A Universal Nonparametric Test for Detecting Changes in Trend*

Chuan Goh†

15 June 2007

*Preliminary and incomplete. Comments welcome.

†Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario, Canada M5S 3G7. Telephone: +1 (416) 978-4964. Facsimile: +1 (416) 978-6713. goh@economics.utoronto.ca; http://www.chuangoh.org. I am grateful to Marine Carrasco, Jean-Marie Dufour, René Garcia, Sílvia Gonçalves, Bill McCausland and Benoît Perron for their hospitality and comments on an earlier version of this paper. They are obviously not responsible for any shortcomings in the current version of this project. This research was supported in part by the Connaught Fund of the University of Toronto.

Abstract

This paper proposes a nonparametric test for detecting changes in the deterministic component of a dynamic time series model. The testing procedure is based on the asymptotic behaviour of a quasi-likelihood ratio statistic similar to the class of procedures proposed by Fan et al. (2001) for assessing the fit of nonparametric regression models in an iid-error setting. In particular, the test proposed here does not involve any asymptotic bias associated with parameter estimates of the stochastic component of the model when the deterministic trend component exhibits structural change. As such, the proposed testing procedure avoids a potentially important source of non-monotonic power.

The proposed test is also shown to be asymptotically optimal for detecting global departures from the null of parameter stability in the presence of iid Gaussian innovations. In particular, the theory of large deviations is used to show that the proposed test is optimal in terms of uniformly maximizing the exponential rate of decrease of type-II error over the set of alternatives subject to a constraint on the minimal exponential rate of decrease of type-I error. This approach is "universal" in the sense that the resulting characterization of inferential efficiency is independent of location in the space of alternatives. In contrast, the more usual framework in which power is evaluated against alternatives that converge to the null of parameter stability in large samples does not generally result in a universal testing procedure.

JEL Classification: C12, C14, C22

KEYWORDS: Structural change, large deviations, limits of experiments

1 Introduction

This paper is concerned with inference regarding the deterministic trend function in a class of dynamic time-series models that are assumed to be trend-stationary. In particular, a nonparametric test is proposed for detecting smooth departures from time-stability in a trend function that both exhibits little scope for non-monotonic power behaviour and—when the innovation terms in the model are assumed to be iid Gaussian—is also unambiguously asymptotically optimal over the space of alternatives exhibiting structural change according to the efficiency criterion of Hoeffding (1965). Other testing procedures that are applicable for testing structural change in deterministic trend functions include those proposed by Bai (1996); Chu and White (1992); Ploberger and Kramer (1996); Kuan and Hornik (1995); Kuan (1998); Vogelsang (1997, 1998, 1999) and by Juhl and Xiao (2005).[1]

It is well-known that testing procedures designed to detect changes in the trend function may exhibit nonmonotonic power in the sense of the power function initially increasing when the alternative hypothesis diverges from the null of parameter stability while later exhibiting decreasing behaviour over a range of the alternative space that is more distant from the null. Evidence presented by Vogelsang (1999) suggests that a major source of nonmonotonicity in the power function derives from the bias in estimates of autoregressive parameters when the deterministic component of the time series exhibits structural change.[2] The test proposed here avoids the problem of bias in estimated AR parameters by exploiting the intuition of Carroll et al. (1997) in the context of partially linear regression models. In particular, it is recognized that the AR parameters differ from the possibly time-varying trend function parameters in being finite-dimensional, and as such an iterative procedure is proposed to estimate both sets of parameters separately under the alternative—the time-varying parameters are estimated by minimizing a locally kernel-weighted criterion function, while estimates of the AR parameters are generated by minimization of a global objective function. Details are supplied in Section 2.

[1]A recent comprehensive review of tests for structural change, as well as of related procedures including testing procedures designed to discriminate between changes in deterministic trend functions and persistence in the stochastic component of a time series can be found in Perron (2006).

[2]In particular, when a particular alternative is not far from the null of trend stability, the bias in the AR parameter estimates is generally small and has little effect on the overall power of the testing procedure. This estimation bias, however, can be substantial when the alternative happens to be far from the null, and may have the effect of reducing power by a large margin.

This paper also examines the question of asymptotic efficiency for the inference problem of detecting changes in a smooth trend function in terms of the large-deviation properties of both type-I and type-II error probabilities against *fixed* alternatives. This perspective is motivated by results indicating that inference procedures justified by superior theoretical peformance when alternatives become indistinguishable from the null in large samples[3] do not necessarily lead the applied researcher to an unambiguous choice of test statistic. In particular, the optimal test functional for asymptotically local alternatives that are scaled to be close to the null may not be the same as that for the case where the scaling is in the opposite direction.[4] Borrowing terminology from the literature on signal detection, the test proposed in this paper is *universal* in the sense that a characterization of its (asymptotic) inferential efficiency against fixed alternatives does not depend on the magnitude of the actual break in the trend function. In particular, the procedure proposed below is shown to be justified in terms of the global efficiency criterion of Hoeffding (1965)—the proposed test is shown to attain a maximal rate of exponential decrease in type-II error probability uniformly over the alternative space subject to a constraint on the minimal exponential rate of decrease in the probability of type-I error.

The present—still preliminary—version of this paper is divided into two parts. Section 2 presents the model and inference problem in detail, and derives the asymptotic distribution of the proposed test statistic. Section 3 characterizes the asymptotic efficiency property of tests based on the quasi-likelihood ratio statistic of Section 2 in terms of large-deviations theory. Proofs appear in the appendix. *Future versions of this paper will contain the results of simulation experiments.*

2 The Model, Inference Problem and Test Statistic

Consider a univariate time series $\{y_t\}$ whose components are separable in a deterministic component δ_t and a stochastic component u_t, i.e.,

$$y_t = \delta_t + u_t,\ t = 1, \ldots, T.$$

[3]Cf. e.g., Andrews and Ploberger (1994).

[4]In the case of a model possibly exhibiting a single structural break, the optimal test statistic was shown by Andrews and Ploberger (1994) to have the form of an average over a suitable class of test statistics evaluated at each possible break date when the normalized magnitude of the break tends to zero. On the other hand, the optimal test functional has an exponential form when the magnitude of the normalized break tends to infinity.

Following recent examples in the literature[5] involving models with nonlinear time trends, the deterministic component is modelled explicitly as a function of the scaled time index $\frac{t}{T}$, to wit:

$$\delta_t \equiv g\left(\frac{t}{T}\right) \equiv \boldsymbol{\gamma}\left(\frac{t}{T}\right)^\top \boldsymbol{\theta}\left(\frac{t}{T}\right),$$

where $\boldsymbol{\gamma}\left(\frac{t}{T}\right)$ denotes a d-vector of deterministic functions and $\boldsymbol{\theta}\left(\frac{t}{T}\right)$ a conformable vector of possibly time-varying parameters.

The stochastic component of $\{y_t\}$ is modelled as an $AR(p)$ process with components

$$u_t = \sum_{i=1}^{p} \rho_i u_{t-i} + \epsilon_t,$$

where it is assumed that the roots of the polynomial equation

$$1 - \rho_1 z - \cdots - \rho_p z^p = 0$$

all lie outside the unit circle and the innovation process $\{\epsilon_t\}$ is iid $WN(0, \sigma^2)$ for some $0 < \sigma^2 < \infty$.

The hypothesis of interest is that no structural change occurs in the deterministic trend function, i.e., that

$$H_0 : \boldsymbol{\theta}\left(\frac{t}{T}\right) \equiv \boldsymbol{\theta}_0$$

for some fixed $\boldsymbol{\theta}_0 \in \mathbb{R}^d$.

Note that the model under consideration may be rewritten as

$$y_t = \boldsymbol{x}_t^\top \boldsymbol{\alpha}\left(\frac{t}{T}\right) + \boldsymbol{z}^\top \boldsymbol{\rho} + \epsilon_t, \qquad (1)$$

where

$$\begin{aligned}
\boldsymbol{x}_t &\equiv \left[\boldsymbol{\gamma}\left(\frac{t}{T}\right)^\top, \boldsymbol{\gamma}\left(\frac{t-1}{T}\right)^\top, \ldots, \boldsymbol{\gamma}\left(\frac{t-p}{T}\right)^\top\right]^\top \\
\boldsymbol{\alpha}\left(\frac{t}{T}\right) &\equiv \left[\boldsymbol{\theta}\left(\frac{t}{T}\right)^\top, \rho_1 \boldsymbol{\theta}\left(\frac{t}{T}\right)^\top, \ldots, \rho_p \boldsymbol{\theta}\left(\frac{t-p}{T}\right)^\top\right]^\top \\
\boldsymbol{z}_t &\equiv (y_{t-1}, \ldots, y_{t-p})^\top \\
\boldsymbol{\rho} &\equiv (\rho_1, \ldots, \rho_p)^\top.
\end{aligned}$$

[5]Cf. e.g., Ripatti and Saikkonen (2001); Saikkonen (2001a,b).

This paper proposes to test the hypothesis $H_0 : \boldsymbol{\theta}\left(\frac{t}{T}\right) \equiv \boldsymbol{\theta}_0$ of parameter stability by exploiting the limiting behaviour of the quasi-likelihood ratio statistic

$$\lambda_T(H_0) \equiv \frac{T-p}{2} \log\left(\frac{CSS_0}{CSS_1}\right),$$

where

$$CSS_0 \equiv \sum_{t=p+1}^{T} \left(y_t - \boldsymbol{x}_t^\top \hat{\boldsymbol{\alpha}}_T^{(0)} - \boldsymbol{z}_t^\top \hat{\boldsymbol{\rho}}_T\right)^2 \quad (2)$$

$$CSS_1 \equiv \sum_{t=p+1}^{T} \left(y_t - \boldsymbol{x}_t^\top \hat{\boldsymbol{\alpha}}_T\left(\frac{t}{T}\right) - \boldsymbol{z}_t^\top \hat{\boldsymbol{\rho}}_T\right)^2, \quad (3)$$

and where $\hat{\boldsymbol{\alpha}}_T^{(0)}$ and $\hat{\boldsymbol{\alpha}}_T(\cdot)$ respectively denote appropriate restricted and unrestricted estimators of $\boldsymbol{\alpha}(\cdot)$ with respect to H_0. Following the usage of Fan et al. (2001), a test of H_0 based on $\lambda_T(H_0)$ will be said to be of the *generalized conditional likelihood ratio* (GCLR) type.[6]

It should be clear that the parameters of (1) may be estimated with least squares when H_0 is assumed to hold and $\{\epsilon_t\}$ is a martingale difference sequence. On the other hand, estimates of the parameters in (1) that are valid in the presence of shifts in trend, however, are more difficult to come by. In particular, least-squares estimates of $\boldsymbol{\rho}$ will be inconsistent if $\boldsymbol{\theta}(\cdot)$ is not a constant function, and may lead to nonmonotonic power behaviour in any resulting tests for changes in trend.[7]

As such, this paper proposes to construct CSS_1 as given earlier in (3) by first estimating $\boldsymbol{\theta}(\cdot)$ and $\boldsymbol{\rho}$ via an iterative procedure, which is described as follows.

The basic idea is first to approximate the parameter $\boldsymbol{\alpha}(\cdot)$ locally with the linear function

$$\boldsymbol{\alpha}(q) \approx \boldsymbol{\alpha}(\cdot) + \boldsymbol{\nabla}\boldsymbol{\alpha}(r)(r-q) \equiv \boldsymbol{a} + \boldsymbol{b}(r-q)$$

for q in a neighbourhood of r. Let $K(\cdot)$ denote a symmetric pdf and $K_h(u) \equiv \frac{1}{h}K\left(\frac{u}{h}\right)$ the usual rescaling of $K(\cdot)$ with a bandwidth h. It is then possible to estimate $\boldsymbol{\theta}(\cdot)$ and $\boldsymbol{\rho}$ via the steps in the following procedure:

[6] In particular, $\hat{\boldsymbol{\alpha}}_T(\cdot)$ is not necessarily assumed to be a nonparametric maximum likelihood estimator, which may not in fact exist in this context. Cf. Fan et al. (2001).

[7] In particular, test power may decrease as the magnitude in the variability of $\boldsymbol{\theta}(\cdot)$ increases (Vogelsang, 1997). Also cf. Perron (1990), Vogelsang (1999).

1. Obtain an initial estimate $\tilde{\boldsymbol{\rho}}$ of $\boldsymbol{\rho}$ using least squares, or perhaps using the Priestley and Chao (1972) detrending procedure described in the context of (1) by Juhl and Xiao (2005).

2. Compute the estimate $\tilde{\boldsymbol{\alpha}}\left(\frac{t}{T}; h, \tilde{\boldsymbol{\rho}}\right) \equiv \tilde{\boldsymbol{a}}$ by minimizing the local criterion function

$$\sum_{t=1}^{T} \left\{ y_t - \boldsymbol{x}_t^\top \left[\boldsymbol{a} + \boldsymbol{b} \left(\frac{t}{T} - q \right) \right] - \boldsymbol{z}_t^\top \tilde{\boldsymbol{\rho}} \right\} K_h \left(\frac{t}{T} - q \right) \tag{4}$$

   with respect to $\boldsymbol{a}$ and $\boldsymbol{b}$. Here h is taken to be a bandwidth that converges to an optimal smoothing parameter appropriate for estimating $\boldsymbol{\rho}$.

3. Update the initial estimate of $\boldsymbol{\rho}$ by minimizing the global criterion function

$$\sum_{t=1}^{T} \left[y_t - \boldsymbol{x}_t^\top \tilde{\boldsymbol{\alpha}} \left(\frac{t}{T}; h, \tilde{\boldsymbol{\rho}} \right) - \boldsymbol{z}_t^\top \boldsymbol{\rho} \right]^2 \tag{5}$$

   with respect to $\boldsymbol{\rho}$.

4. Repeat Steps 2 and 3 until convergence is attained.

5. Fix $\boldsymbol{\rho}$ at its estimated value from Step 4. Compute a final estimate of $\boldsymbol{\alpha}(\cdot)$ given by

$$\hat{\boldsymbol{\alpha}}_T \left(q; h, \hat{\boldsymbol{\rho}}_T \right) \equiv \hat{\boldsymbol{a}}_T,$$

   where $\hat{\boldsymbol{a}}_T$ is obtained by the minimization of (4) with respect to $\boldsymbol{a}$, $\boldsymbol{b}$ and with $\boldsymbol{\rho}$ fixed at the value $\hat{\boldsymbol{\rho}}_T$ given in Step 5. In addition, the bandwidth h used in this step is the estimate of the optimal bandwidth appropriate for estimating $\boldsymbol{\alpha}(\cdot)$ when $\boldsymbol{\rho}$ is known.

6. Calculate final estimates of $\boldsymbol{\theta}(q), \boldsymbol{\theta}\left(q - \frac{1}{T}\right), \ldots, \boldsymbol{\theta}\left(q - \frac{p}{T}\right)$ by scaling the corresponding components of $\hat{\boldsymbol{\alpha}}_T$ by the appropriate elements of $\hat{\boldsymbol{\rho}}_T$.

Note that the proposed iterative estimation procedure exploits the intuition of Carroll et al. (1997) for estimating the parameters of a generalized partially linear single-index model—because the vector $\boldsymbol{\alpha}(\cdot)$ is modelled nonparametrically, consideration of a local criterion or contrast function is natural. On the other hand, efficient estimation of the constant parameter vector $\boldsymbol{\rho}$ requires the use of all available observations, which argues in favour of a global criterion function.

The iterative procedure proposed in the context of the model given in (1) proposes the local estimation of $\boldsymbol{\alpha}(\cdot)$ using the objective function of (4) followed with the estimation of $\boldsymbol{\rho}$ using the global objective function of (5) with the preliminary estimate filling in for $\boldsymbol{\alpha}(\cdot)$. Naturally, the alternation of parametric and nonparametric estimands in this procedure as given above in Steps 2 and 5, respectively, requires two settings for the bandwidth h used in this context. Finally, it should be noted that the function estimator $\hat{\boldsymbol{\alpha}}_T\left(\cdot; h, \hat{\boldsymbol{\rho}}_T\right)$ obtained in Step 5 may be computed over a fine, but fixed, grid of points in the unit interval with linear interpolation used as needed to fill in any remaining gaps.

Prior to deriving an asymptotic expansion for the estimate $\hat{\boldsymbol{\alpha}}\left(\frac{t}{T}; h, \hat{\boldsymbol{\rho}}_T\right) \equiv \hat{\boldsymbol{a}}_T$ of $\boldsymbol{\alpha}\left(\frac{t}{T}\right)$, note the following alternative "one-step" procedure for estimating the model parameters:

Minimize

$$\sum_{t=1}^{T}\left\{y_t - \boldsymbol{x}_t^{\top}\left[\boldsymbol{a} + \boldsymbol{b}\left(\frac{t}{T} - q\right)\right] - \boldsymbol{z}_t^{\top}\boldsymbol{\rho}\right\}^2 K_h\left(\frac{t}{T} - q\right) \tag{6}$$

with respect to $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{\rho}$. Set $\hat{\boldsymbol{\alpha}}\left(\frac{t}{T}; h, \tilde{\boldsymbol{\rho}}\right) \equiv \hat{\boldsymbol{a}}$ to be the resulting estimator of $\boldsymbol{\alpha}$. Next, improve the estimate of $\boldsymbol{\rho}$ by minimizing

$$\sum_{t=1}^{T}\left[y_t - \boldsymbol{x}_t^{\top}\hat{\boldsymbol{\alpha}}\left(\frac{t}{T}; h, \tilde{\boldsymbol{\rho}}\right) - \boldsymbol{z}_t^{\top}\boldsymbol{\rho}\right]^2 \tag{7}$$

with respect to $\boldsymbol{\rho}$.

In what follows, a first-order asymptotic expansion is given for $\hat{\boldsymbol{\alpha}}(q) - \boldsymbol{\alpha}(q)$ as obtained in the one-step procedure just described. Naturally, this expansion leads naturally to arguments for the consistency and asymptotic normality of $\hat{\boldsymbol{\alpha}}(q)$. It can be shown that this expansion continues to hold when the local smoothing in (6) is carried out under the assumption that $\boldsymbol{\rho}$ is known. The reason for this adaptive quality is that the minimization of (7) yields an estimate of $\boldsymbol{\rho}$ with a parametric rate of convergence. Both of these claims will be sketeched out as follows:

First, consider the limiting distribution of the nonparametric estimate $\hat{\boldsymbol{\alpha}}\left(\frac{t}{T}; h, \tilde{\boldsymbol{\rho}}\right)$. Set $c_T \equiv \frac{1}{Th}$, where $h \to 0$, $Th \to \infty$ as $T \to \infty$. Define

$$\tilde{\boldsymbol{\rho}}^* \equiv \begin{bmatrix} c_T^{-1}\left(\hat{\boldsymbol{a}} - \boldsymbol{\alpha}(q)\right) \\ c_T^{-1}h\left(\hat{\boldsymbol{b}} - \boldsymbol{\alpha}'(q)\right) \\ c_T^{-1}\left(\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho}\right) \end{bmatrix}$$

$$
\begin{aligned}
\boldsymbol{X}_t^* &\equiv \begin{bmatrix} \boldsymbol{x}_t \\ \frac{1}{h}\left(\frac{t}{T}-q\right)\boldsymbol{x}_t \\ \boldsymbol{z}_t \end{bmatrix} \\
\bar{\mu}_t &\equiv \boldsymbol{x}_t^T\boldsymbol{\alpha}(q) - \boldsymbol{\rho}^\top \boldsymbol{z}_t + \boldsymbol{x}_t^\top \boldsymbol{\alpha}'\left(\frac{t}{T}-q\right) \\
Q_T(\boldsymbol{\rho}^*) &\equiv h\sum_{t=p+1}^{T}\left[\left(y_t-\bar{\mu}_t-c_T\boldsymbol{X}_t^{*\top}\boldsymbol{\rho}^*\right)^2-\left(y_t-\bar{\mu}_t\right)^2\right]K_h\left(\frac{t}{T}-q\right).
\end{aligned}
$$

Note that $\tilde{\boldsymbol{\rho}}^*$ minimizes $Q_T(\boldsymbol{\rho}^*)$ if $\left(\hat{\boldsymbol{a}}^\top, \hat{\boldsymbol{b}}^\top, \tilde{\boldsymbol{\rho}}\right)^\top$ minimizes the local criterion function in (6).

The limiting behaviour of $\tilde{\boldsymbol{\rho}}^*$ naturally follows from that of the corresponding criterion function $Q_T(\boldsymbol{\rho}^*)$. In particular, since

$$
\begin{aligned}
& \left(y_t-\bar{\mu}_t-c_T\boldsymbol{X}_t^{*\top}\boldsymbol{\rho}^*\right)^2 \\
\approx\ & \left(y_t-\bar{\mu}_t\right)^2 - 2c_T\boldsymbol{X}_t^{*\top}\boldsymbol{\rho}^*(y_t-\bar{\mu}_t) + c_T^2\boldsymbol{\rho}^{*\top}\boldsymbol{X}_t^*\boldsymbol{X}_t^{*\top}\boldsymbol{\rho}^*,
\end{aligned}
$$

we have

$$
\begin{aligned}
Q_T(\boldsymbol{\rho}^*) &= -2hc_T\sum_{t=p+1}^{T}(y_t-\bar{\mu}_t)K_h\left(\frac{t}{T}-q\right)\boldsymbol{X}_t^{*\top}\boldsymbol{\rho}^* \\
&\quad + hc_T^2\boldsymbol{\rho}^{*\top}\left[\sum_{t=p+1}^{T}\boldsymbol{X}_t^*\boldsymbol{X}_t^{*\top}K_h\left(\frac{t}{T}-q\right)\right]\boldsymbol{\rho}^* + o_p(1) \\
&\equiv -\boldsymbol{A}_T(q)\boldsymbol{\rho}^* + \boldsymbol{\rho}^{*\top}\boldsymbol{B}(q)\boldsymbol{\rho}^* + o_p(1),
\end{aligned}
$$

where

$$
\boldsymbol{B}(q) \equiv plim\left[hc_T^2\sum_{t=p+1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t^\top K_h\left(\frac{t}{T}-q\right)\right].
$$

Convexity of the criterion function yields

$$
\boldsymbol{\rho}^* = \boldsymbol{B}^{-1}(q)\boldsymbol{A}_T(q) + o_p(1),
$$

and the asymptotic normality of $\tilde{\boldsymbol{\rho}}^*$ follows naturally from that of $\boldsymbol{A}_T(q)$. Now recall the following "folk" lemma:

Lemma 1. *Let $C \subset \mathbb{R}^d$, $D \subset \mathbb{R}^p$ be compact. Let $f : D \times C \to \mathbb{R}$ be a continuous function in both arguments. Suppose $\hat{\boldsymbol{\theta}}(\boldsymbol{x}) \in C$ is continuous in $\boldsymbol{x} \in D$ and also uniquely minimizes $f(\boldsymbol{x}, \boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}) \in C$ maximize $f_n(\boldsymbol{x}, \boldsymbol{\theta})$. If*

$$\sup_{\boldsymbol{\theta} \in C, \boldsymbol{x} \in D} |f_n(\boldsymbol{x}, \boldsymbol{\theta}) - f(\boldsymbol{x}, \boldsymbol{\theta})| \to 0,$$

then

$$\sup_{\boldsymbol{x} \in D} \left| \hat{\boldsymbol{\theta}}_n(\boldsymbol{x}) - \hat{\boldsymbol{\theta}}(\boldsymbol{x}) \right| \to 0$$

as $n \to \infty$.

Applying Lemma 1 yields the uniformity result

$$\sup_{q \in D} \left\| \tilde{\boldsymbol{\rho}}^*(q) - \boldsymbol{B}^{-1} \boldsymbol{A}_T(q) \right\| \xrightarrow{p} 0, \tag{8}$$

where in this case D denotes a compact subset of $(0, 1)$. Taking the first $2(p+1)$ elements of the vectors in (8) yields

$$\sup_{q \in D} \left\| \hat{\boldsymbol{\alpha}}(q) - \boldsymbol{\alpha}(q) - \frac{1}{T^2 h} \sum_{t=p+1}^{T} \boldsymbol{W}_t K_h \left(\frac{t}{T} - q \right) \right\| = o_p(c_T),$$

where $\boldsymbol{W}_t$ indicates the first $2(p+1)$ elements of $2(y_t - \bar{\mu}_t) \boldsymbol{B}^{-1} \boldsymbol{X}_t^*$.

Now let

$$\begin{aligned} \hat{\boldsymbol{\zeta}} &\equiv \sqrt{T} \left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho} \right) \\ \hat{\mu}_t &\equiv \boldsymbol{x}_t^\top \hat{\boldsymbol{\alpha}} \left(\frac{t}{T} \right) + \boldsymbol{\rho}^\top \boldsymbol{z}_t \\ \mu_t &\equiv \boldsymbol{x}_t^\top \boldsymbol{\alpha} \left(\frac{t}{T} \right) + \boldsymbol{z}_t^\top \boldsymbol{\rho}. \end{aligned}$$

Note that $\hat{\boldsymbol{\zeta}}$ minimizes

$$Q_T(\boldsymbol{\zeta}) \equiv \sum_{t=p+1}^{T} \left[\left(y_t - \hat{\mu}_t - \frac{1}{\sqrt{T}} \boldsymbol{z}_t^\top \boldsymbol{\zeta} \right)^2 - (y_t - \hat{\mu})^2 \right].$$

Taking a Taylor expansion, we have

$$\begin{aligned} Q_T(\boldsymbol{\zeta}) &= -\frac{2}{\sqrt{T}} \sum_{t=p+1}^{T} (y_t - \hat{\mu}_t) \boldsymbol{z}_t^\top \boldsymbol{\zeta} + \boldsymbol{\zeta}^\top \left(\frac{1}{T} \sum_{t=p+1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t^\top \right) \boldsymbol{\zeta} + o_p(1) \\ &\equiv -\frac{2}{\sqrt{T}} \sum_{t=p+1}^{T} (y_t - \hat{\mu}_t) \boldsymbol{z}_t^\top \boldsymbol{\zeta} + \boldsymbol{\zeta}^\top \boldsymbol{C} \boldsymbol{\zeta} + o_p(1), \end{aligned}$$

where $\boldsymbol{C} \equiv plim\frac{1}{T}\sum_{t=p+1}^{T} \boldsymbol{z}_t\boldsymbol{z}_t^\top$.

By convexity of the criterion function, therefore,

$$\hat{\boldsymbol{\zeta}} = \boldsymbol{C}^{-1}\hat{\boldsymbol{C}}_T + o_p(1),$$

where

$$\hat{\boldsymbol{C}}_T \equiv \frac{2}{\sqrt{T}}\sum_{t=p+1}^{T}(y_t - \hat{\mu}_t)\boldsymbol{z}_t.$$

Naturally, the $\sqrt{T}$-asymptotic normality of $\hat{\boldsymbol{\rho}}$ follows from the limiting behaviour of $\frac{2}{\sqrt{T}}\sum_{t=p+1}^{T}(y_t - \hat{\mu}_t)\boldsymbol{z}_t$. In particular,

$$\begin{aligned}
& \frac{2}{\sqrt{T}}\sum_{t=p+1}^{T}(y_t - \hat{\mu}_t)\boldsymbol{z}_t \\
= \; & \frac{2}{\sqrt{T}}\sum_{t=p+1}^{T}(y_t - \mu_t)\boldsymbol{z}_t + \frac{2}{\sqrt{T}}\sum_{t=p+1}^{T}(\mu_t - \hat{\mu}_t)\boldsymbol{z}_t \\
= \; & \frac{2}{\sqrt{T}}\sum_{t=p+1}^{T}(y_t - \mu_t)\boldsymbol{z}_t + \frac{2}{\sqrt{T}}\sum_{t=p+1}^{T}\boldsymbol{x}_t^\top\left[\frac{1}{T}\sum_{s=p+1}^{T}\boldsymbol{W}_s K_h\left(\frac{s}{T}-\frac{t}{T}\right) + o_p(c_T)\right]\boldsymbol{z}_t \\
= \; & \frac{2}{\sqrt{T}}\sum_{t=p+1}^{T}(y_t - \mu_t)\boldsymbol{z}_t + \frac{2}{T^{\frac{3}{2}}}\sum_{t=p+1}^{T}\sum_{s=p+1}^{T}\boldsymbol{x}_t^\top\boldsymbol{W}_s K_h\left(\frac{t}{T}-\frac{s}{T}\right)\boldsymbol{z}_t + o_p\left(\sqrt{T}c_t\right) \\
= \; & \frac{2}{\sqrt{T}}\sum_{t=p+1}^{T}\left[y_t - \mu_t + \frac{1}{T}\boldsymbol{x}_t^\top\sum_{s=p+1}^{T}\boldsymbol{W}_s K_h\left(\frac{t}{T}-\frac{s}{T}\right)\right]\boldsymbol{z}_t + o_p(1) \\
\equiv \; & \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\boldsymbol{\Omega}\left(\boldsymbol{x}_t, \boldsymbol{z}_t, \frac{t}{T}\right) + o_p(1),
\end{aligned}$$

which satisfies a central limit theorem for martingale differences (*to be checked*).

Therefore

$$\hat{\boldsymbol{\zeta}} = \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\boldsymbol{C}_2^{-1}\boldsymbol{\Omega}\left(\boldsymbol{x}_t, \boldsymbol{z}_t, \frac{t}{T}\right) + o_p(1).$$

Now consider the least-squares estimate $\hat{\boldsymbol{\alpha}}_T^{(0)}$ of $\boldsymbol{\alpha}(\cdot)$ under the hypothesis H_0 : $\boldsymbol{\alpha}\left(\frac{t}{T}\right) \equiv \boldsymbol{\alpha}_0$ for some constant $\boldsymbol{\alpha}_0 \in \mathbb{R}^{d(p+1)}$. In particular, note that $T^{\frac{3}{2}}\left(\hat{\boldsymbol{\alpha}}_T^{(0)} - \boldsymbol{\alpha}_0\right) =$

$O_p(1)$, with

$$\hat{\boldsymbol{\alpha}}_T^{(0)} = \boldsymbol{\alpha}_0 + \frac{1}{T^3}\sum_{s=p+1}^{T} \boldsymbol{x}_s \epsilon_s.$$

Combining this with previous results and recalling that

$$CSS_0 \equiv \sum_{t=p+1}^{T} \left(y_t - \boldsymbol{x}_t^\top \hat{\boldsymbol{\alpha}}_T^{(0)} - \boldsymbol{z}_t^\top \hat{\boldsymbol{\rho}}_T \right)^2$$

and

$$CSS_1 \equiv \sum_{t=p+1}^{T} \left(y_t - \boldsymbol{x}_t^\top \hat{\boldsymbol{\alpha}}_T \left(\frac{t}{T}\right) - \boldsymbol{z}_t^\top \hat{\boldsymbol{\rho}}_T \right)^2,$$

we obtain the following representations:

$$\begin{aligned} CSS_0 - CSS_1 &= 2\sum_{t=p+1}^{T} \epsilon_t \boldsymbol{x}_t^\top \left[\sum_{s=p+1}^{T} \epsilon_s \left(\frac{2}{T^2 h} \tilde{\boldsymbol{W}}_s K_h \left(\frac{s}{T} - \frac{t}{T} \right) - \frac{1}{T^3} \boldsymbol{x}_s \right) \right] \\ &\quad + \frac{1}{T^6} \sum_{t=p+1}^{T} \boldsymbol{x}_t^\top \left(\sum_{r,s} \boldsymbol{x}_r \epsilon_r \epsilon_s \boldsymbol{x}_s^\top \right) \boldsymbol{x}_t \\ &\quad - \frac{4}{T^4 h^2} \sum_{t=p+1}^{T} \boldsymbol{x}_t^\top \left[\sum_{r,s} \epsilon_r \epsilon_s K_h \left(\frac{r}{T} - \frac{t}{T} \right) K_h \left(\frac{s}{T} - \frac{t}{T} \right) \tilde{\boldsymbol{W}}_r \tilde{\boldsymbol{W}}_s^\top \right] \boldsymbol{x}_t \\ &\quad + O_p \left(\frac{1}{\sqrt{T}} + c_T \right) \qquad (9) \\ \frac{1}{T-p} CSS_1 &= \sigma^2 \left(1 + O_p \left(c_T \right) \right) + O_p \left(\frac{1}{\sqrt{T}} \right). \qquad (10) \end{aligned}$$

Noting that

$$\lambda_T(H_0) \equiv \frac{T-p}{2} \log \left(\frac{CSS_0}{CSS_1} \right) \approx \frac{T-p}{2} \cdot \frac{CSS_0 - CSS_1}{CSS_1},$$

it follows from (9)–(10) that the GCLR statistic $\lambda_T(H_0)$ is approximable by a generalized quadratic form, for which a central limit theorem is available.[8] In particular, the asymptotic normality of $\lambda_T(H_0)$ as $T \to \infty$ may be deduced using the same methods of proof used in Fan et al. (2001). (*Statement of theorem to follow*)

[8]Cf. De Jong (1987).

3 Large-Deviations Optimality of the GCLR Statistic when the Innovations are iid Gaussian

This section involves the characterization of an efficient test of $H_0 : \boldsymbol{\theta}\left(\frac{t}{T}\right) \equiv \boldsymbol{\theta}_0$ against fixed alternatives. Recall the setup of (1) above—in particular, the reformulation of the original model as

$$y_t = \boldsymbol{x}_t^\top \boldsymbol{\alpha}\left(\frac{t}{T}\right) + \boldsymbol{z}_t^\top \boldsymbol{\rho} + \epsilon_t,$$

where

$$\begin{aligned}
\boldsymbol{x}_t &\equiv \left[\boldsymbol{\gamma}\left(\frac{t}{T}\right)^\top, \boldsymbol{\gamma}\left(\frac{t-1}{T}\right)^\top, \ldots, \boldsymbol{\gamma}\left(\frac{t-p}{T}\right)^\top\right]^\top \\
\boldsymbol{\alpha}\left(\frac{t}{T}\right) &\equiv \left[\boldsymbol{\theta}\left(\frac{t}{T}\right)^\top, \rho_1\boldsymbol{\theta}\left(\frac{t-1}{T}\right)^\top, \ldots, \rho_p\boldsymbol{\theta}\left(\frac{t-p}{T}\right)^\top\right]^\top \\
\boldsymbol{z}_t &\equiv (y_{t-1}, \ldots, y_{t-p})^\top \\
\boldsymbol{\rho} &\equiv (\rho_1, \ldots, \rho_p)^\top,
\end{aligned}$$

where the d-variate deterministic trend function $\boldsymbol{\gamma}(\cdot)$ satisfies smoothness and moment conditions that will be made explicit below.

Here it is explicitly assumed that the innovations $\{\epsilon_t\}$ are iid $N(0, \sigma^2)$. In what follows, let $\boldsymbol{\eta}(\cdot) \equiv (\boldsymbol{\alpha}(\cdot)^\top, \boldsymbol{\rho}^\top)^\top \in \mathbb{R}^{d(p+1)+p}$ denote the overall parameter vector (after concentrating out the σ^2). Let $\boldsymbol{\eta} \in \Theta$, where the parameter space Θ is taken to be a bounded Hölder class.

It is easy to see that up to a constant term, the conditional profile log-likelihood of the model given the initial observations $\{y_1, \ldots, y_p\}$ has the form

$$l(\boldsymbol{\eta}) = -\frac{T-p}{2}\log(CSS),$$

where

$$\begin{aligned}
CSS &\equiv \sum_{t=p+1}^{T}\left(y_t - \boldsymbol{x}_t^\top \boldsymbol{\alpha}\left(\frac{t}{T}\right) - \boldsymbol{z}_t^\top \boldsymbol{\rho}\right)^2 \\
&\equiv \sum_{t=p+1}^{T}\left(y_t - \boldsymbol{w}_t^\top \boldsymbol{\eta}\left(\frac{t}{T}\right)\right)^2
\end{aligned}$$

$$= 2\sum_{t=p+1}^{T} \boldsymbol{w}_t^{\top}\boldsymbol{\eta}\left(\frac{t}{T}\right)y_t + \sum_{t=p+1}^{T} \boldsymbol{\eta}^{\top}\left(\frac{t}{T}\right)\boldsymbol{w}_t\boldsymbol{w}_t^{\top}\boldsymbol{\eta}\left(\frac{t}{T}\right) + \sum_{t=p+1}^{T} y_t^2.$$

Divide the parameter space Θ into disjoint subsets Θ_0 and Θ_1 associated with H_0 and H_1, respectively. Now take $P_{T,\boldsymbol{\eta}}$ to be the conditional distribution of $\boldsymbol{y}_T \equiv (y_{p+1}, \ldots, y_T)^{\top}$ given $\{y_1, \ldots, y_p\}$ for a fixed value of $\boldsymbol{\eta}(\cdot) \in \Theta$. Let $P_T \equiv P_{T,\boldsymbol{\eta}_0}$ be a measure dominating $P_{T,\boldsymbol{\eta}}$.[9] Now define $\mathcal{M}$ to be the set of all probability measures $P_{T,\boldsymbol{\eta}}$ dominated by P_T. It follows that the conditional log-likelihood ratio statistic for H_0 has the form

$$\log \frac{dP_{T,\boldsymbol{\eta}}}{dP_T}(\boldsymbol{y}_T) = \frac{T-p}{2}\log\left(\frac{CSS_0}{CSS_1}\right),$$

where CSS_0 and CSS_1 denote restricted and unrestricted conditional sums of squares, respectively. Set

$$\Xi_{T,\boldsymbol{\eta}} \equiv \frac{1}{T-p}\log\frac{dP_{T,\boldsymbol{\eta}}}{dP_T} = \frac{1}{2}\log\left(\frac{CSS_0}{CSS_1}\right)$$

and define the corresponding conditional log-likelihood ratio *process* on Θ:

$$\Xi_{T,\Theta} \equiv \{\Xi_{T,\boldsymbol{\eta}} \in \mathbb{R} : \boldsymbol{\eta} \in \Theta\} \in \mathbb{R}^{\Theta}.$$

Let $\mathcal{L}(\Xi_{T,\Theta}|\, P_T)$ denote the distribution of $\Xi_{T,\Theta}$ under the dominating measure P_T. Endow $\mathbb{R}^{\Theta}$ with the product topology so that on the Borel σ-field of $\mathbb{R}^{\Theta}$, $\Xi_{T,\Theta}$ is taken to be a $\mathbb{R}^{\Theta}$-valued random element.

The idea here is to consider the *large-deviations (LD) convergence* of $\Xi_{T,\Theta}$, i.e., the convergence of LD-probabilities associated with $\mathcal{L}(\Xi_{T,\Theta}|\, P_T)$.[10] In this connection, we recall the notion of a *large-deviations principle*:[11]

Let $\{Q_T : T \geq 1\}$ be a sequence of probability measures on the Borel σ-field of a Hausdorff topological space S.

Definition 1 (Large-deviations principle (LDP)). *$\{Q_T : T \geq 1\}$ is said to obey the LDP with rate function $I : S \to [0, \infty]$ iff for all closed sets $G \subset S$,*

$$\limsup_{T\to\infty} \frac{1}{T}\log Q_T(G) \leq -\inf_{x\in G} I(x),$$

[9] i.e., $P_{T,\boldsymbol{\eta}} << P_T$, for some $\boldsymbol{\eta}_0 \in \Theta_0$.

[10] Compare with the more familiar notion of *weak convergence* of the conditional log-likelihood ratio process (Le Cam, 1986).

[11] Cf. e.g., Varadhan (1984).

and for all open sets $H \subset S$,

$$\liminf_{T\to\infty} \frac{1}{T} \log Q_T(H) \geq -\inf_{x\in H} I(x).$$

Moreover, $I^{-1}([0,a]) \subset S$ is compact for each $a \geq 0$.

Now suppose S is a metric space. We recall a result of Puhalskii (1993) to the effect that $\{Q_T : T \geq 1\}$ obeys the LDP iff

$$\lim_{T\to\infty} \left[\int_S f(x) dQ_T\right]^{\frac{1}{T}} = \sup_{x\in S} f(x) \exp(-I(x)) \tag{11}$$

for all non-negative bounded continuous functions f on S.

Note that if condition (11) holds for all non-negative, bounded and continuous functions f on S, then $\{Q_T : T \geq 1\}$ *LD-converges* to $\exp(-I(x))$, and we write $Q_T \stackrel{l.d.}{\to} \exp(-I(x))$.

Now conditional on the first p observations, it is possible to regard the situation considered in this paper as a sequence $\{\mathcal{E}_T : T \geq 1\}$ of statistical experiments

$$\mathcal{E}_T \equiv \left(\mathbb{R}^{T-p}, \sigma\left\{\mathbb{R}^{T-p}\right\}, P_{T,\boldsymbol{\eta}} : \boldsymbol{\eta} \in \Theta\right\}.$$

Given the assumption of a dominating measure P_T, call $\langle \mathcal{E}_T, P_T\rangle_{T\geq 1}$ a sequence of dominated statistical experiments.

This leads to the following definition of Puhalskii and Spokoiny (1998).

Definition 2 (Dominated LDP)**.** *The sequence $\langle \mathcal{E}_T, P_T\rangle_{T\geq 1}$ of dominated experiments is said to obey the dominated LDP iff*

1. $\{\mathcal{L}(\Xi_{T,\Theta} | P_T) : T \geq 1\}$ *obeys the LDP with some rate function* $I : \mathbb{R}^\Theta \to [0,\infty]$

2. *for every $\boldsymbol{\eta} \in \Theta$, $\Xi_{T,\boldsymbol{\eta}}$ satisfies*

$$\lim_{M\to\infty} \limsup_{T\to\infty} E_T^{\frac{1}{T}} \left[\exp(T\Xi_{T,\boldsymbol{\eta}}) 1(\Xi_{T,\boldsymbol{\eta}} > M)\right] = 0,$$

where $E_T^{\frac{1}{T}}[\cdot] \equiv (E_T[\cdot])^{\frac{1}{T}}$.

Note that part 2 of the definition is an "exponential tightness condition" that ensures that the lower bounds of any minimax risk function associated with relevant statistical decisions to not depend on one's choice of a particular dominating measure.

It remains to recall one more fundamental result from the theory of large deviations. In particular, it is noted that the *contraction principle*[12] ensures that the LDP is preserved by continuous mappings. This ensures the following result:

Lemma 2. $\langle \mathcal{E}_T, P_T \rangle_{T \geq 1}$ *obeys the dominated LDP.*

Proof. See Appendix A. □

The LDP for $\langle \mathcal{E}_T, P_T \rangle_{T \geq 1}$ is now shown to imply the existence of lower bounds for minimax risk functions associated with certain statistical decisions. General conditions under which these lower bounds are tight will also be studied.[13]

First, let $\mathcal{D}$ denote a Hausdorff topological space whose Borel σ-field is taken to be the decision space. For each $\boldsymbol{\eta} \in \Theta$, let $W_{\boldsymbol{\eta}} \equiv \{W_{\boldsymbol{\eta}}(r) : r \in \mathcal{D}\}$ be non-negative and lower semi-continuous functions on $\mathcal{D}$. The family $\{W_{\boldsymbol{\eta}} : \boldsymbol{\eta} \in \Theta\}$ will be referred to as the relevant set of *loss functions*. Let $\mathcal{R}_T$ denote the set of all measurable mappings $r_T : \mathbb{R}^{T-p} \to \mathcal{D}$. As such, $\mathcal{R}_T$ is the set of all *decision functions* with values in $\mathcal{D}$. The (logarithmic) *LD risk* of a decision $r_T \in \mathcal{R}_T$ in the experiment $\mathcal{E}_T$ is denoted by

$$R_T(r_T) \equiv \sup_{\boldsymbol{\eta} \in \Theta} \frac{1}{T} \log E_{\boldsymbol{\eta}} \left[W_{\boldsymbol{\eta}}(r_T) \right].$$

Note that $\liminf_{T\to\infty} R_T(r_T)$ describes the exponential convergence rate of the LD risk associated with the sequence of decisions $\{r_T\}$.

The assumption is made that the loss functions $W_{\boldsymbol{\eta}}$ under consideration are *level-compact*, i.e., that they satisfy the conditions of the following definition (e.g., Strasser, 1985, Definition 6.3):

Definition 3. *A function* $f : U \to \mathbb{R}$ *on a topological space* U *is level-compact iff it is bounded from below and the level sets* $\{u \in U : f(u) \leq \alpha\}$ *are compact for all* $\alpha < \sup_{u \in U} f(u)$.

[12]Cf. e.g., Dembo and Zeitouni (1998, Thm. 4.2.1).

[13]It is noted that the level of generality adopted here is slightly greater than what is strictly required for the purposes of this paper in order to facilitate extensions to other statistical problems in the context of the maintained model.

Remark 1. *For U Hausdorff, a level-compact function is lower-semicontinuous. In addition, the supremum of a family of level-compact functions on U is also level-compact.*

Remark 2. *Level-compact functions also attain infima on closed sets. In the proof of Lemma 2, it was shown that $\Xi_{T,\boldsymbol{\eta}} = \xi_{\boldsymbol{\eta}}\left(\hat{F}_{T|p}^{(p+1)}\right)$ for all $\boldsymbol{\eta} \in \Theta$, where $\xi_{\boldsymbol{\eta}}$ is a continuous functional on $\mathcal{F}^{(p+1)}$. The lower bound of the logarithmic LD-risk function for level-compact losses is given in the following theorem.*

Theorem 1. *Under the maintained assumptions regarding the model under consideration here,*

$$\liminf_{T\to\infty} \int_{r_T \in \mathcal{R}_T} \sup_{\boldsymbol{\eta}\in\Theta} R_T(r_T) \geq R^*,$$

where for $\mathcal{F}^{(p+1)}$ as given in the proof of Lemma 2,

$$R^* \equiv \sup_{F\in\mathcal{F}^{(p+1)}} \sup_{\boldsymbol{\eta}\in\Theta} \inf_{r_T\in\mathcal{R}_T} \left\{\log W_{\boldsymbol{\eta}}(r_T) + \xi_{\boldsymbol{\eta}}(F) - I_{DZ}(F)\right\}.$$

Proof. This is essentially a specialization of the proof of Puhalskii and Spokoiny (1998, Theorem 3.1), and is therefore omitted. □

Now consider the statistical problem of interest in this paper, namely, that of testing the hypothesis of parameter stability. The relevant decision space $\mathcal{D}$ is given by $\mathcal{D} \equiv \{0, 1\}$. Equip $\mathcal{D}$ with the discrete topology and for any decision (i.e., "test") r, denote the events $\{r = 0\}$ and $\{r = 1\}$ as "non-rejection" and "rejection", respectively, of H_0.

The usual loss functions in this context are of course indicators for making the wrong decision, i.e.,

$$W_{\boldsymbol{\eta}}(r) \equiv 1\left(\boldsymbol{\eta} \notin \Theta_r\right),\ r = 0, 1,$$

where $W_{\boldsymbol{\eta}}(1)$ and $W_{\boldsymbol{\eta}}(0)$ indicate error of the first and second kind, respectively. The specific objective here is to find a sequence of tests that maximizes the exponential rate of decay of type-II error probability subject to a constraint on the minimal exponential rate of decrease of type-I error. In other words, it is desired to find a sequence that *minimizes*

$$\liminf_{T\to\infty} \sup_{\boldsymbol{\eta}\in\Theta_1} \frac{1}{T} \log P_{\boldsymbol{\eta}}\left[r_T = 0\right]$$

uniformly over $\boldsymbol{\eta} \in \Theta_1$ subject to

$$\liminf_{T\to\infty} \sup_{\boldsymbol{\eta}\in\Theta_0} \frac{1}{T} \log P_{\boldsymbol{\eta}} \left[r_T = 1\right] \leq -K$$

for some given $K > 0$. In other words, it is desired to find a sequence of tests satisfying a generalized Neyman-Pearson criterion.[14]

In this connection, a Chernoff-type loss function involving the balancing of both type-I and type-II errors is considered. In particular, for some constant $\pi \in (0, 1)$, consider the loss function

$$W^{\pi}_{\boldsymbol{\eta}}(r) \equiv \pi 1\left(\boldsymbol{\eta} \in \Theta_0,\, r = 1\right) + (1 - \pi) 1\left(\boldsymbol{\eta} \in \Theta_1,\, r = 0\right). \tag{12}$$

It is clear that this actually implies a higher level of generality than the generalized Neyman-Pearson framework just presented. On the other hand, it is easy to see that varying the weighting constant π in the unit interval generates the family of generalized Neyman-Pearson tests in this context.[15]

A tight lower bound to the logarithmic LD-risk function involving $W^{\pi}_{\boldsymbol{\eta}}(\cdot)$ is available as a consequence of Theorem 1. First recall the following definition of a functional smoothness class as used in Robinson (1988):

Definition 4. *$\mathcal{G}(\mu, \alpha)$, where $\alpha > 0$, $\mu > 0$, is the class of all functions $g : \mathbb{R}^q \to \mathbb{R}$ such that*

1. *$g(\cdot)$ is $(m-1)$-times partially differentiable with $m - 1 \leq \mu \leq m$;*

2. *there exists an $\eta > 0$ such that*

$$\sup_{y\in\phi_{z\eta}} \frac{|g(y) - g(z) - Q_g(y, z)|}{|y - z|^{\mu}} \leq G_g(z)$$

for all z, $\phi_{z\eta} \equiv \{y : |y - z| < \eta\}$, where $Q_g = 0$ when $m = 1$, and Q_g is otherwise an $(m-1)$th-degree homogeneous polynomial in $y - z$ with coefficients equal to the partial derivatives of g at z of orders 1 through $m - 1$;

3. *$g(z)$, each of its partial derivatives of order $m - 1$ and less, and $G_g(z)$ have finite moments of order α.*

[14] Cf. e.g., Hoeffding (1965) or Zeitouni and Gutman (1991) for special cases.

[15] Alternatively, interpret π as a Lagrange multiplier in the appropriate constrained optimization problem.

Now let $\bar{\mathcal{G}}(\mu, \alpha)$ denote the subset of $\mathcal{G}(\mu, \alpha)$ consisting of functions that are uniformly bounded on $[0, 1]$.

Assume that $\Theta \subset \bar{\mathcal{G}}(2, 4)$. Note that by the Arzelà-Ascoli Theorem, $\bar{\mathcal{G}}(2, 4)$ is compact in the space $C[0, 1]$ equipped with the uniform topology. As such, for any bandwidth sequence $\{h\}$, with $h \equiv h_T \to 0$ and $Th_T \to \infty$ as $T \to \infty$,

$$\limsup_{T\to\infty} \sup_{\boldsymbol{\eta}\in\bar{\mathcal{G}}(2,4)} \left\| \Xi_{T,\hat{\boldsymbol{\eta}}} - \hat{\xi}_{\boldsymbol{\eta},h_T}(F) \right| = 0$$

for all $F \in \mathcal{F}^{(p+1)}$, and where $\xi_{\hat{\boldsymbol{\eta}},h_T}\left(\hat{F}^{(p+1)}_{T|p}\right)$ denotes the GCLR test statistic. In addition, the following uniform version of the exponential tightness condition in the definition of the dominated LDP for $\langle \mathcal{E}_T, P_T \rangle_{T\geq 1}$ holds:

$$\lim_{M\to\infty} \limsup_{T\to\infty} \sup_{\boldsymbol{\eta}\in\bar{\mathcal{G}}(2,4)} E_T^{\frac{1}{T}} \left[\exp\left(T\Xi_{T,\boldsymbol{\eta}}\right) 1\left(\Xi_{T,\boldsymbol{\eta}} > M\right)\right] = 0.$$

An adaptation of the proof of Puhalskii and Spokoiny (1998, Theorem 3.1) shows that these facts imply that the lower bound in Theorem 1 is both tight and attainable by a suitable LD-efficient decision—namely, one based on the GCLR statistic. These conclusions are essentially the statement of Theorem 2 below.

Note the maximum logarithmic LD-risk function associated with the Chernoff-type loss function given above in (12):

$$R_{\pi,T}(r_T) \equiv \max\left\{ \sup_{\boldsymbol{\eta}\in\Theta_0} \frac{1}{T} \log\left[\pi P_{\boldsymbol{\eta}}\left[r_T = 1\right]\right], \sup_{\boldsymbol{\eta}\in\Theta_1} \frac{1}{T} \log\left[(1-\pi) P_{\boldsymbol{\eta}}\left[r_T = 0\right]\right] \right\}.$$

An adaptation of Puhalskii and Spokoiny (1998, Theorems 3.1 & 4.3) yields the following result:

Theorem 2. *Under the maintained assumptions regarding the model under consideration here,*

$$\liminf_{T\to\infty} \inf_{r_T\in\mathcal{R}_T} R_{\pi,T}(r_T) \geq R^*_\pi,$$

where

$$R^*_\pi \equiv \sup_{F\in\mathcal{F}^{(p+1)}} \min\left\{ \log\pi + \sup_{\boldsymbol{\eta}\in\Theta_0} \left[\xi_{\boldsymbol{\eta}}(F) - I_{DZ}(F)\right], \log(1-\pi) + \sup_{\boldsymbol{\eta}\in\Theta_1} \left[\xi_{\boldsymbol{\eta}}(F) - I_{DZ}(F)\right] \right\}.$$

Moreover, the GCLR test is LD-efficient in the sense that for $h \equiv h_T$ such that $h_T \to 0$, $Th_T \to \infty$ as $T \to \infty$,

$$\lim_{T\to\infty} R_{\pi,T}\left(\hat{r}^*_{T,h}\right) = R^*_\pi,$$

where

$$\hat{r}^*_{T,h} \equiv 1\left[\xi_{\hat{\boldsymbol{\eta}},h}\left(\hat{F}^{(p+1)}_{T|p}\right) > \log\left(\frac{\pi}{1-\pi}\right)\right].$$

Proof. The proof is essentially identical to that of Puhalskii and Spokoiny (1998, Theorem 4.3). Note, however, that the compactness of $\bar{\mathcal{G}}(2,4)$ implies that the sub-problem

$$\min\left\{\log\pi + \sup_{\boldsymbol{\eta}\in\Theta_0}\xi_{\boldsymbol{\eta}}(F), \log(1-\pi) + \sup_{\boldsymbol{\eta}\in\Theta_1}\xi_{\boldsymbol{\eta}}(F)\right\},$$

where $F \in \mathcal{F}^{(p+1)}$ is fixed, has a solution of the form

$$\hat{r}_F \equiv 1\left[\sup_{\boldsymbol{\eta}\in\Theta_1}\xi_{\boldsymbol{\eta}}(F) > \log\left(\frac{\pi}{1-\pi}\right) + \sup_{\boldsymbol{\eta}\in\Theta_0}\xi_{\boldsymbol{\eta}}(F)\right]. \tag{13}$$

Note also that compactness of the parameter space and consistency of the estimator $\hat{\boldsymbol{\eta}}$ described earlier in Section 2 yield the uniformity result

$$\limsup_{T\to\infty}\sup_{\boldsymbol{\eta}\in\bar{\mathcal{G}}(2,4)}\left|\xi_{\hat{\boldsymbol{\eta}},h}(F) - \xi_{\boldsymbol{\eta}}(F)\right| = 0$$

for all $F \in \mathcal{F}^{(p+1)}$ and bandwidths $h \equiv h_T$ such that $h_T \to 0$, $Th_T \to \infty$ as $T \to \infty$. As such, the GCLR-based test proposed here inherits the LD-efficiency property of the conditional likelihood-ratio test given in (13). □

A Proof of Lemma 2

Note that under P_T,

$$\Xi_{T,\boldsymbol{\eta}} = \frac{1}{2}\log\left(\frac{CSS_0}{CSS_1}\right) \approx \frac{1}{2}\cdot\frac{CSS_0 - CSS_1}{CSS_1} = \frac{1}{2}\cdot\frac{\frac{1}{\sigma^2(T-p)}(CSS_0 - CSS_1)}{\frac{1}{\sigma^2(T-p)}CSS_1}.$$

We have for some $\boldsymbol{\eta}_0 \in \Theta_0$ that

$$\frac{1}{(T-p)\sigma^2}\left(CSS_0 - CSS_1\right)$$

$$
\begin{aligned}
= \quad & \frac{2}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\boldsymbol{w}_t^\top\boldsymbol{\eta}\left(\frac{t}{T}\right)\left(y_t-\boldsymbol{w}_t^\top\boldsymbol{\eta}_0\right)+\frac{2}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\boldsymbol{w}_t^\top\boldsymbol{\eta}\left(\frac{t}{T}\right)\boldsymbol{w}_t^\top\boldsymbol{\eta}_0 \\
& -\frac{1}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\boldsymbol{\eta}^\top\left(\frac{t}{T}\right)\boldsymbol{w}_t\boldsymbol{w}_t^\top\boldsymbol{\eta}\left(\frac{t}{T}\right)-\frac{2}{\sigma^2(T-p)}\sum_{t=p+1}^{T}\boldsymbol{w}_t^\top\boldsymbol{\eta}_0\left(y_t-\boldsymbol{w}_t^\top\boldsymbol{\eta}_0\right) \\
& -\frac{2}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\boldsymbol{w}_t^\top\boldsymbol{\eta}_0\boldsymbol{w}_t^\top\boldsymbol{\eta}_0+\frac{1}{\sigma^2(T-p)}\sum_{t=p+1}^{T}\boldsymbol{\eta}_0^\top\boldsymbol{w}_t\boldsymbol{w}_t^\top\boldsymbol{\eta}_0 \\
= \quad & \frac{2}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\boldsymbol{w}_t^\top\boldsymbol{\eta}\left(\frac{t}{T}\right)\epsilon_t+\frac{2}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\boldsymbol{w}_t^\top\boldsymbol{\eta}\left(\frac{t}{T}\right)\boldsymbol{w}_t^\top\boldsymbol{\eta}_0 \\
& -\frac{1}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\boldsymbol{\eta}\left(\frac{t}{T}\right)^\top\boldsymbol{w}_t\boldsymbol{w}_t^\top\boldsymbol{\eta}\left(\frac{t}{T}\right)-\frac{2}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\boldsymbol{w}_t^\top\boldsymbol{\eta}_0\epsilon_t \\
& -\frac{1}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\boldsymbol{\eta}_0^\top\boldsymbol{w}_t\boldsymbol{w}_t^\top\boldsymbol{\eta}_0 \\
= \quad & \frac{2}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\boldsymbol{w}_t^\top\left(\boldsymbol{\eta}\left(\frac{t}{T}\right)-\boldsymbol{\eta}_0\right)\epsilon_t \\
& -\frac{1}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\left(\boldsymbol{\eta}\left(\frac{t}{T}\right)-\boldsymbol{\eta}_0\right)^\top\boldsymbol{w}_t\boldsymbol{w}_t^\top\left(\boldsymbol{\eta}\left(\frac{t}{T}\right)-\boldsymbol{\eta}_0\right);
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{1}{(T-p)\sigma^2}CSS_1 \\
= \quad & \frac{1}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\left[y_t-\boldsymbol{w}_t^\top\boldsymbol{\eta}\left(\frac{t}{T}\right)\right]^2 \\
= \quad & \frac{1}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\left[\epsilon_t-\boldsymbol{w}_t^\top\left(\boldsymbol{\eta}\left(\frac{t}{T}\right)-\boldsymbol{\eta}_0\right)\right]^2 \\
= \quad & \frac{1}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\epsilon_t^2-\frac{2}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\boldsymbol{w}_t^\top\left(\boldsymbol{\eta}\left(\frac{t}{T}\right)-\boldsymbol{\eta}_0\right)\epsilon_t \\
& +\frac{1}{(T-p)\sigma^2}\sum_{t=p+1}^{T}\left(\boldsymbol{\eta}\left(\frac{t}{T}\right)-\boldsymbol{\eta}_0\right)^\top\boldsymbol{w}_t\boldsymbol{w}_t^\top\left(\boldsymbol{\eta}\left(\frac{t}{T}\right)-\boldsymbol{\eta}_0\right).
\end{aligned}
$$

Now define for $k\in[0,1]$ and $q\geq 1$ the empirical distribution

$$
\hat{F}_{T|p}^{(q)}(\epsilon_1,\ldots,\epsilon_q,k)\equiv\frac{1}{T-p}\sum_{t=p+1}^{\lfloor Tk\rfloor}\prod_{s=1}^{q}1\left(\epsilon_{t-s+1}\leq\epsilon_s\right)
$$

and set

$$\bar{g}\left(\frac{t}{T}, \epsilon_{t-1}, \dots, \epsilon_{t-p}\right) \equiv \bar{w}_t^\top \left(\boldsymbol{\eta}\left(\frac{t}{T}\right) - \boldsymbol{\eta}_0\right)\frac{1}{\sigma}.$$

Since $\bar{g}\left(\cdot, \epsilon_{t-1}, \dots, \epsilon_{t-p}\right)$ is continuous (by assumption) we have for large T the approximation

$$\begin{aligned}
& \frac{1}{(T-p)\sigma^2}\left(CSS_0 - CSS_1\right) \\
\approx\; & 2\int_0^1 \overbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}}}^{p+1} \bar{g}\left(k, \epsilon_1, \dots, \epsilon_p\right)\epsilon_{p+1}\hat{F}^{(p+1)}_{T|p}\left(d\epsilon_1, \dots, d\epsilon_{p+1}, dk\right) \\
& - \int_0^1 \overbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}}}^{p+1} \bar{g}^2\left(k, \epsilon_1, \dots, \epsilon_p\right)\hat{F}^{(p)}_{T|p}\left(d\epsilon_1, \dots, d\epsilon_p, dk\right) \\
\equiv\; & \xi_1\left(\hat{F}^{(p+1)}_{T|p}\right). \qquad (14)
\end{aligned}$$

Set $\mathcal{F}^{(q)}$ to be the space of all distribution functions $F \equiv F(\boldsymbol{\epsilon}, k)$ on $\mathbb{R}^q \times [0,1]$. Endow $\mathcal{F}^{(q)}$ with the weak topology. Let $\mathcal{F}^{(q)}_0$ denote the subset of $\mathcal{F}$ consisting of functions that are both absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^q \times [0,1]$ and also have densities $p_k(\boldsymbol{\epsilon})$ such that $\int_{\mathbb{R}^q} p_k(\boldsymbol{\epsilon})d\boldsymbol{\epsilon} = 1$ for all $k \in [0,1]$. From a result of Dembo and Zajic (1995), the sequence $\left\{\mathcal{L}\left(\hat{F}^{(q)}_{T|p}\middle| P_T\right) : T \geq 1\right\}$ obeys the LDP in $\mathcal{F}^{(q)}$ with rate function

$$I_{DZ}(F) \equiv \begin{cases} \int_0^1 \int_{\mathbb{R}^q} \log \frac{p_k(\boldsymbol{\epsilon})}{p(\boldsymbol{\epsilon})} p_k(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon} dk, & \text{if } F \in \mathbb{F}^{(q)}_0; \\ \infty, & \text{otherwise.} \end{cases}$$

We have for large T

$$\begin{aligned}
& \frac{1}{(T-p)\sigma^2} CSS_1 \\
\approx\; & \frac{1}{\sigma^2}\int_0^1 \hat{\epsilon}^2 \hat{F}^{(1)}_{T|p}(d\epsilon, dk) - 2\int_0^1 \overbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}}}^{p+1} \bar{g}(k, \epsilon_1, \dots, \epsilon_p)\epsilon_{p+1}\hat{F}^{(p+1)}_{T|p}(d\epsilon_1, \dots, d\epsilon_{p+1}, dk) \\
& + \int_0^1 \overbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}}}^{p} \bar{g}^2(\epsilon_1, \dots, \epsilon_p, k)\hat{F}^{(p)}_{T|p}(d\epsilon_1, \dots, d\epsilon_p, dk) \\
\equiv\; & \xi_2\left(\hat{F}^{(p+1)}_{T|p}\right),
\end{aligned}$$

a continuous functional in $\mathcal{F}^{(p+1)}$. Similarly, $\xi_1(\cdot)$ in (14) above is also a continuous functional in $\mathcal{F}^{(p+1)}$. Noting that $\frac{1}{(T-p)\sigma^2}\sum_{t=p+1}^T \epsilon_t^2 \xrightarrow{p} 1$ and that $\int_0^1 \overbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}}}^{p} \bar{g}^2(\epsilon_1, \dots, \epsilon_p, k)dk \geq 0$, the contraction principle yields the LDP for $\left\{\mathcal{L}\left(\Xi_{T,\Theta}\middle| P_T\right) : T \geq 1\right\}$.

Now consider the verification of part 2 of the definition of the dominated LDP for $\langle \mathcal{E}_T, P_T \rangle_{T \geq 1}$. Since

$$1\left(\Xi_{T,\boldsymbol{\eta}} > M\right) \leq \exp\left(T\Xi_{T,\boldsymbol{\eta}}\right)\exp(-TM),$$

we have

$$E_T^{\frac{1}{T}}\left[\exp\left(T\Xi_{T,\boldsymbol{\eta}}\right)1\left(\Xi_{T,\boldsymbol{\eta}} > M\right)\right] \leq \exp\left(-M\right)E_T^{\frac{1}{T}}\left[\exp\left(2T\Xi_{T,\boldsymbol{\eta}}\right)\right]$$

for all $\boldsymbol{\eta} \in \Theta$. But

$$E_T^{\frac{1}{T}}\left[\exp\left(2T\Xi_{T,\boldsymbol{\eta}}\right)\right] \approx \left\{E_T\left[\exp\left(\frac{T\left(CSS_0 - CSS_1\right)}{CSS_1}\right)\right]\right\}^{\frac{1}{T}} = O(1),$$

since, as noted above in Section 2, $\frac{T(CSS_0 - CSS_1)}{CSS_1}$ is asymptotically normal under H_0.

It follows that

$$\begin{aligned} E_T^{\frac{1}{T}}\left[\exp\left(T\Xi_{T,\boldsymbol{\eta}}\right)1\left(\Xi_{T,\boldsymbol{\eta}} > M\right)\right] &\leq \exp(-M)\left\{E_T\left[\exp\left(\frac{T(CSS_0 - CSS_1)}{CSS_1}\right)\right]\right\}^{\frac{1}{T}} \\ &\to 0 \end{aligned}$$

as $M \to \infty$. This completes the proof.

References

Andrews, D. W. K., and W. Ploberger (1994) 'Optimal tests when a nuisance parameter is present only under the alternative.' *Econometrica* 62, 1383–1414

Bai, J. (1996) 'Testing for parameter constancy in linear regressions: An empirical distribution approach.' *Econometrica* 64, 597–622

Carroll, R. J., J. Fan, I. Gijbels, and M. P. Wand (1997) 'Generalized partially linear single-index models.' *Journal of the American Statistical Association* 92, 477–489

Chu, C., and H. White (1992) 'A direct test for changing trend.' *Journal of Business and Economic Statistics* 10, 289–299

De Jong, P. (1987) 'A central limit theorem for generalized quadratic forms.' *Probability Theory and Related Fields* 75, 261–277

Dembo, A., and O. Zeitouni (1998) *Large Deviations Techniques and Applications,* second ed. (New York: Springer)

Dembo, A., and T. Zajic (1995) 'Large deviations: From empirical mean and measure to partial sums process.' *Stochastic Processes and their Applications* 57, 191–224

Fan, J., C. Zhang, and J. Zhang (2001) 'Generalized likelihood ratio statistics and Wilks phenomenon.' *Annals of Statistics* 29, 153–193

Hoeffding, W. (1965) ‘Asymptotically optimal tests for multinomial distributions.’ *Annals of Mathematical Statistics* 36, 369–400

Juhl, T., and Z. Xiao (2005) ‘A nonparametric test for changing trends.’ *Journal of Econometrics* 127, 179–199

Kuan, C. M. (1998) ‘Tests for changes in models with polynomial trends.’ *Journal of Econometrics* 84, 75–92

Kuan, C. M., and K. Hornik (1995) ‘The generalized fluctuation test: A unifying view.’ *Econometric Reviews* 14, 135–161

Le Cam, L. (1986) *Asymptotic Methods in Statistical Decision Theory* (New York: Springer-Verlag New York)

Perron, P. (1990) ‘Testing for a unit root in a time series with changing mean.’ *Journal of Business and Economic Statistics* 8, 153–162

__ (2006) ‘Dealing with structural breaks.’ In *Palgrave Handbook of Econometrics,* ed. K. Patterson and T. C. Mills, vol. 1: Econometric Theory (Basingstoke, Hampshire, U.K.: Palgrave Macmillan) pp. 278–352

Ploberger, W., and W. Kramer (1996) ‘A trend-resistant test for structural change based on OLS residuals.’ *Journal of Econometrics* 70, 175–186

Priestley, M., and M. Chao (1972) ‘Non-parametric function fitting.’ *Journal of the Royal Statistical Society, Series B (Methodological)* 34, 386–392

Puhalskii, A. (1993) ‘On the theory of large deviations.’ *Theory of Probability and its Applications* 38, 490–497

Puhalskii, A., and V. Spokoiny (1998) ‘On large-deviation efficiency in statistical inference.’ *Bernoulli* 4, 203–272

Ripatti, A., and P. Saikkonen (2001) ‘Vector autoregressive processes with nonlinear time trends in cointegrating relations.’ *Macroeconomic Dynamics* 5, 577–597

Robinson, P. M. (1988) ‘Root-n-consistent semiparametric regression.’ *Econometrica* 56, 931–954

Saikkonen, P. (2001a) ‘Consistent estimation in cointegrated vector autoregressive processes with nonlinear time trends in cointegrating relations.’ *Econometric Theory* pp. 296–326

__ (2001b) ‘Statistical inference in cointegrated vector autoregressive processes with nonlinear time trends in cointegrating relations.’ *Econometric Theory* pp. 327–356

Strasser, H. (1985) *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory* (Berlin: Walter de Gruyter)

Varadhan, S. R. S. (1984) *Large Deviations and Applications* (Philadelphia: Society for Industrial and Applied Mathematics)

Vogelsang, T. J. (1997) 'Wald-type tests for detecting breaks in the trend function of a dynamic time series.' *Econometric Theory* 13, 818–849

— (1998) 'Trend function hypothesis testing in the presence of serial correlation.' *Econometrica* 66, 123–148

— (1999) 'Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series.' *Journal of Econometrics* 88, 283–299

Zeitouni, O., and M. Gutman (1991) 'On universal hypotheses testing via large deviations.' *IEEE Transactions on Information Theory* 37, 285–290