

Efficient Semiparametric Detection of Changes in Trend

Chuan Goh*

8 June 2008

*The author is grateful to Don Andrews, Stephen Cosslett, Robert de Jong, Jean-Marie Dufour, Joon Park, Peter Phillips and Shinichi Sakata for their comments on an earlier version of this paper. As is customary in footnotes of this sort, the “usual disclaimer” is inserted. The author really *does* mean it. This research was supported in part by the Connaught Fund of the University of Toronto. Address correspondence to: Department of Economics, University of Toronto, 150 St. George St., Toronto, ON, Canada, M5S 3G7. goh@economics.utoronto.ca, <http://www.chuangoh.org>.

Abstract

This paper proposes a test for parameter variation in the context of a dynamic time-series model that is stationary about a deterministic trend function. The trend function is taken under the null to be piecewise Lipschitz-continuous, and the approach taken explicitly seeks an appropriately defined semiparametric power bound against contiguous alternatives to the null of parameter stability. This is done by deriving the efficient influence function of a least dispersed regular estimator of the total variation of each component of the trend parameter. This derivation is then used in the construction of a Wald-type test for parameter variation based on an asymptotically linear estimator of the vector of total variations whose influence function coincides with the efficient influence function. The resulting testing procedure is effectively an omnibus specification test for the adequacy of a model involving stationary fluctuations about a deterministic trend that evolves smoothly in intervals between the occurrences of an at most finite number of break points. The utility of this procedure to detect discrete changes or continuous variation in the trend parameter as well as unit-root alternatives is illustrated in simulation experiments. This paper also includes an application examining the adequacy of a linear trend-stationary specification with infrequent trend breaks for the historical evolution of U.S. real output.

JEL Classification: C12, C14, C22

KEYWORDS: Structural change, trend-stationary processes, nonparametric regression, efficient influence function

1 Introduction

A particularly important class of model in econometric time-series analysis involves nonstationary processes characterized by stationary fluctuations about a deterministic trend function. In particular, suppose that data Y_1, \dots, Y_T are observed from the model

$$Y_t = \mathbf{d} \left(\frac{t}{T} \right)^\top \boldsymbol{\gamma} \left(\frac{t}{T} \right) + u_t, \quad t = 1, \dots, T, \quad (1)$$

where u_t is generated by a mean-zero stationary process. The deterministic component in (1) involves both a known function $\mathbf{d} : [0, 1] \rightarrow \mathbb{R}^k$ and an unknown parameter $\boldsymbol{\gamma} : [0, 1] \rightarrow \mathbb{R}^k$. The specification of “trend-stationarity” in (1) has generally been associated with a traditional “Keynesian” view of macroeconomic fluctuations alleging that the dynamic behaviour of most macroeconomic series is well described by stationary fluctuations about a smoothly evolving deterministic trend.¹ Beginning with the seminal study of Nelson and Plosser (1982), challenges to this traditional view have typically argued that the evolution of most macroeconomic series is better described by “difference-stationarity”, or the behaviour of a unit-root process. In this paradigm, the trend function in (1) is random, and one-off shocks to the system have permanent effects. This challenge to the traditional conception of macroeconomic fluctuations has been influential in theoretical macroeconomics, as well as serving to inspire the development of an extensive empirical and methodological literature in econometrics.²

The “segmented trend” model of Gallant and Fuller (1973) provides a formulation intermediate between the trend-stationary and difference-stationary paradigms. In this case, the trend function is neither a smooth function of time as in the trend-stationary view nor characterized by breaks every period, as would be the case if the data were realizations of a unit-root process. In particular, a segmented trend model involves dynamic behaviour similar to a trend-stationary model between

¹Cf. e.g., Kydland and Prescott (1980); Blanchard (1981).

²Empirical studies essentially corroborating the results of Nelson and Plosser (1982) using the same Dickey–Fuller methodology but different datasets include Stulz and Wasserfallen (1985); Wasserfallen (1986) and Perron (1988). The notion of dynamic path dependence inherent in the difference-stationarity paradigm features in the model proposed by Blanchard and Summers (1986) to describe “hysteresis” in European unemployment in the 1980s as well as in real business cycle models of macroeconomic fluctuations (e.g., King et al., 1991). The theoretical literature on unit-root testing and the behaviour of statistical models involving integrated variables is vast. Recent surveys on unit-root testing and cointegration analysis can be found in Haldrup and Jansson (2006) and Johansen (2006), respectively.

the occurrence of infrequent trend breaks, which in the context of (1) show up as discontinuities in one or more components of $\mathbf{d}(\cdot)$. Segmented trend-stationarity may be a more realistic reflection of the actual evolution of macroeconomic series than what one would get with a unit-root process when trend breaks are interpreted as the consequence of infrequent permanent shocks.³ The statistical question treated in this paper is the development of a procedure to test whether the trend-function parameter γ in (1) is in fact time-varying when the deterministic component \mathbf{d} exhibits at most a finite number of breaks at time periods selected by the researcher. As such, this paper develops a lack-of-fit test for the adequacy of a given segmented-trend model.

In this connection, the parameter of interest is taken to be the k -vector $\chi(\gamma)$, whose j th component is given by

$$\chi_j(\gamma) \equiv \int_0^1 |\gamma'_j(s)| ds,$$

i.e., the total variation of the j th component of γ over the unit interval.⁴ Note that unit-root behaviour in the dependent variable is associated with unpredictable changes each period in one or more components of the trend parameter, which translates into a value of $\sum_{j=1}^k \chi_j(\gamma) = \infty$. On the other hand, the existence of both a finite number of discrete breaks or smooth continuous change in one or more components of γ would naturally be associated with $\sum_{j=1}^k \chi_j(\gamma) \in (0, \infty)$. Under the null of parameter stability, on the other hand, each of the additive components $d_j(\frac{t}{T}) \gamma_{j0}$ ($j \in \{1, \dots, k\}$) of the trend function evolves smoothly as a function of time between any two consecutive breakpoints in d_j , if in fact such breakpoints have been “built-in” by the researcher as the presumed reflection of a belief in the occurrence at certain periods of significant exogenous shocks having permanent effects.

The approach taken in this paper involves the development from first principles of a method for testing the hypothesis that $\chi(\gamma) = \mathbf{0}$ for all γ in the space of \mathbb{R}^k -valued functions on the unit interval. The resulting test is effectively an omnibus

³Potential examples of such infrequent shocks having permanent effects include natural disasters, abrupt policy changes and large sudden movements in asset prices. Cf. Perron (1989); Rappoport and Reichlin (1989); Perron (1990) and Perron and Wada (2006).

⁴The vector consisting of the $L_2(\text{Leb}[0, 1])$ -norm of each component of γ may also be used. (Here $\text{Leb}[0, 1]$ denotes Lebesgue measure on the unit interval.) Total variation is used here because it imposes less of a penalty on regions of the unit interval where $|\gamma'_j(s)|$ is large, thus allowing for a commensurately greater degree of “roughness” in γ that is still compatible with a decision in favour of a segmented trend-stationary specification in (1).

lack-of-fit test for segmented trend stationarity with power against alternatives involving single or multiple unmodelled breaks in the trend function, as well as alternatives involving continuous unmodelled structural change or the presence of unit roots. The test proposed here will also be shown to be efficient in the sense of asymptotically attaining a localized uniform power bound against contiguous alternatives to the null of parameter stability.

The literature on detecting time variation in the trend-function parameters is vast, and is comprehensively surveyed in Perron (2006).⁵ The proposal presented in this paper is distinguished by the fact that it is explicitly designed—in the absence of strong assumptions imposed on the data-generating mechanism for the stochastic component—to attain a relevant semiparametric efficiency bound for local alternatives to the null of trend-parameter stability belonging to an appropriately defined tangent space.

In what follows, an efficient semiparametric detection procedure for time variation in the trend-function parameter is described. The test is based on the asymptotic behaviour of an efficient semiparametric estimator of the total variation of the components of the trend parameter $\gamma(\cdot)$ over the unit interval. In particular, this Wald-type test statistic is shown to be both regular over the parameter space as well as asymptotically linear with influence function equal to the efficient influence function. A description of this basic idea in the specific context of the model given in (1) appears in Section 3. This description is preceded in Section 2 by a more general discussion of the semiparametric efficiency criterion used in this paper. Details on constructing a feasible efficient test statistic are provided in Section 4. Simulation evidence regarding how the feasible testing procedure described in Section 4 performs in small samples is provided in Section 5, while Section 6 gives the results of applications of the technique developed here to assessing the adequacy of a linear trend-stationary specification with infrequent trend breaks to the historical evolution of real GDP in the United States. Section 7 concludes. Proofs of most of the theoretical results given in Sections 3 and 4 are collected in the appendix.

Notation and terminology

This section summarizes for convenience certain notational conventions and definitions that are used extensively in the remainder of this paper.

⁵Recent proposals for detecting structural change in trend functions include those of Chu and White (1992); Kuan and Hornik (1995); Bai (1996); Ploberger and Kramer (1996); Vogelsang (1997); Kuan (1998); Vogelsang (1998, 1999); Juhl and Xiao (2005) and Wu and Zhao (2007).

1. For a vector $\mathbf{x} \in \mathbb{R}^k$, $\|\mathbf{x}\|$ denotes the Euclidean norm of \mathbf{x} . For a random vector \mathbf{X} taking values in \mathbb{R}^k , $\|\mathbf{X}\| \equiv \left(E \left[\|\mathbf{X}\|^2\right]\right)^{\frac{1}{2}}$.
2. Consider a measurable space (X, \mathcal{B}) . Let $\{P_0, P_1, P_2, \dots\}$ be a collection of probability measures defined on (X, \mathcal{B}) dominated by some σ -finite measure μ , and let $\{p_0, p_1, p_2, \dots\}$ denote the corresponding collection of densities with respect to μ . For $P \in \{P_0, P_1, P_2, \dots\}$, define $\dot{P} \equiv \{h \in L_2(P) : \int h dP = 0\}$.

For each $n \geq 1$, consider the mapping $P_n \rightarrow s\left(\frac{1}{\sqrt{n}}\right)$, where $s\left(\frac{1}{\sqrt{n}}\right) \equiv \sqrt{p_n}$. Also define $s(0) \equiv \sqrt{p_0}$. Suppose that $s(\cdot)$ is Fréchet-differentiable in $L_2(\mu)$ at $s(0)$, i.e., that there exists a linear operator $\dot{s}_0 : [0, 1] \rightarrow \mathbb{R}$ such that for every sequence of positive numbers $\{\epsilon_n\} \rightarrow 0$ and a sequence $\{d_n\}$ with each $|d_n| < \infty$ with

$$\begin{aligned} \left\| \frac{s(\epsilon_n d_n) - s(0)}{\epsilon_n} - \dot{s}_0(d_n) \right\| &= \int \left(\frac{s(\epsilon_n d_n) - s(0)}{\epsilon_n} - \dot{s}_0(d_n) \right)^2 d\mu \\ &\rightarrow 0. \end{aligned}$$

It follows that for $\epsilon_n \equiv \frac{1}{\sqrt{n}}$, $d_n \equiv 1$ we have the conventional formulation of *differentiability in quadratic mean*, i.e.,

$$\int [\sqrt{n}(\sqrt{p_n} - \sqrt{p_0}) - \dot{s}_0(1)]^2 d\mu \rightarrow 0.$$

By Bickel et al. (1993, Example 3.2.1), $\dot{s}_0(1) = \frac{1}{2}h\sqrt{p_0}$ for some $h \in \dot{P}$, and in this case the sequence $\{P_n : n \geq 1\}$ has a *tangent* h at P_0 .

3. Now consider a linear operator T between Banach spaces $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$. First suppose that A is a subset of V and that A has an associated *tangent space* \dot{A} , i.e., that for every $d \in \dot{A}$ there is a corresponding *local sequence* $\{d_n\}$ with $\|d_n - d\|_V \rightarrow 0$ and a real-valued sequence $\{\epsilon_n\} \rightarrow 0$ such that $a + \epsilon_n d_n \in A$ for every $n \geq 1$ and every $a \in A$. If for some $a \in A$ there is a linear operator $\dot{T}_a : A \rightarrow W$ such that for every $\{\epsilon_n\} \rightarrow 0$ and $d \in \dot{A}$ with corresponding local sequence $\{d_n\} \rightarrow d$ we have

$$\left\| \frac{T(a + \epsilon_n d_n) - T(a)}{\epsilon_n} - \dot{T}_a(d_n) \right\|_W \rightarrow 0, \quad (2)$$

then the linear operator T is said to be *pathwise-* (or *Hadamard-*) *differentiable* along \dot{A} at a .

In what follows, consideration is focused on a slightly simpler situation implied by (2), namely the special case where $\epsilon_n = \frac{1}{\sqrt{n}}$ in (2).

4. A function g defined on an interval I of the real line is said to be of *bounded variation* if

$$\sup \left\{ \sum_i |g(t_i) - g(t_{i-1})| \right\} < \infty,$$

where the supremum is taken over all partitions $\{\dots < t_{i-1} < t_i < \dots\}$ of I .

2 Efficient semiparametric estimation and testing

The detection procedure proposed in this paper for time variation in the trend-function parameter is derived explicitly from an efficient estimator of the total variation of the trend parameter over the unit interval. This section is expository in nature and presents in general terms the efficiency criterion adopted in this paper, which is associated with minimum dispersion of regular estimators of a parameter of interest. The theoretical results stated here are fairly well-known, and as such are presented without proof.⁶

Let (X, \mathcal{B}) again denote a measurable space. let $\mathcal{P} \equiv \{P_\theta : \theta \in \Theta\}$ be a family of probability measures defined on (X, \mathcal{B}) . Let κ be a functional between Θ and \mathbb{R}^m . The initial statistical question concerns how well one can estimate $\kappa(\theta_0)$ for some $\theta_0 \in \Theta$ given sequences of observations generated by $\{P_{\theta_n} : n \geq 1\} \subset \mathcal{P}$.

Assume that the parameter space Θ is a subset of a Banach space $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$. Let $\dot{\Theta}$ be the tangent space corresponding to Θ . The following two basic assumptions are made:

Assumption 1. For every $\delta \in \dot{\Theta}$ with associated local sequence $\{\delta_n\}$, the collection $\left\{P_{\theta_0 + \frac{1}{\sqrt{n}}\delta_n}\right\}$ has tangent $\dot{P}_{\theta_0}[\delta]$ at θ_0 , where $\dot{P}_{\theta_0}[\cdot]$ denotes a linear operator on \mathcal{H} into the the space

$$\dot{P}_{\theta_0} \equiv \left\{ h \in L_2(P_{\theta_0}) : \int h dP_{\theta_0} = 0 \right\}.$$

Assumption 2. The functional κ is pathwise-differentiable along $\dot{\Theta}$ at θ_0 , i.e., there exists a linear operator $\dot{\kappa}_{\theta_0} : \mathcal{H} \rightarrow \mathbb{R}^m$ such that for every $\delta \in \dot{\Theta}$,

$$\left\| \sqrt{n} \left(\kappa \left(\theta_0 + \frac{1}{\sqrt{n}}\delta \right) - \kappa(\theta_0) \right) - \dot{\kappa}_{\theta_0}(\delta) \right\|_{\mathcal{H}} \rightarrow 0 \quad (3)$$

as $n \rightarrow \infty$.

Note the following definitions for convenience:

Definition 1. 1. The closed linear span of $\left\{ \dot{P}_{\theta_0}[\delta] : \delta \in \dot{\Theta} \right\}$ is called the **tangent space** of \mathcal{P} and will be denoted by \mathcal{T} .

⁶Further details can be obtained in Pfanzagl and Wefelmeyer (1982); Begun et al. (1983) and in the monograph of Bickel et al. (1993).

2. An **estimator** is a sequence $\{\kappa_n : n \geq 1\}$ such that for every $n \geq 1$, κ_n is a measurable function on X^n into \mathcal{R}^m .
3. An estimator $\{\kappa_n\}$ is said to be **(locally) regular** at θ_0 if there is a distribution Q_0 such that for every $\delta \in \Theta$,

$$\mathcal{L} \left(\sqrt{n} \left(\kappa_n - \kappa \left(\theta_0 + \frac{1}{\sqrt{n}} \delta \right) \right) \middle| P_{\theta_0 + \frac{1}{\sqrt{n}} \delta}^n \right) \rightarrow Q_0,$$

where Q_0 does not depend on δ .

4. An estimator $\{\kappa_n\}$ is said to be **asymptotically linear** at θ_0 with **influence function** $\psi_0 : X \rightarrow \mathbb{R}^m$ if

$$|\psi_0(\cdot)| \in L_2(P_{\theta_0}); \quad (4)$$

$$\int \psi_0 dP_{\theta_0} = \mathbf{0}; \quad (5)$$

$$\kappa_n = \kappa(\theta_0) + \frac{1}{n} \sum_{i=1}^n \psi_0(X_i) + o_p \left(n^{-\frac{1}{2}} \right); \quad (6)$$

and

$$\sqrt{n} \left(\kappa_n - \kappa(\theta_0) - \frac{1}{n} \sum_{i=1}^n \psi_0(X_i) \right) \xrightarrow{p} \mathbf{0}, \quad (7)$$

where the convergences in (6) and (7) are with respect to $(P_{\theta_0})^n$ -probability.

We are led to the following fundamental result:

Theorem 1 (Convolution Theorem; e.g., Bickel et al. (1993, Thm. 3.3.2)). *Suppose that Assumptions 1 and 2 hold and that there exists an element $\psi_{\theta_0} \in \mathcal{T}^m$ such that*

$$\int \psi_{\theta_0} \dot{P}_{\theta_0}[\delta] dP_{\theta_0} = \dot{\kappa}_{\theta_0}(\delta) \quad (8)$$

for every $\delta \in \dot{\Theta}$ and

$$\{\alpha^\top \psi_{\theta_0} : \alpha \in \mathbb{R}^m\} \quad (9)$$

is a subset of the closure of the tangent space \mathcal{T} of \mathcal{P} , i.e., $\{\dot{P}_{\theta_0}[\delta] : \delta \in \dot{\Theta}\}$.

Then for $\Psi_0 \equiv \int \psi_{\theta_0} \psi_{\theta_0}^\top dP_{\theta_0}$,

1. The limiting distribution Q_0 of a locally regular estimator at θ_0 is a convolution of $N(\mathbf{0}, \Psi_0)$ and some other distribution M , i.e.,

$$Q_0 = N(\mathbf{0}, \Psi_0) * M.$$

2. An estimator $\{\kappa_n\}$ is locally regular at θ_0 with limiting distribution $N(\mathbf{0}, \Psi_0)$ iff $\{\kappa_n\}$ is asymptotically linear and has influence function ψ_{θ_0} at θ_0 .

Remark 1. The condition (8) of Theorem 1 essentially imposes the requirement of local regularity at $\theta = \theta_0$ on the underlying statistical model $\mathcal{P} \equiv \{P_\theta : \theta \in \Theta\}$.⁷ If the parameter of interest $\kappa(\theta)$ is represented as a functional $\nu(P_\theta)$ on \mathcal{P} , a necessary condition of (8) is the pathwise differentiability of both $\kappa(\theta)$ and its equivalent representation $\nu(P_\theta)$ along $\dot{\Theta}$ at θ_0 and along $\dot{\mathcal{P}}$ at P_{θ_0} , respectively. Regular estimators of an interest parameter $\kappa(\theta)$ are not possible if $\kappa(\theta)$ is pathwise differentiable but the underlying model \mathcal{P} is irregular.⁸

The Convolution Theorem gives rise to a notion of estimator efficiency contained in the following:

Definition 2. 1. The map $\psi_{\theta_0} : X \rightarrow \mathbb{R}^m$ in Theorem 1 is called the **efficient influence function** for $\kappa(\theta_0)$.

2. An asymptotically linear estimator $\{\kappa_n\}$ is said to be **efficient** for $\kappa(\theta_0)$ if its influence function at $\theta = \theta_0$ is equal to the efficient influence function ψ_{θ_0} .

The notion of asymptotic efficiency contained in Definition 2 for (locally) regular and asymptotically linear estimators of a parameter $\kappa(\theta_0)$ can be linked closely to a notion of optimality for tests of hypotheses regarding $\kappa(\theta_0)$. In particular, testing procedures that are asymptotically optimal in an appropriately defined sense can be constructed from asymptotically efficient estimators of $\kappa(\theta_0)$. In this connection, suppose that the interest parameter $\kappa(\theta_0)$ is scalar-valued, and consider the problem of testing the hypothesis $H_0 : \kappa(\theta_0) \leq 0$ against the alternative $H_1 : \kappa(\theta_0) > 0$. Following the usual convention, let z_α denote the $(1-\alpha)$ -quantile of a standard normal distribution, and also suppose the existence of a functional $\nu : \mathcal{P} \rightarrow \mathbb{R}$ that enables an alternative representation of the parameter of interest, i.e.,

$$\kappa(\theta_0) \equiv \nu(P_{\theta_0}).$$

⁷Cf. Bickel et al. (1993, Theorem 3.3.1 and surrounding discussion.)

⁸Cf. van der Vaart (1991) and van der Vaart (1998, Theorem 25.32).

We consider the power of an arbitrary test of H_0 against arbitrary sequence of local alternatives given by $\left\{P_{\theta_0 + \frac{1}{\sqrt{n}}\delta}\right\}$, where $\delta \in \dot{\Theta}$. The regularity requirement at θ_0 of the Convolution Theorem (i.e., (8)) is assumed to be satisfied at each element of the sequence $\left\{P_{\theta_0 + \frac{1}{\sqrt{n}}\delta}\right\}$. This ensures that for every element δ of the tangent set $\dot{\Theta}$ such that

$$\int \psi_{\theta_0} \dot{P}_{\theta_0}[\delta] dP_{\theta_0} = \dot{\kappa}_{\theta_0}(\delta) > 0, \quad (10)$$

the corresponding local alternative $P_{\theta_0 + \frac{1}{\sqrt{n}}\delta}$ is in fact contained in H_1 , since the implication of pathwise differentiability of $\kappa(\theta) = \nu(P_\theta)$ at $\theta = \theta_0$ gives

$$\nu\left(P_{\theta_0 + \frac{1}{\sqrt{n}}\delta}\right) = \frac{1}{\sqrt{n}} \int \psi_{\theta_0} \dot{P}_{\theta_0}[\delta] dP_{\theta_0} + o\left(\frac{1}{\sqrt{n}}\right)$$

if in fact $\kappa(\theta_0) = \nu(P_{\theta_0}) = 0$. The power envelope of an arbitrary test of H_0 against sequences $\left\{P_{\theta_0 + \frac{1}{\sqrt{n}}\delta}\right\}$ satisfying (10) is analyzed in the following theorem:

Theorem 2 (e.g., van der Vaart (1998, Thm. 25.44)). *Suppose the conditions of Theorem 1 are satisfied, and suppose $\kappa(\theta_0) = 0$. Consider the sequence of power functions $\{\pi_n(P_\theta) : n \geq 1\}$ ($\theta \in \theta$) corresponding to an arbitrary sequence of level- α tests of $H_0 : \tilde{\psi}(P_{\theta_0}) \leq 0$. Then for every $\delta \in \dot{\Theta}$ satisfying*

$$\int \psi_{\theta_0} \dot{P}_{\theta_0}[\delta] dP_{\theta_0} = \dot{\kappa}_{\theta_0}(\delta) > 0,$$

we have

$$\limsup_{n \rightarrow \infty} \pi_n\left(P_{\theta_0 + \frac{1}{\sqrt{n}}\delta}\right) \leq 1 - \Phi\left(z_\alpha - \frac{\int \psi_{\theta_0} \dot{P}_{\theta_0}[\delta] dP_{\theta_0}}{\sqrt{\Psi_0}}\right), \quad (11)$$

where $\Psi_0 \equiv \int \psi_{\theta_0}^2 dP_{\theta_0}$.

An immediate consequence of Theorem 2 is that a ‘‘Wald-type’’ test based on an efficient estimator of $\kappa(\theta_0)$ will be ‘‘asymptotically locally uniformly most powerful’’ in the sense that its power function will asymptotically attain the upper bound given in (11):

Corollary 1 (e.g., van der Vaart (1998, Thm. 25.45)). *Suppose that the estimator $\{\kappa_n\}$ is (locally) regular at $\theta = \theta_0$ with a limiting distribution given by $Q_0 \sim N(0, \Psi_0)$. Let $\{\tau_n^2\}$ be a consistent sequence of estimators for the asymptotic variance of κ_n , i.e.,*

$$\tau_n^2 \xrightarrow{p} \Psi(\theta)$$

where the convergence is in P_{θ_0} -probability. Then for every $\delta \in \dot{\Theta}$,

$$\lim_{n \rightarrow \infty} P_{\theta_0 + \frac{1}{\sqrt{n}}\delta} \left(\frac{\sqrt{n}\kappa_n}{\tau_n} \geq z_\alpha \right) = 1 - \Phi \left(z_\alpha - \frac{\int \psi_{\theta_0} \dot{P}_{\theta_0}[\delta] dP_{\theta_0}}{\sqrt{\Psi_0}} \right).$$

As such, a test that rejects H_0 whenever $\frac{\sqrt{n}\kappa_n}{\tau_n} \geq z_\alpha$ has size α and attains the power bound given in Theorem 2. In what follows, it is shown that conditions (8) and (9) hold for the parameter of interest in the setting considered here, namely the basic model given above in (1). As shown in Section 4, this leads naturally to a feasible Wald-type testing procedure for the hypothesis of stability applied to either a scalar-valued trend-function parameter or to a linear combination of the elements of a vector-valued trend parameter—in both cases the power functions of these procedures will asymptotically attain the corresponding bound spelled out in general terms in Corollary 1. The testing procedure proposed in Section 4 will presumably also possess asymptotic invariance properties in the case of vector-valued hypotheses involving the components of a multivariate trend parameter, which in turn lends itself to an analysis of its asymptotic optimality from the point of view of a maximin approach. Although this is presumed to be of interest in certain situations, it does not seem likely that a test satisfying a generally convincing asymptotic optimality criterion exists when constructing tests of nonscalar restrictions on a multidimensional trend parameter.

3 Efficient estimation of the trend parameter

As presented in the introduction, the basic assumption is that a stretch $\mathbf{Y}_T \equiv (Y_1, \dots, Y_T)^\top$ of observations is generated by the model

$$Y_t = \mathbf{d} \left(\frac{t}{T} \right)^\top \gamma \left(\frac{t}{T} \right) + u_t, \quad t = 1, \dots, T,$$

where $\mathbf{d}(\cdot)$ is a known \mathbb{R}^k -valued function on $[0, 1]$, $\gamma(\cdot)$ is an unknown \mathbb{R}^k -valued function on $[0, 1]$ and u_t is a stochastic component assumed to belong to a

mean-zero stationary process $\{u_t\}$.⁹ The deterministic component \mathbf{d} of the trend function is assumed to satisfy the conditions of Assumption 3, to wit:

Assumption 3. *Each component of the deterministic portion \mathbf{d} of the trend function is bounded and belongs to the set $PL[0, 1]$ of piecewise-Lipschitz continuous functions on $[0, 1]$ with a finite number of jumps, i.e., if d_j ($j \in \{1, \dots, k\}$) is the j th component of \mathbf{d} , then for some $\delta \in [0, \infty) \cap \mathbb{Z}$, there exists a partition $0 = t_0 < t_1 < \dots < t_\delta < t_{\delta+1} = 1$ such that d_j is Lipschitz-continuous on each of the intervals $[t_i, t_{i+1})$ ($i = 0, \dots, \delta$) and the jumps $d_j(t_i) - d_j(t_i-) \neq 0$ for all $1 \leq i \leq \delta$. Here $d_j(t_i-) \equiv \lim_{s \uparrow t_i} d_j(s)$.*

Remark 2. *Assumption 3 is effectively designed to allow for indicator functions to be incorporated into \mathbf{d} .*

The stochastic component of the model is assumed to be generated according to the mechanism

$$u_t = H(\dots, \epsilon_{t-1}, \epsilon_t), \quad (12)$$

where $\{\epsilon_t : t \in \mathbb{Z}\}$ is an iid process and $H(\cdot)$ is unspecified but measurable. A deliberately weak short-range dependence condition on the error process $\{u_t\}$ is assumed, and is set out as follows:

Assumption 4. 1. *For an iid sequence $\{\epsilon_t : t \in \mathbb{Z}\}$, we have*

$$u_t = H(\dots, \epsilon_{t-1}, \epsilon_t),$$

for a measurable function H such that u_t is a well-defined random variable with mean zero and finite variance.

2. $E[|u_t|^4] < \infty$.

3. *For an iid copy ϵ'_t of ϵ_t , define*

$$u_t^* \equiv H(\dots, \epsilon_{-1}, \epsilon'_0, \epsilon_1, \dots, \epsilon_{t-1}, \epsilon_t).$$

Then

$$\sum_{t=1}^{\infty} t (E[|u_t - u_t^*|^4])^{\frac{1}{4}} < \infty \quad (13)$$

⁹The “weak-trend” scaling by $\frac{t}{T}$ adopted here follows a common approach in the econometric literature when dealing with models having nonlinear time trends. Cf. e.g., Phillips and Hansen (1990); Park and Hahn (1999); Ripatti and Saikkonen (2001); Saikkonen (2001a,b); Juhl and Xiao (2005); Wu and Zhao (2007).

From part 1 of Assumption 4 it follows that $\{u_t\}$ is strictly stationary and ergodic.¹⁰ It is also *causal* in the sense that u_t does not depend on future values ϵ_s of the innovation process, where $s > t$. The class of error processes subsumed by the specification given in Assumption 4 is obviously large and includes both linear and nonlinear time-series processes commonly adopted in applied work. The generality of the specification of the error process covered by the conditions of Assumption 4 is made tractable by means of the strong invariance principle recently established by Wu (2007). In this connection, define the partial-sum process of regression errors $\{S_t : t = 1, 2, \dots\}$, where $S_t \equiv \sum_{s=1}^t u_s$. Wu (2007) established that under the short-range dependence condition given in part 3 of Assumption 4,¹¹ there exists a standard Brownian motion \mathbb{B} that uniformly approximates S_t . In particular,

$$\max_{t \leq T} |S_t - \sigma \mathbb{B}(t)| = o_{a.s.} \left(T^{\frac{1}{4}} \log T \right), \quad (14)$$

where $\sigma^2 \equiv \sum_{t=-\infty}^{\infty} E[u_0 u_t]$ denotes the long-run variance.¹²

Combining the provisions of Assumption 4 and the strong invariance principle of Wu (2007) allows one to approximate the large-sample behaviour of suitably

¹⁰Cf. e.g., Taniguchi and Kakizawa (2000, Thm. 1.3.3.).

¹¹Note that the expression on the left-hand side of condition (13) can be interpreted as a measure of the extent to which the innovation at time $t = 0$ is capable of predicting future expected values of the regression error u_t . Cf. Wu (2005).

¹²It was shown by Wu and Shao (2004) that for some $r \in (0, 1)$, the condition $\|u_t - u_t^*\| = O(r^T)$ holds for many nonlinear error processes, which implies the short-range dependence condition (13) for these processes. In the perhaps most familiar case where the errors are drawn from an ARMA process, i.e.,

$$u_t - \sum_{i=1}^p \psi_i u_{t-i} = \sum_{j=1}^q \theta_j \epsilon_{t-j},$$

for $(\psi_1, \dots, \psi_p, \theta_1, \dots, \theta_q)^\top \in \mathbb{R}^{p+q}$, we have the well-known fact that if each root of the polynomial

$$z^p - \sum_{i=1}^p \psi_i z^{p-i} = 0$$

has modulus less than one, then

$$u_t = \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i},$$

where each α_i satisfies $|\alpha_i| = O(r^i)$ for some $r \in (0, 1)$. The condition (13) is accordingly easy to verify for the case of stationary ARMA processes.

weighted partial sums of the regression errors, i.e., statistics of the form

$$\sum_{t=1}^T \omega_{t,T} u_t,$$

with that of the analogous statistic

$$\sum_{t=1}^T \omega_{t,T} Z_t,$$

where $\{Z_t\}$ is an iid sequence of $N(0, \sigma^2)$ random variables, with σ^2 the long-run variance $\sigma^2 \equiv \sum_{t=-\infty}^{\infty} E[u_0 u_t]$. It is in this sense that the original regression model given above in (1) can be reduced to the rather more prosaic model

$$\tilde{Y}_t = \mathbf{d} \left(\frac{t}{T} \right)^\top \gamma \left(\frac{t}{T} \right) + Z_t \quad (15)$$

with iid $N(0, \sigma^2)$ errors. This idea is used to analyze the asymptotic behaviour of the efficient test statistic described in Section 4.

The parameter space for the model generating the observations \mathbf{Y}_T is given by the set $\Theta \equiv \Gamma \times \mathcal{G} \times \mathcal{F}$, where Γ is the space of \mathbb{R}^k -valued functions on the unit interval, \mathcal{G} is the model for the joint distribution of the “initial condition” $(\dots, u_{t-2}, u_{t-1})$ and \mathcal{F} is the model for the conditional distribution of u_t given $\sigma\{\dots, u_{t-2}, u_{t-1}\}$, i.e., the σ -algebra generated by the history of the regression error process viewed from time t . Denote a fixed point θ in the parameter space by $\theta \equiv (\gamma, G^{(t)}, F^{(t|t-1)})$, and let $P_{(\gamma, G^{(t)}, F^{(t|t-1)})}$ denote the joint distribution of $(\dots, u_{t-2}, u_{t-1}, Y_t)$. The parameter of interest is the vector each of whose components is the *total variation* of the corresponding component of γ , i.e., a functional of the form

$$\kappa(\theta) \equiv \chi(\gamma) \equiv \left(\int_0^1 |\gamma'_1(s)| ds, \dots, \int_0^1 |\gamma'_k(s)| ds \right)^\top. \quad (16)$$

The statistical question of interest in this paper is the development of a sensible procedure for inferring whether or not $\chi(\gamma) = \mathbf{0}$ under the sequence

$$\left\{ P_{(\gamma, G^{(T)}, F^{(T|T-1)})} : T \geq 1 \right\}.$$

Let $\theta_0 \equiv (\gamma_0, G_0^{(t)}, F_0^{(t|t-1)})$ denote a point in Θ where $\chi(\gamma) = \mathbf{0}$, and let $\dot{\Gamma}$, $\dot{\mathcal{G}}$ and $\dot{\mathcal{F}}$ denote the tangent spaces of Γ , \mathcal{G} and \mathcal{F} , respectively. Note that the parameter of interest $\chi(\gamma)$, viewed as a functional on Γ into \mathbb{R}^k , is pathwise differentiable along $\dot{\Gamma}$ at γ_0 with derivative

$$\dot{\chi}(\mathbf{a}) = \left(\int_0^1 |\gamma_1'(s)| a_1(s) ds, \dots, \int_0^1 |\gamma_k'(s)| a_k(s) ds \right)^\top,$$

where $\mathbf{a}(s) \equiv (a_1(s), \dots, a_k(s))^\top$ is a point in $\dot{\Gamma} \subset (L_2(\text{Leb}[0, 1]))^k$, where $\text{Leb}[0, 1]$ denotes Lebesgue measure on the unit interval. As such, Assumption 2 is easily seen to be satisfied.

The conditional distribution $F_0^{(t|t-1)}$ is assumed to have zero mean, finite variance and finite Fisher information for location:

Assumption 5. $F_0^{(t|t-1)}$ is absolutely continuous with respect to Lebesgue measure with corresponding density

$$f_0^{(t|t-1)}(u_t) \equiv f_0(u_t | \sigma\{\dots, \epsilon_{t-2}, \epsilon_{t-1}\}).$$

$F_0^{(t|t-1)}$ satisfies $\int u dF_0^{(t|t-1)}(u) = 0$ and $\sigma_0^2 \equiv \int u^2 dF_0^{(t|t-1)}(u) < \infty$.

In addition, for

$$l \equiv -\frac{f_0^{(t|t-1)'}}{f_0^{(t|t-1)}} \mathbf{1}_{\{f_0^{(t|t-1)} > 0\}},$$

we have

$$J \equiv \int l^2 dF_0^{(t|t-1)} < \infty.$$

Assumption 5 enables the straightforward verification of Assumption 1 using methods similar to those given in for example Hájek and Sidák (1967, pp. 210–214) or Bickel et al. (1993, Section 3.2):

Lemma 1. Suppose that $\{F_\eta^{(t|t-1)}\}$ is a curve in \mathcal{F} with tangent c at F_0 and that $\{G_\eta^{(t)}\}$ is a curve in \mathcal{G} with tangent b at $G_0^{(t)}$. Then if $\mathbf{a} \in \dot{\Gamma}$ has a corresponding local sequence $\{\gamma_\eta\} \rightarrow \gamma_0$, the curve $\left\{ P(\gamma_\eta, G_\eta^{(t)}, F_\eta^{(t|t-1)}) \right\}$ has tangent

$$\mathbf{d}(s)^\top \mathbf{a}(s) l(u) + b(u_0) + c(u),$$

with respect to $\left\{ P(\gamma_0, G_0^{(t)}, F_0^{(t|t-1)}) \right\}$, for each $s \in [0, 1]$, $u_0 \sim G_0$, $u \sim F_0^{(t|t-1)}$.

The verification of Assumption 1 is straightforward from Lemma 1, and it is easily seen that for $\delta \equiv (\delta_1, \delta_2, \delta_3) \in \dot{\Gamma} \times \dot{\mathcal{G}} \times \dot{\mathcal{F}}$, the operator \dot{P}_{θ_0} in Assumption 1 is given by

$$\dot{P}_{\theta_0}[\delta] = \mathbf{d}(s)^\top \boldsymbol{\delta}_1(s) + \delta_2(u_0) + \delta_3(u),$$

and the tangent space \mathcal{T} of $\mathcal{P} = \left\{ P_{(\gamma, G^{(t)}, F^{(t|t-1)})} : \gamma \in \Gamma, G^{(t)} \in \mathcal{G}, F^{(t|t-1)} \in \mathcal{F} \right\}$ is the closed linear span of $\left\{ \dot{P}_{\theta_0}[\delta] : \delta \in \dot{\Gamma} \times \dot{\mathcal{G}} \times \dot{\mathcal{F}} \right\}$.

In order to reduce the notational complexity involved in deriving the efficient influence function for $\kappa(\theta_0)$, let

$$l_*(u_t) \equiv \frac{u_t}{\sigma_0^2} \quad (17)$$

$$J_* \equiv \frac{1}{\sigma_0^2}, \quad (18)$$

where $\sigma_0^2 \equiv \int u_t^2 dF_0^{(t|t-1)}(u_t)$, the variance of $F_0^{(t|t-1)}$. Let

$$\boldsymbol{\gamma}_0(\cdot) \equiv (\gamma_{01}(\cdot), \dots, \gamma_{0k}(\cdot))^\top,$$

and also define

$$\begin{aligned} \mathbf{v}(\cdot) &\equiv (|\gamma'_{01}(\cdot)|, \dots, |\gamma'_{0k}(\cdot)|)^\top; \\ \bar{\mathbf{v}} &\equiv \left(\int_0^1 |\gamma'_{01}(s)| ds, \dots, \int_0^1 |\gamma'_{0k}(s)| ds \right)^\top; \\ \mathbf{v}_0(\cdot) &\equiv \mathbf{v}(\cdot) - \bar{\mathbf{v}}; \\ \Delta &\equiv \frac{J - J_*}{J_*}; \\ \mathbf{w}(\cdot) &\equiv \frac{1}{J} (\mathbf{v}(\cdot) + \Delta \bar{\mathbf{v}}); \\ \bar{\mathbf{w}} &\equiv \frac{1}{J} \int_0^1 (\mathbf{v}(s) + \Delta \bar{\mathbf{v}}) ds; \\ \mathbf{w}_0(\cdot) &\equiv \mathbf{w}(\cdot) - \bar{\mathbf{w}}. \end{aligned}$$

Note that

$$\bar{\mathbf{w}} = \frac{1}{J} (\bar{\mathbf{v}} + \Delta \bar{\mathbf{v}}) = \left(\frac{1 + \Delta}{J} \right) \bar{\mathbf{v}},$$

and that

$$\mathbf{w}_0(\cdot) = \frac{1}{J} (\mathbf{v}(\cdot) - \bar{\mathbf{v}}).$$

Note that J_* as given in (18) is guaranteed to be positive by virtue of Assumption 5. This in turn guarantees the existence of the efficient influence function for $\kappa(\theta_0)$ given in the following:

Theorem 3. *Given Assumption 5, the efficient influence function for $\kappa(\theta_0)$ is given by*

$$\begin{aligned}\psi_{\theta_0}^* &\equiv \mathbf{v}_0(s) \frac{l(u_t)}{J} + \bar{\mathbf{v}} \frac{l_*(u_t)}{J_*} \\ &= \mathbf{w}_0(s) l(u_t) + \bar{\mathbf{w}} l_*(u_t)\end{aligned}$$

for $s \in [0, 1]$ and $u_t \sim F_0^{(t|t-1)}$.

Proof. Let $\mathbf{a} \in \dot{\Gamma}$, $b \in \dot{\mathcal{G}}$ and $c \in \dot{\mathcal{F}}$. We have

$$\begin{aligned}& \int_0^1 E \left[\psi_{\theta_0}^* (b(u_0) + \mathbf{d}(s)^\top \mathbf{a}(s) l(u_t) + c(u_t)) \right] ds \\ &= \int_0^1 E \left[\mathbf{w}_0(s) l(u_t) b(u_0) + \mathbf{w}_0(s) \mathbf{d}(s)^\top \mathbf{a}(s) l^2(u_t) + \mathbf{w}_0(s) c(u_t) l(u_t) \right. \\ &\quad \left. + \bar{\mathbf{w}} l_*(u_t) b(u_0) + \bar{\mathbf{w}} l_*(u_t) \mathbf{d}(s)^\top \mathbf{a}(s) l(u_t) + \bar{\mathbf{w}} l_*(u_t) c(u_t) \right] ds \\ &= \int_0^1 \left(\mathbf{w}_0(s) E[l(u_t) b(u_0)] + \mathbf{w}_0(s) \mathbf{d}(s)^\top \mathbf{a}(s) E[l^2(u_t)] + \mathbf{w}_0(s) E[c(u_t) l(u_t)] \right. \\ &\quad \left. + \bar{\mathbf{w}} E[l_*(u_t) b(u_0)] + \bar{\mathbf{w}} \mathbf{d}(s)^\top \mathbf{a}(s) E[l_*(u_t) l(u_t)] + \bar{\mathbf{w}} E[l_*(u_t) c(u_t)] \right) ds \\ &= \int_0^1 \left(\mathbf{w}_0(s) \mathbf{d}(s)^\top \mathbf{a}(s) E[l^2(u_t)] + \bar{\mathbf{w}} E[l_*(u_t) c(u_t)] \right) ds,\end{aligned}$$

where the expectation is with respect to the product of the joint distribution of $(\dots, u_{t-2}, u_{t-1})^\top$ and the conditional distribution of u_t with respect to $\sigma\{\dots, u_{t-2}, u_{t-1}\}$. When $\theta = \theta_0$, we have $\bar{\mathbf{w}} = \left(\frac{1+\Delta}{J}\right) \bar{\mathbf{v}} = \mathbf{0}$ and $\mathbf{w}_0(s) = \frac{1}{J} \mathbf{v}(s)$, so

$$\begin{aligned}& \int_0^1 \left(\mathbf{w}_0(s) \mathbf{d}(s)^\top \mathbf{a}(s) E[l^2(u_t)] + \bar{\mathbf{w}} E[l_*(u_t) c(u_t)] \right) ds \\ &= \int_0^1 \mathbf{w}_0(s) \mathbf{d}(s)^\top \mathbf{a}(s) E[l^2(u_t)] ds \\ &= \int_0^1 \mathbf{v}(s) \mathbf{d}(s)^\top \mathbf{a}(s) ds \\ &= \dot{\chi}(\mathbf{a}),\end{aligned}$$

which verifies condition (8) in the statement of Theorem 1. For $\boldsymbol{\lambda} \in \mathbb{R}^k$, $\boldsymbol{\lambda}^\top \boldsymbol{\psi}_{\theta_0}^*$ is clearly in the closed linear span of \mathcal{T} , and it follows that $\boldsymbol{\psi}_{\theta_0}^*$ is indeed the efficient influence function for $\boldsymbol{\kappa}(\theta_0) = \boldsymbol{\chi}(\gamma_0)$. \square

Efficient inference regarding the parameter of interest $\boldsymbol{\chi}(\gamma)$ can be conducted by way of the following reformulation of the convolution theorem.¹³

Theorem 4. 1. *The limiting distribution Q_0 of any estimator of $\boldsymbol{\chi}(\gamma_0)$ regular at $\gamma_0 \in \Gamma$ is a convolution of $N(0, E[\boldsymbol{\psi}_{\theta_0}^* \boldsymbol{\psi}_{\theta_0}^{*\top}])$ and some other distribution M , i.e.,*

$$Q_0 = N(0, E[\boldsymbol{\psi}_{\theta_0}^* \boldsymbol{\psi}_{\theta_0}^{*\top}]) * M.$$

2. *An estimator $\{\hat{\boldsymbol{\chi}}_T\}$ of $\boldsymbol{\chi}(\gamma_0)$ is regular at γ_0 with limiting distribution $Q_0 = N(0, E[\boldsymbol{\psi}_{\theta_0}^* \boldsymbol{\psi}_{\theta_0}^{*\top}])$ iff the sequence $\{\hat{\boldsymbol{\chi}}_T\}$ has influence function $\boldsymbol{\psi}_{\theta_0}^*$ at $\theta_0 = (\gamma_0, G_0^{(t)}, F_0^{(t|t-1)})$.*

In particular, an estimator of $\boldsymbol{\chi}(\gamma_0)$ is said to be *efficient* at $\gamma_0 \in \Gamma$ iff it is asymptotically linear with influence function $\boldsymbol{\psi}_{\theta_0}^*$ at θ_0 , as given above in the statement of Theorem 3. In this paper it is proposed to test the hypothesis of time invariance of the trend parameter γ via an efficient estimator $\hat{\boldsymbol{\chi}}_T$ of the interest parameter $\boldsymbol{\chi}(\gamma_0)$ as given above in (16). Rejection is associated with $\sqrt{T} \boldsymbol{\iota}_k^\top \hat{\boldsymbol{\chi}}_T$ being significantly greater than zero, where $\boldsymbol{\iota}_k \equiv (1, \dots, 1)^\top$, the unit vector in \mathbb{R}^k . The next section describes the construction of an efficient estimator of the parameter of interest.

¹³Adaptive estimation of $\boldsymbol{\chi}(\gamma_0)$ in the presence of $G_0^{(t)}$ and $F_0^{(t|t-1)}$ is also possible. This follows from

$$\begin{aligned} & \int_0^1 E[\boldsymbol{\psi}_{\theta_0}^*(b(u_0) + c(u_t))] ds \\ &= \int_0^1 E[\boldsymbol{w}_0(s)l(u_t)b(u_0) + \boldsymbol{w}_0(s)l(u_t)c(u_t) + \bar{\boldsymbol{w}}l_*(u_t)b(u_0) + \bar{\boldsymbol{w}}l_*(u_t)c(u_t)] ds \\ &= \int_0^1 (\boldsymbol{w}_0(s)E[l(u_t)b(u_0)] + \boldsymbol{w}_0(s)E[l(u_t)c(u_t)]) ds \\ &= \mathbf{0} \end{aligned}$$

and Bickel et al. (1993, Corollary 3.4.3).

4 Construction of an efficient test statistic

This section of the paper is concerned with the problem of constructing an efficient estimator of parameter of interest $\chi(\gamma_0)$ as given above in (16). As indicated above, this approach is largely concerned with constructing an estimate of the efficient influence function $\psi_{\theta_0}^*$ as given in the statement of Theorem 3. In this connection, the general approach developed by Schick (1986), Klaassen (1987) and Schick (1987) will be followed.

Consider the ‘‘Priestley-Chao’’ estimator $\hat{\gamma}_T$ of the trend-function parameter γ given by

$$\hat{\gamma}_T(s) \equiv (\mathbf{d}(s)\mathbf{d}(s)^\top)^{-1} \mathbf{d}(s) \sum_{t=1}^T w_{l,t,T}(s) Y_t, \quad (s \in (0, 1)), \quad (19)$$

where for a positive sequence $\{h_l\}$ such that $h_l = h_{lT} \rightarrow 0$ with $Th_{lT} \rightarrow \infty$ as $T \rightarrow \infty$,

$$w_{l,t,T}(s) \equiv \frac{1}{Th_l} K \left(\frac{1}{h_l} \left(\frac{t}{T} - s \right) \right). \quad (20)$$

The estimator $\hat{\gamma}_T(s)$ is accordingly obtained by a rescaling of the familiar Priestley-Chao estimate of the value of the trend function $\mathbf{d}(s)^\top \gamma(s)$ at $s \in (0, 1)$.¹⁴

The vector of derivatives $\gamma'(s) \equiv (\gamma'_1(s), \dots, \gamma'_k(s))^\top$ of the trend parameter can also be estimated using the ‘‘Priestley-Chao’’ approach. In particular, for a sequence of bandwidths $\{h_d\}$ possibly different from $\{h_l\}$ and satisfying $h_d = h_{dT} \rightarrow 0$ with $Th_{dT}^3 \rightarrow \infty$ as $T \rightarrow \infty$, we may estimate $\gamma'(s)$ with what is essentially the derivative with respect to $s \in (0, 1)$ of the expression in (19), to wit,

$$\hat{\gamma}'_T(s) \quad (21)$$

$$\equiv \left[-2 (\mathbf{d}(s)\mathbf{d}(s)^\top)^{-2} \mathbf{d}(s)\mathbf{d}'(s)^\top \mathbf{d}(s) + (\mathbf{d}(s)\mathbf{d}(s)^\top)^{-1} \mathbf{d}'(s) \right] \sum_{t=1}^T w_{d,t,T}(s) Y_t \\ + (\mathbf{d}(s)\mathbf{d}(s)^\top)^{-1} \mathbf{d}(s) \sum_{t=1}^T w'_{d,t,T}(s) Y_t, \quad (22)$$

where $\mathbf{d}'(s)$ denotes the vector whose components are the derivatives of the corresponding components of $\mathbf{d}(s)$, and where for $w_{d,t,T}(s)$ denoting the expression

¹⁴Cf. Priestley and Chao (1972).

in (20) with h_d in place of h_l ,

$$w'_{d,t,T}(s) \equiv \frac{d}{ds} w_{d,t,T}(s).$$

The asymptotic properties of the estimators given in (19) and (22) are derived under various conditions regarding \mathbf{d} , h_l , h_d and $K(\cdot)$, as well as under the various general conditions imposed on the regression error process $\{u_t\}$ in Assumption 4.

It is noted that the estimators of γ and of γ' given above in (19) and (22), respectively, are major ingredients in the construction of the efficient test statistic described here. As such, assumptions regarding the deterministic trend component \mathbf{d} , the bandwidth sequences $\{h_{lT}\}$ and $\{h_{dT}\}$ as well as the smoothing kernel $K(\cdot)$ are stated:

Assumption 6. 1. $h_{lT} \rightarrow 0$ and $Th_{lT} \rightarrow \infty$ as $T \rightarrow \infty$.

2. $h_{dT} \rightarrow 0$ and $Th_{dT}^3 \rightarrow \infty$ as $T \rightarrow \infty$.

Assumption 7. The smoothing kernel $K : \mathbb{R} \rightarrow \mathbb{R}$ used in the construction of the estimators in (19) and (22) is Borel-measurable, possesses a first-order derivative of bounded variation over the real line, and satisfies

1. $K(z) \geq 0$ for every $z \in \mathbb{R}$;

2. $\int K(z) dz = 1$;

3. $\int K^2(z) dz < \infty$.

Assumptions 4, 3, 6 and 7 lead to results set out in Appendix A.1 regarding the uniform rate of convergence of $\hat{\gamma}_T(\cdot)$ and of $\hat{\gamma}'_T(\cdot)$ over the unit interval. These results are in turn used in the proof of the main result of this section.

Given the estimators $\hat{\gamma}_T(\cdot)$ and $\hat{\gamma}'_T(\cdot)$ given in (19) and (22), respectively, define the following:

$$\hat{u}_{t,T} \equiv Y_t - \mathbf{d} \left(\frac{t}{T} \right)^\top \hat{\gamma}_T \left(\frac{t}{T} \right); \quad (23)$$

$$\hat{\mathbf{v}}_T(s) \equiv (|\hat{\gamma}'_{T1}(s)|, \dots, |\hat{\gamma}'_{Tk}(s)|)^\top; \quad (24)$$

$$\bar{\mathbf{v}}_T \equiv \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{v}}_T \left(\frac{t}{T} \right); \quad (25)$$

$$\delta_{t,T} \equiv \mathbf{d} \left(\frac{t}{T} \right)^\top \left(\hat{\gamma}_T \left(\frac{t}{T} \right) - \gamma \left(\frac{t}{T} \right) \right); \quad (26)$$

$$\hat{f}_{UT}(u) \equiv \frac{1}{T} \sum_{t=1}^T \frac{1}{a_T} k \left(\frac{u - \hat{u}_{t,T}}{a_T} \right); \quad (27)$$

$$\hat{l}_{UT}(u) \equiv -\frac{\hat{f}'_{UT}(u)}{\hat{f}_{UT}(u) + b_T}; \quad (28)$$

$$\hat{J}_T \equiv \frac{1}{T} \sum_{t=1}^T \hat{l}_{UT}(\hat{u}_{t,T}); \quad (29)$$

where $\{a_T\}$, $\{b_T\}$ are positive sequences tending to zero as $T \rightarrow \infty$. The kernel function $k(\cdot)$ employed in $\hat{f}_{UT}(u)$ and $\hat{f}'_{UT}(u)$ is assumed to satisfy the following condition:

Assumption 8. k is symmetric, three times continuously differentiable, satisfies $\int u^2 k(u) du < \infty$ and

$$|k^{(i)}(u)| \leq Ck(u)$$

for some positive constant C , and any $u \in \mathbb{R}$ and $i \in \{1, 2, 3\}$.

Remark 3. The logistic density and the Epanechnikov (1969) kernel both satisfy Assumption 8.

The quantities in (23)–(29) are used to construct the estimator

$$\hat{\chi}_T \equiv \frac{1}{T} \sum_{t=1}^T \left[\hat{\mathbf{v}}_T \left(\frac{t}{T} \right) + \hat{J}_T^{-1} \left(\hat{\mathbf{v}}_T \left(\frac{t}{T} \right) - \bar{\mathbf{v}}_T \right) \hat{l}_{UT}(\hat{u}_{t,T}) + \bar{\mathbf{v}}_T \hat{u}_{t,T} \right]. \quad (30)$$

In what follows, it is shown that $\hat{\chi}_T$ is asymptotically linear with influence function $\psi_{\theta_0}^*$ at $\theta = \theta_0$ and is therefore efficient for $\chi(\gamma_0)$. In particular, we give conditions such that

$$\hat{\chi}_T = \frac{1}{T} \sum_{t=1}^T \left[\mathbf{v} \left(\frac{t}{T} \right) + J^{-1} \left(\mathbf{v} \left(\frac{t}{T} \right) - \bar{\mathbf{v}} \right) l_U(u_t) + \bar{\mathbf{v}} u_t \right] + o_{\theta_T} \left(T^{-\frac{1}{2}} \right), \quad (31)$$

where the notation $C_T = o_{\theta_T}(r_T)$ indicates that $r_T^{-1}C_T$ converges to zero in P_{θ_T} -probability, for $\{\theta_T\}$ a sequence in Θ and $\{r_T\}$ a sequence of positive numbers. In particular, $\{\theta_T\}$ will be taken to be a local sequence about the point $\theta_0 = (\gamma_0, G_0^{(t)}, F_0^{(t|t-1)}) \in \Theta$.

A central feature of the proof of (31) involves showing in a number of different cases that suitable estimates $\hat{\mathbf{L}}_T$ of a functional \mathbf{L}_T satisfy

$$\sqrt{T} \sum_{t=1}^T \left(\hat{\mathbf{L}}_T(u_t) - \mathbf{L}_T(u_t) \right) = o_{\theta_T}(1). \quad (32)$$

In particular, the preliminary results given below as Lemmas 2–5 involve showing (32) for several different settings of $\hat{\mathbf{L}}_T$ and \mathbf{L}_T . It is assumed throughout that the stochastic component of the model given in (1) satisfies Assumption 4. The proofs of Lemmas 2–5 involve combining the strong invariance principle of Wu (2007) with the machinery of Schick (1987, Lemma 3.1). In this connection, suppose that $\tilde{\mathbf{L}}_T$ is a suitable estimate of the functional of interest \mathbf{L}_T .¹⁵ Then for $\{Z_t\}$ denoting as before an iid $N(0, \sigma^2)$ process discussed earlier, we can write

$$\begin{aligned} & \sqrt{T} \sum_{t=1}^T \left(\hat{\mathbf{L}}_T(u_t) - \mathbf{L}_T(u_t) \right) \\ = & \sqrt{T} \sum_{t=1}^T \left(\hat{\mathbf{L}}_T(u_t) - \tilde{\mathbf{L}}_T(Z_t) \right) + \sqrt{T} \sum_{t=1}^T \left(\tilde{\mathbf{L}}_T(Z_t) - \mathbf{L}_T(Z_t) \right) \\ & + \sqrt{T} \sum_{t=1}^T \left(\mathbf{L}_T(Z_t) - \mathbf{L}_T(u_t) \right) \\ \equiv & \Delta_{T1} + \Delta_{T2} + \Delta_{T3}. \end{aligned} \quad (33)$$

In the context of Lemmas 2–5, the Δ_{T1} and Δ_{T3} quantities will be analyzed using the invariance principle of Wu (2007), while the behaviour of Δ_{T2} , given that it involves the iid process $\{Z_t\}$, will be analyzed using the approach of Schick (1987, Lemma 3.1).¹⁶

For some constant $r_l \in (0, 1)$, set $h_l \equiv h_{lT} \propto T^{-r_l}$ to be the bandwidth used in the construction $\hat{\gamma}_T$ given above in (19), and for some $r_d \in (0, 1)$ set $h_d \equiv h_{dT} \propto T^{-r_d}$ to be the corresponding quantity used in the construction of $\hat{\gamma}'_T$ in (22). Let $r \equiv \min \{r_l, r_d\}$.

Also let

$$A_T \equiv \max_{0 \leq s \leq 1} \|\hat{\mathbf{v}}_T(s)\| + 1, \quad (34)$$

¹⁵The estimate $\tilde{\mathbf{L}}_T$ may coincide with $\hat{\mathbf{L}}_T$.

¹⁶The modification of Schick (1987, Lemma 3.1) suitable for the purpose of analyzing the convergence of Δ_{T2} appears as Lemma 11 in Appendix A.2.

and let F_U denote the marginal distribution of u_t . Lemmas 2–5 are preliminary results required to show (31). Their proofs appear in Appendix B.

Lemma 2. *Suppose Assumptions 3, 4, 6, 7 and 8 hold. Then the following hold:*

1. *If $a_T^{-2}T^{-\frac{3}{2}-2r} \rightarrow 0$, then*

$$\frac{1}{T} \sum_{t=1}^T \int \left(\hat{l}'_{UT}(u - \delta_{t,T}) - \hat{l}'_{UT}(u) \right) dF_U(u) = o_{\theta_T} \left(T^{-\frac{1}{2}} \right).$$

2. *If $A_T a_T^{-2} T^{-\frac{3}{2}-2r} \rightarrow 0$, then*

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) \int \left(\hat{l}_{UT}(u - \delta_{t,T}) - \hat{l}_{UT}(u) + \hat{l}'_{UT}(u) \delta_{t,T} \right) dF_U(u) \\ &= o_{\theta_T} \left(T^{-\frac{1}{2}} \right). \end{aligned}$$

Lemma 3. *Suppose Assumptions 3, 4, 6, 7 and 8 hold. Then if $a_T^{-4}T^{-1-2r} \rightarrow 0$ and $A_T^2 (T^{-1}a_T^{-4}b_T^{-2} + T^{-2-r}a_T^{-5}b_T^{-1}) \rightarrow 0$,*

$$\frac{1}{T} \sum_{t=1}^T \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) \left\{ \hat{l}_{UT}(\hat{u}_{t,T}) - \int \hat{l}_{UT}(u - \delta_{t,T}) dF_U(u) - \hat{l}_{UT}(u_t) \right\} = o_{\theta_T} \left(T^{-\frac{1}{2}} \right).$$

Lemma 4. *Suppose Assumptions 3, 4, 6, 7 and 8 hold. If $T^{-3-3r}a_T^{-6} \rightarrow 0$, $T^{-2\alpha}a_T^{-4} \rightarrow 0$, $T^{-4-4r}a_T^{-7}b_T^{-1} \rightarrow 0$ and $T^{-3-2r}a_T^{-6}b_T^{-2} \rightarrow 0$,*

$$\frac{1}{T} \sum_{t=1}^T \left(\hat{l}'_{UT}(\hat{u}_{t,T}) - \int \hat{l}'_{UT}(u - \delta_{t,T}) dF_U(u) \right) = o_{\theta_T} \left(T^{r+\frac{1}{2}} \right).$$

Lemma 5. *Suppose Assumptions 3, 4, 6, 7 and 8 hold. Then*

$$\frac{1}{T} \sum_{t=1}^T \bar{\mathbf{v}}_T \hat{u}_{t,T} = \bar{\mathbf{v}} \frac{1}{T} \sum_{t=1}^T u_t + o_{\theta_T} \left(T^{-\frac{1}{2}} \right).$$

Lemmas 2–5 imply the central result of this section.

Theorem 5. *Suppose Assumptions 3, 4, 5, 6, 7 and 8 hold. If the conditions*

$$a_T^{-4}T^{-1-2r} \rightarrow 0 \quad (35)$$

$$A_T a_T^{-2} T^{-\frac{3}{2}-2r} \rightarrow 0 \quad (36)$$

$$A_T^2 T^{-1} a_T^{-4} b_T^{-2} \rightarrow 0 \quad (37)$$

$$A_T^2 T^{-2-2r} a_T^{-5} b_T^{-1} \rightarrow 0 \quad (38)$$

are satisfied, then the estimator $\hat{\chi}_T$ given in (30) satisfies (31).

Proof. Lemmas 2 and 4 yield

$$\begin{aligned} \hat{J}_T &\equiv \sum_{t=1}^T \hat{l}_{UT}(\hat{u}_{t,T}) \\ &= \frac{1}{T} \sum_{t=1}^T \int \hat{l}_{UT}(u - \delta_{t,T}) dF_U(u) + o_{\theta_T}(T^{r+\frac{1}{2}}) \\ &= \frac{1}{T} \sum_{t=1}^T \int \hat{l}_{UT}(u) dF_U(u) + o_{\theta_T}(T^{r+\frac{1}{2}}) \\ &= \int \hat{l}_{UT}(u) dF_U(u) + o_{\theta_T}(T^{r+\frac{1}{2}}). \end{aligned}$$

This result, along with Lemmas 2 and 3, can be used to show

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^T \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) \left\{ \hat{l}_{UT}(\hat{u}_{t,T}) - \int \hat{l}_{UT} dF_U - \hat{J}_T \delta_{t,T} \right\} \\ &= \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) \left\{ \hat{l}_{UT}(\hat{u}_{t,T}) - \int \hat{l}_{UT} dF_U - \delta_{t,T} \int \hat{l}_{UT} dF_U \right\} + o_{\theta_T}(T^{-\frac{1}{2}}) \\ &= \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) \left\{ \int \hat{l}_{UT}(u - \delta_{t,T}) dF_U + l_U(u_t) \right\} \\ &\quad - \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) \int (\hat{l}_{UT} + \delta_{t,T} \hat{l}_{UT}) dF_U + o_{\theta_T}(T^{-\frac{1}{2}}) \\ &= \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) l_U(u_t) + o_{\theta_T}(T^{-\frac{1}{2}}) \end{aligned}$$

This result, along with the uniform $\sqrt{Th_d^3} = T^{\frac{1}{2} - \frac{3r_d}{2}}$ -consistency of $\hat{\mathbf{v}}_T(\cdot)$ to $\mathbf{v}(\cdot)$,¹⁷ yields

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{J}_T} \left(\hat{\mathbf{v}}_T \left(\frac{t}{T} \right) - \bar{\mathbf{v}}_T \right) \left(\hat{l}_{UT}(\hat{u}_{t,T}) + \hat{J}_T \delta_{t,T} \right) \\
&= \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{J}_T} \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) \left\{ \hat{l}_{UT}(\hat{u}_{t,T}) + \hat{J}_T \delta_{t,T} \right\} - \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{J}_T} \bar{\mathbf{v}}_T \left\{ \hat{l}_{UT}(\hat{u}_{t,T}) + \hat{J}_T \delta_{t,T} \right\} \\
&\quad + o_{\theta_T}(1) \\
&= \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{J}_T} \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) l_U(u_t) + \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{J}_T} \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) \int \hat{l}_{UT} dF_U - \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{J}_T} \bar{\mathbf{v}}_T l_U(u_t) \\
&\quad - \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{J}_T} \bar{\mathbf{v}}_T \int \hat{l}_{UT} dF_U + o_{\theta_T} \left(T^{-\frac{1}{2}} \right) \\
&= \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{J}_T} \left\{ \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) - \bar{\mathbf{v}}_T \right\} l_U(u_t) + o_{\theta_T} \left(T^{-\frac{1}{2}} \right) \\
&= \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{J}_T} \left\{ \mathbf{v} \left(\frac{t}{T} \right) - \bar{\mathbf{v}} \right\} l_U(u_t) + o_{\theta_T} \left(T^{-\frac{1}{2}} \right).
\end{aligned}$$

Appealing to Lemmas 5 and 9, we have that

$$\begin{aligned}
& \hat{\chi}_T - \frac{1}{T} \sum_{t=1}^T \left[\mathbf{v} \left(\frac{t}{T} \right) + \frac{1}{\hat{J}_T} \left(\hat{\mathbf{v}}_T \left(\frac{t}{T} \right) - \bar{\mathbf{v}}_T \right) l_U(u_t) + \bar{\mathbf{v}}_T u_t \right] \\
&= \frac{1}{T} \sum_{t=1}^T \left[\hat{\mathbf{v}}_T \left(\frac{t}{T} \right) - \mathbf{v} \left(\frac{t}{T} \right) \right] + \frac{1}{T} \sum_{t=1}^T \left\{ \frac{1}{\hat{J}_T} \left[\hat{\mathbf{v}}_T \left(\frac{t}{T} \right) - \bar{\mathbf{v}}_T \right] \hat{l}_{UT}(\hat{u}_{t,T}) \right\} \\
&\quad - \frac{1}{T} \sum_{t=1}^T \left\{ \frac{1}{\hat{J}_T} \left[\mathbf{v} \left(\frac{t}{T} \right) - \bar{\mathbf{v}} \right] l_U(u_t) \right\} + \frac{1}{T} \sum_{t=1}^T [\bar{\mathbf{v}}_T \hat{u}_{t,T} - \bar{\mathbf{v}} u_t] \\
&= \frac{1}{T} \sum_{t=1}^T \left\{ \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) - \mathbf{v} \left(\frac{t}{T} \right) + \left[\hat{\mathbf{v}}_T \left(\frac{t}{T} \right) - \bar{\mathbf{v}} \right] \delta_{t,T} \right\} + o_{\theta_T} \left(T^{-\frac{1}{2}} \right) \\
&= o_{\theta_T} \left(T^{-\frac{1}{2}} \right).
\end{aligned}$$

□

¹⁷Cf. Lemma 8 in Appendix A.1.

Combining Theorem 4 with Corollary 1, an efficient testing procedure for the null of parameter stability follows straightforwardly from Theorem 5. Setting

$$\hat{\Psi}_T \equiv \frac{1}{T} \sum_{t=1}^T \hat{\psi}_{t,T} \hat{\psi}_{t,T}^\top$$

where

$$\hat{\psi}_{t,T} \equiv \hat{J}_T^{-1} \left(\hat{\mathbf{v}}_T \left(\frac{t}{T} \right) - \bar{\mathbf{v}}_T \right) \hat{l}_{UT}(\hat{u}_{t,T}) + \bar{\mathbf{v}}_T \hat{u}_{t,T},$$

we have that for $\boldsymbol{\nu}_k \equiv (1, \dots, 1)^\top \in \mathbb{R}^k$, a reasonable asymptotically level- α test of $H_0 : \boldsymbol{\nu}_k^\top \boldsymbol{\chi}(\gamma_0) = 0$ against $H_1 : \boldsymbol{\nu}_k^\top \boldsymbol{\chi}(\gamma_0) > 0$ can be obtained by rejecting the hypothesis whenever

$$\frac{\frac{1}{\sqrt{T}} \boldsymbol{\nu}_k^\top \sum_{t=1}^T \hat{\psi}_{t,T}}{\sqrt{\boldsymbol{\nu}_k^\top \hat{\Psi}_T \boldsymbol{\nu}_k}} \quad (39)$$

exceeds the $(1 - \alpha)$ -quantile of a $N(0, 1)$ -random variate. By Corollary 1 this procedure is uniformly most powerful against contiguous alternatives.

The next section examines the finite-sample performance of the Wald-type test given in (39).

5 Numerical evidence

This section presents the results of a modest simulation exercise examining the size and power performance in small samples of the efficient test for trend stability based on the asymptotically normal statistic in (39). Simulated observations were drawn from the process given by

$$\left. \begin{aligned} Y_t &= \gamma \left(\frac{t}{T} \right) + u_t \\ u_t &= \rho u_{t-1} + \epsilon_t, \end{aligned} \right\} t = 1, \dots, T; \quad (40)$$

where $\{\epsilon_t\}$ is set to be iid $N(0, 1)$ throughout. In each experiment conducted, the kernel function $k(\cdot)$ in (27) was taken to be the logistic density, while the Priestley-Chao estimates of the trend parameter and its derivative were constructed using a standard normal kernel. Sample sizes of $T \in \{100, 200, 300\}$ were employed for the purposes of comparing size and power performance in a series of Monte Carlo experiments. Each experiment involved 1000 replications.

Table 1 contains the results of the first set of experiments, which were intended to examine the empirical size of the proposed testing procedure when the model

given in (40) was simulated under the null of a time-invariant trend. In each of the experiments summarized in Table 1, the autoregressive coefficient was set to $\rho \equiv \frac{1}{2}$, while the trend parameter was set to be identically zero, i.e., $\gamma(\cdot) \equiv 0$. Various settings for the three bandwidths used in the construction of the test statistic were experimented with, subject to the rate constraints imposed by Theorem 5. In particular, the bandwidths used were of the form $h_{lT} = h_{dT} = h_T \equiv c_h T^{-\frac{2}{5}}$, $a_T = c_a T^{-\frac{2}{9}}$ and $b_T = c_b T^{-\frac{2}{9}}$, where the leading constants c_h , c_a and c_b range over the set $\{0.5, 1.0, 1.5\}$ as indicated in Table 1.¹⁸ A glance at Table 1 reveals the sensitivity of the empirical size of the test to bandwidth choice, although setting the leading constants to $c_h = c_a = c_b = .5$ seems to deliver a proportion of incorrect rejections over the range of sample sizes considered that is fairly close to the level desired.

The small-sample power of the proposed testing procedure is examined in a series of four simulations. Like the series of experiments reported in Table 1, each of these simulations also involved 1000 replications and a nominal level of 5%. The sample sizes employed were each fixed at $T = 200$. Based on the results reported in Table 1, the bandwidths used in each experiment were set to $h_{dT} = h_{lT} = h_T .5 T^{-\frac{2}{5}}$ and $a_T = b_T = .5 T^{-\frac{2}{9}}$, which correspond roughly to the row of Table 1 having the most accurate empirical sizes for the nominal level of 5%.

The first experiment examining finite-sample power involves data that were generated using the specification given in (40) with $\rho = \frac{1}{2}$, but where the trend parameter was set to have a one-time break halfway through the sampling interval, i.e.,

$$\gamma\left(\frac{t}{T}\right) = 1\left(\frac{t}{T} > .5\right) \gamma_0, \quad (41)$$

for $\gamma_0 \geq 0$. The break magnitude γ_0 is taken to be a constant ranging over a grid of 50 equally-spaced points in the interval $[0, 1]$. A glance at the curve plotted in Figure 1 indicates that the power performance of the test appears to be quite satisfactory.

¹⁸It should be clear that the conditions of Theorem 5 are fairly loose with respect to the permissible rates of decay of bandwidths that one might consider reasonable for constructing the test statistic. The rates of decay used in h_T , a_T were simply set to be twice the optimal rates from the point of view of minimizing the mean integrated squared error of pointwise estimates of a regression function and the derivative of a density, respectively. Note that the ‘‘asymptotic undersmoothing’’ engaged in here reflects the role of the nonparametric estimators as preliminary ingredients in estimates of nonparametric functionals. Cf. the unifying theory of Goldstein and Messer (1992).

The power performance of the proposed testing procedure in small samples also appears to be quite satisfactory in three further experiments. Figure 2 summarizes the results of an experiment involving two breaks in the trend function. In particular, the trend used is that given by

$$\gamma\left(\frac{t}{T}\right) = \left(\frac{1}{2} \times 1\left(.25 < \frac{t}{T} < .75\right) + 1\left(\frac{t}{T} \geq .75\right)\right) \gamma_0,$$

where as in (41), the parameter γ_0 ranges over a grid of 50 equally-spaced points in the unit interval. In this experiment, as in the next experiment summarized in Figure 3, the data-generating process for the errors is that given above in (40) with $\rho = \frac{1}{2}$.

The simulations summarized by Figure 3 investigate the power performance of the test when the trend function undergoes a continuous linear change in the middle half of the sampling interval. In particular, the trend function used is given by

$$\gamma\left(\frac{t}{T}\right) = \left\{1\left(.25 < \frac{t}{T} < .75\right) \left[2\left(\frac{t}{T} - \frac{1}{4}\right)\right] + 1\left(\frac{t}{T} \geq .75\right)\right\} \gamma_0,$$

where as before γ_0 ranges across a lattice of 50 equally-spaced points in $[0, 1]$.

The last experiment examining the power performance of the proposed test involves using the test to detect the existence of a stochastic trend. As such, the specification given above in (40) was used, but with $\gamma\left(\frac{t}{T}\right) \equiv 0$ and the autoregressive coefficient ρ taking successive values in a grid of 50 equally-spaced points in the interval $\left[\frac{1}{2}, 1\right]$. As indicated in Figure 4, power performance is satisfactory until the point where ρ reaches a value of about .80. Power is subject to nonmonotonicities for ρ lying approximately between .80 and .97, but increases sharply as ρ increases beyond .97.

Figures 1-4 suggest that the proposed testing procedure has generally satisfactory power against a large class of alternatives.

6 Empirical example: Models of U.S. real output

This section serves both to illustrate the practical applicability of the efficient testing procedure proposed in this paper and to provide further casual evidence concerning the sensitivity of the test's performance to the implementation of the Priestley-Chao estimates of the trend parameter and its derivative. In this connection data describing the dynamic behaviour of the level of U.S. real output

will be analyzed. The hypothesis considered will be the adequacy of a linear trend-stationary specification with a single break in mean. In particular, the break relates to one of two “stylized facts”, the validity of which will essentially be examined using the methodology proposed here. These stylized facts involve the persistent effects on trend of the onset of the Great Depression in 1929, which is widely held to have led to a dramatic reduction in the *level* of economic activity, and the start of the OPEC oil embargo in 1973, which is believed to have led to a somewhat less dramatic reduction in the *growth rate* of economic activity from its previous trend path.¹⁹

In this connection, two time series will be analyzed. The first series considered is the annual series of real Gross National Product from 1909 to 1970 appearing in the dataset used by Nelson and Plosser (1982).²⁰ The test proposed in this paper will be used to assess the adequacy of a linear trend-stationary specification with a single break in level just after 1929, i.e., the model given by

$$Y_t = \gamma_{01} + d_2 \left(\frac{t}{T} \right) \gamma_{02} + \frac{t}{T} \gamma_{03} + u_t,$$

where $d_2(\cdot)$ is unity for all observations corresponding to 1930 or later and zero for all observations corresponding to years 1909 through 1929. The standard normal kernel was used to construct the estimates of $\gamma(\cdot)$ and $\gamma'(\cdot)$ given above in (19) and (22), respectively, while the function $k(\cdot)$ appearing in (27) was set to be the Epanechnikov (1969) kernel. As was done in the simulations summarized in Figures 1–4, the bandwidths a_T and b_T used in (27) and (28) were set to $a_T = b_T = .5T^{-\frac{2}{9}}$.

Various settings of bandwidth were tried out when implementing the Priestley-Chao estimates given in (19) and (22). In particular, the estimate of γ given in (19) was implemented with a bandwidth given by $h_l = c_l T^{-\frac{2}{5}}$, where $c_l \in \{.125, 1.0, 8.0\}$. Similarly, the estimate of γ' given in (22) used the bandwidth $h_d = c_d T^{-\frac{2}{7}}$, where $c_d \in \{.125, 1.0, 8.0\}$. As indicated in Figures 5 and 6, the quality of the fit provided by the Priestley-Chao estimates to their respective estimands is very sensitive to the particular bandwidth used.

The Priestley-Chao estimates are of course ingredients in the construction of the proposed test statistic given above in (39). The effect of the bandwidths used

¹⁹Cf. e.g., Perron (1989); Perron and Wada (2006).

²⁰The data are in the form of natural logarithms of the measurements appearing in the source indicated in Nelson and Plosser (1982, note 10). All fourteen historical macroeconomic time series analyzed by Nelson and Plosser (1982) may be downloaded from <http://www.spatial-econometrics.com/data/contents.html>.

in their implementation on the proposed test is summarized in Table 2, which displays the value of the Wald statistic in (39) as a function of the bandwidth scaling constants c_l and c_d . It is clear that the test is highly sensitive to the bandwidth used to construct the estimate of $\gamma(\cdot)$, while the particular bandwidth used to construct $\hat{\gamma}'_T(\cdot)$ as given in (22) appears to have little effect on the conclusion of the proposed test. Of the various combinations tried, the setting $h_l = .125T^{-\frac{2}{5}}$ appears to provide the best nonparametric fit to the data, as is clear from Figure 5. Using $h_l = .125T^{-\frac{2}{5}}$ also leads to the same conclusion obtained by Perron (1989) using a different method, namely, that real U.S. GNP is stationary about a linear trend when one allows for a one-time break in level just after 1929.

The second series considered consists of seasonally adjusted quarterly Gross Domestic Product from 1947 through to the end of 2007 in billions of chained 2000 dollars.²¹ In this case the model being validated involves linear trend-stationarity with a single break in slope just after the fourth quarter of 1973, i.e.,

$$Y_t = \gamma_{01} + \frac{t}{T} \left(\gamma_{02} + d_3 \left(\frac{t}{T} \right) \gamma_{03} \right) + u_t,$$

where $d_3(\cdot)$ is unity for all observations corresponding to the fourth quarter of 1973 or later and zero for all observations corresponding to the first quarter of 1947 through to third-quarter 1973, inclusive. The test statistic in (39) was constructed using the same combinations of bandwidths and choices of kernels as was described above in relation to the Nelson and Plosser (1982) real GNP series. Table 3 indicates that in the case of the postwar real GDP series, the conclusion of the test based on the Wald statistic given in (39) continues to be highly sensitive to the bandwidth used to implement the Priestley-Chao estimate of the trend parameter, while remaining quite insensitive to the bandwidth used to estimate $\gamma'(\cdot)$. A glance at Figure 7 indicates that as was the case for annual real GNP, the bandwidth $h_l = .125T^{-\frac{2}{5}}$ provides a good nonparametric fit for the evolution of quarterly postwar real GDP. Its use in the construction of the test statistic in (39), however, does not lead one to conclude that postwar real GDP is well described by linear trend-stationarity with a single break in slope around the onset of the 1973 oil crisis.²² That said, it seems plausible from the pattern of the results displayed in Figure 5–Table 2 and Figure 7–Table 3 that further “undersmoothing” of the Priestley-Chao estimate of γ will have the effect of pushing the realized value of

²¹The data used are the natural logarithms of the measurements downloaded from <http://www.bea.gov>.

²²This is at odds with the conclusions of Perron (1989) and Perron and Wada (2006).

the proposed test statistic below the .95-quantile of the standard normal distribution. It seems clear that further research on bandwidth selection in this context is needed.

7 Conclusion

This paper has presented a lack-of-fit test for segmented-trend stationarity based on an efficient estimator of the vector whose components are the total variation norms of the corresponding components of the trend-function parameter. The data-generating process of the stochastic disturbances under the null is largely unspecified. The proposed testing procedure is shown to be locally uniformly most powerful in large samples against deviations from the null of stability of the trend parameter. As such, the test is capable of detecting all manner of conceivable breaks in trend that have not already been explicitly modelled by the researcher. Monte Carlo simulations reported in Section 5 show that the proposed test is potentially correctly sized and powerful against discrete breaks and continuous changes in trend, as well as against the alternative of a unit root process. As underscored by the empirical results presented in Section 6, however, these good qualities appear to be particularly dependent on a suitable choice of bandwidth used to implement the preliminary nonparametric estimate of the trend parameter given in (19). A practical rule for the selection of the bandwidth to be used in this context is clearly worthy of further research.

Another potential topic for further investigation involves the conversion of the test for trend stationarity proposed here into a test of the unit-root hypothesis. In particular, this would entail deriving the limiting distribution of the statistic given in (39) under the assumption that the data are drawn from an $I(1)$ process. Evidence against the unit-root hypothesis would naturally be associated with small realized values of the statistic given in (39). A unit-root test of this nature would complement the specification test for segmented-trend stationarity proposed in this paper by providing the researcher with yet another method of distinguishing between processes characterized by what are effectively unpredictable trend breaks every period from those where the trend changes only infrequently.

A Further preliminary results

This appendix gathers together various preliminary results needed in the proofs of Lemmas 2–5 and of Theorem 5.

A.1 Rates of convergence of the Priestley-Chao estimator

This appendix section is concerned with the uniform rates of convergence over the unit interval of the estimators of $\gamma(\cdot)$ and of $\gamma'(\cdot)$ given in (19) and (22), respectively.

For $\mu\left(\frac{t}{T}\right) \equiv \mathbf{d}\left(\frac{t}{T}\right)^\top \boldsymbol{\gamma}\left(\frac{t}{T}\right)$ with \mathbf{d} satisfying the conditions of Assumption 3, let

$$Y_t = \mu\left(\frac{t}{T}\right) + u_t \quad (42)$$

denote the basic model under consideration. Assuming that $\{u_t\}$ satisfies the conditions of Assumption 4, the strong invariance principle of Wu (2007) applies, and it is possible to approximate the original dependent variables $\{Y_t\}$ with $\{\tilde{Y}_t\}$ drawn from the model

$$\tilde{Y}_t \equiv \mu\left(\frac{t}{T}\right) + Z_t, \quad (43)$$

where $\{Z_t\}$ be an iid $N(0, \sigma^2)$ process with $\sigma^2 \equiv \sum_{t=-\infty}^{\infty} E[u_0 u_t]$.

Consider the estimator of Priestley and Chao (1972) applied to the model given in (43), to wit:

$$\tilde{\mu}_T(s) \equiv \sum_{t=1}^T w_{1,t,T}(s) \tilde{Y}_t \quad (s \in (0, 1)) \quad (44)$$

where $w_{l,t,T}(\cdot)$ is as in (20) above. Similarly, for $s \in (0, 1)$, let

$$\tilde{\mu}'_T(s) \equiv \sum_{t=1}^T w'_{d,t,T}(s) \tilde{Y}_t, \quad (45)$$

denote the Priestley-Chao estimator of the derivative of $\mu(s)$ where $w'_{d,t,T}(s)$ denotes the derivative of the expression given in (20) but with bandwidth h_d in place of h_l . The following result under the conditions of Assumptions 3, 6 and 7 is an easy consequence of the sort of arguments appearing in e.g., Priestley and Chao (1972, Section 4):

Lemma 6. *Under the conditions of Assumptions 3, 6 and 7 we have for $s \in (0, 1)$ that*

1. $\tilde{\mu}_T(s)$ is consistent for $\mu(s)$ in (43) at a $\sqrt{Th_l}$ rate;
2. $\tilde{\mu}'_T(s)$ is consistent for the derivative $\mu'(s)$ of $\mu(s)$ in (43) at a $\sqrt{Th_d^3}$ rate.

Now consider the Priestley-Chao procedure applied to the original model (42). In particular, suppose that for $s \in (0, 1)$, $\mu(s)$ and $\mu'(s)$ are estimated by

$$\hat{\mu}_T(s) \equiv \sum_{t=1}^T w_{l,t,T}(s) Y_t \quad (46)$$

and

$$\hat{\mu}'_T(s) \equiv \sum_{t=1}^T w'_{d,t,T}(s) Y_t, \quad (47)$$

respectively. Assumption 4 and the strong invariance principle of Wu (2007) allow the asymptotic behaviour of $\hat{\mu}_T$ and $\hat{\mu}'_T$ to be uniformly almost surely approximated on the unit interval by that of $\tilde{\mu}_T$ and $\tilde{\mu}'_T$, respectively:

Lemma 7. *Under the conditions of Assumptions 3, 4, 6 and 7 we have*

1. $\max_{0 \leq s \leq 1} |\hat{\mu}_T(s) - \tilde{\mu}_T(s)| = o_{a.s.} \left(T^{-\frac{3}{4}} h_l^{-1} \log T \right);$
2. $\max_{0 \leq s \leq 1} |\hat{\mu}'_T(s) - \tilde{\mu}'_T(s)| = o_{a.s.} \left(T^{-\frac{3}{4}} h_d^{-2} \log T \right).$

Proof. 1. The bounded variation of $K(\cdot)$ gives

$$\begin{aligned} \Omega_{lT}(s) &\equiv |w_{l,1,T}(s)| + \sum_{j=2}^T |w_{l,j,T}(s) - w_{l,j-1,T}(s)| \\ &= O\left(\frac{1}{Th_l}\right), \end{aligned}$$

while the embedding of (14) above yields the bound

$$|\hat{\mu}_T(s) - \tilde{\mu}_T(s)| = o_{a.s.} \left(\Omega_{lT}(s) T^{\frac{1}{4}} \log T \right) = o_{a.s.} \left(T^{-\frac{3}{4}} h_l^{-1} \log T \right),$$

which can be extended to a uniform approximation:

$$\max_{0 \leq s \leq 1} |\hat{\mu}_T(s) - \tilde{\mu}_T(s)| = o_{a.s.} \left(T^{-\frac{3}{4}} h_l^{-1} \log T \right).$$

2. Setting $\Omega'_{dT}(s) \equiv |w'_{d,1,T}(s)| + \sum_{j=2}^T |w'_{d,j,T}(s) - w'_{d,j-1,T}(s)|$, the bounded variation of $K'(\cdot)$ yields the bound

$$\Omega'_{dT}(s) = O\left(\frac{1}{Th_d^2}\right).$$

As such, (14) can be exploited to yield the uniform approximation

$$\max_{0 \leq s \leq 1} |\hat{\mu}'_T(s) - \tilde{\mu}'_T(s)| = o_{a.s.} \left(T^{-\frac{3}{4}} h_d^{-2} \log T \right).$$

□

In this paper interest is of course centred more on the trend-function parameter $\gamma(\cdot)$ and the derivatives of its components. Let

$$\mathbf{v}(\cdot) \equiv (|\gamma'_1(\cdot)|, \dots, |\gamma'_k(\cdot)|)^\top, \quad (48)$$

i.e., the vector whose components are the magnitudes of the derivatives of the corresponding components of $\gamma(\cdot)$. For $\hat{\gamma}_T(\cdot)$ as given above in (19) and $\hat{\mathbf{v}}_T(\cdot)$ as given in (24), we have the following uniform rates of convergence:

Lemma 8. *Under the conditions of Assumptions 3, 4, 6 and 7 we have*

1. $\max_{0 \leq s \leq 1} \|\hat{\gamma}_T(s) - \gamma(s)\| = O_p\left(T^{-\frac{1}{2}} h_l^{-\frac{1}{2}}\right);$
2. $\max_{0 \leq s \leq 1} \|\hat{\nu}_T(s) - \nu(s)\| = O_p\left(T^{-\frac{3}{2}} h_d^{-\frac{3}{2}}\right).$

Proof. Note that the pointwise rates of convergence in Lemma 6 can be made to hold uniformly over $s \in (0, 1)$ by virtue of the bounded variation of $K(\cdot)$ and of its derivative $K'(\cdot)$. As such,

$$\max_{0 \leq s \leq 1} |\tilde{\mu}_T(s) - \mu(s)| = O_p\left(T^{-\frac{1}{2}} h_l^{-\frac{1}{2}}\right) \quad (49)$$

and

$$\max_{0 \leq s \leq 1} |\tilde{\mu}'_T(s) - \mu'(s)| = O_p\left(T^{-\frac{1}{2}} h_d^{-\frac{3}{2}}\right). \quad (50)$$

Combining (49) and (50) with the conclusions of Lemma 7 yields the desired conclusion. \square

A.2 Further lemmas

This appendix section collects various preliminary results needed in the proofs of Lemmas 2–5 and of Theorem 5. In what follows, given a generic statistic $\hat{\mathbf{W}}_T$ obtained from the basic model given above in (1), the statistic of the form $\tilde{\mathbf{W}}_T$ will denote the procedure implied by $\hat{\mathbf{W}}_T$, but applied to the iid normal-error model given in (15). As such, $\tilde{\nu}_T(s) \equiv |\tilde{\gamma}'_T(s)|$, $\tilde{u}_{t,T} \equiv Y_t - \mathbf{d}\left(\frac{t}{T}\right)^\top \tilde{\gamma}_T\left(\frac{t}{T}\right)$, etc. Also let

$$\tilde{\delta}_{t,T} \equiv \mathbf{d}\left(\frac{t}{T}\right)^\top \left(\tilde{\gamma}_T\left(\frac{t}{T}\right) - \gamma\left(\frac{t}{T}\right) \right),$$

and let

$$F_Z(z) \equiv \Phi\left(\frac{z}{\sigma}\right)$$

and

$$f_Z(z) \equiv \frac{1}{\sigma} \phi\left(\frac{z}{\sigma}\right)$$

denote the distribution and density functions, respectively, of the regression error appearing in (15).

Lemma 9. *Suppose Assumptions 3, 4, 6, 7 and 8 hold. Then for $\nu(\cdot)$ as given in (48),*

$$\frac{1}{T} \left[\hat{\nu}_T\left(\frac{t}{T}\right) - \nu\left(\frac{t}{T}\right) + \left(\hat{\nu}_T\left(\frac{t}{T}\right) - \bar{\nu} \right) \delta_{t,T} \right] = o_{\theta_T}\left(T^{-\frac{1}{2}}\right).$$

Proof. We have

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \left[\hat{\nu}_T\left(\frac{t}{T}\right) - \nu\left(\frac{t}{T}\right) + \left(\hat{\nu}_T\left(\frac{t}{T}\right) - \bar{\nu} \right) \delta_{t,T} \right] \\ & \leq \max_{0 \leq s \leq 1} |\mathbf{d}(s)^\top (\hat{\gamma}_T(s) - \gamma(s))| \left\| \frac{1}{T} \sum_{t=1}^T \hat{\nu}_T\left(\frac{t}{T}\right) - \bar{\nu} \right\| + \left\| \frac{1}{T} \sum_{t=1}^T \left[\hat{\nu}_T\left(\frac{t}{T}\right) - \nu\left(\frac{t}{T}\right) \right] \right\|. \end{aligned}$$

Appealing to Lemma 8, note that

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{v}}_T \left(\frac{t}{T} \right) - \bar{\mathbf{v}} \\
&= \frac{1}{T} \sum_{t=1}^T \left[\hat{\mathbf{v}}_T \left(\frac{t}{T} \right) - \mathbf{v} \left(\frac{t}{T} \right) \right] + \frac{1}{T} \sum_{t=1}^T \mathbf{v} \left(\frac{t}{T} \right) - \bar{\mathbf{v}} \\
&= o_{\theta_T} \left(T^{-\frac{3}{2}} h^{-\frac{3}{2}} \right) + O \left(T^{-\frac{1}{2}} \right) \\
&= O_{\theta_T} \left(T^{-\frac{1}{2}} \right).
\end{aligned}$$

The result follows from a further appeal to Lemma 8. \square

For $x \in \mathbb{R}$ and $\mathbf{z} \equiv (z_1, \dots, z_T)^\top$, write

$$\tilde{l}_{ZT}(x) \equiv \tilde{l}_{ZT}(x, \mathbf{z}) \equiv - \frac{a_T^{-2} \sum_{t=1}^T k' \left(\frac{x-z_t}{a_T} \right)}{b_T + a_T^{-1} \sum_{t=1}^T k \left(\frac{x-z_t}{a_T} \right)},$$

Let

$$\tilde{l}_{ZT}^{(i)}(x, \mathbf{z}) \equiv \frac{\partial^i}{\partial x^i} \tilde{l}_{ZT}(x, \mathbf{z}). \tag{51}$$

We have the following.

Lemma 10. *Given Assumption 8, there exists a positive constant c_0 such that for every $x \in \mathbb{R}$ and $\mathbf{z}, \mathbf{z}' \in \mathbb{R}^T$ and $i \in \{0, 1, 2\}$, the following bounds hold:*

$$\left| \tilde{l}_{ZT}^{(i)}(x, \mathbf{z}) \right| \leq \frac{c_0}{a_T^{i+1}} \tag{52}$$

$$\left| \tilde{l}_{ZT}^{(i)}(x, \mathbf{z}) - \tilde{l}_{ZT}^{(i)}(x, \mathbf{z}') \right| \leq \frac{c_0}{a_T^{3+i} b_T} \sum_{t=1}^T \min \{ a_T, \|\mathbf{z}'_t - \mathbf{z}_t\| \} \tag{53}$$

$$\left| \tilde{l}_{ZT}^{(i)}(x, \mathbf{z}) - \tilde{l}_{ZT}^{(i)}(x, \mathbf{z}') \right|^2 \leq \frac{c_0}{a_T^{5+2i} b_T} \sum_{t=1}^T \|\mathbf{z}'_t - \mathbf{z}_t\|^2. \tag{54}$$

Proof. Follows directly from the definitions. \square

The following modification of Schick (1987, Lemma 3.1) is needed in order to prove Lemmas 3–5.

Lemma 11. *Let $\mathbf{Z}^{(T)} \equiv (Z_1, \dots, Z_T)^\top$. For each pair (t, T) of positive integers with $1 \leq t \leq T$, let $\mathbf{L}_{t,T} : \mathbb{R} \times \mathbb{R}^T \rightarrow \mathbb{R}^k$ be a measurable function. Let*

$$\tilde{\mathbf{L}}_{t,T}(\cdot) \equiv \boldsymbol{\lambda}_{t,T} \left(\cdot, \mathbf{Z}^{(T)} \right)$$

be an estimate of $\mathbf{L}_{t,T}(\cdot, \mathbf{Z}^{(T)})$, where $\lambda_{t,T} : \mathbb{R} \times \mathbb{R}^T \rightarrow \mathbb{R}^k$ is a measurable function. Also let

$$\bar{\mathbf{L}}_{t,T}(z) \equiv E \left[\tilde{\mathbf{L}}_{t,T}(z) \middle| Z_1, \dots, Z_{t-1}, Z_{t+1}, \dots, Z_T \right].$$

Suppose that the following conditions hold:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \int \tilde{\mathbf{L}}_{t,T}(z) dF_Z(z) = o_p(1) \quad (55)$$

$$\frac{1}{T} \sum_{t=1}^T E \left[\int \left\| \tilde{\mathbf{L}}_{t,T}(z) - \mathbf{L}_{t,T}(z, \mathbf{Z}^{(T)}) \right\|^2 dF_Z(z) \right] = o(1) \quad (56)$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\tilde{\mathbf{L}}_{t,T}(Z_t) - \bar{\mathbf{L}}_{t,T}(Z_t) \right) = o_p(1) \quad (57)$$

$$\sum_{t=1}^T E \left[\int \left\| \tilde{\mathbf{L}}_{t,T}(z) - \bar{\mathbf{L}}_{t,T}(z) \right\|^2 dF_Z(z) \right] = o(1). \quad (58)$$

Then

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\tilde{\mathbf{L}}_{t,T}(Z_t) - \mathbf{L}_{t,T}(Z_t, \mathbf{Z}^{(T)}) \right) = o_p(1).$$

Proof. We write

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\tilde{\mathbf{L}}_{t,T}(Z_t) - \mathbf{L}_{t,T}(Z_t, \mathbf{Z}^{(T)}) \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^T (\mathbf{B}_{t1} + \mathbf{B}_{t2} + \mathbf{B}_{t3}),$$

where

$$\begin{aligned} \mathbf{B}_{t3} &\equiv \int \bar{\mathbf{L}}_{t,T}(z) dF_Z(z) \\ \mathbf{B}_{t1} &\equiv \tilde{\mathbf{L}}_{t,T}(Z_t) - \bar{\mathbf{L}}_{t,T}(Z_t) \\ \mathbf{B}_{t2} &\equiv \bar{\mathbf{L}}_{t,T}(Z_t) - \mathbf{L}_{t,T}(Z_t, \mathbf{Z}^{(T)}) - \mathbf{B}_{t3}. \end{aligned}$$

By (57), we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{B}_{t1} = o_p(1).$$

By (55) and (58), we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{B}_{t3} = o_p(1).$$

As such, it remains to show that

$$E \left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{B}_{t2} \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{B}_{t2} \right)^\top \right] = \frac{1}{T} \sum_{t=1}^T E [\mathbf{B}_{t2} \mathbf{B}_{t2}^\top] + \frac{1}{T} \sum_{t \neq s} E [\mathbf{B}_{t2} \mathbf{B}_{s2}^\top] \rightarrow 0.$$

By (56) and (58) and the nature of conditional variances, we have

$$\begin{aligned} \left\| \frac{1}{T} \sum_{t=1}^T E [\mathbf{B}_{t2} \mathbf{B}_{t2}^\top] \right\| &\leq \frac{1}{T} \sum_{t=1}^T \left\| E \left[\left(\bar{\mathbf{L}}_{t,T}(Z_t) - \mathbf{L}_{t,T}(Z_t, \mathbf{Z}^{(T)}) \right) \left(\bar{\mathbf{L}}_{t,T}(Z_t) - \mathbf{L}_{t,T}(Z_t, \mathbf{Z}^{(T)}) \right)^\top \right] \right\| \\ &\rightarrow 0. \end{aligned}$$

Let

$$\bar{\mathbf{B}}_{t2}^{(s)} \equiv E [\mathbf{B}_{t2} | Z_1, \dots, Z_{s-1}, Z_{s+1}, \dots, Z_T].$$

(Clearly, $\bar{\mathbf{B}}_{s2}^{(s)} = \mathbf{0}$ with probability one.)

It is possible to show that for $t \neq s$,

$$E [\bar{\mathbf{B}}_{t2}^{(s)} \mathbf{B}_{s2}^\top] = E [\mathbf{B}_{t2} \bar{\mathbf{B}}_{s2}^{(t)\top}] = E [\bar{\mathbf{B}}_{t2}^{(s)} \bar{\mathbf{B}}_{s2}^{(t)\top}] = \mathbf{0}.$$

This, along with the Cauchy-Schwarz inequality, the nature of conditional variances and (58) yields

$$\begin{aligned} &\frac{1}{T} \sum_{t \neq s} \left\| E [\mathbf{B}_{t2} \mathbf{B}_{s2}^\top] \right\| \\ &= \frac{1}{T} \sum_{t \neq s} \left\| E \left[\left(\mathbf{B}_{t2} - \bar{\mathbf{B}}_{t2}^{(s)} \right) \left(\mathbf{B}_{s2} - \bar{\mathbf{B}}_{s2}^{(t)} \right)^\top \right] \right\| \\ &\leq \frac{1}{T} \sum_{t \neq s} \left\| E \left[\left(\mathbf{B}_{t2} - \bar{\mathbf{B}}_{t2}^{(s)} \right) \left(\mathbf{B}_{t2} - \bar{\mathbf{B}}_{t2}^{(s)} \right)^\top \right] \right\| \\ &\leq \frac{1}{T} \sum_{t \neq s} E \left[\int \left\| \tilde{\mathbf{L}}_{t,T}(z) - \bar{\mathbf{L}}_{s,T}(z) \right\|^2 dF_Z(z) \right] \\ &\rightarrow \mathbf{0}. \end{aligned}$$

This concludes the proof. \square

The next preliminary result requires the use of the following quantities. For $s \in (0, 1)$, let $\tilde{\mu}_{T,-k}(s)$ denote the Priestley-Chao estimator applied to $\mu(s)$ in (43) without the use of the k th observation, i.e.,

$$\tilde{\mu}_T(s) \equiv \sum_{t \neq k} w_{l,t,T}(s) \tilde{Y}_t,$$

where as before

$$w_{l,t,T}(s) \equiv \frac{1}{Th_l} K \left(\frac{1}{h_l} \left(\frac{t}{T} - s \right) \right).$$

Define

$$\begin{aligned} N_{T1} &\equiv \frac{1}{T} \sum_{t=1}^T E \left[\left(\tilde{\mu}_T \left(\frac{t}{T} \right) - \mu \left(\frac{t}{T} \right) \right)^2 \right]; \\ N_{T2} &\equiv \sum_{t=1}^T E \left[\left(\tilde{\mu}_T \left(\frac{t}{T} \right) - \tilde{\mu}_{T,-t} \left(\frac{t}{T} \right) \right)^2 \right]; \end{aligned}$$

$$N_{T3} \equiv \frac{1}{T} \sum_{t=1}^T \sum_{k \neq t} E \left[\left(\tilde{\mu}_T \left(\frac{t}{T} \right) - \tilde{\mu}_{T,-k} \left(\frac{t}{T} \right) \right)^2 \right].$$

Note that

$$N_{T1} = O \left(T^{-(1+r)} \right) \quad (59)$$

$$N_{T2} = O \left(T^{-(1+2r)} \right) \quad (60)$$

$$N_{T3} = O \left(T^{-(1+2r)} \right). \quad (61)$$

Conditions (59)–(61) are used in the lemma that follows. For brevity, let $\alpha \equiv 1+r$ and $\alpha_* \equiv 1+2r$, and let

$$\tilde{\mathbf{Y}}_{-t,T} \equiv (\tilde{Y}_1, \dots, \tilde{Y}_{t-1}, \tilde{Y}_{t+1}, \dots, \tilde{Y}_T)^\top.$$

Lemma 12. *Suppose that $T a_T^4 b_T^2 \rightarrow \infty$. Then for $i \in \{0, 1\}$, define*

$$\begin{aligned} L_{T1}^{(i)} &\equiv \frac{1}{T} \sum_{t=1}^T E \left[\sup_{y \in \mathbb{R}} \left(\tilde{l}_{ZT}^{(i)} \left(y - \tilde{\mu}_T \left(\frac{t}{T} \right) \right) - \tilde{l}_{ZT}^{(i)} \left(y - \tilde{\mu}_{T,-t} \left(\frac{t}{T} \right) \right) \right)^2 \right]; \\ L_{T2} &\equiv \frac{1}{T^2} \sum_{t=1}^T E \left[\sup_{z \in \mathbb{R}} \left(\tilde{l}_{ZT} \left(z + \mu \left(\frac{t}{T} \right) - \tilde{\mu}_{T,-t} \left(\frac{t}{T} \right) \right) - \tilde{l}_{ZT}(z) \right)^2 \right]; \\ L_{T3}^{(i)} &\equiv \frac{1}{T^2} \sum_{t=1}^T \sum_{k \neq t} E \left[\left(\tilde{l}_{ZT}^{(i)} \left(\tilde{Y}_t - \tilde{\mu}_{T,-t} \left(\frac{t}{T} \right) \right) - \tilde{l}_{ZT}^{(i)} \left(\tilde{Y}_t - E \left[\tilde{\mu}_{T,-t} \left(\frac{t}{T} \right) \mid \tilde{\mathbf{Y}}_{-t,T} \right] \right) \right)^2 \right]; \\ L_{T4} &\equiv E \left[\int \left(\tilde{l}_{ZT}(z) - l_Z(z) \right)^2 dF_Z(z) \right]; \end{aligned}$$

where $\tilde{l}_{ZT}^{(i)}$ is as given above in (51).

Then

$$L_{T1}^{(i)} = T^{-1} a_T^{-2i} O_p \left(a_T^{-4} T^{-\alpha_*} + T^{-1} a_T^{-4} b_T^{-2} + a_T^{-5} b_T^{-1} T^{-2\alpha} \right) \quad (62)$$

$$L_{T2} = T^{-1} O_p \left(T^{-2} a_T^{-4} b_T^{-2} + a_T^{-5} b_T^{-1} T^{-2\alpha} \right) \quad (63)$$

$$L_{T3}^{(i)} = T^{-1} a_T^{-2i} O_p \left(T^{-1} a_T^{-4} b_T^{-2} + a_T^{-5} b_T^{-1} T^{-2\alpha} \right) \quad (64)$$

$$L_{T4} = O \left(a_T^{-5} b_T^{-1} T^{-2\alpha} \right). \quad (65)$$

Proof. The proof is given in detail in the following appendix section, i.e., Appendix A.3. \square

A.3 Proof of Lemma 12

The proof of Lemma 12 requires the following two lemmas, the first of which is a result of Schick (1987).

Lemma 13. Under Assumption 8 on a kernel function $k(\cdot)$, if for $z \in \mathbb{R}$ we have

$$f_T(z) \equiv \int f(z - a_T t) k(t) dt,$$

then

$$\begin{aligned} \Sigma_{T1} &\equiv E \left[\int \left(\tilde{l}_{ZT}(z) + \frac{f'_T(z)}{f_T(z) + b_T} \right)^2 dF_Z(z) \right] = O(T^{-1} a_T^{-4} b_T^{-2}); \\ \Sigma_{T2} &\equiv \int \left(\frac{f'_T(z)}{b_T + f_T(z)} - \frac{f'_Z(t)}{f_Z(t)} \right)^2 f_Z(t) dt \rightarrow 0. \end{aligned}$$

The next lemma is an easy consequence of conditions (59)–(61).

Lemma 14. Define

$$\begin{aligned} M_{T1} &\equiv \frac{1}{T^2} \sum_{t=1}^T \sum_{k \neq t} E \left[(Z_t - \tilde{u}_{t,T,-k})^2 \right]; \\ M_{T2} &\equiv \frac{1}{T} \sum_{t=1}^T \sum_{k \neq t} E \left[(\tilde{u}_{t,T} - \tilde{u}_{t,T,-k})^2 \right]; \\ M_{T3} &\equiv \frac{1}{T^2} \sum_{t=1}^T \sum_{k \neq t} \sum_{l \neq t} \sum_{k \neq l} E \left[(\tilde{u}_{l,T,-k} - E[\tilde{u}_{l,T,-k} | Z_1, \dots, Z_{t-1}, Z_{t+1}, \dots, Z_T])^2 \right]. \end{aligned}$$

Then

$$\begin{aligned} M_{T1} &= O(T^{-2\alpha}); \\ M_{T2} &= O(T^{-\alpha_*}); \\ M_{T3} &= O(T^{-\alpha_*}). \end{aligned}$$

Proof. We have

$$M_{T1} \leq \frac{1}{T} \sum_{t=1}^T E \left[(Z_t - \tilde{u}_{t,T})^2 \right] \leq N_{T1},$$

while

$$M_{T3} \leq M_{T2} = N_{T3}. \quad \square$$

Returning to the argument of Lemma 12 proper, appeal to results (52)–(54) in Lemma 10 to deduce that there exists a positive constant C such that

$$L_{T1}^{(i)} \leq C a_T^{-2i} (T^{-1} a_T^{-4} N_{T2} + T^{-2} a_T^{-4} b_T^{-2} + T^{-1} a_T^{-5} b_T^{-1} M_{T2}); \quad (66)$$

$$L_{T2} \leq C (T^{-1} a_T^{-4} N_{T1} + T^{-3} a_T^{-4} b_T^{-2} + T^{-1} a_T^{-5} b_T^{-1} M_{T1}); \quad (67)$$

$$L_{T3}^{(i)} \leq C a_T^{-2i} (T^{-1} a_T^{-4} N_{T3} + T^{-2} a_T^{-4} b_T^{-2} + T^{-1} a_T^{-5} b_T^{-1} M_{T3}); \quad (68)$$

$$L_{T4} \leq C (a_T^{-5} b_T^{-1} N_{T1} + \Sigma_{T1} + \Sigma_{T2}). \quad (69)$$

Conclusions (62)–(65) follow from (66)–(69), Lemma 13, Lemma 14 and conditions (59)–(61). This concludes the proof of Lemma 12.

B Proofs of Lemmas 2–5

B.1 Proof of Lemma 2

Note the following.

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^T \int \left(\hat{l}'_{UT}(u - \delta_{t,T}) - \hat{l}'_{UT}(u) \right) dF_U(u) \\
= & \frac{1}{T} \sum_{t=1}^T \left[\int \left(\hat{l}'_{UT}(u - \delta_{t,T}) - \hat{l}'_{UT}(u) \right) dF_U(u) - \int \left(\tilde{l}'_{ZT}(z - \tilde{\delta}_{t,T}) - \tilde{l}'_{ZT}(z) \right) dF_Z(z) \right] \\
& + \frac{1}{T} \sum_{t=1}^T \int \left(\tilde{l}'_{ZT}(z - \tilde{\delta}_{t,T}) - \tilde{l}'_{ZT}(z) \right) dF_Z(z). \tag{70}
\end{aligned}$$

The first term in (70) is seen to be bounded above by

$$\begin{aligned}
& \max_{1 \leq t \leq T} \left| \delta_{t,T} - \tilde{\delta}_{t,T} \right| \left| \int \hat{l}'_{UT}(u) f'_U(u) du - \int \tilde{l}'_{ZT} f'_Z(z) dz \right| \\
= & o_{a.s.} \left(T^{-\frac{3}{4} + 2r} \log T \right) \cdot o_{a.s.} \left(T^{-\frac{3}{4}} \log T \cdot T^{-1} a_T^{-2} \right) \\
= & o_{a.s.} \left(T^{-\frac{1}{2}} \right).
\end{aligned}$$

Now consider the second term in (70). We note that for

$$\tilde{A}_T \equiv \max_{0 \leq s \leq 1} \|\tilde{\mathbf{v}}_T(s)\| + 1,$$

we have

$$\max_{1 \leq t \leq T} \left\| \tilde{\mathbf{v}}_T \left(\frac{t}{T} \right) \right\| = O_p \left(\tilde{A}_T \right).$$

Let J_Z denote the Fisher information for location corresponding to F_Z . Note that it is finite, and therefore that

$$\int |f_Z(z+d) - f_Z(z)| dz \leq |d| \sqrt{J_Z}.$$

Combining this with (52) in Lemma 10, we have

$$\begin{aligned}
& \left| \int \left(\tilde{l}'_{ZT}(z-d) - \tilde{l}'_{ZT}(z) \right) f_Z(z) dz \right| \\
\leq & \int \left| \tilde{l}'_{ZT}(z) \right| |f_Z(z+d) - f_Z(z)| dz \\
\leq & c_0 a_T^{-2} |d| \sqrt{J_Z}.
\end{aligned}$$

The second term in (70) is accordingly $o_{\theta_T} \left(T^{-\frac{1}{2}} \right)$.

$$\begin{aligned}
& \left| \int \left(\tilde{l}_{ZT}(z-d) - \tilde{l}_{ZT}(z) + d\tilde{l}'_{ZT}(z) \right) f_Z(z) dz \right| \\
& \leq |d| \int_0^1 \left| \int \left(\tilde{l}'_{ZT}(z-\lambda d) - \tilde{l}'_{ZT}(z) \right) f_Z(z) dz \right| d\lambda \\
& \leq |d| \int_0^1 \int \left| \tilde{l}'_{ZT}(z) (f_Z(z+\lambda d) - f_Z(z)) \right| dz d\lambda \\
& \leq c_0 a_T^{-2} d^2 \sqrt{J_Z}.
\end{aligned}$$

The demonstration of the second part of Lemma 2 is similar.

B.2 Proof of Lemma 3

The proof is organized along the lines of showing the convergence of the quantities corresponding to Δ_{T1} , Δ_{T2} and Δ_{T3} in (33) above. The focus here is on showing the convergence of Δ_{T2} , since the convergence of Δ_{T1} and Δ_{T3} is easily shown. In particular, we show that Δ_{T2} converges by applying Lemma 11 with

$$\begin{aligned}
\tilde{\mathbf{L}}_{t,T}(z) &\equiv \tilde{\mathbf{v}}_T \left(\frac{t}{T} \right) \left\{ \tilde{l}_{ZT}(z) - \int \tilde{l}_{ZT}(z - \tilde{\delta}_{t,T}) dF_Z(z) \right\}; \\
\mathbf{L}_{t,T}(z) &\equiv \tilde{\mathbf{v}}_T \left(\frac{t}{T} \right) l_Z(z).
\end{aligned}$$

It follows that conditions (55)–(58) need to be shown. This is done in sequence.

B.2.1 Verification of condition (55):

Applying Lemma 8 to the model in (15) at the appropriate location, and recalling the definition of L_{T2} in Lemma 12, we have

$$\begin{aligned}
& \left\| E \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \int \tilde{\mathbf{L}}_{t,T}(z) dF_Z(z) \right] \right\|^2 \\
& \leq \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T E \left[\tilde{\mathbf{v}}_T \left(\frac{t}{T} \right) \left\{ \tilde{l}_{ZT}(z) dF_Z(z) - \int \tilde{l}_{ZT}(z - \tilde{\delta}_{t,T}) dF_Z(z) \right\} \right] \right\|^2 \\
& \leq \frac{2}{T} \left\{ \sum_{t=1}^T E \left[\left\| \tilde{\mathbf{v}}_T \left(\frac{t}{T} \right) - \mathbf{v} \left(\frac{t}{T} \right) \right\|^2 \right] + \sum_{t=1}^T \left\| \mathbf{v} \left(\frac{t}{T} \right) \right\|^2 \right\} \times T^2 L_{T2} \\
& = 2T \{ O(T^{3r}) + O(T) \} \times T^{-1} O(T^{-2} a_T^{-4} b_T^{-2} + a_T^{-5} b_T^{-1} T^{-2-2r}) \\
& = O(T) \times O(T^{-2} a_T^{-4} b_T^{-2} + a_T^{-5} b_T^{-1} T^{-2-2r}) \\
& = O(T^{-1} a_T^{-4} b_T^{-2} + a_T^{-5} b_T^{-1} T^{-1-2r}) \\
& = o(1).
\end{aligned}$$

B.2.2 Verification of condition (56):

We have

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^T E \left[\int \left\| \tilde{\mathbf{L}}_{t,T}(z) - \mathbf{L}_{t,T}(z) \right\|^2 dF_Z(z) \right] \\
& \leq \frac{1}{T} \left\{ \sum_{t=1}^T E \left[\left\| \tilde{\mathbf{v}}_T \left(\frac{t}{T} \right) - \mathbf{v} \left(\frac{t}{T} \right) \right\|^2 \right] + \sum_{t=1}^T \left\| \mathbf{v} \left(\frac{t}{T} \right) \right\|^2 \right\} \\
& \quad \times \sum_{t=1}^T E \left[\left(\int \tilde{l}_{ZT}(z) dF_Z(z) - \int \tilde{l}_{ZT}(z - \tilde{\delta}_{t,T}) dF_Z(z) - \int l_Z(z) dF_Z(z) \right)^2 \right].
\end{aligned}$$

Note that

$$\tilde{l}_{ZT}(z - \tilde{\delta}_{t,T}) = \tilde{l}_{ZT}(z) - \tilde{\delta}_{t,T} \tilde{l}'_{ZT}(z) + O_p(\tilde{\delta}_{t,T}^2).$$

As such,

$$\begin{aligned}
& \int \left(\tilde{l}_{ZT}(z) - l_Z(z) \right) dF_Z(z) - \int \tilde{l}_{ZT}(z - \tilde{\delta}_{t,T}) dF_Z(z) \\
& \approx \tilde{\delta}_{t,T} \int \tilde{l}'_{ZT}(z) dF_Z(z) - \int l_Z(z) dF_Z(z) \\
& = \tilde{\delta}_{t,T} \int \tilde{l}'_{ZT}(z) dF_Z(z).
\end{aligned}$$

Therefore with an appeal to Lemma 8 applied to the model in (15) we have

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^T E \left[\int \left\| \tilde{\mathbf{L}}_{t,T}(z) - \mathbf{L}_{t,T}(z) \right\|^2 dF_Z(z) \right] \\
& \leq T^{-1} \{ O(T^{2-3r}) + O(T) \} \times \sum_{t=1}^T E \left[\tilde{\delta}_{t,T}^2 \left(\int \tilde{l}'_{ZT}(z) dF_Z(z) \right)^2 \right] \\
& = O(T^{1-3r} + 1) \times O(T^{-1+r}) \\
& = O(T^{-2r} + T^{-1+r}) \\
& = o(1).
\end{aligned}$$

B.2.3 Verification of condition (57):

Noting that

$$Z_t = \tilde{Y}_t - \mu \left(\frac{t}{T} \right) = \tilde{Y}_t - \tilde{\mu}_{T,-t} \left(\frac{t}{T} \right) + O_p(T^{-1+r}),$$

we have

$$\left\| E \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\tilde{\mathbf{L}}_{t,T}(Z_t) - \bar{\mathbf{L}}_{t,T}(Z_t) \right) \right] \right\|^2$$

$$\begin{aligned}
&\leq 2 \sum_{t=1}^T E \left[\left\| \tilde{\mathbf{v}}_T \left(\frac{t}{T} \right) - E \left[\tilde{\mathbf{v}}_T \left(\frac{t}{T} \right) \middle| \tilde{Y}_1, \dots, \tilde{Y}_{t-1}, \tilde{Y}_{t+1}, \dots, \tilde{Y}_T \right] \right\|^2 \right] \\
&\quad \times \frac{1}{T} \sum_{t=1}^T E \left[\tilde{l}_{ZT}^2 \left(\tilde{Y}_t - \mu \left(\frac{t}{T} \right) \right) \right] + 2T^2 L_{T3}^{(0)} \times \frac{1}{T} \sum_{t=1}^T E \left[\left\| \tilde{\mathbf{v}}_T \left(\frac{t}{T} \right) \right\|^2 \right] \\
&= o(1) + T \times T^{-1} O \left(T^{-1} a_T^{-4} b_T^{-2} + a_T^{-5} b_T^{-1} T^{-2-2r} \right) O \left(T^{3r} \right) \\
&= O \left(T^{-1+3r} a_T^{-4} b_T^{-2} + a_T^{-5} b_T^{-1} T^{-2+r} \right) \\
&= o(1),
\end{aligned}$$

where $L_{T3}^{(0)}$ is as in Lemma 12 and the conclusion of Lemma 8 as applied to the model in (15) is used in the appropriate location.

B.2.4 Verification of condition (58):

This follows the same approach as was used to show condition (57).

The conditions of Lemma 11 hold. This concludes the proof.

B.3 Proof of Lemma 4

The proof follows the same pattern as was used in the proof of Lemma 3. In particular, the convergence of the quantities corresponding to Δ_{T1} , Δ_{T2} and Δ_{T3} in (33) above is shown. The focus here, as in the proof of Lemma 3, is on showing the convergence of Δ_{T2} , since the convergence of Δ_{T1} and of Δ_{T3} is straightforward.

As such, we show that Δ_{T2} converges by applying Lemma 11 with

$$\begin{aligned}
\tilde{L}_{t,T}(z) &\equiv \frac{1}{\sqrt{T}} \left(\tilde{l}_{ZT}(z) - \int \tilde{l}_{ZT}(z - \tilde{\delta}_{t,T}) dF_Z(z) \right); \\
L_{t,T}(z) &\equiv 0.
\end{aligned}$$

Conditions (55)–(58) will be shown in sequence.

B.3.1 Verification of condition (55):

Recalling the definition of $L_{T1}^{(1)}$ in Lemma 12, we have

$$\begin{aligned}
&\left(E \left[\frac{1}{\sqrt{T}} \left| \sum_{t=1}^T \int \tilde{L}_{t,T}(z) dF_Z(z) \right| \right] \right)^2 \\
&\leq \frac{2}{T} L_{T1}^{(1)} \\
&= O \left(T^{-2} a_T^{-2} \{ a_T^{-4} T^{-1-2r} + T^{-1} a_T^{-4} b_T^{-2} + a_T^{-5} b_T^{-1} T^{-2-2r} \} \right) \\
&= O \left(a_T^{-6} T^{-3-2r} + T^{-3} a_T^{-6} b_T^{-2} + a_T^{-7} b_T^{-1} T^{-4-2r} \right) \\
&= o(1).
\end{aligned}$$

B.3.2 Verification of condition (56):

This is similar to the verification of condition (55).

B.3.3 Verification of condition (57):

As was done above we note that

$$Z_t = \tilde{Y}_t - \mu \left(\frac{t}{T} \right) = \tilde{Y}_t - \tilde{\mu}_{T,-t} \left(\frac{t}{T} \right) + O_p(T^{-1+r}),$$

and get

$$\begin{aligned} & \left(E \left[\frac{1}{\sqrt{T}} \left| \sum_{t=1}^T \left(\tilde{L}_{t,T}(Z_t) - \bar{L}_{t,T}(Z_t) \right) \right| \right] \right)^2 \\ & \leq L_{T3}^{(1)} \\ & = O(T^{-1} a_T^{-2} \{ T^{-1} a_T^{-4} b_T^{-2} + a_T^{-5} b_T^{-1} T^{-2-2r} \}) \\ & = O(T^{-2} a_T^{-6} b_T^{-2} + a_T^{-7} b_T^{-1} T^{-3-2r}) \\ & = o(1), \end{aligned}$$

where $L_{T3}^{(1)}$ is as given in the statement of Lemma 12.

B.3.4 Verification of condition (58):

This is similar to the verification of condition (57).

The conditions of Lemma 11 have been shown to hold. This concludes the proof.

B.4 Proof of Lemma 5

We again use the framework of (33) above and focus on showing the convergence of Δ_{T2} . In order to do this, we apply Lemma 11 with

$$\begin{aligned} \tilde{\mathbf{L}}_{t,T}(z) & \equiv \frac{z}{\sqrt{T}} \bar{\mathbf{v}}_T; \\ \mathbf{L}_{t,T}(z) & \equiv z \bar{\mathbf{v}}, \end{aligned}$$

where

$$\bar{\mathbf{v}} \equiv \left(\int_0^1 |\gamma'_1(s)| ds, \dots, |\gamma'_k(s)| ds \right)^\top.$$

The four conditions of Lemma 11 are shown in sequence.

B.4.1 Verification of condition (55):

We have

$$\begin{aligned} \left\| E \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \int \tilde{\mathbf{L}}_{t,T}(z) dF_Z(z) \right] \right\|^2 &\leq \frac{2}{T} \left\{ E \left[\|\bar{v}_T - \bar{v}\|^2 \right] + \|\bar{v}\|^2 \right\} \int z^2 dF_Z(z) \\ &= o(1), \end{aligned}$$

where use was made of Lemma 8 as applied to the model given in (15).

B.4.2 Verification of condition (56):

This is similar to the verification of condition (55).

B.4.3 Verification of conditions (57) and (58):

Easy. This concludes the proof.

Table 1: Empirical size at a nominal level of 5%

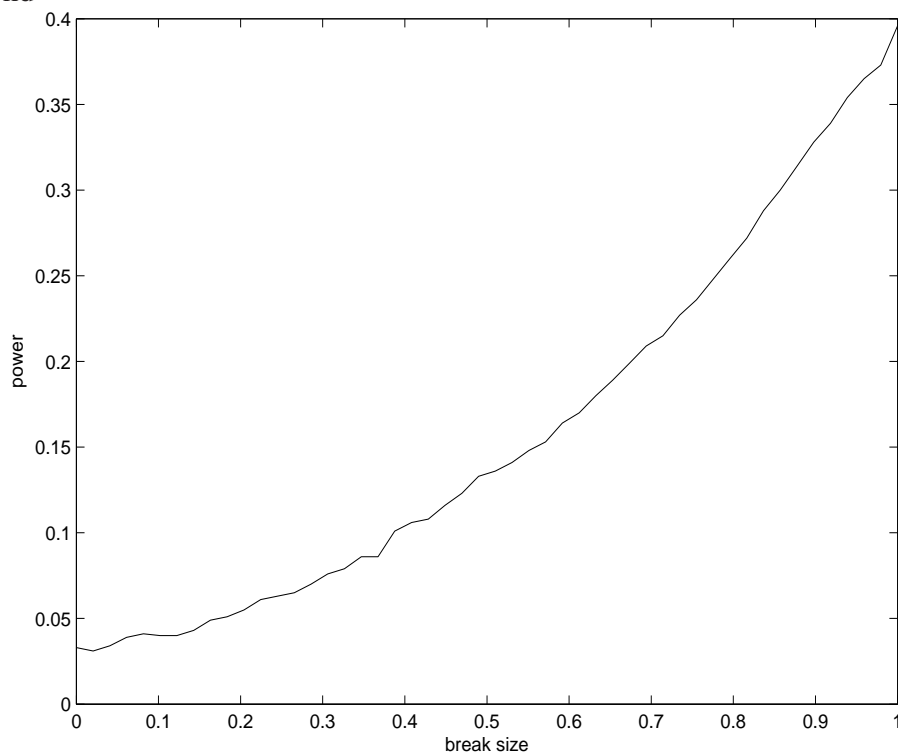
| Bandwidth | | | Sample size | | |
|-----------|-------|-------|-------------|-----------|-----------|
| c_h | c_a | c_b | $T = 100$ | $T = 200$ | $T = 300$ |
| 0.5 | 0.5 | 0.5 | 0.076 | 0.043 | 0.062 |
| 1.0 | 1.0 | 1.0 | 0.176 | 0.085 | 0.022 |
| 1.5 | 1.5 | 1.5 | 0.014 | 0.167 | 0.036 |
| 1.0 | 0.5 | 0.5 | 0.169 | 0.054 | 0.010 |
| 0.5 | 1.0 | 0.5 | 0.099 | 0.094 | 0.068 |
| 0.5 | 0.5 | 1.0 | 0.051 | 0.138 | 0.027 |
| 1.5 | 0.5 | 0.5 | 0.009 | 0.044 | 0.089 |
| 0.5 | 1.5 | 0.5 | 0.029 | 0.061 | 0.180 |
| 0.5 | 0.5 | 1.5 | 0.243 | 0.171 | 0.023 |
| 0.5 | 1.0 | 1.0 | 0.124 | 0.126 | 0.094 |
| 0.5 | 1.0 | 0.5 | 0.125 | 0.171 | 0.076 |
| 1.0 | 1.0 | 0.5 | 0.165 | 0.099 | 0.059 |
| 1.5 | 1.0 | 1.0 | 0.031 | 0.361 | 0.146 |
| 1.0 | 1.5 | 1.0 | 0.090 | 0.039 | 0.075 |
| 1.0 | 1.0 | 1.5 | 0.002 | 0.195 | 0.007 |
| 0.5 | 1.5 | 1.5 | 0.014 | 0.096 | 0.060 |
| 1.5 | 0.5 | 1.5 | 0.100 | 0.249 | 0.319 |
| 1.5 | 1.5 | 0.5 | 0.052 | 0.006 | 0.049 |
| 1.0 | 1.5 | 1.5 | 0.032 | 0.039 | 0.069 |
| 1.5 | 1.0 | 1.5 | 0.108 | 0.056 | 0.222 |
| 1.5 | 1.5 | 1.0 | 0.011 | 0.007 | 0.128 |
| 0.5 | 1.0 | 1.5 | 0.109 | 0.094 | 0.107 |
| 0.5 | 1.5 | 1.0 | 0.105 | 0.073 | 0.033 |
| 1.0 | 0.5 | 1.5 | 0.058 | 0.003 | 0.063 |
| 1.0 | 1.5 | 0.5 | 0 | 0.018 | 0.084 |
| 1.5 | 0.5 | 1.0 | 0.118 | 0 | 0 |
| 1.5 | 1.0 | 0.5 | 0.105 | 0.069 | 0.066 |

Notes to Table 1:

1. Normal $AR(1)$ -error design in (40) with $\gamma\left(\frac{t}{T}\right) \equiv 0$, $\rho = \frac{1}{2}$ and 1000 replications.
2. Bandwidths h_T , a_T and b_T are scaled by constants c_h , c_a and c_b , respectively. In particular,

$$\begin{aligned} h_T &= c_h T^{-\frac{2}{5}}; \\ a_T &= c_a T^{-\frac{2}{5}}; \\ b_T &= c_b T^{-\frac{2}{5}}. \end{aligned}$$

Figure 1: Power performance at a nominal level of 5% against a one-time break in trend



Notes to Figure 1:

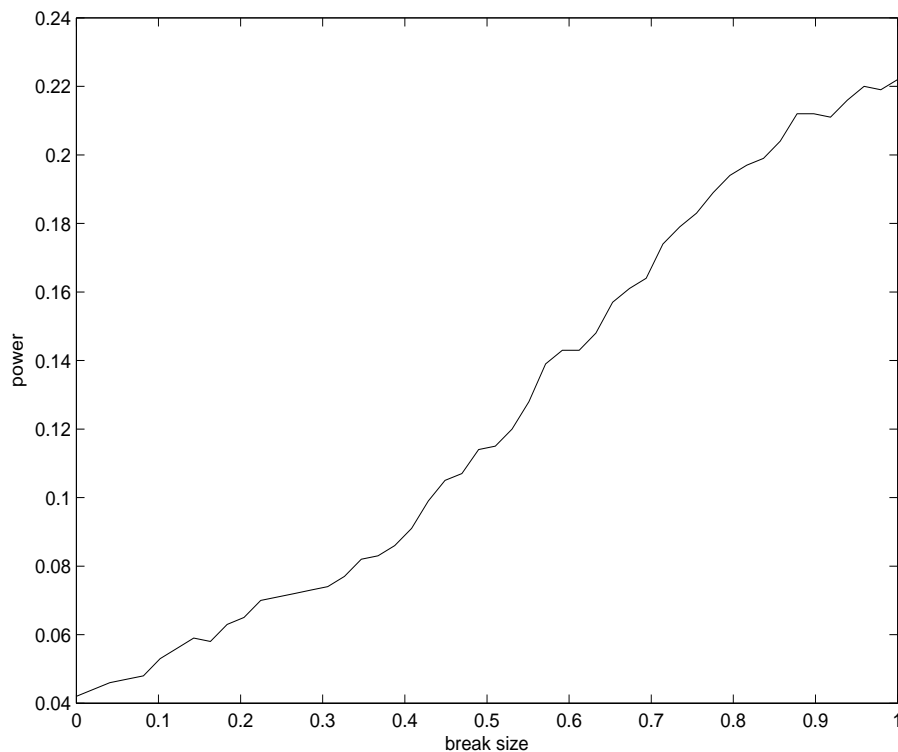
1. Normal $AR(1)$ -error design in (40) with $\rho = \frac{1}{2}$, $T = 200$ and 1000 replications.
2. The trend function is given by

$$\gamma\left(\frac{t}{T}\right) = 1\left(\frac{t}{T} > .5\right) \gamma_0,$$

where the break size γ_0 takes values in a grid of 50 equally-spaced points in the interval $[0, 1]$.

3. The bandwidths used are $h_T = .5T^{-\frac{2}{5}}$, $a_T = b_T = .5T^{-\frac{2}{9}}$, where $T = 200$.

Figure 2: Power performance at a nominal level of 5% against two breaks in trend



Notes to Figure 2:

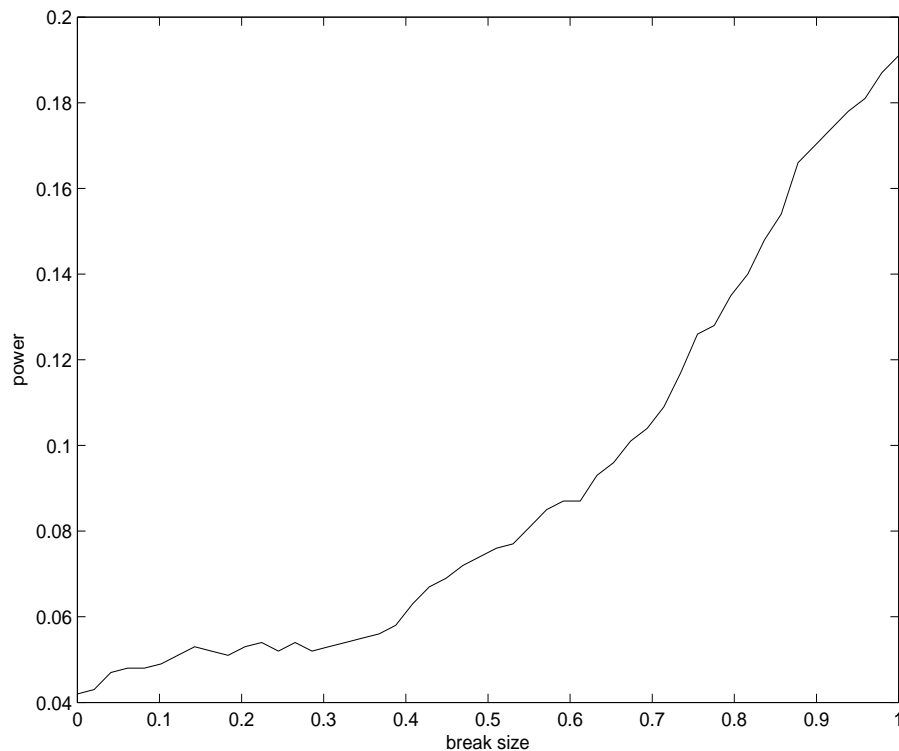
1. Normal $AR(1)$ -error design in (40) with $\rho = \frac{1}{2}$, $T = 200$ and 1000 replications.
2. The trend function is given by

$$\gamma\left(\frac{t}{T}\right) = \left(\frac{1}{2} \times 1\left(.25 < \frac{t}{T} < .75\right) + 1\left(\frac{t}{T} \geq .75\right)\right) \gamma_0,$$

where γ_0 takes values in a grid of 50 equally-spaced points in the interval $[0, 1]$.

3. The bandwidths used are $h_T = .5T^{-\frac{2}{5}}$, $a_T = b_T = .5T^{-\frac{2}{9}}$, where $T = 200$.

Figure 3: Power performance at a nominal level of 5% against a continuous linear change in mean



Notes to Figure 3:

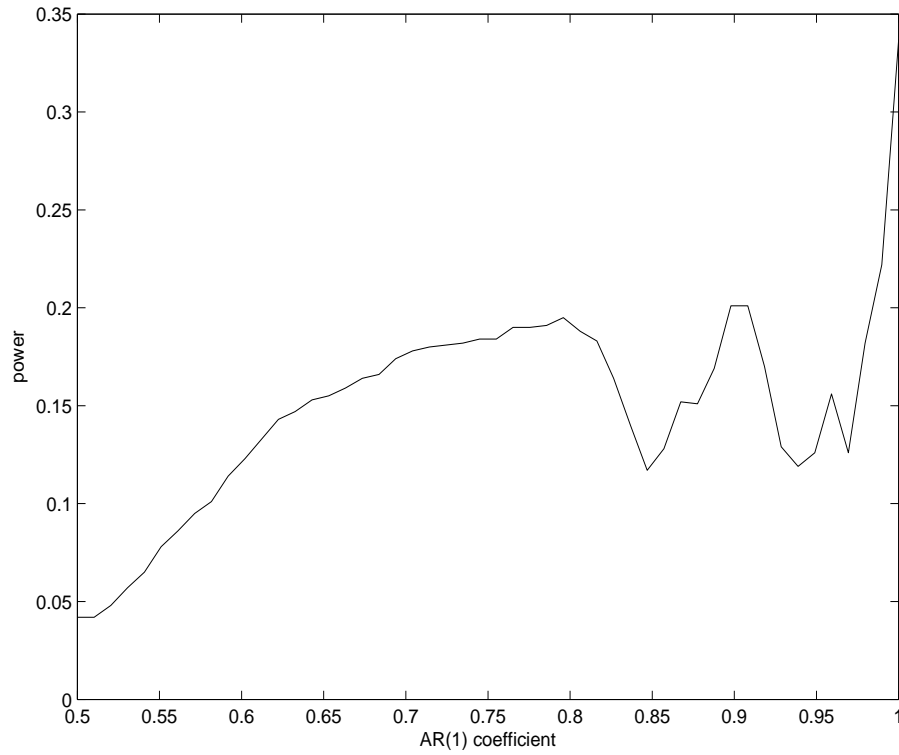
1. Normal $AR(1)$ -error design in (40) with $\rho = \frac{1}{2}$, $T = 200$ and 1000 replications.
2. The trend function is given by

$$\gamma\left(\frac{t}{T}\right) = \left\{ 1\left(.25 < \frac{t}{T} < .75\right) \left[2\left(\frac{t}{T} - \frac{1}{4}\right) \right] + 1\left(\frac{t}{T} \geq .75\right) \right\} \gamma_0,$$

where γ_0 takes values in a grid of 50 equally-spaced points in the interval $[0, 1]$.

3. The bandwidths used are $h_T = .5T^{-\frac{2}{5}}$, $a_T = b_T = .5T^{-\frac{2}{5}}$, where $T = 200$.

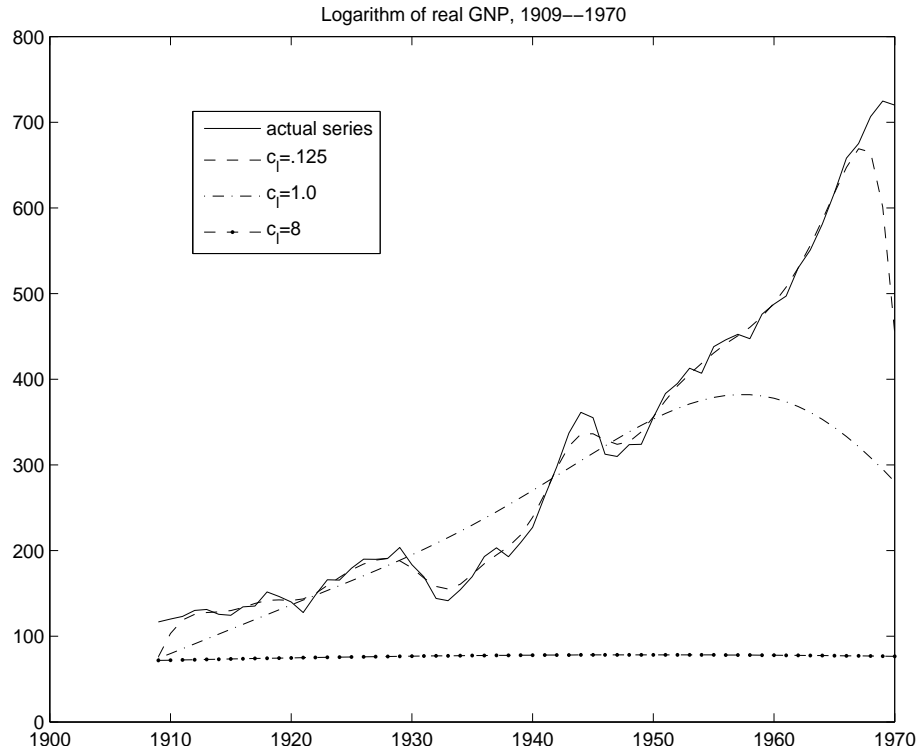
Figure 4: Power performance at a nominal level of 5% against increasing persistence in the data-generating process



Notes to Figure 4:

1. Normal $AR(1)$ -error design in (40) with $\gamma\left(\frac{t}{T}\right) \equiv 0$, $T = 200$ and 1000 replications.
2. The autoregressive coefficient in the error process takes values in a grid of 50 equally-spaced points in the interval $[\cdot 5, 1]$.
3. The bandwidths used are $h_T = \cdot 5T^{-\frac{2}{5}}$, $a_T = b_T = \cdot 5T^{-\frac{2}{5}}$, where $T = 200$.

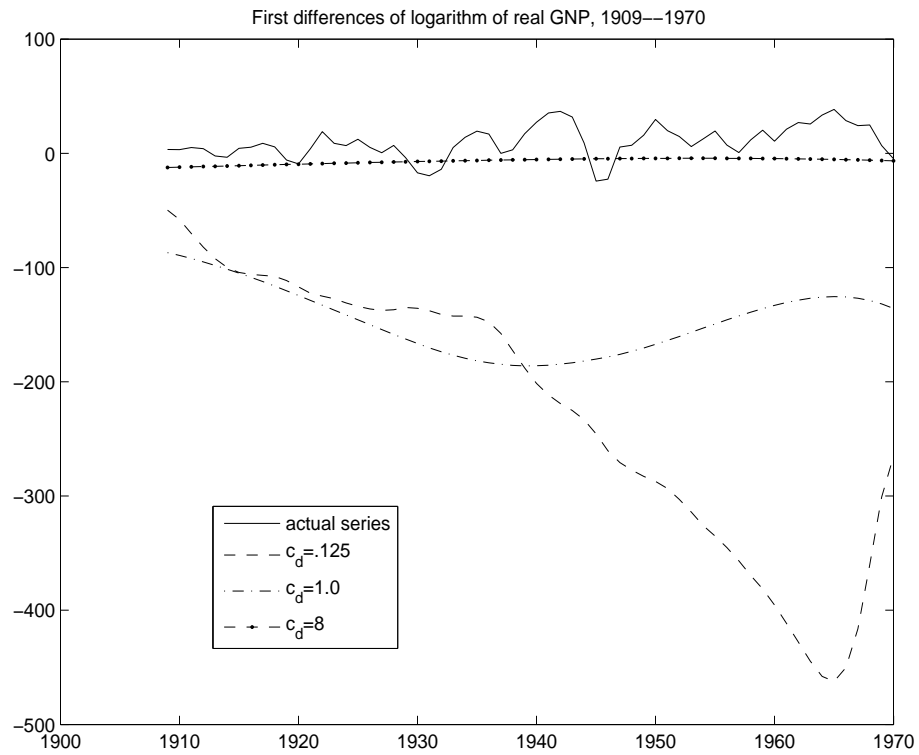
Figure 5: Nonparametric fits of the Nelson and Plosser (1982) real GNP series



Notes to Figure 5:

1. The data used were annual measurements of natural logarithms of real GNP for the United States from 1909 to 1970. The source of the data is indicated in Nelson and Plosser (1982, note 10).
2. The broken lines indicate various implementations of the basic Priestley and Chao (1972) estimator given in (46) applied to the real GNP series with different bandwidth settings. In particular, the bandwidth used is given by $h_l = c_l T^{-\frac{2}{5}}$, where $c_l \in \{.125, 1.0, 8.0\}$. The standard normal kernel was used throughout.

Figure 6: Nonparametric fits of the first differences of the Nelson and Plosser (1982) real GNP series



Notes to Figure 6:

1. The broken lines indicate various implementations of the Priestley-Chao estimator given in (47) applied to the first differences of the log real GNP series used by Nelson and Plosser (1982).
2. The bandwidth used is given by $h_d = c_d T^{-\frac{2}{7}}$, where $c_d \in \{.125, 1.0, 8.0\}$. The standard normal kernel was used throughout.

Table 2: Empirical results for the Nelson and Plosser (1982) real GNP series

| Bandwidth | | Wald statistic |
|-----------|-------|----------------|
| c_l | c_d | |
| .125 | .125 | 1.5572 |
| .125 | 1.0 | 1.5586 |
| .125 | 8.0 | 1.5571 |
| 1.0 | .125 | 3.3145 |
| 1.0 | 1.0 | 3.3142 |
| 1.0 | 8.0 | 3.3145 |
| 8.0 | .125 | 6.2229 |
| 8.0 | 1.0 | 6.2228 |
| 8.0 | 8.0 | 6.2229 |

Notes to Table 2:

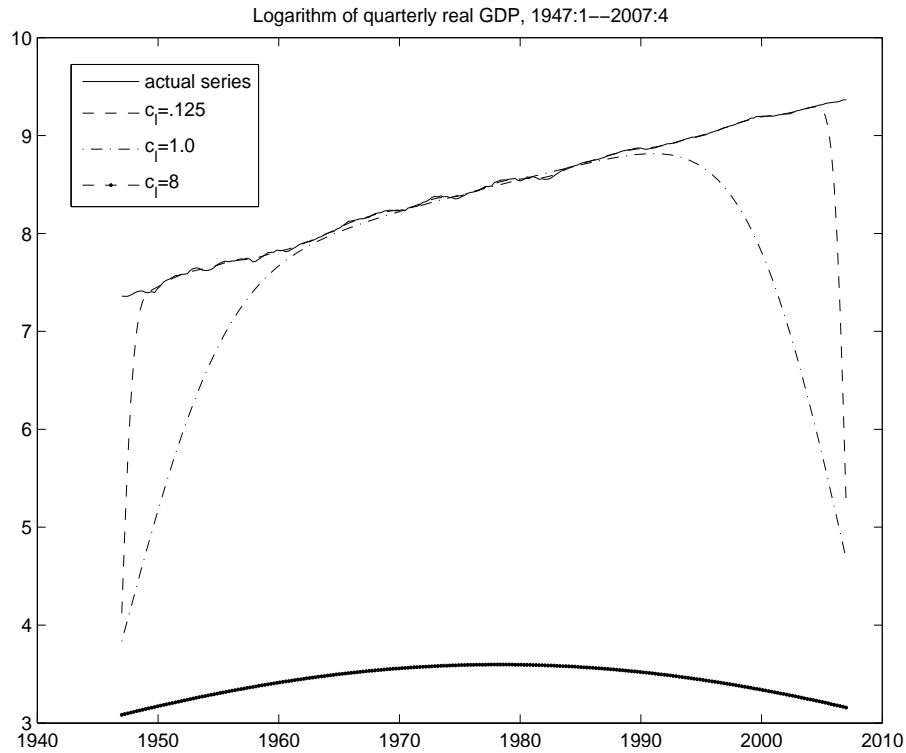
1. The model being validated is that of linear trend-stationarity with a single break in level just after 1929, i.e.,

$$Y_t = \gamma_{01} + d_2 \left(\frac{t}{T} \right) \gamma_{02} + \frac{t}{T} \gamma_{03} + u_t,$$

where $d_2(\cdot)$ is unity for all observations corresponding to 1930 or later and zero for all observations corresponding to years 1909 through 1929. Cf. Perron (1989, §5).

2. The bandwidths used are given by $h_l = c_l T^{-\frac{2}{5}}$, $h_d = c_d T^{-\frac{2}{7}}$ with $c_l, c_d \in \{.125, 1.0, 8.0\}$, and $a_T = b_T = .5T^{-\frac{2}{5}}$.
3. The standard normal kernel was used to construct the Priestley-Chao estimates of the trend parameters and their derivatives, while the Epanechnikov (1969) kernel was used to construct the estimate given in (27).

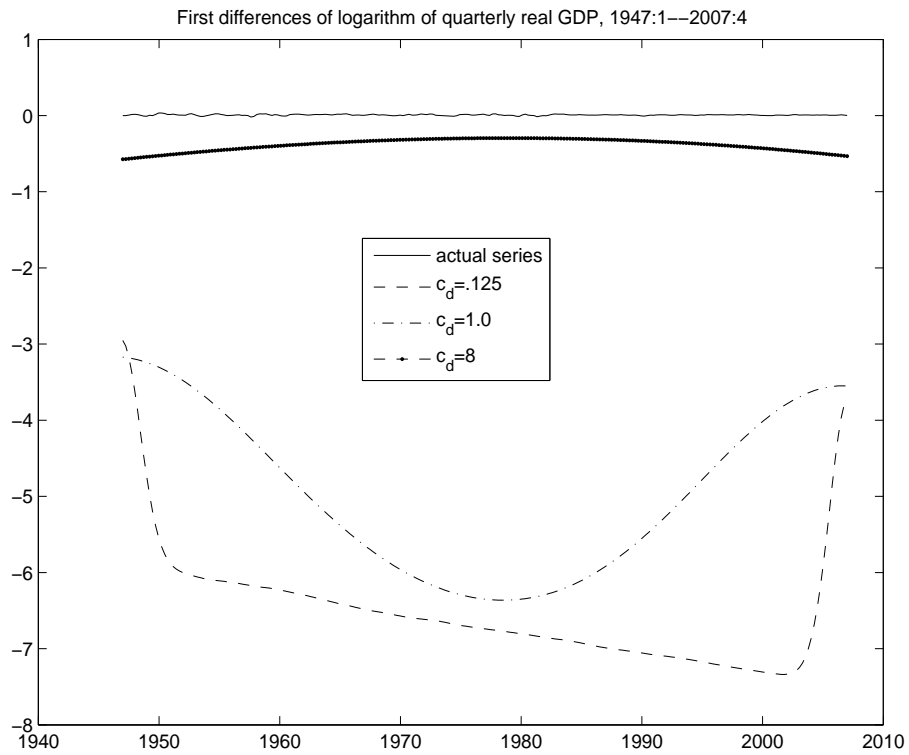
Figure 7: Nonparametric fits of quarterly postwar real GDP



Notes to Figure 7:

1. The data used were quarterly measurements of natural logarithms of GDP in billions of chained 2000 dollars for the United States between 1947 and 2007, inclusive. The original observations were seasonally adjusted at annual rates. The data were obtained from <http://www.bea.gov>.
2. The broken lines indicate various implementations of the basic Priestley and Chao (1972) estimator given in (46) applied to the real GDP series with different bandwidth settings. In particular, the bandwidth used is given by $h_l = c_l T^{-\frac{2}{5}}$, where $c_l \in \{.125, 1.0, 8.0\}$. The standard normal kernel was used throughout.

Figure 8: Nonparametric fits of the first differences of quarterly postwar GDP



Notes to Figure 8:

1. The broken lines indicate various implementations of the Priestley-Chao estimator given in (47) applied to the first differences of the logarithms of quarterly postwar GDP series considered in Figure 7.
2. The bandwidth used is given by $h_d = c_d T^{-\frac{2}{7}}$, where $c_d \in \{.125, 1.0, 8.0\}$. The standard normal kernel was used throughout.

Table 3: Empirical results for quarterly postwar real GDP

| Bandwidth | | Wald statistic |
|-----------|-------|----------------|
| c_l | c_d | |
| .125 | .125 | 2.8356 |
| .125 | 1.0 | 2.7976 |
| .125 | 8.0 | 2.8739 |
| 1.0 | .125 | 8.3581 |
| 1.0 | 1.0 | 8.2383 |
| 1.0 | 8.0 | 8.1927 |
| 8.0 | .125 | 15.5274 |
| 8.0 | 1.0 | 15.5279 |
| 8.0 | 8.0 | 15.5251 |

Notes to Table 3:

1. The model being validated is that of linear trend-stationarity with a single break in slope just after the first quarter of 1973, i.e.,

$$Y_t = \gamma_{01} + \frac{t}{T} \left(\gamma_{02} + d_3 \left(\frac{t}{T} \right) \gamma_{03} \right) + u_t,$$

where $d_3(\cdot)$ is unity for all observations corresponding to 1973:1 or later and zero for all observations corresponding to quarters between 1947:1 and 1973:1, inclusive. Cf. Perron (1989, §5).

2. The bandwidths used are given by $h_l = c_l T^{-\frac{2}{5}}$, $h_d = c_d T^{-\frac{2}{7}}$ with $c_l, c_d \in \{.125, 1.0, 8.0\}$, and $a_T = b_T = .5T^{-\frac{2}{5}}$.
3. The standard normal kernel was used to construct the Priestley-Chao estimates of the trend parameters and their derivatives, while the Epanechnikov (1969) kernel was used to construct the estimate given in (27).

References

- Bai, J. (1996) 'Testing for parameter constancy in linear regressions: An empirical distribution approach.' *Econometrica* 64, 597–622
- Begun, J., W. Hall, W.-M. Huang, and J. Wellner (1983) 'Information and asymptotic efficiency in parametric-nonparametric models.' *Annals of Statistics* 11, 432–452
- Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models* (Baltimore: The Johns Hopkins University Press)
- Blanchard, O. J. (1981) 'What is left of the multiplier-accelerator?' *American Economic Review* 71(2): Papers and Proceedings of the Ninety-Third Annual Meeting of the American Economic Association, 150–54. ed. R. W. Clower, G. W. Harrison and W. St. John
- Blanchard, O. J., and L. H. Summers (1986) 'Hysteresis and the European unemployment problem.' In *NBER Macroeconomics Annual 1986*, ed. S. Fischer (Cambridge, Mass.: MIT Press) pp. 15–77
- Chu, C., and H. White (1992) 'A direct test for changing trend.' *Journal of Business and Economic Statistics* 10, 289–299
- Epanechnikov, V. A. (1969) 'Nonparametric estimates of a multivariate probability density.' *Theory of Probability and Applications* 14, 153–158
- Gallant, A. R., and W. A. Fuller (1973) 'Fitting segmented polynomial regression models whose join points have to be estimated.' *Journal of the American Statistical Association* 68, 144–147
- Goldstein, L., and K. Messer (1992) 'Optimal plug-in estimators for nonparametric functional estimation.' *Annals of Statistics* 20, 1306–1328
- Hájek, J., and Z. Sidák (1967) *Theory of Rank Tests* (New York: Academic Press)
- Haldrup, N., and M. Jansson (2006) 'Improving size and power in unit root testing.' In *Palgrave Handbook of Econometrics*, ed. K. Patterson and T. C. Mills, vol. 1: Econometric Theory (Basingstoke, Hampshire, U.K.: Palgrave Macmillan) pp. 252–277
- Johansen, S. (2006) 'Cointegration: An overview.' In *Palgrave Handbook of Econometrics*, ed. K. Patterson and T. C. Mills, vol. 1: Econometric Theory (Basingstoke, Hampshire, U.K.: Palgrave Macmillan) pp. 540–577
- Juhl, T., and Z. Xiao (2005) 'A nonparametric test for changing trends.' *Journal of Econometrics* 127, 179–199
- King, R., C. Plosser, J. Stock, and M. Watson (1991) 'Stochastic trends and economic fluctuations.' *American Economic Review* 81(4), 819–840
- Klaassen, C. A. J. (1987) 'Consistent estimation of the influence function of locally asymptotically linear estimators.' *Annals of Statistics* 15, 1548–1562

- 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
- Kydland, F. E., and E. C. Prescott (1980) 'A competitive theory of fluctuations and the feasibility and desirability of stabilization policy.' In *Rational Expectations and Economic Policy*, ed. S. Fischer (Chicago: University of Chicago Press)
- Nelson, C. R., and C. I. Plosser (1982) 'Trends and random walks in macroeconomic time series: Some evidence and implications.' *Journal of Monetary Economics* 10, 139–162
- Park, J. Y., and S. B. Hahn (1999) 'Cointegrating regressions with time-varying coefficients.' *Econometric Theory* 15, 664–703
- Perron, P. (1988) 'Trends and random walks in macroeconomic time series: Further evidence from a new approach.' *Journal of Economic Dynamics and Control* 12, 297–332
- (1989) 'The Great Crash, the oil price shock, and the unit root hypothesis.' *Econometrica* 57, 1361–1401
- (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
- Perron, P., and T. Wada (2006) 'Let's take a break: Trends and cycles in U.S. real GDP.' Department of Economics, Boston University
- Pfanzagl, J., and W. Wefelmeyer (1982) *Contributions to a General Asymptotic Statistical Theory* (Berlin: Springer-Verlag)
- Phillips, P. C. B., and B. E. Hansen (1990) 'Statistical inference in instrumental variables regression with $I(1)$ processes.' *Review of Economic Studies* 57, 99–125
- 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
- Rappoport, P., and L. Reichlin (1989) 'Segmented trends and non-stationary time series.' *Economic Journal* 99 (Conference 1989), 168–177
- Ripatti, A., and P. Saikkonen (2001) 'Vector autoregressive processes with nonlinear time trends in cointegrating relations.' *Macroeconomic Dynamics* 5, 577–597

- 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
- Schick, A. (1986) 'On asymptotically efficient estimation in semi-parametric models.' *Annals of Statistics* 14, 1139–1151
- (1987) 'A note on the construction of asymptotically linear estimators.' *Journal of Statistical Planning and Inference* 16, 89–105
- Stulz, R. M., and W. Wasserfallen (1985) 'Macroeconomic time series, business cycles and macroeconomic policies.' In *Carnegie–Rochester Conference Series on Public Policy*, ed. K. Brunner and A. H Meltzer, vol. 22: Understanding Monetary Regimes pp. 9–54
- Taniguchi, M., and Y. Kakizawa (2000) *Asymptotic Theory of Statistical Inference for Time Series* (New York: Springer-Verlag New York)
- van der Vaart, A. (1991) 'On differentiable functionals.' *Annals of Statistics* 19, 178–204
- van der Vaart, A. W. (1998) *Asymptotic Statistics* (Cambridge, U.K.: Cambridge University Press)
- 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
- Wasserfallen, W. (1986) 'Non-stationarities in macro-economic time series.' *Canadian Journal of Economics* 19, 498–510
- Wu, W. B. (2005) 'Nonlinear system theory: Another look at dependence.' *Proceedings of the National Academy of Sciences* 102, 14150–14154
- (2007) 'Strong invariance principles for dependent random variables.' *Annals of Probability* 35, 2294–2320
- Wu, W. B., and X. Shao (2004) 'Limit theorems for iterated random functions.' *Journal of Applied Probability* 41, 425–436
- Wu, W. B., and Z. Zhao (2007) 'Inference of trends in time series.' *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 69, Part 3, 391–410