

ii)

$$J_\beta(s) = \frac{1}{T}\sum_{t=1}^{[sT]} 4\frac{\dot{\xi}_t^{\beta}(\cdot|\beta_0, 0)\dot{\xi}_t^{\beta}(\cdot||\beta_0, 0)'}{\xi_t(\cdot|\beta_0, 0)^2} \longrightarrow sJ_\beta$$

for some positive definite nonrandom $k \times k$ matrix function J_β and for any $s \in [0, 1]$ and $J_\beta(1)$ is positive definite for all t

If the error density is twice differentiable, $\dot{\xi}_t^{\beta}(\cdot|\beta_0, \beta_t) = \frac{1}{2}\dot{m}_t(X_t)\dot{\ell}_t^g$ where $\dot{\ell}_t^g$ is the first derivative of the log of $g(\cdot)$. Part (i) of Condition 8, called *quadratic mean differentiability (QMD)*, is weak enough to be satisfied by a wide variety of densities and strong enough to deliver the approximation similar to the Taylor expansion. Under *QMD*, Chapter 2 suggests a second order local approximation for Condition 6 random parameter models. Let's define M_e as $M_e = I_T - \frac{1}{T}e'e$ where I_T is a $T \times T$ identity matrix, and e is a $T \times 1$ vector of ones. The following lemma gives the local approximation of the integrand in (III.3.3).

Lemma 6 *Let* $\dot{\ell}^{\beta} = (\dot{\ell}^{\beta}_1(\beta_0), \ldots, \dot{\ell}^{\beta}_T(\beta_0))$ *where* $\dot{\ell}^{\beta}_t(\beta_0) = 2\frac{\dot{\xi}^{\beta}_t(\cdot|\beta_0,0)}{\xi_t(\cdot|\beta_0,0)}$. *Under Condition 6 to 8, the integrand of* (III.3.3)*, denoted as* L_T*, is equivalent to*

$$L_T = (1 + o_p(1))exp\left[\dot{\ell}^{\beta\prime}(M_e \otimes I_k)\beta - \frac{1}{2}\beta'(M_e \otimes J_\beta)\beta\right] \qquad \text{(III.3.5)}$$

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 parameter, and the null and the alternative distribution is contiguous, which is shown in Theorem 1). Using Lemma 5, it can be shown that LR_T is asymptotically equivalent to

$$\widetilde{LR}_T = \int exp\left[\dot{\ell}^{\beta\prime}(M_e \otimes I_k)\beta - \frac{1}{2}\beta'(M_e \otimes J_\beta)\beta\right] d\nu_\beta \qquad \text{(III.3.6)}$$

I suggest an asymptotically efficient test statistic, denoted as $B(\Omega)$, which is asymptotically equivalent to an increasing transformation of $\widetilde{LR}_T$. Let $\Omega^* = J_\beta^{\frac{1}{2}}\Omega J_\beta^{\frac{1}{2}}$. 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 $a_i > 0, \forall i$. The first derivative normalized to have unit variance and zero covariance can be written as $\dot{\ell}^{\beta *}(\beta_0) = (I_T \otimes P'J^{-1/2})\dot{\ell}^{\beta}(\beta_0)$ or $\dot{\ell}^{\beta *}_t(\beta_0) = P'J^{-1/2}\dot{\ell}^{\beta}_t(\beta_0)$. Furthermore, define $\dot{\ell}^{\beta *}_{t,i}$ to be the i^{th} element of $\dot{\ell}^{\beta *}_t(\beta_0)$ and $\zeta^{\beta}_i(\beta_0, J_\beta)$ to be the vector of the partial sum of $\dot{\ell}^{\beta *}_{t,i}$, i.e. j^{th} element of ζ^{β}_i to be $\sum_{t=1}^{j}\dot{\ell}^{\beta *}_{t,i}$. The test statistic I suggest is

$$B(\Omega) = \sum_{i=1}^{k} \zeta_i^{\beta\prime}\left[\frac{T^2}{a_i^2}I_T + FM_eF'\right]^{-1}\zeta_i^{\beta} \qquad \text{(III.3.7)}$$

where F is a $T \times T$ lower triangular matrix in which all the nonzero elements are ones. By matrix manipulation it can be shown that $B(\Omega) = \frac{1}{2}\ln\widetilde{LR}_T +$

constant, if the distribution of β is normal with zero mean and variance $\frac{1}{T^2}FF'\otimes\Omega$. Consequently, $B(\Omega)$ is proved to be asymptotically efficient if $\widetilde{LR}_T$ with any specific ν_β for Condition 6 parameter process converges in probability to the same test function under both the null and the alternative hypotheses.

Let $\phi_T(Z|\Omega)$ be a critical function for testing breaking processes. That is, $\phi_T(Z|\Omega)$ is a $[0,1]$ valued function determined by Z. I consider asymptotically α-significant tests, i.e. $\lim_{T\to\infty}\int\phi_T(Z|\Omega)f_0(Z|\beta_0)dZ = \alpha$. The power function of the test is defined as $\int\phi_T(Z|\Omega)f_1(Z|\beta_0)dZ$. The following theorem implies that the test function $B(\Omega)$ provides the asymptotic power envelope under known error distribution.

Theorem 5 *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. $\Psi(\Omega) = \underline{\lim}\int\int\psi_T(Z|\Omega)f(Z|\beta_0,\beta)d\nu_\beta dZ$. Suppose the error distribution $g(\epsilon_t|\beta_0,\beta_t)$ is known. Under Conditions 6 to 8, the test $B(\Omega)$ satisfies*

$$\overline{\lim}_{T\to\infty}\int\int\phi_T(Z|\Omega)f_1(Z|\beta_0,\beta)d\nu_\beta dZ \le \Psi(\Omega)$$

The test function $B(\Omega)$, however, is not feasible because it is a function of unknown nuisance parameters β_0 and J_β. They should be replaced by their maximum likelihood estimators in order to make $B(\Omega)$ feasible. When constructing efficient tests, we need the following.

Condition 9 *Under H_0, the likelihood function satisfies.*

$$T^{-1/2}\sum_{t=1}^{[sT]}\dot{\ell}_t^\beta(\beta_0+T^{-1/2}h) = T^{-1/2}\sum_{t=1}^{[sT]}\dot{\ell}_t^\beta(\beta_0) - sK(\beta_0)h + o_p(1) \qquad \text{(III.3.8)}$$

Condition 9 is similar to the notion of *regular score* in the sense of Hall and Mathiason (1990) and weaker than Nyblom (1989). This condition is generally satisfied when an asymptotic normal MLE does exist, and can be extended to other cases. The following lemma indicates that the plug-in version of $B(\Omega)$, denoted as $B(\Omega, \hat{\beta}, \hat{J}_\beta)$ attains the asymptotic power envelope.

Lemma 7 *Suppose there exist a $\sqrt{T}$-consistent estimator $\hat{\beta}$ of β_0 and a consistent estimator $\hat{J}_\beta$ of J_β. Under Condition 6 to 9, the following holds both under H_0 and H_1*

$$B(\Omega, \hat{\beta}, \hat{J}_\beta) = B(\Omega) + o_p(1)$$

Lemma 5 in Chapter 2 shows that the asymptotic null distribution of $B(\Omega)$ is

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

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

III.4 Asymptotically Optimal Tests in Parametric Submodels

The optimal test considered in the previous section assumes that the error distribution is correctly specified, which is generally infeasible in practice. Sections 4 and 5 extend the previous results by investigating asymptotically efficient tests under unknown error distribution. This relaxation modifies the model in the previous section into the semiparametric one with a real valued parametric component $\theta = (\beta_0', \beta_1', \ldots, \beta_T')' \in R^{k(T+1)}$, and a single nonparametric component $g \in \mathcal{G}$ which denotes the unknown distribution of the error term, where $\mathcal{G}$ is a specified set of density functions.

In this section, I assume that the error density is known to belong to a specific parametric family of distribution indexed by finite dimensional parameters. A familiar case is testing partial structural breaks in which only the part of the parameters are suspected to have structural breaks while the others remain constant. Another case occurs when testing stability of the coefficient of a linear regression model in which, instead of standard Gaussian error term, the error term is *iid* from a more generalized distribution, such as an asymmetric exponential family (Fernandez and Steel (1998)) with unknown skewness and kurtosis parameters.

The true set of conditional densities of y_t is characterized as a parametric family $\mathcal{P}_\eta = \{F_t(y|\eta) : \eta \in R^q\}$ with dominating measure μ and corresponding densities $f_t(y|\eta) = dF_t(y|\eta)/dy$ such that $g(\cdot) = f_t(y|\eta)$. The model with this parametrization $\mathcal{P} = \{P_{\theta,\eta} : \theta \in R^{k(T+1)}, \eta \in R^s\}$ is called a parametric submodel. For the convenience, I consider a parametric submodel with a single unknown nuisance parameter $\eta \in \mathcal{R}$. The extension to the finite dimensional case is straightforward.

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 the local parameter space $\mathcal{H}_\theta$ 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 the tests that have the invariant asymptotic size regardless of h. This type of tests, which is called an *asymptotically similar test*, is defined as below.

Definition 6 *Let $\phi_T(Z)$ be the test function for the breaking processes and $f_0(Z|h)$ be the null distribution of Z given h. The test function $\phi_T(Z)$ is asymptotically similar at η_0 if for a fixed $\alpha > 0$,*

$$\lim_{T\to\infty} \int \phi_T(Z) f_0(Z|h) dZ \leq \alpha \qquad \textit{for every } h \tag{III.4.1}$$

Note that the asymptotic size restriction is imposed for every value of 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 in the likelihood ratio of the model. The likelihood ratio function associated with $\mathcal{P}_\eta$ is written as

$$LR_T^S = \int \prod_{t=1}^{T} \frac{g(\epsilon_t|\beta_0 + \beta_t, \eta_0 + \frac{1}{\sqrt{T}}h)}{g(\epsilon_t|\beta_0, \eta_0)} d\nu_\beta \tag{III.4.2}$$

Analogous to the parametric model case, we need a differentiability condition for the density $f(\cdot|\eta)$ in order to get the LAQ of the integrand in LR_T^S. The following condition is the modified version of the QMD in Condition 8.

Condition 10 *Let* $\xi_t^S(\cdot|\beta_t,\eta)$ *be the square root of the error density,* $g(\epsilon_t|\beta_t,\eta)$ *and* b *be the* $(k+1)\times 1$ *vector. Define* $\theta_t^\eta = (\beta_t',\eta)'$*, and* $\theta_0^\eta = (0',\eta_0)'$*. Under* H_0*,*

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

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

ii)

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

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

$\{\dot{\xi}_t^S(\cdot,\theta_0^\eta)\}$ is still a function of β_0 but I suppress the dependency for the purpose of convenience. Lemma 8 gives locally asymptotic quadratic approximation of the integrand in LR_T^S.

Lemma 8 *Let's define* $\dot{\ell}_t^\eta = 2\frac{\dot{\xi}_t^\eta(\theta_0^\eta)}{\xi_t^S(\theta_0^\eta)}$*, and* $J_\eta = 4E\left[\left(\frac{\dot{\xi}_t^\eta(\theta_0^\eta)}{\xi_t^S(\theta_0^\eta)}\right)^2\right]$*. Under Condition 6, 7, and 10, the integrand of* (III.4.2)*, denoted as* L_T^S*, is equivalent to*

$$L_T^S = (1+o_p(1))\exp\left[\dot{\ell}^{\beta'}(M_e\otimes I_k)\beta - \frac{1}{2}\beta'(M_e\otimes J_\beta)\beta\right]\cdot\exp\left[\frac{h}{\sqrt{T}}\sum_{t=1}^{T}\dot{\ell}_t^\eta - \frac{h^2}{2}J_\eta\right] \quad \text{(III.4.3)}$$

Now using (III.4.3), it can be shown that the likelihood ratio function is asymptotically equivalent to

$$\widetilde{LR}_T^S = \int exp\left[\dot{\ell}^{\beta\prime}(M_e \otimes I_k)\beta - \frac{1}{2}\beta'(M_e \otimes J_\beta)\beta\right] d\nu_\beta \cdot \exp\left[\frac{h}{\sqrt{T}}\sum_{t=1}^{T}\dot{\ell}_t^\eta - \frac{h^2}{2}J_\eta\right] \tag{III.4.4}$$

Note that the integral part in (III.4.4) is the same as the likelihood ratio function in the parametric model, except the first derivative $\dot{\ell}^\beta$ depends on the value of the nuisance parameter 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.

I use the method of limits of experiments to derive the asymptotic power envelope. An experiment can be regarded as a synonym of a probability model. The 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. Using the results in the previous section, and functional central limit theorem, the asymptotic null distribution of $log(LR_T^S)$ is

$$log(LR_T^S) \to^d \Lambda^S(\Omega, h) = \Gamma + \Lambda(\Omega) + hW^\eta(1) - \frac{h^2}{2}J_\eta \tag{III.4.5}$$

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

The best power can be achieved if the nuisance parameter has the true value, i.e. $h = 0$, so that the power envelope can be calculated by maximizing $E\left[\phi(Z)exp\left(\Lambda^S(\Omega, 0)\right)\right]$. This type of power envelope, however, does not give the practical implication in the sense that the power under $h = 0$ does not correctly reflect the problem of unknown h. It is therefore sensible to derive the power envelope under certain nonzero h by maximizing $E\left[\phi(Z)exp\left(\Lambda^S(\Omega, h)\right)\right]$. This power envelope is generally less than that of the previous section, because the latter does not achieve true error distribution. However, Theorem 7 below shows the interesting result that both power envelopes are identical in this set-up.

The intuition is as follows; the LR function (III.4.4) and its asymptotic counterpart (III.4.5) are factored into two parts, containing the measure of the parameter of interest, ν_β and the perturbation of the nuisance parameter, h. The power of the test with respect to the size of the break is determined only by the first part, and the asymptotic size restriction is imposed only to the second part. Therefore, the test based on the first part is expected to provide the power envelope, while it avoids the size dependency of unknown h. Since the first part $\Lambda(\Omega)$ is equivalent to the best limit test function in parametric models, it is possible to construct a test based on $\Lambda(\Omega)$, that has the same asymptotic power as the power envelope under known error distribution. Let's define the limit power function Ψ^S as

$$\Psi^S(\Omega) = E\left[1_{\{\Lambda(\Omega) > k_h^\alpha\}} exp\left(\Lambda^S(\Omega, h)\right)\right] \qquad \text{(III.4.6)}$$

where k_h^α is the continuous function that ensures the test function has asymptotic size-α. The following theorem proves the argument.

Theorem 7 *Under Conditions 6, 7, and 10, any asymptotic similar test function $\phi_T(Z|\Omega, \eta)$ associated with $\mathcal{P}_\eta$ satisfies*

$$\overline{\lim_{T\to\infty}} \int\int \phi_T(Z|\Omega,\eta) f_1(Z|\theta,\eta) dZ d\nu_\delta \leq \Psi^S(\Omega) = \Psi(\Omega) \qquad \text{(III.4.7)}$$

where $\Psi(\Omega)$ is the asymptotic power envelope in a parametric model defined in Theorem 5.

Theorem 7 implies that it is possible not to lose any power even though we do not know the true value of the nuisance parameter η_0, as the sample size gets large. The main reason for this is because the alternative process, $\beta_t - \bar{\beta}$ 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 the score of the parameter of interest $\dot{\ell}_t^\beta$ only through its effective score function, which is defined as

$$\dot{\ell}_t^{\beta e} = \dot{\ell}_t^\beta - J_{\beta\eta} J_\eta^{-1} \sum_{i=1}^T \dot{\ell}_i^\eta$$

where $J_{\beta\eta} = E[\dot{\ell}_t^\beta \dot{\ell}_t^\eta]$. The effective score function lies on the orthonormal complement of the space spanned by the score of the nuisance parameter, so that $\sum \dot{\ell}_t^{\beta e}$ and $\sum \dot{\ell}_t^\eta$ are asymptotically independent. The likelihood ratio in (III.4.5) is a function of $\dot{\ell}_t^\beta$ through $\dot{\ell}_t^\beta - \sum_{i=1}^T \dot{\ell}_i^\beta$. Subtracting $J_{\beta\eta} J_\eta^{-1} \sum_{i=1}^T \dot{\ell}_i^\eta$ from the first term and adding it to the second term gives that

$$\dot{\ell}_t^{\beta e} - \sum_{i=1}^T \dot{\ell}_i^{\beta e} = \dot{\ell}_t^\beta - \sum_{i=1}^T \dot{\ell}_i^\beta$$

which implies that the test is locally invariant to η. Therefore, not knowing η_0 does not give any loss of asymptotic power. The intuition is similar to Stein's necessary condition for adaptation which is that $J_{\beta\eta}$ is zero. Under this condition, the effective score is always equivalent to the actual score so that the invariance property always holds. The set-up in this section does not 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, $\beta_t - \bar{\beta}$.

The asymptotic power envelope suggested in Theorem 7 is achievable in practice if we have a $\sqrt{T}$-consistent estimator of η_0. Let $B^S(\Omega)$ be the small sample counterpart of $\Lambda(\Omega)$, i.e. $B^S(\Omega)$ is the same as $B(\Omega)$ in (III.3.7) except the first derivative of the log likelihood function with respect to β depends also on the true nuisance parameter η_0, and let $B^S(\Omega, \hat{\eta}_0)$ be the plug-in version of $B^S(\Omega)$. $B^S(\Omega)$ achieves the asymptotic power envelope in Theorem 7 because it is a finite sample counterpart of $\Lambda(\Omega)$. Therefore, it suffices to show that the feasible test statistic $B^S(\Omega, \hat{\eta}_0)$ converges in probability to $B^S(\Omega)$ under both H_0 and H_1. Lemma 9 below proves the argument.

Lemma 9 *Suppose there exist $\sqrt{T}$-consistent estimators $\hat{\eta}$ and $\hat{\beta}$, and a consistent estimator $\hat{J}_\beta$. Assume that $\dot{\ell}_t^\beta(\eta)$ satisfies condition 4) for both η_0 and β_0. Under Conditions 6, 7, and 10*

$$|B^S(\Omega; \hat{\eta}) - B^S(\Omega)| \longrightarrow 0 \quad \textit{in probability under } H_0 \textit{ and } H_1 \tag{III.4.8}$$

An important implication of Lemma 9 is that it is better to use an error distribution which is more general than normal. Note that the asymptotic power function of $B(\Omega)$ is an increasing function of $\Omega^* = J_\beta^{\frac{1}{2}}\Omega J_\beta^{\frac{1}{2}}$ which is proportional to the Fisher information of the error distribution. Accordingly, the power envelope is strictly increasing in the Fisher information. Suppose that the true error density is in a generalized exponential family, i.e. the error density has the form as,

$$g(\epsilon_t) = A(\eta)\exp\left[B(\eta)|\epsilon_t|^\eta\right] \tag{III.4.9}$$

where $\eta > 1/2$ and $A(\eta)$ and $B(\eta)$ are decided to satisfy $\int_{-\infty}^{\infty} g(\epsilon)d\epsilon = 1.$. Normal density is a special case of (III.4.9) when $\eta = 2$. The fisher information of this type of density ranges $[1, \infty]$ where it is one when $g(\epsilon_t)$ is normal and ∞ when $\eta = 1/2$. Therefore, the Fisher information has the minimal if we use normal distribution and would be increased if we use any other $g(\epsilon_t)$ than normal. Consequently, any non-Gaussian density in (III.4.9) would have a higher asymptotic power envelope than normal. Since the asymptotic power envelope is attainable with $\sqrt{T}$-consistent estimator of η, we may get a significant power gains by using (III.4.9) rather than normal density whenever $\eta \neq 2$. Figure III.1 presents asymptotic power envelopes for various value of the Fisher information in (III.4.9), where the bottom line represents Gaussian case. It shows a large increase in power, which justifies the use of a test with non-Gaussian error density.

III.5 Asymptotically Optimal Tests in Semiparametric Models

The previous section investigates an optimal test under which finite numbers of nuisance parameters in the error distribution g 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 the error distribution g is entirely unknown. Rather than allowing for the unknown error distribution to be fully nonparametric, I give a mild restriction that the error distribution is parameterized by an infinite dimensional unknown nuisance parameter. 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 can be defined to be the infimum of the power envelope of all submodels. The previous section shows that every parametric submodel has the same asymptotic power envelope, $\Psi^S(\Omega)$, that is equivalent to that under known error distribution. It implies that the power envelope of the semiparametric models would also be equivalent to $\Psi^S(\Omega)$. Unlike the previous section, however, the $\sqrt{T}$-consistent estimator for the infinite dimensional nuisance parameter is generally not available. Hence, the plug-in version of the efficient test $B^{PS}(\Omega;\hat{\eta})$ is inappropriate in this set-up.

This problem is similar to that of adaptive estimation, which is originally proposed by Bickel (1982). The adaptive estimator is defined as the estimator constructed without knowledge of g but is asymptotically as efficient as any well-behaved estimator that relies on knowledge of g. This idea has been extended to a standard testing problem to show that the definition and the method of adaptive estimation can be directly employed in a standard testing problem; if a model satisfies the condition for the adaptive estimation and thereby the adaptive estimator exists, we can construct asymptotically efficient test statistics.

This duality between estimation and test holds only when the model satisfies LAN approximation. However, LAN is not available in out set-up, and there is no parameter of interest to be estimated, so that the duality is not applicable. It has not been considered whether the duality holds in a more general model where LAN condition is not satisfied. Instead, Jansson (2006) suggests working with a notion of adaptation that depends only on the model under consideration and makes no reference to any other particular type of inference. Accordingly, we get similar inference to that of the adaptation in this testing problem, by looking back to the likelihood ratio in a parametric model in (III.3.5). The purpose is to find a feasible test statistic $B^*(\Omega)$ which converges in probability to $B(\Omega)$ both under the null and the alternative hypothesis. Based on (III.3.5), it implies that there exist estimators $\{\hat{\dot{\ell}}_t^\beta\}$ and $\hat{J}_\beta$ which satisfy

$$\begin{aligned} \sum_{t=1}^{T}(\beta_t - \frac{1}{T}\sum_{i=1}^{T}\beta_i)\hat{\dot{\ell}}_t^\beta &= \sum_{t=1}^{T}(\beta_t - \frac{1}{T}\sum_{i=1}^{T}\beta_i)\dot{\ell}_t^\beta + o_p(1) \\ \hat{J}_\beta &= J_\beta + o_p(1) \end{aligned} \qquad \text{(III.5.1)}$$

for all $\{\beta_t\}$ in Condition 6. The objective of this section is to show the existence of the estimators that satisfy (III.5.1), and to demonstrate that it provides the existence of an efficient test function. A possible construction of the efficient estimator is to use a kernel estimation method. Using data and the consistent estimator of β_0, compute the residuals $\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_T$ with $\tilde{\epsilon}_t = \epsilon(y_1, \ldots, y_t, X_1, \ldots, X_T, \widehat{\beta})$ for $t = 1, \ldots, T$. A kernel density estimator is defined as for all e in a small neighborhood of each value of $\tilde{\epsilon}_t$

$$\hat{f}_T(e;\tilde{\epsilon}_1,\ldots,\tilde{\epsilon}_T) = \frac{1}{(T-1)a_T}\sum_{i\neq t} k\left(\frac{e-\tilde{\epsilon}_i}{a_T}\right) \quad \text{(III.5.2)}$$

$$\hat{f}'_T(e;\tilde{\epsilon}_1,\ldots,\tilde{\epsilon}_T) = \frac{1}{(T-1)a_T^2}\sum_{i\neq t} k'\left(\frac{e-\tilde{\epsilon}_i}{a_T}\right) \quad \text{(III.5.3)}$$

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 $\int z^2 k(z)dz < \infty$. (See Schick (1993).) The score estimator is defined as

$$\hat{\dot{\ell}}_t^{\beta}(\tilde{\epsilon}_t;\tilde{\epsilon}_1,\ldots,\tilde{\epsilon}_T) = \frac{\hat{f}'_T(\tilde{\epsilon}_t;\tilde{\epsilon}_1,\ldots,\tilde{\epsilon}_T)}{b_T + \hat{f}_T(\tilde{\epsilon}_t;\tilde{\epsilon}_1,\ldots,\tilde{\epsilon}_T)} \quad \text{(III.5.4)}$$

$$\hat{J}_\beta = \frac{1}{T}\sum_{t=1}^{T} \hat{\dot{\ell}}_t^{\beta}(\tilde{\epsilon}_t)\hat{\dot{\ell}}_t^{\beta}(\tilde{\epsilon}_t)' \quad \text{(III.5.5)}$$

where $\{b_T\}$ is a sequence of constants such that $(Ta_T^3 b_T)^{-1} \to 0$. Note that $\{\hat{\dot{\ell}}_t^{\beta}\}$ uses the entire sample data. Most existing research splits the sample period and uses only the observations in one sample period to estimate $\{\hat{\dot{\ell}}_t^{\beta}\}$ of the other split sample. They use the method 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 thereby to produce a better power. Schick (1987) suggests a general condition to use the whole data. Koul and Schick (1997) use all data in adaptively estimating nonlinear time series models under additional conditions on the boundeness of $\dot{m}(\cdot)$ and the memory property of $\{X_T\}$. The method in this section is generally similar to them, and Conditions 6 and 7 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 $\psi_T(Z|\Omega) = 1_{[B^* > k_\alpha]}$ where k_α is the continuous function satisfying $E_0[\psi_T(Z|\Omega)] = \alpha$ and $B^*(\Omega)$ as

$$B^*(\Omega) = \sum_{i=1}^{k} \hat{\zeta}_i' \left[\frac{a_i^2}{T^2} I_T - F M_e F' \right]^{-1} \hat{\zeta}_i \qquad \text{(III.5.6)}$$

where $\hat{\zeta}_i = (\hat{\zeta}_{i,1}, \ldots, \hat{\zeta}_{i,T})'$, $\hat{\zeta}_{i,j} = \sum_{t=1}^{j} \hat{\dot{\ell}}_{t,i}^{\beta *}$, and $\hat{\dot{\ell}}_{t,i}^{\beta *}$ is the i_{th} element of $\hat{\dot{\ell}}_t^{*\beta}$. Let $\Psi^*(\Omega)$ be the asymptotic power function of $B^*(\Omega)$ i.e.

$$\Psi^*(\omega) = \underline{\lim}_{T\to\infty} \int \int \psi_T(Z|\Omega) f_1(Z|\eta) dZ d\nu_\delta$$

The following theorem shows that we can construct an asymptotically efficient test based on (III.5.4) and (III.5.5), without further strict conditions.

Theorem 8 *Under Condition 6 to 9, any asymptotically similar test $\phi(Z|\Omega)$ associated with $\mathcal{S}$ satisfies*

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

where $\Psi(\Omega)$ is the asymptotic power envelope in a parametric model defined in Theorem 5

Theorem 8 indicates that the asymptotic power function based on $B^*(\Omega)$ provides the asymptotic power envelope in a semiparametric model, and $B^*(\Omega)$ is adaptive in the sense that its asymptotic power function attains the asymptotic power envelope when the error distribution is known. This property provides the main argument of this chapter: The knowledge of the error distribution is asymptotically irrelevant for conducting an optimal test under mild conditions suggested in this chapter. As sample size gets larger, there's little loss of power by not using the test based on the correctly specified error distribution. Note that I do not restrict the set of error distribution to a symmetric case, while most research including Jansson (2006) requires the symmetry. The proposed algorithm to construct an asymptotically efficient test statistic is as follows.

Step 1) Estimate β_0 under H_0 denoted as $\hat{\beta}$. Any method of the estimation such as QMLE, M-estimation, and GLS is possible as long as $\hat{\beta}$ is $\sqrt{T}-$consistent. Calculate the residuals $\tilde{\epsilon}_t = y_t - m(X_t, \hat{\beta})$

Step 2) Estimate the error density and its derivatives by using (III.5.3). Bandwidth a_T can be chosen by the optimal window width method. Calculate the estimates of $\hat{\dot{\ell}}_t^{\beta}$ and $\hat{J}_\beta$ and thereby $\hat{\dot{\ell}}_t^{\beta *} = \hat{J}_\beta^{-\frac{1}{2}} \hat{\dot{\ell}}_t^{\beta}$:

$$\hat{\dot{\ell}}_t^{\beta} = \dot{m}(X_t, \hat{\beta}_0) \frac{\hat{f}_T'(e, \tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_T)}{b_T + \hat{f}_T(e, \tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_T)} \tag{III.5.7}$$

$$\hat{J}_\beta = \frac{1}{T} \sum_{t=1}^{T} \dot{m}(X_t, \hat{\beta}_0) \dot{m}(X_t, \hat{\beta}_0)' \left(\frac{\hat{f}_T'(e, \tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_T)}{b_T + \hat{f}_T(e, \tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_T)} \right)^2 \tag{III.5.8}$$

where b_T can be chosen to be small, but large enough to eliminate the technical difficulty caused by a too small denominator in (III.5.7). Denote i^{th} elements of $\{\hat{\dot{\ell}}_t^{\beta *}\}$ by $\{\hat{\dot{\ell}}_{t,i}^{\beta *}\}$, $i = 1, \ldots, k$.

Step 3) For each $\{\hat{\dot{\ell}}_{t,i}^{\beta *}\}$, generate a new variable, $\hat{w}_{t,i} = r\hat{w}_{t-1,i}^{*} + \Delta \hat{\dot{\ell}}_{t,i}^{\beta *}$ and $\hat{w}_{t,1} = \hat{\dot{\ell}}_{t,1}^{\beta *}$.

Step 4) Regress $\{\hat{w}_{t,i}\}$ on $\{r_{a_i}^t\}$ for each i to get the sum of squared residuals where $r_a = 1 - aT^{-1}$. Sum all of those over $i = 1, \ldots, k$.

Step 5) Multiply this sum by r, and subtract it from $\sum_{i=1}^{k}\sum_{t=1}^{T}\hat{\ell}_{t,i}^{\hat{\beta}*}$.

III.6 Comparative Simulation Study

This section examines the performance of the asymptotically efficient test $B^*(\Omega)$ in finite samples through Monte Carlo experiments. I consider the simple linear regression model as below.

$$y_t = X_t'(\beta_0 + \beta_t) + \epsilon_t \quad t = 1, \ldots, T \tag{III.6.1}$$

where y_t is a scalar, X_t, β_0 and β_t are $k \times 1$ vectors, $\{y_t, X_t\}$ are observed, β_0, β_t are unknown, and $\{X_t\}$ are assumed to be exogenous and satisfy Condition 7 with $E[X_t X_t'] = \Sigma_X$. ϵ_t is *iid* from a unknown distribution but satisfies Conditions 7, and 10, that is the error distribution is independent of β_t and differentiable in quadratic mean. Therefore, (III.6.1) satisfies conditions in this chapter and $B^*(\Omega)$ can be used as an asymptotically efficient test statistic.

β_0 can be estimated simply by OLS which is $\sqrt{T}$-consistent in this set-up. 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 $0.001 \times a^{1/3}$.

I perform the Monte Carlo simulation to calculate the empirical sizes and the powers of the test under various error distributions. Five different error distributions are designed, which are listed below.

A1) Standard Normal Distribution

A2) Symmetric Laplace Distribution

A3) Asymmetric Laplace Distribution with skewed parameter= 0.2

A4) Student t-distribution with $\nu = 4$ degree of freedom

A5) Mixture of two standard Normal distributions with mean 2, and -2, respectively

The small sample sizes and powers are compared with those of $SupF$ test, Andrews and Ploberger (1994)'s test ($ExpLM$), Nyblom (1989)'s test (Nyb), and the test in Chapter 2 ($B(\Omega)$) . $SupF$ and $ExpLM$ are designed for single structural break processes. Nyb considers martingale processes which include a single break with random occurrence and the random walk process. $B(\Omega)$ considers the same breaking processes with this chapter, but assumes the error distribution is known. $B(\Omega)$ is reduced to Elliott and Müller (2006)'s test if the error distribution is normal. I set up these tests based on Gaussian error distribution, because it is most widely applied. Therefore, these tests might have the best powers in A1 but lose some powers under other distributions. Following Andrews et al. (1996) and Bai and Perron (1998), I choose a 5% trimming for $SupF$ test and 2% trimming for $ExpLM$ test. $B^*(\Omega)$ and $B(\Omega)$ are not feasible because Ω is generally unknown. I choose a specific Ω as $\Omega^* = 100 \times I$ followed by Chapter 2, and Elliott and Müller (2006). Hence, the tests are point optimal and there might be some loss of power when the true Ω is not Ω^*.

I consider the simple regression model with univariate X_t with a constant term where $\{X_t\}$ are generated from the AR(1) model with *iid* Gaussian error. I consider 30 combinations of 3 different critical levels (1%, 5%, and 10%), 2 sample sizes (100, and 200), and 5 error distributions to compute the empirical sizes. In calculating small sample powers, four different types of breaking processes are considered: single break, multiple breaks (2 and 4 times) and random walk parameters. Five thousand replications are generated for each distribution and sample size.

Table III.7 shows the experimental result of the empirical sizes. The small sample sizes performance of $B^*(\Omega)$ is fairly good in distribution A1-A4. However, it has size distortions when the error distribution is bimodal. For example, it has a size of 10%, while the actual size is 5% when $T = 100$. The gap becomes moderated as sample size gets larger, but still not negligible when T is increased to 200. The degree of the size distortion depends on the choice of bandwidth, a_T. It has empirical sizes close to the actual ones if a_T is chosen to be small so that the estimated density is smooth not to clearly identify bimodality. However, it costs a loss of small sample power. Therefore, the problem of the efficient choice of bandwidth is still in question. Other tests have good size properties in all distributions, except the Exp-LM test. The asymptotic efficient test function $B^*(\Omega)$ has little gain in size performance in finite samples.

The selected results of the simulated small sample powers are shown in figures III.2 to III.5. The powers of all six tests are close to each other when the error distribution is unimodal and symmetric. Figure III.2 shows that $B^*(\Omega^*)$ has similar powers to the others even when they correctly identify the error distribution as Gaussian. It implies that $B^*(\Omega)$ is little outperformed by the existing tests based on Gaussian distribution, even in the worst case. Figure III.3 shows that in t-distribution, $B^*(\Omega)$ performs the best against multiple breaks and random walk parameter. However, the power gaps between $B^*(\Omega)$ and others are small. Unlike the large sample case (figure III.1), 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 chapter. 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 III.4 shows that $B^*(\Omega^*)$ performs the best when the error distribution is skewed and the gaps become larger as the number of breaks increase. The gaps are relatively bigger than previous distributions. This may imply that power property depends more on the skewness rather than the tail behavior. The power gaps become fairly consequential in bimodal error distribution, as shown in figure III.5. $B^*(\Omega^*)$ 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 $B^*(\Omega^*)$. The degree of the improvement depends on the modality and the skewness, rather than the tail behavior.

III.7 Conclusion

Parameter instability is of central importance in time series models. This chapter has an advancement in that it suggests an asymptotically optimal test by using little information about the underlying distribution and unstable parameter process. Adaptation has shown to be possible in this nonstandard testing problem, which makes the knowledge of the error distribution inappropriate. It implies that an attempt to find a well-fitted error distribution is asymptotically useless 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 might be too discretionary. By avoiding the sample-split method, the test $B^*(\Omega)$ also shows good power performance even in small samples.

III.A Proofs

III.A.A Proof of Lemma 6

Differentiability in quadratic mean, (III.3.4), implies that ξ_t^1 is expanded by

$$\xi_t^1 = \xi_t^0 + \frac{1}{T}\delta_t^{*\prime}\dot{\xi}_t^0 + r_t \tag{III.A.1}$$

where $\delta_t^* = \delta_t - \frac{1}{T}\sum_{i=1}^T \delta_i$, $E[(\frac{r_t}{\xi_t^0})^2] = o_p(\|(d_t/T)\|^2)$. By using (III.A.1), the square root of the integrand of the LR statistics in (III.3.3) can be written as,

$$\begin{aligned}
\sqrt{L_T} &= \prod_{t=1}^T \left(\frac{\xi_t^1}{\xi_t^0}\right) \qquad \text{(III.A.2)}\\
&= \prod_{t=1}^T \left(\frac{\xi_t^1 - \xi_t^0}{\xi_t^0} + 1\right)\\
&= \prod_{t=1}^T \left(\frac{1}{T}\delta_t^{*\prime}\frac{\dot{\xi}_t^0}{\xi_t^0} + \frac{r_t}{\xi_t^0} + 1\right)\\
&= \prod_{t=1}^T (1+\eta_t)
\end{aligned}$$

where $\eta_t = \frac{1}{T}\delta_t^{*\prime}\frac{\dot{\xi}_t^0}{\xi_t^0} + R_t$ 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 $\sum_{t=1}^T \eta_t^2 = O_p(1)$. Hence Lemma 1 is proved by showing

1. $\sum_{t=1}^T \eta_t = \frac{1}{2T}\sum_{t=1}^T \delta_t^{*\prime} \dot{l}_t^{\beta} - \frac{1}{8T^2}\sum_{t=1}^T \delta_t^{*\prime} J_\beta \delta_t^* + o_p(1)$
2. $\sum_{t=1}^T \eta_t^2 = \frac{1}{4T^2}\sum_{t=1}^T \delta_t^{*\prime} J_\beta \delta_t^* + o_p(1)$
3. $\max_t |\eta_t| = o_p(1)$

These conditions are the same as those in the proof of Lemma 1 in Chapter 2 except that each η_t is now the function of whole alternative parameters $(\delta_1', \ldots, \delta_T')'$. Since δ_t is independent of $\dot{\ell}_t^{\beta}$ under the null hypothesis, condition (1) to (3) can be proved through the same way as in Chapter 2. ♢

III.A.B Proof of Theorem 5, Lemma 7

The proof of Theorem 1) and Lemma 7 is not much different from the proofs of Theorem 1) and 2) in Chapter 2. I skip the proof. ♢

III.A.C Proof of Lemma 8

Let $\theta_t = (\delta_t', \sqrt{T}\eta)'$, $\dot{\ell}_t^{\theta} = (\dot{\ell}_t^{\beta\prime}, \dot{\ell}_t^{\eta})'$ and $J = \begin{pmatrix} J_\beta & J_{\beta\eta} \\ J_{\beta\eta}' & J^{\eta} \end{pmatrix}$ where $J_{\beta\eta} = E[\dot{\ell}_t^{\beta}\dot{\ell}_t^{\eta}]$. By Lemma 1) of Chapter 2. The integrand, L_T^S can be written as

$$L_T^S = (1 + o_p(1))exp\left[\frac{1}{T}\sum_{t=1}^T \theta_t'\dot{\ell}_t^{\theta} - \frac{1}{2T^2}\sum_{t=1}^T \theta_t' J \theta_t\right] \tag{III.A.3}$$

The second order term in the exponential of (III.A.3) is rewritten as

$$\frac{1}{2T^2}\sum_{t=1}^{T}\theta_t' J\theta_t = \frac{1}{2T^2}\sum_{t=1}^{T}\delta_t' J_\beta \delta_t + \frac{1}{2T}\sum_{t=1}^{T}\eta' J_\beta \eta + \frac{1}{2T^{\frac{3}{2}}}\sum_{t=1}^{T}\eta' J_{\beta\eta'}\beta_t$$

Since $\sum_{t=1}^{T}\delta_t = 0$ by Condition 6 (iv), the last term is a zero vector. Consequently replacing δ_t by $\delta_t - \sum_{i=1}^{T}\delta_i$ completes the proof. $\diamondsuit$

III.A.D Proof of Theorem 7

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 $\phi_{T''} \rightarrow^d \phi$ as $T'' \rightarrow \infty$ 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

$$\lim_{T''\rightarrow\infty} E\left[\phi_{T''}(Z_T)\right] \longrightarrow^d E\left[\phi(S_\beta, W_\eta) exp[\Lambda^S]\right] \qquad \text{(III.A.4)}$$

(III.A.4) 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

$$\begin{aligned} \Psi(\Omega) &= E\left[1_{\{\Lambda > k_\alpha\}} \exp[\Lambda]\right] \\ \Psi^S(\Omega, h) &= E\left[1_{\{\Lambda > k_\alpha\}} \exp[\Lambda^S]\right] \qquad \text{(III.A.5)} \end{aligned}$$

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

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

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

where $S_\beta = (\int W_\beta' dW_\epsilon - \int W_\beta' W_\epsilon(1), \int W_\beta' W_\beta - (\int W_\beta)'(\int W_\beta))$, W_β 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 $J_{\beta\eta}$. Note that W_η has zero covariance with $\int W_\beta' dW_\epsilon - \int W_\beta' W_\epsilon(1)$ so that W_η is independent of S_β and normal with zero mean and variance J_η. Consequently,

$$\begin{aligned}E &\left[\exp\left[hW_\eta - \frac{h^2}{2} J_\eta\right] |S_\beta\right] \\ &= \int \exp\left[hW_\eta - \frac{h^2}{2} J_\eta\right] \exp\left[-\frac{1}{2} W_\eta' J_\eta^{-1} W_\eta\right] dW_\eta \\ &= \int \exp\left[-\frac{1}{2}(W_\eta - hJ_\eta)' J_\eta^{-1} (W_\eta - hJ_\eta)\right] dW_\eta = 1 \qquad \text{(III.A.6)}\end{aligned}$$

Consequently, we get

$$\Psi^S(\Omega, h) = E\left[1_{\{\Lambda > k_\alpha\}} \exp[\Lambda]\right] = \Psi(\Omega) \qquad \text{(III.A.7)}$$

which completes the proof. ⋄

III.A.E Proof of Lemma 9

Let's rewrite $\hat{\eta}$ as $\hat{\eta} = \eta_0 + T^{-\frac{1}{2}} W_T + T^{-\frac{1}{2}} o_p(1)$

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

$$
\begin{aligned}
T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} \widehat{J}_1^{-\frac{1}{2}} \dot{\ell}_t^{\beta}(\hat{\eta}) &= T^{-1/2} \sum_{t=1}^{[sT]} \widehat{J}^{\beta-\frac{1}{2}} \dot{\ell}_t^{\beta}(\eta_0 + T^{-1/2} W_T) + o_p(1) \\
&= T^{-1/2} \sum_{t=1}^{[sT]} \widehat{J}^{\beta-\frac{1}{2}} \dot{\ell}_t^{\beta}(\eta_0) - sK(\eta_0) W_T + o_p(1)
\end{aligned}
\tag{III.A.8}
$$

Since $W_T(1)$ is constant for all $t \leq T$, it can be easily proved by showing that the test statistic $B^{PS}(\eta_0, \Omega)$ doesn't change for the transformation from $\{\dot{\ell}_i^{\beta}(\eta_0)\}$ to $\{\dot{\ell}_i^{\beta}(\eta_0) + c\}$ where c is the $T \times 1$ vector of constants. Note that $M_e \dot{\ell}_i^{\beta *}(\eta_0) = M_e \left[\dot{\ell}_i^{\beta *}(\eta_0) + c\right]$. Lemma 5 of Chapter 2 shows that $M_e - G_{a_i} = M_e[M_e + K_{ai}^{-1}]^{-1} M_e$ where $K_{ai} = a_i^2 \left(\frac{FF'}{T^2}\right)^{-1}$. Under H_0,

$$
\begin{aligned}
B^{PS}(\beta_0, \Omega) &= \sum_{i=1}^{k} \dot{\ell}_i^{\beta *'}(\eta_0) \{M_e - G_{a_i}\} \dot{\ell}_i^{\beta *}(\eta_0) \\
&= \sum_{i=1}^{k} \dot{\ell}_i^{\beta *'}(\eta_0) M_e [M_e + K_{ai}^{-1}]^{-1} M_e \dot{\ell}_i^{\beta *}(\eta_0) \\
&= \sum_{i=1}^{k} \left[\dot{\ell}_i^{\beta *'}(\eta_0) + c\right] M_e [M_e + K_{ai}^{-1}]^{-1} M_e \left[\dot{\ell}_i^{\beta *}(\eta_0) + c\right] \\
&= \sum_{i=1}^{k} \dot{\ell}_i^{\beta *'}(\hat{\eta}_0) \{M_e - G_{a_i}\} \dot{\ell}_i^{\beta *}(\hat{\eta}_0) + o_p(1) \\
&= B^{PS}(\hat{\eta}, \Omega) + o_p(1)
\end{aligned}
\tag{III.A.9}
$$

which shows the asymptotic equivalency under the null hypothesis. The asymptotic equivalency under the alternative hypothesis can be proved if the alternative distribution is contiguous to the null distribution which means
$E[\exp(\Lambda^S(\Omega))] = 1$.

$$\begin{aligned} E\left[\exp[\Lambda^S]\right] &= E\left[\exp[\Lambda]\exp\left[hW_\eta - \frac{h^2}{2}J_\eta\right]\right] \\ &= E\left[\exp[\Lambda]E\left[\exp\left[hW_\eta - \frac{h^2}{2}J_\eta\right]|S_\beta\right]\right] \\ &= E\left[\exp[\Lambda]\cdot 1\right] = 1 \end{aligned} \tag{III.A.10}$$

The third equality comes from (III.A.6) and the last equality is by Theorem 1), which completes the proof. $\diamond$

III.A.F proof of Theorem 8

Let's define the two test statistics.

$$\widetilde{LR}_T = \int exp\left[\dot{\ell}^{\beta'}(M_e \otimes I_k)\beta - \tfrac{1}{2}\beta'(M_e \otimes J_\beta)\beta\right] d\nu_\delta \simeq B(\Omega) \tag{III.A.11}$$

$$\widehat{LR}_T = \int exp\left[\hat{\dot{\ell}}^{\beta'}(M_e \otimes I_k)\beta - \tfrac{1}{2}\beta'(M_e \otimes \hat{J}_\beta)\beta\right] d\nu_\delta \simeq B^*(\Omega) \tag{III.A.12}$$

Theorem 8 is proven by showing that $P[|LR_T - \widetilde{LR}_T| > \epsilon] \to 0$ under both the null and the alternative hypothesis. Since LR_T is contiguous as shown in the proof of Theorem 1, it suffices to show it only under the null hypothesis. Throughout the proof, I assume that β_0 is known. The asymptotic invariancy of replacing β_0 by $\hat{\beta}$ has already been shown in Lemma 7. For $0 < M < \infty$, define

$$\widetilde{LR}_T(M) = \int exp\left[\dot{\ell}^{\beta\prime}(M_e \otimes I_k)\beta - \frac{1}{2}\beta'(M_e \otimes J_\beta)\beta\right] \mathbf{1}\{\|\delta\| < \sqrt{T}M\} d\nu_\delta \tag{III.A.13}$$

$$\widehat{LR}_T = \int exp\left[\hat{\dot{\ell}}^{\beta\prime}(M_e \otimes I_k)\beta - \frac{1}{2}\beta'(M_e \otimes J_\beta)\beta\right] \mathbf{1}\{\|\delta\| < \sqrt{T}M\} d\nu_\delta \tag{III.A.14}$$

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

$$\begin{aligned} P[|\widetilde{LR}_T - \widehat{LR}_T| > 3\epsilon] \leq\ & P[|\widetilde{LR}_T - \widetilde{LR}_T(M)| > \epsilon] \quad (i) \\ +\ & P[|\widehat{LR}_T - \widehat{LR}_T(M)| > \epsilon] \quad (ii) \\ +\ & P[|\widetilde{LR}_T(M) - \widehat{LR}_T(M)| > \epsilon] \quad (iii) \end{aligned} \tag{III.A.15}$$

Therefore, it suffices to show that each term of (III.A.15) converges to zero, respectively.

Proof of (i):

$$\begin{aligned} \left|\widetilde{LR}_T - \widetilde{LR}_T(M)\right| &= \int \widetilde{L}_T d\nu_\beta - \int_{\|\sqrt{T}\beta\|<M} \widetilde{L}_T d\nu_\delta \\ &= c \cdot \exp\left[\frac{1}{2}\dot{\ell}^{\beta\prime}\{M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda^{-1}\}^{-1}\dot{\ell}^{\beta\prime}\right] \\ &\times \int (2\pi)^{-\frac{k(T-1)}{2}} \left|(M_e \otimes J_\beta) + K^{-1}\right|^{\frac{1}{2}} \\ &\times \exp\left[-\frac{1}{2}\left(\beta - \{(M_e \otimes J_\beta) + K^{-1}\}(M_e \otimes I_k)\dot{\ell}^{\beta}\right)' \right. \\ &\left. \times \{(M_e \otimes J_\beta) + K^{-1}\}\left(\beta - \{(M_e \otimes J_\beta) + K^{-1}\}(M_e \otimes I_k)\dot{\ell}^{\beta}\right)\right] \\ &= c \cdot \exp\left[\frac{1}{2}B(\beta_0, J_\beta, \Omega)\right] \int_{\|\sqrt{T}\beta\|>M} d\nu\beta \end{aligned} \tag{III.A.16}$$

The first term on the last equation is $O_p(1)$ by (III.3.9), 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 (ii):

$$\begin{aligned}
\left|\widehat{LR}_T - \widehat{LR}_T(M)\right| &= \int \widehat{L}_T d\nu_\beta - \int_{\|\sqrt{T}\beta\|<M} \widehat{L}_T d\nu_\beta \\
&= (1+o_p(1))c \cdot \exp\left[\frac{1}{2}\hat{\dot{\ell}}^{*\prime}\{M_e \otimes I_k + (\frac{FF'}{T^2})^{-1} \otimes \Lambda^{-1}\}^{-1}\hat{\dot{\ell}}^{*\prime}\right] \\
&\quad \times \int (2\pi)^{-\frac{k(T-1)}{2}} \left|(M_e \otimes J_\beta) + K^{-1}\right|^{\frac{1}{2}} \\
&\quad \times \exp[-\frac{1}{2}\left(\beta - \{(M_e \otimes J_\beta) + K^{-1}\}(M_e \otimes I_k)\hat{\dot{\ell}}\right)' \\
&\quad \times \{(M_e \otimes J_\beta) + K^{-1}\}\left(\beta - \{(M_e \otimes J_\beta) + K^{-1}\}(M_e \otimes I_k)\hat{\dot{\ell}}\right)]d\nu\beta \\
&= c \cdot \exp\left[\frac{1}{2}B(\hat{\beta}_0, \hat{J}_\beta, \Omega)\right] \int_{\|\sqrt{T}\beta\|>M} d\nu\beta \qquad \text{(III.A.17)}
\end{aligned}$$

The first term on the last equation is $O_p(1)$ by (III.4.5), and the second term can be made arbitrarily small by taking M large by Condition 1). In consequence, $P[|\widehat{LR}_T - \widehat{LR}_T(M)| > \epsilon]$ can be made arbitrarily small for all T large by taking M sufficiently large.

Proof of (iii): Let's define

$$\begin{aligned}
\widetilde{L}_T(M) &= \exp\left[\dot{\ell}^{\beta'}(M_e \otimes I_k)\beta - \frac{1}{2}\beta'(M_e \otimes J_\beta)\beta\right] \cdot \mathbf{1}\{\|\sqrt{T}\beta\| \le M\} \\
&= \widetilde{L}_T(\beta) \cdot \mathbf{1}\{\|\sqrt{T}\beta\| \le M\} \\
\widehat{L}_T(M) &= \exp\left[\hat{\dot{\ell}}^{\beta'}(M_e \otimes I_k)\beta - \frac{1}{2}\beta'(M_e \otimes J_\beta)\beta\right] \cdot \mathbf{1}\{\|\sqrt{T}\beta\| \le M\} \\
&= \widehat{L}_T(\beta) \cdot \mathbf{1}\{\|\sqrt{T}\beta\| \le M\}
\end{aligned}$$

We need to show that

$$\ln(\widetilde{L}_T) = \ln(\widehat{L}_T) + o_p(1) \tag{III.A.18}$$

so that

$$\widetilde{LR}_T(M) = \int \widetilde{L}_T(M) d\nu\beta = \int (1+o_p(1)\widehat{L}_T(M) d\nu_\beta = \widehat{LR}_T(M) + o_p(1) \tag{III.A.19}$$

For the notational convenience, The proof is done based on univariate β_t, The extension to the vector case is straightforward. Let $\beta_t^* = \beta_t \mathbf{1}\{|\sqrt{T}\beta_t| \le M\}$. Then (III.A.18) is proved by showing that

$$\sum_{t=1}^{T}(\beta_t^* - \frac{1}{T}\sum_{i=1}^{T}\beta_i^*)'\dot{m}(X_t)\hat{\dot{\ell}}^g(\epsilon_t) = \sum_{t=1}^{T}(\beta_t^* - \frac{1}{T}\sum_{i=1}^{T}\beta_i^*)'\dot{m}(X_t)\dot{\ell}^g(\epsilon_t) + o_p(1) \tag{III.A.20}$$

$$\hat{J}_\beta = J_\beta + o_p(1) \tag{III.A.21}$$

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 $\dot{m}(X_t)^* = \dot{m}(X_t)\mathbf{1}\{|\dot{m}(X_t)| \le M_m\}$. 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). The proof of Lemma 4.3 of Schick (1987) implies that if $\sqrt{T}\int \hat{\dot{\ell}}^g(\epsilon)d\epsilon \nrightarrow 0$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{\dot{\ell}}^g(\epsilon_t) - \dot{\ell}^g(\epsilon_t)\right) = \sqrt{T}\int \hat{\dot{\ell}}^g(\epsilon)g(\epsilon)d\epsilon + o_p(1) \tag{III.A.22}$$

(III.A.22) implies that (III.A.20) can be obtained if we have the following

$$\sum_{t=1}^{T}\beta_t^*\dot{m}^*(X_t)\left(\hat{\dot{\ell}}^g(\epsilon_t) - \bar{\dot{\ell}}^g(\epsilon_t)\right) = T\bar{\beta}^*\bar{\dot{m}}^*\int\left(\hat{\dot{\ell}}^g(\epsilon) - \bar{\dot{\ell}}^g(\epsilon)\right)g(\epsilon)d\epsilon + o_p(M) \tag{III.A.23}$$

where $\bar{\beta}^*$ and $\bar{\dot{m}}^*$ are the their sample mean. I first show that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\dot{m}^*(X_t)\left(\hat{\dot{\ell}}^g(\epsilon_t) - \dot{\ell}^g(\epsilon_t)\right) = \sqrt{T}\bar{\dot{m}}^*(X_t)\int \hat{\dot{\ell}}^g(\epsilon)g(\epsilon)d\epsilon + o_p(1) \tag{III.A.24}$$

Theorem 6.2 of Koul and Schick (1997) implies that (III.A.24) holds if for some sequence $< \tau_T >$ of positive integers tending to infinity, the following is satisfied (See pp.269-271)

$$\frac{1}{T}\sum_{1\le l,t\le T}\sum_{|t-l|>\tau_t} E\left(|\dot{m}^*(X_t) - E[\dot{m}^*(X_t)|\epsilon_1,\ldots,\epsilon_{l-1},\epsilon_{l+1},\ldots,\epsilon_T]|^2\right) = o_p(1) \tag{III.A.25}$$

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

$$\frac{1}{T}\sum_{1\le t,l\le T}\sum_{|t-l|>\tau_t} E\left(|\dot{m}^*(X_t)-E[\dot{m}^*(X_t)|\epsilon_1,\ldots,\epsilon_l]|^2\right)=o_p(1) \tag{III.A.26}$$

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

$$\begin{aligned}
& \frac{1}{T}\sum_{1\le t,l\le T}\sum_{|t-l|>\tau_t} E\left(|\dot{m}^*(X_t)-E[\dot{m}^*(X_t)|\epsilon_1,\ldots,\epsilon_l]|^2\right) \quad \text{(III.A.27)}\\
= & \frac{1}{T}\sum_{1\le t,l\le T}\sum_{|t-l|>\tau_t} E\left(|(\dot{m}^*(X_t)-E[\dot{m}^*(X_t)])+\right)\\
& \left(|(E[\dot{m}^*(X_t)]-E[\dot{m}^*(X_t)|\epsilon_1,\ldots,\epsilon_l])|^2\right)\\
\le & \frac{1}{T}\sum_{1\le t,l\le T}\sum_{|t-l|>\tau_t} E\left(|\dot{m}^*(X_t)-E[\dot{m}^*(X_t)]|^2+\right.\\
& \left.E\left(|E[\dot{m}^*(X_t)]-E[\dot{m}^*(X_t)|\epsilon_1,\ldots,\epsilon_l]|^2\right)\right)
\end{aligned}$$

The first term converges is $O_p(T^{-2\alpha})$ because $\frac{1}{T}\sum_{1\le t,l\le T}\sum_{|t-l|>\tau_t} E\left(|\dot{m}^*(X_t)-E[\dot{m}^*(X_t)]|^2\right) < T^{-2\alpha}M_x^2 = O_p(T^{-2\alpha})$. Ibragimov theorem 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 interger less than x. I satisfies (III.A.26). Proving (III.A.23) based on (III.A.24) is equivalent to proving (III.A.24) based on

(III.A.22). Therefore we have only to show that $\sqrt{T}\beta_t^*$ satisfies (III.A.25). Note that β_t^* is independent of $\{\epsilon_t\}$ and by Condition 1) $E\left[\beta_t^* | \epsilon_1, \ldots, \epsilon_{l-1}, \epsilon_{l+1}, \ldots, \epsilon_T\right]$ for all l. Consequently,

$$\begin{aligned} & \frac{1}{T} \sum_{1 \le l, t \le T} \sum_{|t-l| > \tau_t} E\left(\left\| \sqrt{T}\beta_t^* - E[\sqrt{T}\beta_t^* | \epsilon_1, \ldots, \epsilon_{l-1}, \epsilon_{l+1}, \ldots, \epsilon_T] \right\|^2\right) \\ = & \frac{1}{T} \sum_{1 \le l, t \le T} \sum_{|t-l| > \tau_t} E\left(\left\| \sqrt{T}\beta_t^* \right\|^2\right) \\ \le & \frac{1}{T} \sum_{1 \le t, l \le T} \sum_{|t-l| > \tau_t} M = O_p(T^{-2\alpha}) \end{aligned} \tag{III.A.28}$$

which satisfies (III.A.26). Convergence of $\hat{J}_\beta$ is proved by Schick (1987) which completes the proof. $\diamond$

Table III.1: Monte Carlo Estimates of the Empirical Sizes (Linear Equation)

(a) Standard Normal Distribution

	$T = 100$					$T = 200$				
Size	B^*	B	Nyb	*Sup*	AP	B^*	B	Nyb	*Sup*	AP
10%	8.88	9.20	8.84	8.70	14.40	10.58	10.16	10.20	9.60	12.20
5%	4.16	4.84	4.28	4.80	8.02	5.12	4.86	4.90	5.12	7.66
1%	0.76	0.92	0.46	1.46	3.02	0.82	1.04	1.00	1.46	2.38

(b) Symmetric Laplace Distribution

	$T = 100$					$T = 200$				
Size	B^*	B	Nyb	*Sup*	AP	B^*	B	Nyb	*Sup*	AP
10%	9.68	8.88	9.18	8.36	14.52	10.26	10.04	10.08	10.04	13.86
5%	4.70	4.28	3.94	4.98	9.14	5.42	5.08	4.66	5.36	8.28
1%	0.78	0.74	0.74	1.76	4.00	1.26	1.14	0.96	1.76	3.20

(c) Asymmetric Laplace Distribution

	$T = 100$					$T = 200$				
Size	B^*	B	Nyb	*Sup*	AP	B^*	B	Nyb	*Sup*	AP
10%	11.30	8.96	9.16	10.24	15.62	11.80	9.38	9.66	9.94	13.30
5%	6.12	4.64	4.52	6.06	10.46	6.42	4.96	4.72	5.80	8.02
1%	1..48	0.78	0.82	2.42	5.22	1.46	1.20	0.80	1.96	3.42

(d) Student t(4) Distribution

	$T = 100$					$T = 200$				
Size	B^*	B	Nyb	*Sup*	AP	B^*	B	Nyb	*Sup*	AP
10%	9.86	8.50	9.15	9.84	16.58	9.86	9.38	9.68	9.98	13.76
5%	5.54	3.96	4.12	5.64	10.50	4.64	4.72	4.18	5.20	8.02
1%	1.20	0.76	0.66	2.02	4.66	1.00	1.06	0.84	1.94	3.42

(e) Bimodal Distribution

	$T = 100$					$T = 200$				
Size	B^*	B	Nyb	*Sup*	AP	B^*	B	Nyb	*Sup*	AP
10%	10.71	9.49	9.25	8.77	12.94	10.74	10.03	10.18	9.96	12.05
5%	5.35	4.80	4.49	4.68	7.57	5.78	5.35	5.19	5.39	6.82
1%	1.24	0.97	0.87	1.50	2.55	1.39	1.22	0.87	1.53	2.20

note) Sup: Sup-F test, AP: Exp-LM test

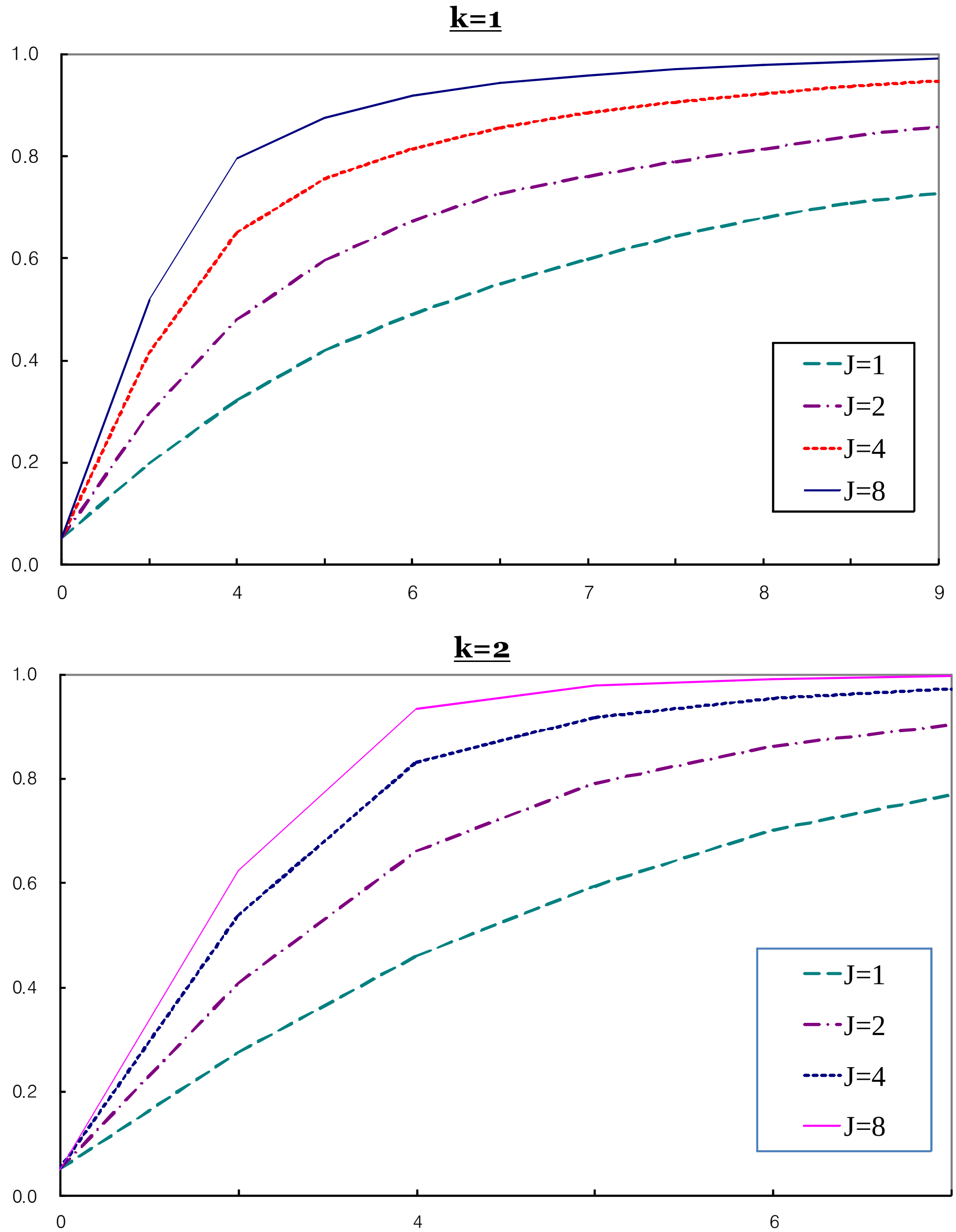

Figure III.1: Asymptotic Power Envelopes for Various Fisher Information

note) Powers are plotted from 10,000 draws using 1,000 standard normal steps to approximate Wiener Processes.

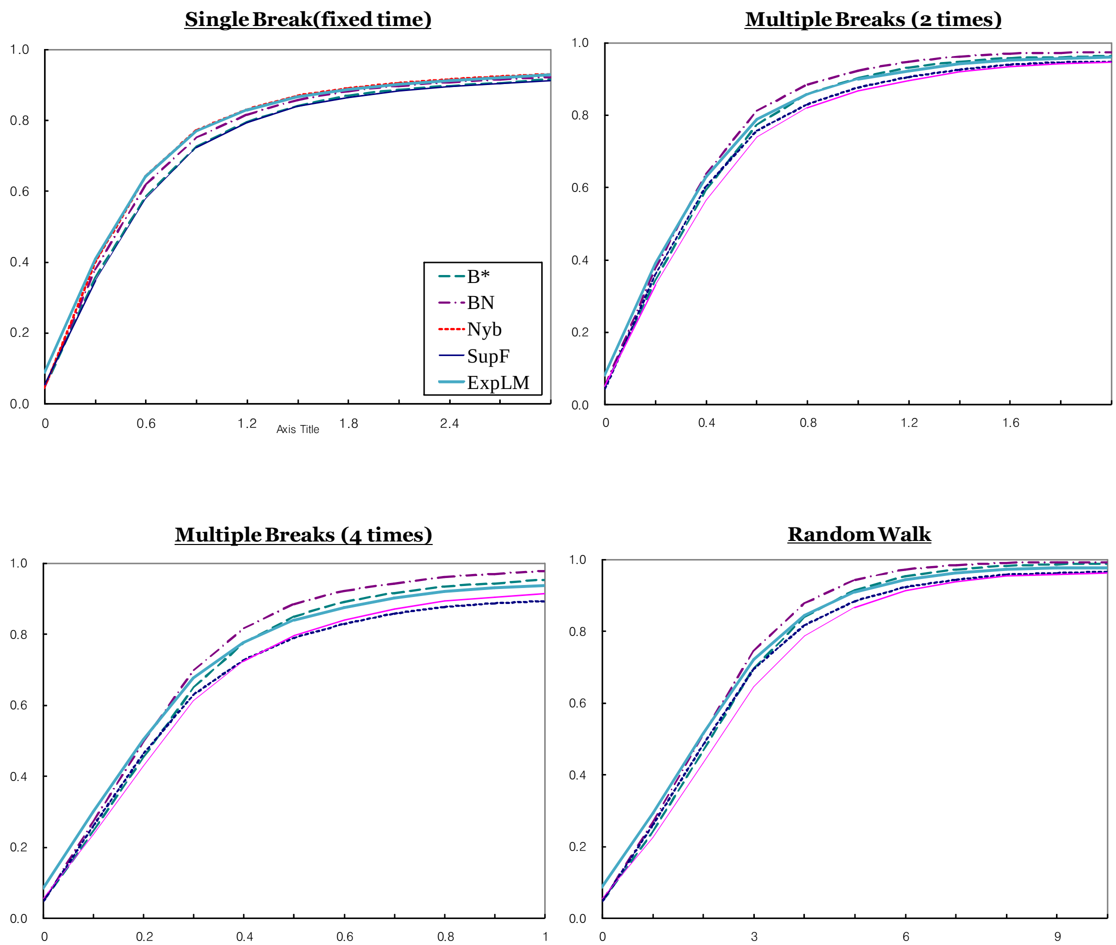

Figure III.2: Small Sample Powers, Linear Regression Model with Gaussian Error, T=100

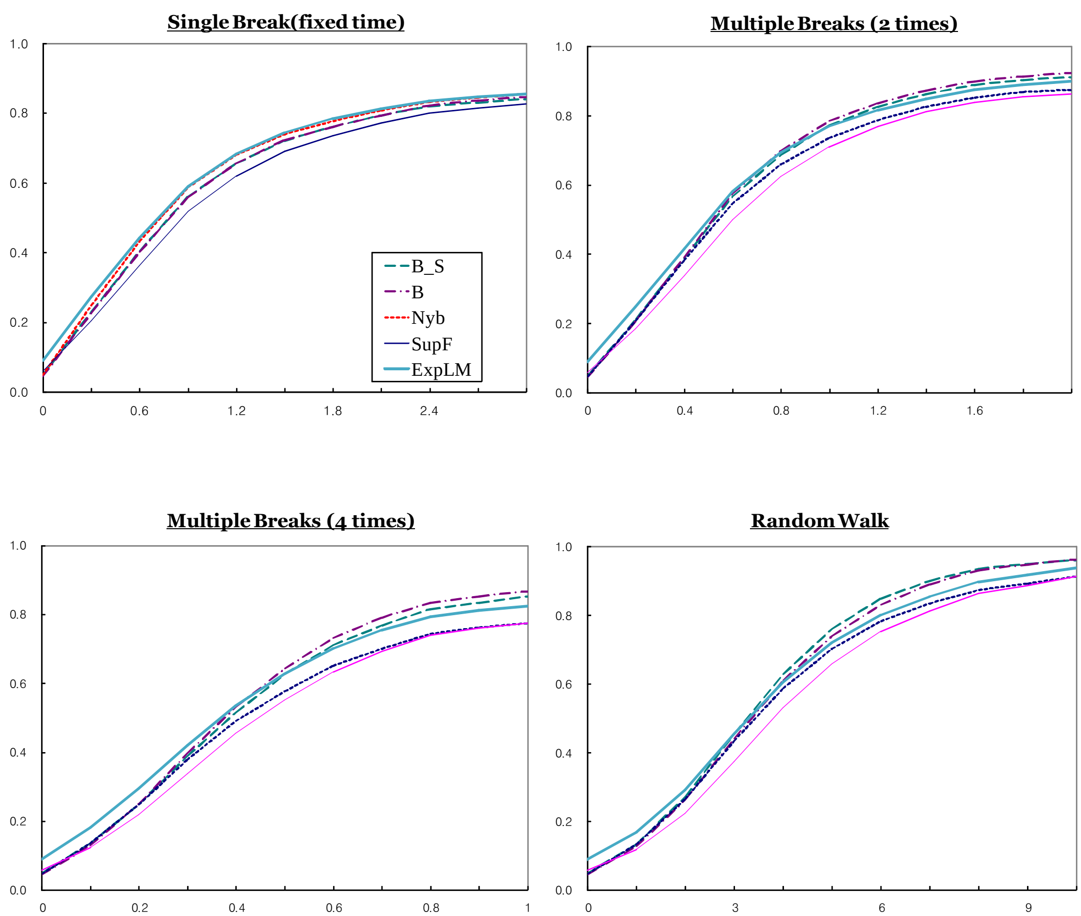

Figure III.3: Small Sample Powers, Linear Regression Model with Student t(4) Error, T=100

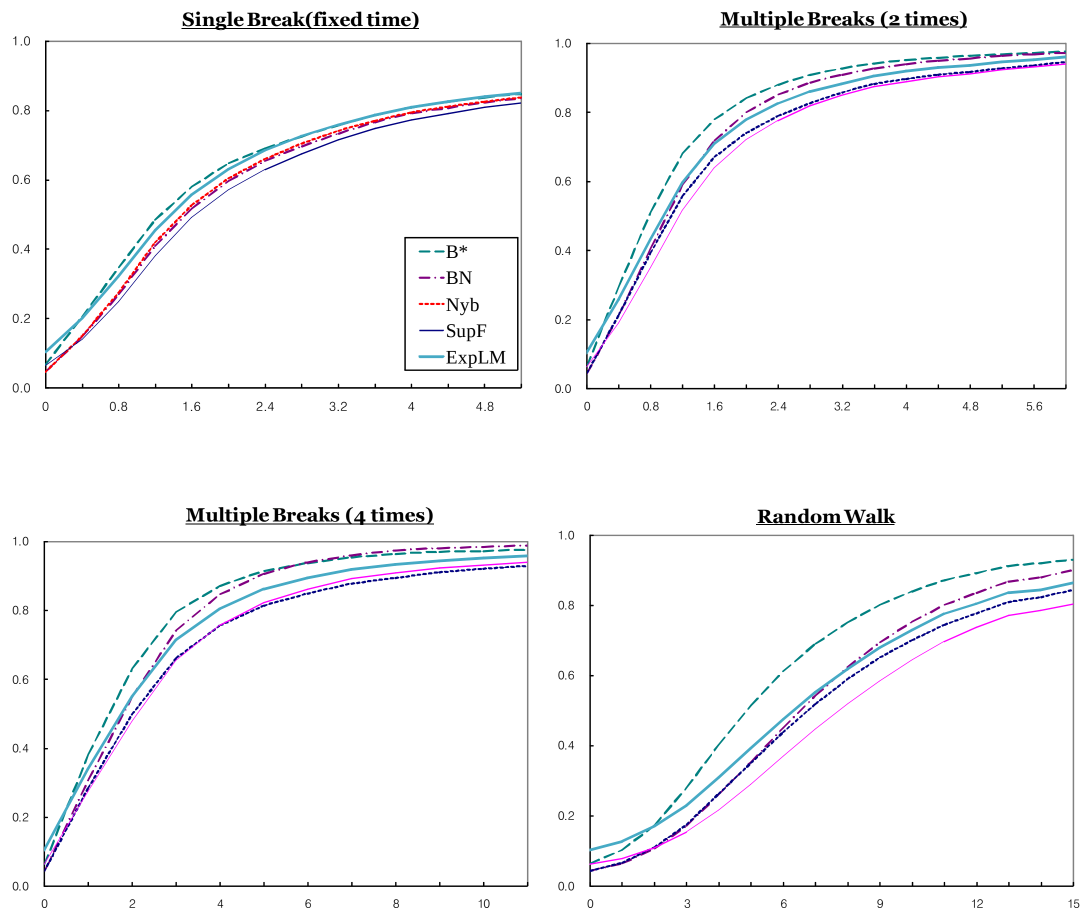

Figure III.4: Small Sample Powers, Linear Regression Model with Asymmetric Laplace Error, T=100

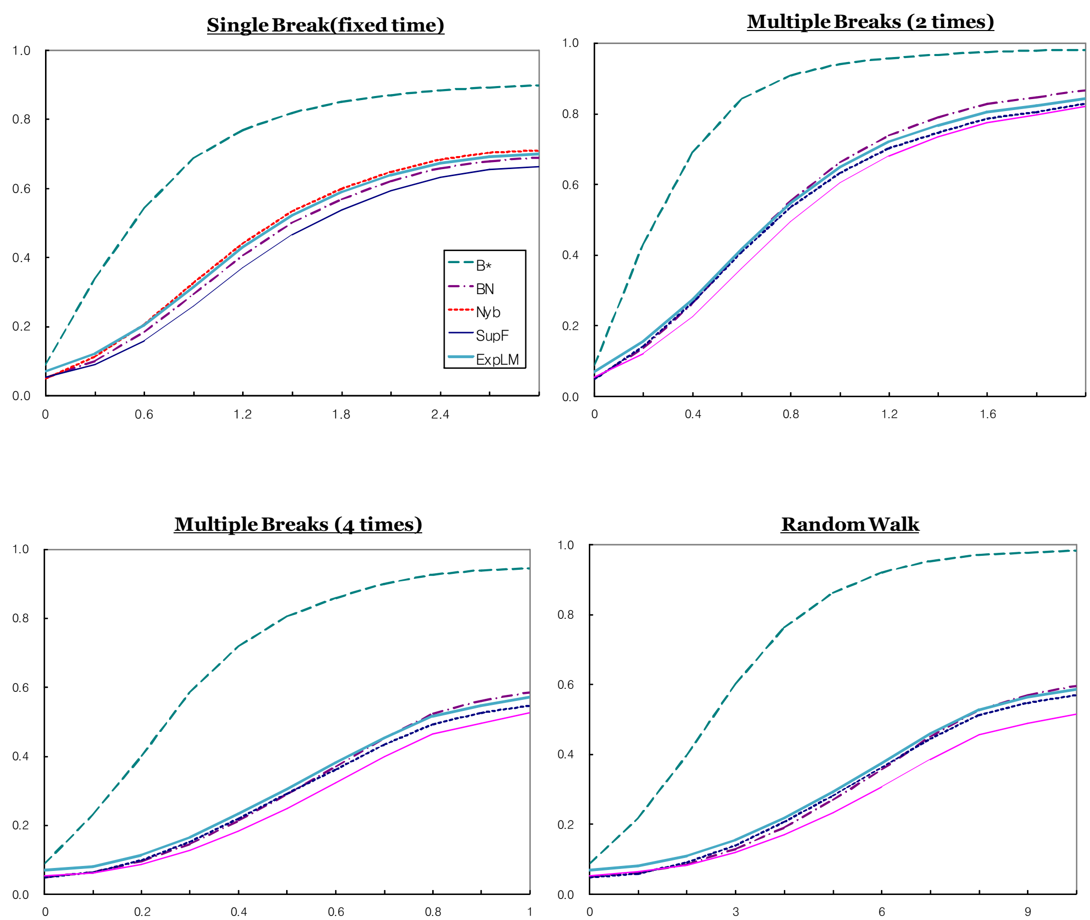

Figure III.5: Small Sample Powers, Linear Regression Model with Bi-Modal Error, T=200

Chapter IV

Testing Parameter Stability in Quantile Models: An Application to U.S. Inflation Process

This chapter considers testing parameter instability in conditional quantile models. The asymptotically optimal parameter instability tests obtained in Chapter 2 and 3 are applied to quantile models both in parametric and semiparametric set-up. In parametric models, Komunjer (2005)'s tick-exponential family of distributions is used as the underlying distribution. The suggested parametric test is still valid even when the error distribution is misspecified. I apply our test statistic to a various quantile model of the U.S. inflation process such as Phillips curve, P-star model, and autoregressive models. The test result shows an evidence of parameter instability in most quantile levels of all models. The semiparametric test rejects the stability even in more recent period with moderate economic volatility. Phillips curve model and autoregressive model have asymmetric test results across quantile levels, implying the asymmetric response of inflation to economic shocks.

IV.1 Introduction

The majority of economic empirical work has focused on conditional mean models. In this type of model, the relationship between X and y is described by how the mean of y changes with X solely. The crucial and convenient assumption for this is that X affects only the mean of the conditional distribution of y. In general, however, covariates X may influence the conditional distribution of the response in many other ways, such as expanding its dispersion as in traditional models of heteroscedasticity, stretching one tail of the distribution, and even inducing multimodality. Explicit investigation of these effects via quantile estimation can provide a more nuanced view of the relationship, and therefore a more informative empirical analysis. In this sense, increasing attention has been devoted to quantile relationships.

Another reason to pay much attention to the quantile estimation method is that the conditional mean model is insufficient to make inferences about the risks of the variable of interest. The measurement and management of risk has been an important issue in finance and macroeconomics. There is no doubt that the quantification of the tradeoff between risk and expected return is one of the main problems in finance, which makes the estimation of risk of central importance. In addition, in many of the central banks, density forecasts of inflation are preferred to point forecasts in the sense that the former contains the uncertainty structure of the forecast. Conditional quantiles have more information than just the conditional mean in that they contain other information about the uncertainty structure of the variable of interest such as skewness, kurtosis and any other factors that determine the shape of the distribution. Therefore, quantile estimation provides a better way to measure risks.

In estimating and forecasting quantiles, it is crucial to investigate whether the model of interest is stable over time. Many economic factors may cause a model to become unstable. Technology shocks, changes in economic policy, and changes in economic regimes such as a shift from a closed economy to an open one are such examples. As long as the instability is not too strong, standard estimation methods are still acceptable. However, in instances of strong instability, such as the nonstationary time varying parameter case, inference using standard methods will be misleading.

This chapter applies the optimal parameter instability tests suggested in Chapter 2 and 3 to linear conditional quantile models. Application of semiparametric optimal test is straightforward in that quantile models can be associated with conditions in Chapter 3. Using parametric optimal tests requires likelihood based models. I use Komunjer (2005)'s tick-exponential family of distributions as the underlying distribution in which the location parameter represents the quantile level. It also provides the quasi maximum likelihood estimate in the sense that the maximum likelihood estimator is still consistent to the true quantile parameter even though the error distribution is misspecified. The quasi maximum likelihood estimation property of tick-exponential distribution allows the test function to have asymptotically correct size property even though the underlying distribution is not tick-exponential.

The test is used to investigate the quantile parameter stability in the U.S. inflation model. Phillips curve, P-star model, and autoregressive models are considered for the testing purpose. The tests result shows an evidence of parameter instability in most quantile levels of all models. Semiparametric test rejects stability even after 1990's. In Phillips curve and autoregressive model, different quantile levels delivers different test results. Considering that the instability is mainly caused by various economic shocks, the non-identical test result implies that the inflation process has asymmetric response to the economic shocks.

IV.2 Notation and Preliminaries

This section considers the basic concept of conditional quantile models and examines the distinction between quantile parameters and mean parameters. Since both parameters have similar role in the inference of the relationship between X and y (how much of the variation of y is explained by that of X), one may argue that test for the constancy of mean parameters is identical to that of quantile parameters. By comparing these two parameters, however, this section shows that the argument is misleading.

Consider a stochastic process $(Y, X) \equiv Z \equiv \{W_t : \Omega \rightarrow \mathbb{R}^{k+1},\ k \in \mathbb{N},\ t = 1, ..., T\}$ 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 X_t is adapted to, i.e. $\mathfrak{F}_t \equiv \sigma(X_1, ..., X_t)$. Define $F_{0,t}(y)$ as the conditional distribution of Y_t, i.e. $\mathrm{F}_{0,t}(y) \equiv P(Y_t \leq y \mid \mathfrak{F}_t)$. Consider the model,

$$y_t = X_t'\beta_t + X_t'\gamma_t\epsilon_t \tag{IV.2.1}$$

where y_t is a scalar, X_t is a $k \times 1$ vector, β_t and γ_t are $k \times 1$ vectors of parameters. ε_t is an error term independent of X_t from a distribution with quantile function $\mathrm{Q}_\alpha(\varepsilon)$. Note that β_t and γ_t may not be constant. β_t is called the mean parameter and $X_t'\gamma_t$ represents the heteroscedasticity of the error term. The assumption that the conditional variances are linear to X might be too restrictive. But this linear scale model of heteroscedasticity is an important special case of the general class of models with linear conditional quantile functions. It subsumes many models of systematic heteroscedasticity which have appeared in the econometrics literature: Goldfeld and Quandt (1965)'s model($\sigma(x) = \sigma x_k$) is a special case when $\gamma_t = \sigma e$, and Harvey (1976) and Godfrey (1978)'s multiplicative heteroscedasticity model can be considered as a case when $\sigma(x) = x\gamma + o(||\gamma||)$.

We are interested in the αth quantile of the distribution of Y_t conditional on the information $\mathfrak{F}_t$. Denoting this as Q_α, it is defined as

$$Q_\alpha(y_t \mid \mathfrak{F}_t) \equiv \inf_{v \in R} \{v : F_{0,t}(v) > \alpha\}$$
$$\text{or if } F_{0,t} \text{ is continuous,} \quad Q_\alpha(y_t \mid \mathfrak{F}_t) \equiv F_{0,t}^{-1}(\alpha) \qquad \text{(IV.2.2)}$$

In words, conditional quantile $Q_\alpha(y_t)$ is the value that the probability of y_t being less than this value is α. The conditional quantile of y_t in the model (IV.2.1) is then simply,

$$Q_\alpha(y_t \mid \mathfrak{F}_t) \equiv X_t'\beta_t + X_t'\gamma_t \cdot Q_\alpha(\varepsilon_t) \qquad \text{(IV.2.3)}$$

From equation (IV.2.3), the αth conditional quantile of y_t can be expressed as a linear function of X_t,

$$Q_\alpha(y_t \mid \mathfrak{F}_t) \equiv X_t'[\beta_t + \gamma_t Q_\alpha(\varepsilon_t)] = X_t'\beta_{\alpha,t} \qquad \text{(IV.2.4)}$$

where $\beta_{\alpha,t} = \beta_t + \gamma_t Q_\alpha(\epsilon_t)$. Hence, the quantile parameters are determined not only by the mean parameters β, but also by the scale parameters γ and other factors that may affect the shape of the conditional distribution Q_α such as skewness and kurtosis. This property depends on whether the error term is heteroscedastic or not. Under homoscedasticity, all elements of γ_t are zero except that of the constant term. If this is the case, equation (IV.2.3) reduces to $X_t'\beta \, + \, Q_\alpha(\varepsilon_t)$, which implies

$$\beta_{\alpha,t} \;=\; \beta_t \;+\; (Q_\alpha(\varepsilon_t), 0, ..., 0)' \tag{IV.2.5}$$

Hence, β_α is equivalent to the parameters of the conditional mean except for the constant term. As noted in the introduction, however, the heteroscedasticity is general in the model of econometrics. For example, in the analysis of a household budget, residuals from the regression model exhibit variance increasing with household income. In finance, GARCH is a widely used method in modeling the financial relationship, in which $\sigma_t(X_t)^2$ is defined as $\alpha_0 + \alpha_1 y_{t-1}^2$.

In this regard, mean parameter constancy does not provide enough information about quantile parameter constancy. Even if mean parameters are constant ($\beta_t = \bar{\beta}$), quantile parameters may vary over time due to either scale parameter (γ_t) or other factors(Q_α). Consequently, in testing the hypothesis for the parameters of quantile such as tests for structural breaks, treating quantile parameter as having the same testing information as mean parameters may be misleading.

IV.3 Testing Parameter Stability in Quantile Models

This section describes methods to perform tests for quantile parameter instability. I use both test statistic considered in Chapter 2 and Chapter 3; one is based on parametric likelihood function (B) and the other is based on the semi-parametric setup (B^*). The suggested test function is

$$B(\hat{\Omega}) = \sum_{i=1}^{k} \zeta_i'(\widehat{\beta}, \widehat{J}_1)' \left[\frac{a_i^2}{T^2} I_T - F M_e F' \right]^{-1} \zeta_i(\widehat{\beta}, \widehat{J}_1) \qquad \text{(IV.3.1)}$$

where ζ_i be the vector of the partial sum of the first derivative of the log likelihood function. $M_e = I_T - \frac{1}{T} ee'$, $F = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \dots & \dots & 1 \end{pmatrix}$, and $\widehat{\beta}$ and $\widehat{J}_1$ are the maximum likelihood estimators under H_0. In parametric circumstances, ζ_i can be calculated from the underlying density. In semiparametric set-up, it is calculated from the nonparametric estimation such as kernel estimation. Application of Semiparametric Test is straightforward because the conditions for the semiparametric optimal tests considered in Chapter 3 are possible to be associated with quantile models. Under iid assumption quantile restriction on the error term simply replaces the zero mean moment condition to identify the model. Since I do not impose any other restrictions on the error term such as the symmetry around zero, the semiparametric setup in Chapter 3 can be directly applied to quantile models.

In parametric test, a likelihood based model is required to obtain the test function. The most researches on quantile models have focused on non-likelihood based estimation which makes the use of the existing LR, LM and Wald type test statistics difficult. The majority of the works have used a quantile regression framework; Koenker and Bassett (1978, 1982), Powell (1986), Koenker and Zhao (1996) are the examples. Recently, Komunjer (2005) gives a way to use LR type test by suggesting a class of likelihood functions in which there exist a quantile parameter and the maximum likelihood estimator of it is QMLE, i.e. the estimator is consistent to the true quantile parameter even though the error distribution is misspecified. The class of likelihood function, called tick-exponential family, has the form

$$\varphi_t^\alpha = \exp\left[-(1-\alpha)\{a_t(X_t\beta) - b_t(y)\}1_{\{y<X_t\beta\}} + \alpha\{a_t(X_t\beta) - c_t(y)\}1_{\{y\geq X_t\beta\}}\right] \tag{IV.3.2}$$

where $a_t(\cdot)$ is continuously differentiable and $a_t(\cdot), b_t(\cdot), c_t(\cdot)$ are functions such that (i) φ_t^α is a probability density, i.e. $\int_{\mathbb{R}} \varphi_t^\alpha dy = 1$; (ii) $X_t\beta$ is the α-conditional quantile of φ_t^α. For a given value of probability α the density function φ_t^α in (IV.3.2) is exponential by parts where the two parts have different slopes, proportional to $1-\alpha$ and α, respectively. The special case is an asymmetric type of a *Laplace distribution* by defining $a_t(X_t\beta) = [1/(\alpha(1-\alpha))]X_t\beta$ and $b_t(y) = c_t(y) = [1/(\alpha(1-\alpha))]y$, which is defined as,

$$\varphi_t^\alpha = \exp[\frac{1}{\alpha}(y_t - X_t\beta)1_{\{y<X_t\beta\}} - \frac{1}{1-\alpha}(y_t - X_t\beta)1_{\{y\geq X_t\beta\}}] \tag{IV.3.3}$$

The distribution (IV.3.3) is proportional to an asymmetric slope function and is reduced to the Laplace distribution when α is 0.5. This type of likelihood function, as proposed by Komunjer (2005), is advantageous for analyzing conditional quantile models. First, parameters in (IV.3.2) represent conditional quantile parameters. In distributions with parameters as functions of mean, variance, and other moments (such as the normal and t-distribution), we need to re-parameterize in order to present them as a function of quantile parameters. As shown before, this re-parametrization makes the quantile parameter a function of the shape of the distribution. This hinders the use of the condition that the distribution is stable across breaks. However, since the parameters of the likelihood in (IV.3.2) represent the conditional quantile, the shape of the distribution may stay stable over different variations of parameter β. As a result, we can easily construct the likelihood ratio test statistic.

Another merit of this function is that maximum likelihood estimators based on this likelihood are consistent even if the likelihood function is misspecified. Komunjer (2005) shows that the QMLE is consistent if any form of the tick-exponential type likelihood function is used for MLE. Koenker and Bassett (1978)'s quantile regression method is regarded as MLE when the likelihood function is a Laplace form of (IV.3.2). This property implies that, as pointed out by White (1982), even though the test statistic is not optimal due to the misspecification, the test statistic is still reasonable if we correct some problems. Later in this section, we will show that using (IV.3.3) makes $B(\hat{\Omega})$ valid even though the true distribution is not. Although (IV.3.2) is not differentiable at points where either $y_t = X_t'\bar{\beta}$ or $y_t = X_t'\bar{\beta} + X_t'\beta_t$, it is easy to show that (IV.3.2) is quadratic mean differentiable so that Lemma 2 in Chapter 2 is applicable in this case. This chapter considers Laplace type case of (IV.3.3) as the underlying distribution because its MLE is equivalent to the widely known Koenker and Bassett (1978)'s Quantile Regression Estimator and it is easy to compute the test statistic $B(\hat{\Omega})$. Any other type of (IV.3.3) would be a simple expansion of the Laplace type case. Now the *Hellinger derivative* and the fisher information of the Laplace type case can be written as,

$$\begin{aligned} \dot{\ell}_t &= \frac{1}{1-\alpha} X_t - \frac{1}{\alpha(1-\alpha)} X_t 1_{\{y_t < X_t \beta_\alpha\}} \\ J &= \left[\frac{1}{(1-\alpha)^2} + \frac{1-2\alpha}{\alpha(1-\alpha)^2} \right] E[X_t X_t'] \end{aligned} \qquad \text{(IV.3.4)}$$

The maximum likelihood estimation of β_α is estimated by solving a linear programming problem such as

$$\min_z c'z \qquad \text{subject to: } Az = y,\ z \geq 0 \qquad \text{(IV.3.5)}$$

where $A = (X, -X, I_T, -I_T)$, $y = (y_1, ..., y_T)'$, $z = (\beta^{+\prime}, \beta^{-\prime}, u^{+\prime}, u^{-\prime})'$, $c = (0', 0', \theta \cdot e', (1-\theta) \cdot e')'$, $X = (X_1, ..., X_T)$, and e is an T×1 vector of ones. subscript + and - are defined as for any a, $a^+ = max[a, 0]$ and $a^- = -min[a, 0]$. Buchinsky (1997a) and Koenker and Park (1996) suggest its solution methods and show the uniqueness of the solution. traditionally there are two methods to solve the linear programming problem. One is to travel from vertex to vertex along the edges of the polyhedral constraint set, choosing at each vertex the path of steepest descent, until we arrive at the optimum. The other is to take Newton steps from the interior of a deformed version of the constraint set toward boundary. Recently, Buchinsky (1997b) suggests GMM estimation by letting the first order conditions of minimization problem (IV.3.5) be the moment conditions. In order for the likelihood to be applied to the test statistics $B(\hat{\Omega})$, it should satisfies the conditions described in the previous section. The following Lemma shows that it suffices

Lemma 10 *Suppose a quantile model satisfies Condition 2 in chapter 2 in which the error distribution is a tick-exponential in* (IV.3.3)*. Then the maximum likelihood estimator using the methods described in the above satisfies Condition 4 where* $K(\bar{\beta}) = \frac{1}{\alpha(1-\alpha)} \frac{1}{T} \sum_{t=1}^{T} E\left[X_t X_t' \varphi_t^0(y, X_t)\right]$.

Therefore the test statistic, $B(\hat{\Omega})$, using the score and the Fisher information in (IV.3.4) has the asymptotic distribution given in (III.3.9). The test statistic is *point-optimal* when the true distribution is (IV.3.3). However, it might be argued that the distributional assumption is too strict to be used to macroeconomic or financial applications. As Komunjer (2005) noted, (IV.3.3) is proposed to be used as a QMLE. Therefore the test statistic is required to be valid even in misspecified cases. As noted above, MLE of (IV.3.3) has a nice property that the estimators are consistent even when the true distribution is not. This property makes it easy to show that $\frac{1}{T} \sum_{t=i}^{[sT]} J^{-1/2} \dot{\ell}_t$ in (IV.3.4) is still converging to *Brown-*

ian bridge. The following lemma shows the test function based on the asymptotic laplace distribution is asymptotically valid.

Lemma 11 *Let $B_L(\hat{\Omega})$ be the test statistic $B(\hat{\Omega})$ under* (IV.3.3) *where $\dot{\ell}_t$ and J is defined in* (IV.3.4). *Suppose a quantile model satisfies Condition 2) in Chapter 2 with the underlying distribution $\{f(\epsilon_t)\}$ which is not necessarily tick-exponential. The asymptotic distribution of $B_L(\hat{\Omega})$ under the null hypothesis of constant parameters is the same as Lemma 5 in Chapter 2.*

IV.4 Quantile Models in Inflation Process

Quantile models of inflation is arousing more attention for both economics and policy making. It is importantly used as an alternative method of density forecasts which are being increasingly used in practice. (See Tay and Wallis (2000) for a survey of application in macroeconomics and finance.) Point forecasts, namely the central tendency of the forecasts, are currently the most widely used methods, but are being increasingly criticized in that they contain no description of the associated uncertainty.

Density forecasts of inflation is estimates of the probability distribution of its possible future values. They provide a description of forecast uncertainty, and act as supplement to the point forecast in that the point forecast is considered as the central points of ranges of forecast uncertainty. Since density forecasts estimate a complete description of the uncertainty structure, they can be seen to provide information on all possible intervals and quantiles. However, density forecasts generally require the function form of the density to be specified or complicate nonparametric estimation of the density which sometimes has poor forecasting power in relatively small samples. Quantile forecast can avoid fully nonparametric method and has the advantage of not requiring the density and will thus be more robust to certain types of misspecification such as tail behavior of the distribution. Sometimes it is enough to obtain a finite levels of quantile for certain forecasting purposes. In addition, even though the purpose is the density forecast, Thompson and Miller (1986) show that a natural way to summarize the predictive distribution is by presenting selected quantiles.

An example of quantile type forecasts of inflation is the U.S. Survey of Professional Forecasts (SPF), known as the ASA-NBER survey. In this survey, forecasters are asked not only to report the point forecast and forecast horizons, but also to attach the density forecasts for inflation and output growth. In each case, a number of bins, in which the future value of output/inflation might fall are preassigned, and each survey respondents is asked to report their associated forecast probabilities. The forecast is thus represented as a histogram on a preassigned grid which is associated with the inverse of forecast quantiles.

The second example is the Bank of England Monetary Policy Committee's density forecast of inflation, known as the inflation fan chart. Figure IV.1 shows the inflation fan chart. The density forecast is represented graphically as a set of prediction intervals, covering 10, 20, ...,90 percentiles of probability distribution. The lighter shades are for the bands further from the mode (or the median). Since the distribution becomes increasingly dispersed, the quantiles fan out as the forecast horizon increases.

The original Bank of England's fan chart chooses the mode of the density forecast as its preferred central projection. Thus the central tendency is apart

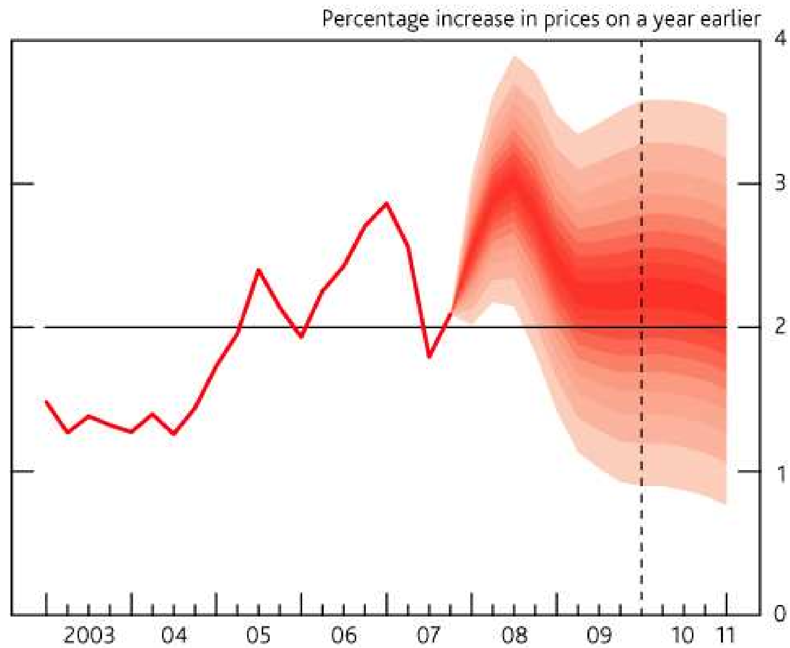

Figure IV.1: Bank of England Inflation Fan Chart

from the median when the predicted distribution is asymmetric. It implies that the graphed prediction intervals do not coincide with quantiles so that, for example, 90 percent prediction intervals is not formed by 5 percent and 95 percent quantiles under asymmetry of distribution. Wallis (1999) gets around the problem by suggesting an alternative fan chart based on central prediction intervals.

The fan chart is analytically drawn by choosing a particular probability density function. Once the values have been assigned to the underlying parameters in the density function, probabilities can be readily calculated. The density function Bank of England uses is the skewed version of the normal distribution, called two piece normal, in which an additional parameter describing the asymmetry of the distribution is introduced. However, it is well known that tails of normal distribution is too thin to adequately describe inflation and other macroeconomic data. In addition, the two piece normal density has only 3 parameters to describe the whole distribution. Consequently, each eighteen quantile levels are determined by 3 parameters.

The problem can be overcome either by using nonparametric forecasting of density (Fujiwara and Koga (2002)) or by directly forecasting quantile. Taylor and Bunn (1999) apply quantile regression approach to generating forecast intervals of various macroeconomic data to show that the quantile forecasting method is encouraging.

Bank of England's fan chart is the combination of model based point forecast and the Monetary Policy Committee (MPC)'s subjective probability forecast. Once the point forecast and the variance are estimated based on the macroeconomic model, MPC judges whether the uncertainty will be increased and whether the upward and downward risks are balanced. Thus the fan chart explicitly allowed the structural breaks at the time of the forecast. However, the model is estimated based on the assumption that the parameters are constant, leading to a conflict that an acknowledged instability at time t is reversed to be stable at any time periods later than t. Cogley et al. (2005) get around this problem by using a Bayesian vector autoregression in which the parameters follow driftless random walks. Vega (2005) also Bayesian method where the parameters are allowed to be non constant but assumes they are stable.

IV.5 Testing Quantile Parameter Stability of U.S. Inflation Process

This section explores the empirical evidence of the instability in quantile models for the U.S. inflation process. Structural break is a widely accepted phenomena in models for inflation, and many researchers have devoted much effort to identifying structural breaks in the inflation processes. Clak and McCracken (2006) find a break in 1982, while Estrella and Fuhrer (2003) suggest another break in 1984. Jouini and Boutahar (2003) find evidence of a structural break in the AR coefficient in 1990. These researches find no break after the 1990's. On the other hand, recent studies such as Stock and Watson (2006), and Atkeson and Ohanian (2001) argue that economic relationships have become unstable, even in recent years, so that the predictive powers of the models for forecasting inflation are still doubtful, even after the reduction of the volatility in the great moderation. Those tests are performed for the mean parameter stability and occasionally the variance stability. No works have been devoted to the instability of inflation quantiles.

The test in this section serves three purposes. First, I analyze whether assuming unstable parameters is justified by the data. The test will provide insights on how to choose a relevant model. Second, I analyze if the commonly used sample splitting method overcomes the instability problem in the presence of structural break. For example, it is generally perceived that around 1982 there were significant changes in macroeconomic behaviors, such as the investment prices' average rate of decline, the conduct of monetary policy, macroeconomic volatility and the regulatory environment. Most studies explicitly considering the break generally make effort to overcome the problem by reestimating the model on split subsamples. Gali et al. (2003), Fisher (2006), Clarida et al. (2000), and Barth and Ramey (2005) suggest to split into pre-Volcker period (post war-1979) and Modern era (1982 to the present). With this insight, I examine whether the split sample method could resolve the instability problem by performing the stability tests to each split sample period. The more recent subsample period (1991-current) is also considered to evaluate the recent debates about whether the inflation forecast models become more stable after 1990s.

Third, I observe whether the breaking processes are identical for all levels of quantile. This will provide an inference on the asymmetry of policy effect in that nonidentical test results implies the different responses of the inflation to policy variables associated with different points on its conditional distribution. It will also give an implication of the quantile forecasting method. Most of the literatures on inflation quantile forecasts, especially the inflation fan chart, are based on specific parametric distributions in which a few number of parameters determine the shape of the distribution. In this case, for any level of quantile, quantile parameters are functions of these few parameters only. Thus a breaking process in one parameter would affect all levels of quantiles. This implies that the test should have identical results at all quantile levels. Consequently, if a model has different test results across different levels of quantiles, it would provide an evidence of inadequacy of the parametric distributional assumption. In addition, different test results will give an inference about potential asymmetry of monetary policy and business cycle. Analysis of different levels of quantiles provides a broader picture of inflation process against a various external and policy shocks. To perform the test, I consider various types of predictive models which are listed below.

Phillips Curve(BPC)	$: \pi_t = \beta_0 + \beta_1(L)\pi_t + \beta_2(L)\hat{g}_t + \epsilon_t$
P-star Model	$: \pi_t = \eta^\alpha \ln\left(\frac{P_{t-1}}{P^*_{t-1}}\right) + \beta_1(L)\pi_t + \epsilon^\alpha_t$
	$: P^*_t = \frac{M_t V^*_t}{y^*_t}$
Autoregressive Model(AR)	$: \pi_t = \beta_0 + \beta_1(L)\pi_t + \epsilon_t$

$\pi_t =$ inflation rate	$\hat{g}_t =$ NAIRU gap
$m_t =$ money supply (M2)	$y_t =$ output (industrial production)

where $X(L)$ represents the lag operator, and P^*, V^* and y^* indicate their longrun equilibrium level, respectively. BPC is a backward-looking statistical Phillips Curve which is believed by many to be the preferred tool for forecasting inflation. (See Stock and Watson (1999).) On the other hand, recent research such as Stock and Watson (2006), and Bachmeier and Swanson (2005) cast doubt on the predictive power of the Phillips Curve model. P-star model has first proposed by Federal Reserve Board as a simple inflation forecasting method. It uses money equation as a longrun relationship of inflation, output, and money in which P-star is interpreted as the longrun price level. However, the forecasting power of P-star model is cast doubt on by many researchers, especially suspected on the stable relationship between inflation and money quantity. (See Bachmeier and Swanson (2005), and Christiano (1989).) The AR model has an advantage and has been used by many in that it is simple and has a good forecasting power.

The data spans the period between January 1962 and October 2007. I also consider a couple of subsamples; Jan.1982-Oct.2007, and Jan.1991 to Oct.2007. The split period is chosen based on the findings of a structural breaks by the current researches. Tests in the later period will provide implications on the stability after the great moderation.

I use urban CPI to calculate inflation, π_t, and industrial production index for the output, y_t. M2 is used for the money supply. Unemployment is the civilian unemployment rate. NAIRU, y_t^* and V_t^* are estimated by Hodrick-Prescott filtering. The lags are determined by Akaike Information Criteria(AIC) and all data are seasonally adjusted.

Table IV.5 shows the test results. The adaptive test $B^*(\Omega)$ shows strong evidence of instability of Phillips Curve in all sample periods, while the test $B(\Omega)$ could not reject the instability in the upper quantile parameters of the sample periods 1962-2007, 1982-2007, and all quantile levels in the recent period (1991-2007). Throughout all the models, the parametric test B has a tendency not to reject the stability compared to $B^*(\Omega)$. Note that $B(\Omega)$ does not have asymptotic optimality

Table IV.1: Test Results: Stability of the U.S. Inflation Process

α (quantile)	B 1962-2007	B 1982-2007	B 1991-2007	B 1962-2007	B 1982-2007	B 1991-2007
	Philips Curve					
0.30	39.11†	51.10†	17.20	57.40†	53.54†	21.43
0.40	32.13**	36.22†	23.85	50.26†	46.72†	28.62**
0.50	28.97*	28.84**	23.99	52.95†	51.67†	26.82
0.60	26.77	27.66	18.64	50.15†	55.35†	32.68**
0.70	21.33	23.45	12.71	60.94†	35.06†	30.99**
	P-star Model					
0.30	25.54**	23.16*	20.69	30.98†	16.94	7.29
0.40	28.36**	14.32	14.24	33.97†	15.04	10.19
0.50	24.25*	16.09	17.07	29.31†	11.56	11.60
0.60	22.09	17.45	17.50	33.64†	16.42	18.50
0.70	17.19	14.97	12.00	40.28†	19.78	17.49
	AR(3) Model					
0.30	28.69**	24.58	15.39	26.45**	24.64*	17.25
0.40	33.02†	26.02**	15.87	31.52†	24.63*	17.53
0.50	28.10**	15.97	14.73	24.74*	25.36**	28.77**
0.60	19.47	19.50	17.73	25.85**	38.47†	43.79†
0.70	15.73	17.25	19.25	32.10†	43.53†	50.04†
	AR(1) Model					
0.30	19.33†	23.76†	24.89†	31.72†	37.69†	53.56†
0.40	18.99†	27.39†	26.28†	31.33†	38.89†	53.60†
0.50	23.92†	30.76†	32.16†	29.96†	36.88†	54.05†
0.60	18.18†	24.34†	22.05†	30.14†	37.48†	53.61†
0.70	17.07†	18.32†	24.79†	30.34†	38.18†	53.52†

note: 1. *, **, † mean that the test rejects parameter stability at 10%, 5%, and 1% significant levels, respectively.

2. the lag lengths are set by AIC.

Table IV.2: Test Results: Mean/Variance Stability

Model	1982-2007	1982-1989	1991-2007
Phillips Curve(BPC)	30.53^{*}	20.72	25.22
P-star Model	44.11^{*}	44.33^{*}	36.53
AR(1)	$25.42^{\dagger}$	13.02^{*}	$32.84^{\dagger}$
AR(3)	24.31	22.17	20.01
Unconditional Vairance	$42.96^{\dagger}$	$15.45^{\dagger}$	$38.77^{\dagger}$

because different quantile levels use different likelihood function although they are all from the same data generating process. Considering that $B^{*}(\Omega)$ is asymptotically optimal, we would expect that the different test results are from the power property of the tests rather than the small sample size distortion, which implies that $B(\Omega)$ result is more favored. This inference becomes clearer if we compare the them with the test of the mean and variance stability. Table IV.2 shows the $B^{*}(\Omega)$ test result for the mean and variance parameters using the same model with the same data. The test could not reject the mean stability in the sample period 1991-2007, which is similar to $B(\Omega)$ test result for quantile instability. But the variance presents instability in all subsample period. As noted in section 2, quantile parameter can be interpreted as the function of not only the mean but also the variance and other distributional behaviors, which implies that any breaks in mean or variance cause breaks in quantiles. In this regards $B^{*}(\Omega)$ test is more associated with the mean-variance tests.

Based on $B^*(\Omega)$ test result, we find evidence that Phillips curve are still unstable even in the ara of great moderation. This result coincides with Atkeson and Ohanian (2001) and Stock and Watson (2006)'s findings that the forecasting power of backward looking type of inflation models is not improved even after 1990's because of instability.

Another finding is that the test does not have identical results across quantiles and it depends on the sample periods. $B(\Omega)$ has a tendency to strongly reject the null hypothesis for parameters of lower quantile levels, while it accept it for higher quantile levels in periods 1962-current, and 1982-current. In 1991-current period, $B^*(\Omega)$ rejects stability in upper quantile levels and accept it in lower quantiles. It has similar test result in AR(3) models. Consequently, the test supports the asymmetry of the inflation response to economic shocks. It also implies that the inflation density forecast methods based on parametric distribution such as inflation fan charts may lose the accuracy of the forecast because, as noted above, assumed stability of the shape of the distribution contradicts the test result.

For P^* models, both $B(\Omega)$ and $B^*(\Omega)$ reject the stability in whole sample period (1962-2007), But they could not reject in more recent subsample periods. However, it should be cautious to admit the forecasting power of P^* model. The main problem of the forecasting power P^* model is the weak relationship between P^* gap and the inflation. But accepting the stability of the model does not necessarily mean that the model shows the close relationship.

The AR(1) model is shown to be most unstable. Both tests reject the stability at 1% significant level. For the AR model based on AIC lag decision (AR(3)), $B(\Omega)$ could not reject the stability in 1991-2007. However, $B^*(\Omega)$ shows evidence of instability in lower quantile while it has opposite test results in upper quantiles.

In summary, I find evidence of instability in all models for the whole sample periods (1962-2007). The evidence of instability after 1990's depends on the selected model. For some models, we find different test result across different quantile levels.

IV.6 Conclusion

This chapter applies the optimal tests for parameter instability to linear conditional quantile models both in parametric and semiparametric setups. Tests functions obtained in Chapter 2 and 3 are shown to be applicable. Tick-exponential distribution allows the parametric test function to be asymptotically valid even under misspecified underlying distribution.

The application of the tests to quantile models for the U.S. inflation process shows evidence of instability of various quantile models. The quantile models are still rejected to be stable even after 1990's, while the conditional mean is shown to be stable in some of the inflation models such as Phillips Curve. The nonidentity of instability test results implies that inadequacy of the density forecast based on parametric density function as well as the asymmetric response of inflation to various economic shocks.

IV.A Proofs

IV.A.A Proof of Lemma 10

Let $\Delta\dot{a}_t = \dot{a}_t(\bar{\beta}+T^{-1/2}\delta)-\dot{a}_t(\bar{\beta})$ and $\Delta 1_t = 1_{[y_t<X_t\bar{\beta}+T^{-1/2}X_t\delta]}-1_{[y_t<X_tb\bar{e}ta]}$. The score function can be written as

$$\dot{\ell}_t(\bar{\beta}+T^{-1/2}\delta) = \dot{\ell}_t(\bar{\beta}) + \Delta\dot{a}_t\dot{\ell}_t(\bar{\beta}) - \dot{a}_t(\bar{\beta})X_t\Delta 1_t + \Delta\dot{a}_tX_t\Delta 1_t \qquad \text{(IV.A.1)}$$

$$\begin{aligned}\Rightarrow T^{-1/2}\sum_{t=1}^{[sT]}\dot{\ell}_t(\bar{\beta}+T^{-1/2}\delta) \;&=\; T^{-1/2}\sum_{t=1}^{[sT]}\dot{\ell}_t(\bar{\beta}) + T^{-1/2}\sum_{t=1}^{[sT]}\Delta\dot{a}_t\dot{\ell}_t(\bar{\beta}) - \\ &\quad T^{-1/2}\sum_{t=1}^{[sT]}\dot{a}_t(\bar{\beta})X_t\Delta 1_t + T^{-1/2}\sum_{t=1}^{[sT]}\Delta\dot{a}_tX_t\Delta 1_t\end{aligned} \qquad \text{(IV.A.2)}$$

Following two Taylor expansions are used to prove the lemma.

$$\begin{aligned}\Delta\dot{a}_t \;&=\; T^{-1/2}\ddot{a}_t(\bar{\beta})X_t'\delta + o_p(\sqrt{T}) \qquad \text{(IV.A.3)}\\ F(T^{-1/2}X_t'\delta) - F(0) \;&=\; T^{-1/2}X_t'\delta f(0) + o_p(\sqrt{T})\end{aligned}$$

By using (IV.A.3), (IV.A.2) can be rewritten as,

$$T^{-1/2}\sum_{t=1}^{[sT]}\dot{\ell}_t(\bar{\beta}+T^{-1/2}\delta) = T^{-1/2}\sum_{t=1}^{[sT]}\dot{\ell}_t(\bar{\beta}) + T^{-1/2}\sum_{t=1}^{[sT]}\ddot{a}_tX_t\dot{\ell}_t(\bar{\beta})'\delta - sK(\bar{\beta})\delta + o_p(1) \qquad \text{(IV.A.4)}$$

The second term converges to zero because

$$
\begin{aligned}
T^{-1}\sum E\left[\ddot{a}_t(\bar{\beta})X_t\dot{\ell}_t(\bar{\beta})\right] &= T^{-1}\sum E\left[(\alpha-1_{[\epsilon_t<0]})\ddot{a}_t(\bar{\beta})X_tX_t'\dot{\ell}_t(\bar{\beta})\right] \\
&= 0
\end{aligned}
$$

which completes the proof. $\diamond$

IV.A.B Proof of Lemma 11

By Lemma 5 in Chapter 2, Lemma 11 is proved if $\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\dot{\ell}_t^* = sW(s)$ where $W(s)$ is a Wiener process. Lemma A.2 of Komunjer (2005) shows that $\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\dot{\ell}_t$ satisfies the conditions to apply the CLT for $\alpha-$mixing sequence. But by Condition 2 (iv) and the condition that ϵ_t is iid, $\{X_t f(\epsilon_t)\}$ is globally covariance stationary with long-run covariance $E[X_tX_t' f(\epsilon_t)^2]$. Therefore, it satisfies the conditions for the FCLT for $\alpha-$mixing sequence (theorem 7.30 in White (2001)), which completes the proof. $\diamond$

References

Akharif, A., and Hallin, M., 2003: Efficient detection of random coefficients in autoregressive models. *The Annals of Statistics*, **31**(2).

Andrews, D., 1993: Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61**(3).

Andrews, D., Lee, I., and Ploberger, W., 1996: Optimal changepoint tests for normal linear regression. *Journal of Econometrics*, **70**(1).

Andrews, D., and Ploberger, W., 1994: Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, **62**(3).

Atkeson, A., and Ohanian, L., 2001: Are phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, **25**(1).

Bachmeier, L., and Swanson, N., 2005: Predicting inflation: Does the quantity theory help? *Economic Inquiry*, **43**(3).

Bai, J., 1996: Testing for parameter constancy in linear regressions: An empirical distributiojn function approach. *Econometrica*, **64**(3).

Bai, J., and Perron, P., 1998: Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**.

Banerjee, M., 2005: Likelihood ratio tests under local alternatives in regular semiparametric models. *Statistica Sinica*, **15**.

Barnett, W. A., Hendry, D. F., Hylleberg, S., Teräsvirta, T., Tjostheim, D., and Würtz, A., 2000: *Nonlinear Econometric Modeling in Time Series*. Cambridge University Press, New York.

Barnett, W. A., Kirman, A., and Salmon, M., 1992: *Nonlinear Dynamics and Economics*. Cambridge University Press, New York.

Barth, M. J., and Ramey, V. A., 2005: Forecasting using relative entropy. *Journal of Money Credit and Banking*, **37**(3).

Benghabrit, Y., and Hallin, M., 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).

Buchinsky, M., 1997a: Methological issues in quantile regression. *Manuscript*.

Buchinsky, M., 1997b: Recent advances in quantile regression models: A practical guideline for empirical research. *The Journal of Human Resources*, **33**.

Choi, S., Hall, W., and Schick, A., 1996: Asymptotically uniformly most powerful tests in parametric and semiparametric models. *The Annals of Statistics*, **24**(2).

Chow, G., 1960: Tests of equality between two sets of coefficients in two linear regressions. *Econometrica*, **28**.

Christiano, L. J., 1989: P*: not the inflation forecaster's holy grail. *Federal Reserve Bank of Minneapolis Quarterly Review*, **13**(4).

Clak, T., and McCracken, M., 2006: The predictive content of the output gap for inflation: Resolving in-sample and out-of-sample evidence. *Journal of Money, Credit, and Banking*, **38**(5).

Clarida, R., Gali, J., and Gertler, M., 2000: Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics*, **115**.

Cogley, T., Morozov, S., and Sargent, T., 2005: Bayesian fan charts for u.k. inflation:forecasting and sources of uncertainty in an evolving monetary system. *Journal of Economic Dynamics and Control*, **29**.

Davies, R., 1977: Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64**(1).

Davies, R., 1987: Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64**(1).

Drost, F. C., and Klaassen, C. A. J., 1997: Efficient estimation in semiparametric garch models. *Journal of Econometrics*, **81**.

Dueker, M., 2005: Dynamic forecasts of qualitative variables: A qual var model. *Journal of Business and Economic Statistics*, **23**(1).

Eichengreen, B., Watson, M., and Grossman, R., 1985: Bank rate policy under the interwar gold standard. *Economic Journal*, 723–745.

Eliasz, P., Stock, J. H., and Watson, M., 2004: Optimal tests for reduced rank time variation in regression coefficients and level variation in the multivariate local level model. *Unpublished*.

Elliott, G., and Müller, U., 2006: Efficients tests for general persistent time variation in regression coefficients. *Review of Economics Studies*, **73**, 907–940.

Engle, R., and Watson, M., 1985: Testing for coefficient stability with a stationary ar(1) alternative. *Review of Economics and Statistics*, **67**(3), 341–346.

Estrella, A., and Fuhrer, J., 2003: Monetary policy shifts and the stability of monetary policy models. *The Review of Economics and Statistics*, **82**(1), 94–104.

Fernandez, C., and Steel, M., 1998: On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, **93**(441), 359–371.

Fisher, J. D., 2006: The dynamic effects of neutral and investment-specific technology shocks. *Journal of Policitcal Economy*, **114**(3), 413–450.

Franzini, L., and Harvey, A., 1983: Testing for deterministic trend and seasonal components in times series models. *Biometrika*, **70**, 673–682.

Fujiwara, I., and Koga, M., 2002: A statistical forecasting method for inflation forecasting. *Bank of Japan Working Paper*, (02-5).

Gali, J., López-Salido, J., and Vallés, J., 2003: Technology shocks and monetary policy: Assessing the fed's performance. *Journal of Monetart Economy*, **50**(3), 723–743.

Godfrey, L., 1978: Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, **46**, 1303–1313.

Goldfeld, S., and Quandt, R., 1965: Some tests for heteroscedasticity. *Journal of American Statistical Association*, **60**(3), 539–547.

Goldfeld, S., and Quandt, R., 1978: Asymptotic tests for the constancy of regressions in the heteroscedastic case. *Econometric Research Program Memorandum*, **229**(3).

Hall, W., and Mathiason, D., 1990: On large-sample estimation and testing in parametric models. *International Statistical Review*, **58**(1), 77–97.

Hallin, M., and Jurečová, J., 1999: Optimal tests for autoregressive models based on autoregression rank scores. *The Annals of Statistics*, **27**(4).

Harvey, A., 1976: Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, **44**, 461–465.

Hawkins, D., 2000: A test for a change point in a parametric model based on a maximal wald-type statistic. *Sankhya*, **49**(2), 368–376.

Jansson, M., 2006: Semiparametric power envelopes for tests of the unit root hypothesis. *UC Berkeley Working Papers*.

Jeganathan, P., 1995: Some aspects of asymptotic theory with applications to time series models. *Econometric Theory*, **11**.

Jouini, J., and Boutahar, M., 2003: Structural breaks in the u.s. inflation process: A further investigation. *Applied Economics Letters*, **10**.

King, M., 1988: Towards a theory of point optimal testing. *Econometric Reviews*, **6**, 169–218.

Koenker, R., and Bassett, G., 1978: Regression quantiles. *Econometrica*, **46**(3), 33–50.

Koenker, R., and Bassett, G., 1982: Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, **50**(3), 42–62.

Koenker, R., and Park, B., 1996: An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, **71**, 265–283.

Koenker, R., and Zhao, Q., 1996: Conditional quantile estimation and inference for arch models. *Econometric Theory*, **12**, 793–813.

Komunjer, I., 2005: Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics*, **128**, 137–164.

Komunjer, I., 2006: Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics*, fortcoming.

Koul, H. L., and Schick, A., 1997: Efficient estimation in nonlinear autoregressive time-series models. *Bernouilli*, **3**(3), 247–277.

Lavielle, M., and Moulines, E., 2000: Least-squares estimation of an unknown number of shifts in time series. *Journal of Times Series Analysis*, **20**(1), 33–60.

LeCam, L., 1970: On the assumptions used to prove asymptotic normality of maximum likelihood estimators. *Annals of Mathematical Statistics*, **41**(1), 802–828.

Lehman, E., and Romano, J., 2005: *Testing Statistical Hypothesis*. Springer-Verlag, New York, 2nd edition.

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 McAleer, M., 2003: On adaptive estimation in nonstationary arma models with garch errors. *The Annals of Statistics*, **31**(2), 642–674.

Murphy, S., and der Vaart, A. V., 1997: Semiparametric likelihood ratio inference. *The Annals of Statistics*, **25**(4), 1471–1509.

Murphy, S., and der Vaart, A. V., 2000: On profile likelihood. *Journal of the American Statistical Association*, **95**(450), 449–465.

Nyblom, J., 1986: Testing for the deterministic linear-trend in time series. *Journal of the American Statistical Association*, **81**, 545–549.

Nyblom, J., 1989: Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, **84**(405), 348–368.

Nyblom, J., and Makeläinen, T., 1983: Comparison of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association*, **78**(405), 856–864.

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

Pfanzagl, J., and Wefeylmeyer, W., 1978: An asymptotically complete class of tests. *W. Wahrsch. verw,. Gebiete*, **45**, 49–72.

Pollard, D., 1998: *Asymptopia*. To be published.

Powell, J., 1986: Censored regression quantiles. *Journal of Econometrics*, **32**, 828–838.

Quandt, R., 1958: The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, **53**(3), 873–880.

Quandt, R., 1960: Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, **55**(3), 324–330.

Rothenberg, T., 1984: Hypothesis testing in linear models when the error coavariance matrix are nonscalar. *Econometrica*, **52**(3), 827–842.

Schick, A., 1987: A note on the construction of asymptotically linear estimators. *Journal of Statistial Planning and Inference*, **16**, 89–105.

Schick, A., 1993: On efficient estimation in regression models. *The Annals of Statistics*, **21**(3), 1486–1521.

Shin, D. W., and So, B. S., 1999: Unit root tests based on adaptive maximum likelihood estimation. *Econometric Theory*, **15**, 1–23.

Shively, T., 1988: An exact test for a coefficient in a time series regression model. *Journal of Time Series Analysis*, **9**, 81–88.

Stein, C., 1956: Efficient nonparametric estimation and testing. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 187–195.

Stock, J., and Watson, M., 1999: Forecasting inflation. *Journal of Monetary Economics*, **44**, 293–335.

Stock, J., and Watson, M., 2006: Why has u.s. inflation become harder to forecast? *Journal of Money, Credit, and Banking*, **Forthcoming**.

Tay, A. S., and Wallis, K. F., 2000: Density forecasting: A survey. *Journal of Forecasting*, **10**, 235–254.

Taylor, J., and Bunn, D., 1999: A quantile regression approach to generating prediction intervals. *Management Science*, **45**(2), 225–237.

Theodossiou, P., 1998: Financial data and the skewed generalized t distribution. *Management Science*, **44**(12), 1650–1661.

Theodossiou, P., 2000: Skewed generalized error distribution of financial assets and option pricing. *Working Paper*.

Thompson, P., and Miller, R., 1986: Sampling the future: a bayesian approach to forecasting from univariate time series models. *Journal of Business and Economics Statistics*, **4**, 427–436.

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

Vega, M., 2005: Policy makers priors and inflation density forecasts. *Economics Working Paper Archive at WUSTL*, **040305**.

Vogelsang, T., 2005: Powerful tests of structural change that are robust to strong serial correlation. *Unpublished manuscript, Cornell Univ.*, 123–148.

Wallis, K., 1999: Asymmetric density forecasts of inflation and the bank of england's fan chart. *National Institute Economic Review*, **Janaury**, 106–112.

White, H., 1982: Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 143–161.

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

Zellner, A., 1962: An efficient method of estimating seemingly unrelated regression and test for aggregation bias. *Journal of the American Statistical Association*, **57**(3), 348–368.