Econometric Theory, **27**, 2011, 992–1025.
doi:10.1017/S0266466610000617

TESTING FOR A SHIFT IN TREND AT AN UNKNOWN DATE: A FIXED-*B* ANALYSIS OF HETEROSKEDASTICITY AUTOCORRELATION ROBUST OLS-BASED TESTS

ÖZGEN SAYGINSOY
Barclays

TIMOTHY J. VOGELSANG
Michigan State University

This paper analyzes tests for a shift in the trend function of a time series at an unknown date based on ordinary least squares (OLS) estimates of the trend function. Inference about the trend parameters depends on the serial correlation structure of the data through the long-run variance (zero frequency spectral density) of the errors. Asymptotically pivotal tests can be obtained by the use of serial correlation robust standard errors that require an estimate of the long-run variance. The focus is on the class of nonparametric kernel estimators of the long-run variance. Tests based on these estimators present two problems for practitioners. The first is the choice of kernel and bandwidth. The second is the well-known overrejection problem caused by strong serial correlation (or a possible unit root) in the errors.We provide solutions to both problems by using the fixed-*b* asymptotic framework of Kiefer and Vogelsang (2005, *Econometric Theory*, 21, 1130–1164) in conjunction with the scaling factor approach of Vogelsang (1998, *Econometrica* 65, 123–148). Our results provide practitioners with a family of OLS-based trend function structural change tests that are size robust to the presence of strong serial correlation or a unit root. Specific recommendations are provided for the tuning parameters (kernel and bandwidth) in a way that maximizes asymptotic integrated power.

1. INTRODUCTION

Many macroeconomic time series can be characterized by shocks fluctuating around a broken trend function. Failure to account for changes in trend function

We thank an editor and two referees for constructive comments and suggestions. We are also grateful for comments and suggestions from participants in the 2nd Annual Granger Time Series Conference (December 2008, Nottingham) at the CIREQ Time Series Conference (December 2005, Montreal) and in seminars at SMU, Cornell University, and Michigan State. Vogelsang acknowledges financial support from the NSF through grant SES-0525707. Address correspondence to Tim Vogelsang, Department of Economics, Michigan State University, 110 Marshall-Adams Hall, East Lansing, MI 48824-1038, USA; e-mail: tjv@msu.edu.

 © Cambridge University Press 2011

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

parameters leads to inconsistent parameter estimates and inaccurate forecasts. Perron (1989, 1990) has shown that the presence of unmodeled breaks in the trend function of a time series can bias unit root tests toward the nonrejection of the unit root hypothesis. In some practical applications, the existence of structural change in the trend function can be of interest itself. For example, the empirical debate that convergence in incomes per capita among U.S. regions has leveled off in the mid-1970s can be explored by modeling the trend function in time series of incomes per capita in each of the U.S. regions as having a slope shift in the mid-1970s.[1]

Related papers on the subject of testing for structural change in the trend function of a univariate time series in the presence of serial correlation include Andrews (1993), Andrews and Ploberger (1994), Bai, Lumsdaine, and Stock (1998), Chen and Tiao (1990), Chu and White (1992), Kramer, Ploberger, and Alt (1988), Vogelsang (1997, 1998a), Perron and Yabu (2009) (PY), and Harvey, Leybourne, and Taylor (2009) (HLT), among others. All of these testing procedures, except Vogelsang (1998a), require a complete specification of the dynamics and/or estimation of the serial correlation parameters.

This paper analyzes tests for a shift in the trend function of a time series that is based on ordinary least squares (OLS) estimates of the trend function. When the errors are stationary, OLS is asymptotically equivalent to generalized least squares (GLS) and is asymptotically efficient (see Grenander and Rosenblatt, 1957). Inference about the trend parameters depends on the serial correlation structure of the data through the long-run variance (zero frequency spectral density) of the errors. Asymptotically pivotal tests based on the OLS estimates can be obtained by the use of serial correlation robust standard errors that require an estimate of the long-run variance. The focus in this paper is on the class of nonparametric kernel estimators of the long-run variance that are more commonly labeled heteroskedasticity autocorrelation (HAC) covariance matrix estimators. Tests based on HAC estimators present two problems for practitioners. The first is the choice of kernel and bandwidth. The second is the well-known overrejection problem caused by strong serial correlation (or a possible unit root) in the errors. We provide solutions to both problems by using the fixed-b asymptotic framework of Kiefer and Vogelsang (2005), as extended to the case of trend function models by Bunzel and Vogelsang (2005), in conjunction with the scaling factor approach of Vogelsang (1998b). The fixed-b asymptotic framework provides an approximation of the HAC estimator that depends on the kernel and bandwidth. Local asymptotic power can then be computed as a function of the kernel and bandwidth, allowing the choice of kernel and bandwidth to be based on the power of the tests. Within a specific group of popular kernels, power optimal kernels and bandwidths can be determined using numerical methods. Power optimal bandwidths are shown to depend on the strength of the serial correlation in the data, leading to a data dependent bandwidth rule. The overrejection problem is addressed using the scaling factor approach, which requires nondegenerate asymptotic null distributions when the errors are both stationary and integrated. The

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

fixed-b approximation delivers the required nondegeneracy (this is in contrast with the standard asymptotic approximation in the integrated error case). Our results provide practitioners with a family of OLS-based trend function structural change tests that are size robust to the presence of strong serial correlation or a unit root.

The remainder of the paper is organized as follows: Section 2 describes the general framework, the assumptions, and the statistics. Four models of structural change in the trend function are analyzed: a shift in level of a nontrending series, a shift in the intercept of a trending series, a shift in the slope of a trending series, and a simultaneous shift in level and slope of a trending series. We consider the case of a single structural change with the break date modeled as unknown. We consider HAC robust Wald statistics with the unknown break date handled using the supremum Wald statistic (Andrews, 1993) and the mean Wald statistic (Andrews and Ploberger, 1994). The scaling factor approach is then described and implemented. Section 3 establishes the asymptotic distributions of test statistics under the fixed-b asymptotic framework and describes methods used in computing asymptotic critical values. In Section 4, local asymptotic power results are provided. Using these results, Sayginsoy (2004) found that the Daniell kernel maximized power within a group of widely used kernels. Focusing on the Daniell kernel, we show that power varies across bandwidth choices and depends on the strength of the serial correlation in the data as indexed by a local-to-unity parameter. For a given local-to-unity parameter, no single bandwidth delivers a test with uniformly highest power because power curves intersect. If integrated power (average power) is used as the criterion, bandwidths are found that maximize integrated power. These power optimal bandwidths depend on the local-to-unity parameter, and this leads to a data dependent bandwidth rule. Section 5 reports the results of a Monte Carlo study that assesses the finite-sample performance of the tests using a feasible version of the data dependent bandwidth rule. The simulations show that the fixed-b asymptotic approximation in conjunction with the scaling factor approach substantially reduces the overrejection problem. Finite-sample power of the tests closely follows the predictions of the asymptotic analysis. Section 6 concludes. Supplemental results and proofs are collected in the Appendix.

2. THE ECONOMETRIC SETUP

2.1. The Models and Assumptions

Consider the following four models for the time series process $\{y_t\}_{t=1}^{T}$:

$$y_t = \mu + \varphi DU_t + u_t, \tag{1}$$

$$y_t = \mu + \phi t + \varphi DU_t + u_t, \tag{2}$$

$$y_t = \mu + \phi t + \gamma DT_t + u_t, \tag{3}$$

$$y_t = \mu + \phi t + \varphi DU_t + \gamma DT_t + u_t, \tag{4}$$

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

where

$$DU_t = \begin{cases} 1, & \text{if } t > T_b \\ 0, & \text{if } t \le T_b \end{cases}, \quad DT_t = \begin{cases} t - T_b, & \text{if } t > T_b \\ 0, & \text{if } t \le T_b \end{cases},$$

and T_b denotes the time of a break, should it be present in the time series $\{y_t\}$. Model (1) allows for a shift in the mean of a time series, whereas model (2) allows for a shift in the intercept of a trending time series. Model (3) allows only for a shift in the slope of a trending time series and thus requires that segments of the trend before and after the break date be joined at the break date. Finally, model (4) allows for a shift in both the intercept and the slope of a trending time series, and therefore the segments of the trend before and after the break date need not be joined.

The results presented in the paper utilize the following assumptions on the error process $\{u_t\}$.

Assumption 1.

$$u_t = \alpha u_{t-1} + \varepsilon_t, \qquad t = 2, 3, \ldots, T,$$

$$u_1 = \varepsilon_1,$$

$$\varepsilon_t = d(L) e_t, \qquad d(L) = \sum_{i=0}^{\infty} d_i L^i, \qquad \sum_{i=0}^{\infty} i\,|d_i| < \infty, \qquad d(1)^2 > 0,$$

where $\{e_t\}$ is a martingale difference sequence with $\sup_t \mathrm{E}\left(e_t^4\right) < \infty$, $\mathrm{E}(e_t | e_{t-1}, e_{t-2}, \ldots) = 0$, $\mathrm{E}\left(e_t^2 \middle| e_{t-1}, e_{t-2}, \ldots\right) = 1$, and L is the lag operator.

The error process $\{u_t\}$ is $I(0)$ when $|\alpha| < 1$. Alternatively, $\{u_t\}$ can be modeled as a nearly $I(1)$ process by defining $\alpha = \left(1 - \overline{\alpha}/T\right)$, where $\overline{\alpha} = 0$ corresponds to the pure $I(1)$ case.[2] The restriction $d(1)^2 > 0$ bounds the spectral density of $\{\varepsilon_t\}$ at frequency zero above zero and serves to eliminate nondegenerate cases. Assumption 1 ensures that the following well-known functional central limit theorem results hold (see, e.g., Phillips, 1987; Phillips and Solo, 1992):

$$T^{-1/2} \sum_{t=1}^{[rT]} u_t \Rightarrow \sigma W(r), \quad \text{if } \{u_t\} \text{ is I(0)}, \tag{5}$$

$$T^{-1/2} u_{[rT]} \Rightarrow d(1) V_{\overline{\alpha}}(r), \quad \text{if } \{u_t\} \text{ is I(1)}, \tag{6}$$

where $[rT]$ denotes the integer part of rT, $r \in [0, 1]$, $\Rightarrow$ denotes weak convergence, $W(r)$ is the standard Wiener process, $\sigma^2 = d(1)^2 (1-\alpha)^{-2}$, and $V_{\overline{\alpha}}(r) = \int_0^r \exp(-\overline{\alpha}(r-s))\, dW(s)$.

It is useful to represent the models in (1)–(4) generically by

$$y_t = \mathbf{f}(t, \lambda)' \boldsymbol{\beta} + u_t, \tag{7}$$

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

where $\mathbf{f}(t,\lambda)$ is a vector of functions of the trends, $\lambda = T_b/T$, $\boldsymbol{\beta}$ is the vector of regressor parameters, and $\{u_t\}$ is the error process with properties as described earlier. For example, if the model under consideration is (4), then $\mathbf{f}(t,\lambda) = [1 \;\; t \;\; DU_t \;\; DT_t]'$ and $\boldsymbol{\beta} = [\mu \;\; \phi \;\; \varphi \;\; \gamma]'$.

It will be useful to further expand (7). Let $\mathbf{f}(t,\lambda) = \left[\mathbf{f}_1(t)', \mathbf{f}_2(t,\lambda)'\right]'$ and $\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\right]'$, where $\mathbf{f}_1(t)$ is the $(k_1 \times 1)$ vector of regressors that does not depend on the break date and $\mathbf{f}_2(t,\lambda)$ is the $(k_2 \times 1)$ vector of regressors that depends on the break date. The terms $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are $(k_1 \times 1)$ and $(k_2 \times 1)$ vectors of regressor parameters corresponding to $\mathbf{f}_1(t)$ and $\mathbf{f}_2(t,\lambda)$, respectively. As an alternative to (7), the models in (1)–(4) can be written as

$$y_t = \mathbf{f}_1(t)'\boldsymbol{\beta}_1 + \mathbf{f}_2(t,\lambda)'\boldsymbol{\beta}_2 + u_t. \tag{8}$$

For example, if the model under consideration is (4), then $\mathbf{f}_1(t) = [1 \; t]'$ and $\mathbf{f}_2(t,\lambda) = [DU_t \;\; DT_t]'$, whereas $\boldsymbol{\beta}_1 = [\mu \;\; \phi]'$ and $\boldsymbol{\beta}_2 = [\varphi \;\; \gamma]'$.

Let T_b^0 denote the true break date. Because in practice the true break date is rarely known, we treat it as unknown. To obtain nondegenerate asymptotic results (see Andrews, 1993), possible break dates, T_b, are restricted to the set $\Lambda^* = \left\{T_b^*, T_b^* + 1, \ldots, T - T_b^*\right\} \subseteq \{2, 3, \ldots, T-2\}$. The models will be estimated using all break dates $T_b \in \Lambda^*$. Note that under the alternative of a break, the model is misspecified except when $T_b = T_b^0$. This feature will complicate the local asymptotic power analysis. Throughout the paper the standard assumptions are made that $\lambda_0 = T_b^0/T$, $\lambda = T_b/T$, and $\lambda^* = T_b^*/T$ remain fixed constants as the sample size increases.[3]

The following additional assumption holds for all four models and is sufficient to establish the main theoretical results of the paper.

Assumption 2.

(a) There exist $(k_1 \times k_1)$ and $(k_2 \times k_2)$ normalization matrices $\boldsymbol{\tau}_{1T}$ and $\boldsymbol{\tau}_{2T}$, respectively, and $(k_1 \times 1)$ and $(k_2 \times 1)$ vectors of functions $\mathbf{F}_1(r)$ and $\mathbf{F}_2(r,l)$ on $[0,1]$, respectively, such that

$$\boldsymbol{\tau}_T \mathbf{f}(t,l) = \begin{bmatrix} \boldsymbol{\tau}_{1T} & \mathbf{0}_{k_1 \times k_2} \\ \mathbf{0}_{k_2 \times k_1} & \boldsymbol{\tau}_{2T} \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(t) \\ \mathbf{f}_2(t,l) \end{bmatrix} = \begin{bmatrix} \mathbf{F}_1(t/T) + o(1) \\ \mathbf{F}_2(t/T, l) + o(1) \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \mathbf{F}_1(r) \\ \mathbf{F}_2(r,l) \end{bmatrix} = \mathbf{F}(r,l),$$

where l generically denotes λ_0 or λ and $\mathbf{0}_{i \times j}$ represents an $(i \times j)$ matrix of zeros;

(b) $\int_0^1 \mathbf{F}(r,l)\,dr < \infty$, and $\det\left[\int_0^1 \mathbf{F}(r,l)\mathbf{F}(r,l)'\,dr\right] > 0$.

2.2. The Statistics

The null hypotheses of interest are $\varphi = 0$ in models (1) and (2), $\gamma = 0$ in model (3), and $\varphi = \gamma = 0$ in model (4). Using the generic representation in (8), the

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

null hypothesis can be written in the form $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}_{q\times 1}$, where $q = 1, 2$ is the number of restrictions as dictated by the given hypothesis of interest.

By the well-known Frisch and Waugh (1933) theorem, the OLS estimator of $\boldsymbol{\beta}_2$ in (8) is given by

$$\widehat{\boldsymbol{\beta}}_2 = \left[\sum_{t=1}^{T} \widetilde{\mathbf{f}}_2(t,\lambda_0)\widetilde{\mathbf{f}}_2(t,\lambda_0)'\right]^{-1} \sum_{t=1}^{T} \widetilde{\mathbf{f}}_2(t,\lambda_0)\,\widetilde{y}_t, \tag{9}$$

where $\left\{\widetilde{\mathbf{f}}_2(t,\lambda_0)\right\}$ and $\left\{\widetilde{y}_t\right\}$ are the residuals from the OLS regressions of $\{\mathbf{f}_2(t, \lambda_0)\}$ and $\{y_t\}$ on $\{\mathbf{f}_1(t)\}$, respectively.

If the true break date T_b^0 is known, the null hypothesis of no break can be tested using the standard HAC robust Wald statistic given by

$$W\left(T_b^0\right) = \widehat{\boldsymbol{\beta}}_2' \left[\sum_{t=1}^{T} \widetilde{\mathbf{f}}_2(t,\lambda_0)\widetilde{\mathbf{f}}_2(t,\lambda_0)'\right] \widehat{\boldsymbol{\beta}}_2 \Big/ \widehat{\sigma}^2, \tag{10}$$

where $\widehat{\sigma}^2$ is a nonparametric estimator of the long-run variance defined as

$$\widehat{\sigma}^2 = \widehat{\gamma}_0 + 2\sum_{j=1}^{T-1} k\left(j/M\right)\widehat{\gamma}_j, \tag{11}$$

where M is the truncation lag or the bandwidth parameter, $\widehat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T} \widehat{u}_t\widehat{u}_{t-j}$ are the sample autocovariances, $\left\{\widehat{u}_t\right\}$ are the OLS residuals from the estimation of (8), and $k(x)$ is a kernel function that is continuous at $x = 0$ and satisfies $k(x) = k(-x)$, $k(0) = 1$, $|k(x)| \leq 1$, and $\int_0^1 k(x)^2\,dx < \infty$. In the case of $I(0)$ errors, if $\widehat{\sigma}^2$ is a consistent estimator of σ^2, then the asymptotic null distribution of $W\left(T_b^0\right)$ is chi-square with q degrees of freedom. This standard approximation is convenient, but it does not reflect the impact of the sampling bias and variability of $\widehat{\sigma}^2$ on the test statistic. This leads, in part, to overrejection problems.

When T_b^0 is unknown, the testing framework falls in a class of tests where a nuisance parameter (in this case, λ_0) is present only under the alternative hypothesis $\left(H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}_{q\times 1}\right)$. In this environment, Andrews and Ploberger (1994) derived a class of optimal tests under regularity conditions that require stationarity of the errors and nontrending regressors. They showed that a composite statistic in that optimal class can be obtained by taking the average of a test statistic computed over a range of possible break dates. Following this approach, define the *Mean-Wald* statistic as

$$MeanW = T^{-1}\sum_{T_b\in\Lambda^*} W(T_b). \tag{12}$$

Note that the optimality result for $MeanW$ only applies to model (1) and does not apply to models (2)–(4) because of the trending regressors. An alternative to the

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

MeanW statistic that has a long history in econometrics is the the *Supremum-Wald* statistic, defined as

$$SupW = \sup_{T_b \in \Lambda^*} W(T_b). \tag{13}$$

Although the *SupW* statistic does not belong to the optimal class of tests for model (1), it can be more powerful than the *MeanW* test in models (2)–(4). A convenient characteristic of the *SupW* statistic is that it can be used to obtain an estimate of the time of the trend break (for details, see Andrews, 1993; Bai, 1994; Perron and Zhu, 2005).

2.3. The Scaled Statistics

It is well known that the accuracy of asymptotic approximations based on the assumption that errors are stationary becomes inadequate when the errors are highly persistent or have a unit root. In the case of *MeanW* and *SupW* the null distributions become skewed to the right as persistence in the errors increases. This leads to overrejections when critical values based on stationary asymptotic results are used. Vogelsang (1998b) proposed a scaling factor approach that scales down a statistic when the errors have a unit root but has no effect asymptotically when the errors are stationary. This is achieved by multiplicatively scaling a statistic by the exponential of a unit root statistic that converges to zero when the errors are stationary. We consider two unit root statistics with this property to construct the scaling factors for *MeanW* and *SupW*. These unit root tests also have the desirable property that they do not require kernel, bandwidth, or lag length choices.[4]

The first unit root statistic, denoted by $J(T_b)$, $T_b \in \Lambda^*$, was proposed by Park (1990) and Park and Choi (1988). The statistic $J(T_b)$ is the standard OLS Wald statistic divided by T for testing the joint hypothesis $\pi_{k_1} = \pi_{k_1+1} = \cdots = \pi_9 = 0$ in the regression model

$$y_t = \mathbf{f}_1(t)'\boldsymbol{\beta}_1 + \mathbf{f}_2(t,\lambda)'\boldsymbol{\beta}_2 + \sum_{i=k_1}^{9} \pi_i t^i + u_t. \tag{14}$$

In particular,

$$J(T_b) = \frac{RSS_R - RSS_U}{RSS_U}, \tag{15}$$

where RSS_U is the residual sum of squares from the estimation of the "unrestricted" regression in (14) and RSS_R is the residual sum of squares from the estimation of the "restricted" regression (8).

The second unit root statistic, denoted by $BG(T_b)$, $T_b \in \Lambda^*$, is the "variance-ratio statistic" of Breitung (2002). The variance-ratio statistic is a Lagrange multiplier–type statistic similar to the statistics proposed by Tanaka (1990),

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

Kwiatkowski, Phillips, Schmidt, and Shin (1992) (KPSS), and Shin and Schmidt (1992) and is defined as

$$BG(T_b) = \frac{T^{-2}\sum_{t=1}^{T}\widehat{S}_t^2}{RSS_R},$$

where $\widehat{S}_t = \sum_{j=1}^{t}\widehat{u}_t$ are the partial sums of the OLS residuals from the estimation of (8).

Both unit root tests are left tail tests and reject for small values. Because the break date is treated as unknown, we consider the statistics

$$J^* = \inf_{T_b\in\Lambda^*} J(T_b), \qquad BG^* = \inf_{T_b\in\Lambda^*} BG(T_b),$$

which are the smallest statistics across all break dates.[5]

Let UR generically denote either J^* or BG^*. Let c_{mean}^{UR} and $c_{\sup}^{UR}$ denote constants. The scaling factors are defined as $\exp\left(-c_{mean}^{UR}\cdot UR\right)$ for $MeanW$ and $\exp\left(-c_{sup}^{UR}\cdot UR\right)$ for $SupW$ statistics. The scaled statistics are then given by

$$MeanW(c_{mean}^{UR}) = MeanW\cdot\exp\left(-c_{mean}^{UR}\cdot UR\right), \tag{16}$$

$$SupW(c_{sup}^{UR}) = SupW\cdot\exp\left(-c_{sup}^{UR}\cdot UR\right). \tag{17}$$

Obviously, setting $c_{mean}^{UR} = 0$ and $c_{sup}^{UR} = 0$ obtains the unscaled versions of the statistics. When the errors are $I(0)$, both J^* and BG^* converge to zero because $J(T_b)$ and $BG(T_b)$ converge to zero for all $T_b \in \Lambda^*$; thus $\exp\left(-c_{mean}^{UR}\cdot UR\right)$ and $\exp\left(-c_{sup}^{UR}\cdot UR\right)$ both converge to one and have no effect on the asymptotic distributions of $MeanW$ and $SupW$. However, when the errors are $I(1)$, J^* and BG^* have nondegenerate limiting distributions under the null hypothesis of no trend break that depend on $\overline{\alpha}$ but are otherwise free of nuisance parameters. Therefore, given a significance level and the local unit root parameter $\overline{\alpha}$, the constants c_{mean}^{UR} and c_{sup}^{UR} can be chosen so that $I(0)$ and $I(1)$ critical values for a given statistic are the same. In other words, given a significance level and $\overline{\alpha}$, the correct choice of constants, c_{mean}^{UR} and c_{sup}^{UR}, renders $MeanW(c_{mean}^{UR})$ and $SupW(c_{sup}^{UR})$ statistics asymptotically size correct whether errors are stationary or integrated with parameter $\overline{\alpha}$.

3. ASYMPTOTIC THEORY

In this section, asymptotic null distribution theory of $MeanW(c_{mean}^{UR})$ and $SupW(c_{\sup}^{UR})$ are established under the fixed-b asymptotic framework. The bandwidth, M, of the long-run error variance estimator in (11) is modeled as a fixed proportion of the sample size by letting $M = bT$, where $b \in (0, 1]$. Some definitions are required before establishing the asymptotic results.

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

DEFINITION 1. *A kernel is labeled Type 1 if $k(x)$ is twice continuously differentiable everywhere and as a Type 2 kernel if $k(x)$ is continuous, twice continuously differentiable everywhere except at $|x| = 1$ and $k(x) = 0$ for $|x| \geq 1$.*

Note that the Bartlett kernel does not classify as a Type 1 or Type 2 kernel and is considered separately. The next definition is used to simplify notation in the asymptotic results that follow.

DEFINITION 2. *Let $k^*(x) = k(x/b)$. Let $k_-^{*\prime}(x)$ denote the first derivative of $k^*(x)$ from the left and let $k^{*\prime\prime}(x)$ denote the second derivative of $k^*(x)$. Define*

$$\widetilde{\boldsymbol{F}}_2(r,\lambda) = \boldsymbol{F}_2(r,\lambda) - \left[\int_0^1 \boldsymbol{F}_2(r,\lambda)\boldsymbol{F}_1(r)'\,dr\right]\left[\int_0^1 \boldsymbol{F}_1(r)\boldsymbol{F}_1(r)'\,dr\right]^{-1}\boldsymbol{F}_1(r),$$

$$\boldsymbol{P}_0(\lambda) = \left[\int_0^1 \widetilde{\boldsymbol{F}}_2(r,\lambda)\widetilde{\boldsymbol{F}}_2(r,\lambda)'\,dr\right]^{-1}\left[\int_0^1 \widetilde{\boldsymbol{F}}_2(r,\lambda)\,dW(r)\right],$$

$$\boldsymbol{P}_1(\lambda) = \left[\int_0^1 \widetilde{\boldsymbol{F}}_2(r,\lambda)\widetilde{\boldsymbol{F}}_2(r,\lambda)'\,dr\right]^{-1}\left[\int_0^1 \widetilde{\boldsymbol{F}}_2(r,\lambda)\,V_{\overline{\alpha}}(r)\,dr\right],$$

$$Q_0(r,\lambda) = W(r) - \int_0^r \widetilde{\boldsymbol{F}}_2(s,\lambda)'\boldsymbol{P}_0(\lambda)\,ds, \quad Q_1(r,\lambda) = \int_0^r V_{\overline{\alpha}}(s)\,ds - \int_0^r \widetilde{\boldsymbol{F}}_2(s,\lambda)'\boldsymbol{P}_1(\lambda)\,ds,$$

$$\Phi_i(b,k,\lambda) = \begin{cases} \int_0^1\int_0^1 -k^{*\prime\prime}(r-s)\,Q_i(r,\lambda)\,Q_i(s,\lambda)\,dr\,ds, & \text{if } k(\cdot) \text{ is Type 1} \\ \int\int_{|r-s|<b} -k^{*\prime\prime}(r-s)\,Q_i(r,\lambda)\,Q_i(s,\lambda)\,dr\,ds & \\ \quad + 2k_-^{*\prime}(b)\int_0^{1-b} Q_i(r+b,\lambda)\,Q_i(r,\lambda)\,dr, & \text{if } k(\cdot) \text{ is Type 2} \\ \frac{2}{b}\left\{\int_0^1 Q_i(r,\lambda)^2\,dr\right. & \\ \quad \left. -\int_0^{1-b} Q_i(r+b,\lambda)\,Q_i(r,\lambda)\,dr\right\}, & \text{if } k(\cdot) \text{ is Bartlett} \end{cases}.$$

The asymptotic distribution of the OLS estimator of $\boldsymbol{\beta}_2$ in (9) has been well established. When the true break date T_b^0 is known, as $T \to \infty$, if $\{u_t\}$ is $I(0)$,

$$T^{1/2}\boldsymbol{\tau}_{2T}^{-1}\left(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right) \Rightarrow \sigma\mathbf{P}_0(\lambda_0), \tag{18}$$

and if $\{u_t\}$ is $I(1)$,

$$T^{-1/2}\boldsymbol{\tau}_{2T}^{-1}\left(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right) \Rightarrow d(1)\mathbf{P}_1(\lambda_0). \tag{19}$$

If the true break date T_b^0 is not known and $T_b \in \Lambda^*$ is used as the break date to estimate the models, then the asymptotic results in (18) and (19) still hold under

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

the null hypothesis of no break in trend, $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}_{q\times 1}$, when λ_0 is replaced with λ.

In the case of $I(1)$ errors, the asymptotic distributions of $MeanW(c^{UR}_{mean})$ and $SupW(c^{UR}_{sup})$ further depend upon the asymptotic null distributions of the unit root statistics that are used to construct the scaling factors. The following lemma establishes the necessary asymptotic results that directly follow from Park (1990), Park and Choi (1988), and Breitung (2002).

LEMMA 1. *Suppose that the true break date, T_b^0, is not known and regressions in (8) are estimated by using $T_b \in \Lambda^*$ as the break date. Further, suppose that Assumptions 1 and 2 hold. Let $\widehat{V}_{\overline{\alpha}}(r,\lambda)$ denote the residuals from the projection of $V_{\overline{\alpha}}(r)$ onto the space spanned by $\boldsymbol{F}(r,\lambda)$ on $[0,1]$. Similarly, let $\check{V}_{\overline{\alpha}}(r,\lambda)$ denote the residuals from the projection of $V_{\overline{\alpha}}(r)$ onto the space spanned by $\left(\boldsymbol{F}(r,\lambda)', r^{k_1}, r^{k_1+1}, \ldots, r^9\right)'$ on $[0,1]$. Then, under the null hypothesis of no break, as $T \to \infty$, if $\{u_t\}$ is $I(0)$,*

$$J(T_b) \Rightarrow 0 \quad \textit{and} \quad BG(T_b) \Rightarrow 0,$$

for all $T_b \in \Lambda^*$, and, if $\{u_t\}$ is $I(1)$,

$$J(T_b) \Rightarrow J_\infty(\lambda) \equiv \frac{\int_0^1 \widehat{V}_{\overline{\alpha}}(r,\lambda)^2\,dr}{\int_0^1 \check{V}_{\overline{\alpha}}(r,\lambda)^2\,dr} - 1,$$

$$BG(T_b) \Rightarrow BG_\infty(\lambda) \equiv \frac{\int_0^1 Q_1(r,\lambda)^2\,dr}{\int_0^1 \widehat{V}_{\overline{\alpha}}(r,\lambda)^2\,dr}.$$

It directly follows from Lemma 1 by the continuous mapping theorem (CMT) that, as $T \to \infty$,

$$J^* = \inf_{T_b\in\Lambda^*} J(T_b) \Rightarrow J^*_\infty \equiv \begin{cases} 0, & \text{if } \{u_t\} \text{ is } I(0) \\ \inf_{\lambda\in[\lambda^*,1-\lambda^*]} J_\infty(\lambda), & \text{if } \{u_t\} \text{ is } I(1) \end{cases},$$

$$BG^* = \inf_{T_b\in\Lambda^*} BG(T_b) \Rightarrow BG^*_\infty \equiv \begin{cases} 0, & \text{if } \{u_t\} \text{ is } I(0) \\ \inf_{\lambda\in[\lambda^*,1-\lambda^*]} BG_\infty(\lambda), & \text{if } \{u_t\} \text{ is } I(1) \end{cases}.$$

Although J^*_∞ and BG^*_∞ depend on $\overline{\alpha}$ when $\{u_t\}$ is $I(1)$ they are otherwise invariant to nuisance parameters.

The following theorem establishes the asymptotic null distributions of the statistics. Parts (i) and (ii) of the theorem follow from Theorem 1 of Bunzel and Vogelsang (2005). Part (iii) of the theorem follows from parts (i) and (ii) by the CMT.

THEOREM 1. *Suppose that the true break date, T_b^0, is not known and regressions in (8) are estimated by using $T_b \in \Lambda^*$as the break date. Further, suppose that Assumptions 1 and 2 hold. Let the bandwidth parameter be a fixed proportion of the sample size, $M = bT$, $b \in (0,1]$. Then, under the null hypothesis $\left(H_0 : \boldsymbol{\beta}_2 = \mathbf{0}_{q\times 1}\right)$, as $T \to \infty$,*

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

(i) $\widehat{\sigma}^2 \Rightarrow \sigma^2\Phi_0(b,k,\lambda)$, if $\{u_t\}$ is $I(0)$; $T^{-2}\widehat{\sigma}^2 \Rightarrow d(1)^2\Phi_1(b,k,\lambda)$, if $\{u_t\}$ is $I(1)$,

(ii) $W(T_b) \Rightarrow P_i(\lambda)'\left[\int_0^1 \widetilde{F}_2(r,\lambda)\,\widetilde{F}_2(r,\lambda)'\,dr\right]P_i(\lambda)\Big/\Phi_i(b,k,\lambda) \equiv W_i^{\infty}(\lambda, b, k)$,

(iii) $MeanW(c_{mean}^{UR}) \Rightarrow \left[\int_{\lambda^*}^{1-\lambda^*} W_i^{\infty}(\lambda,b,k)\,d\lambda\right] \cdot \exp\left(-c_{\text{mean}}^{UR}\cdot UR_{\infty}\right)$,
$SupW(c_{\sup}^{UR}) \Rightarrow \left[\sup_{\lambda\in[\lambda^*,1-\lambda^*]} W_i^{\infty}(\lambda,b,k)\right]\cdot\exp\left(-c_{\sup}^{UR}\cdot UR_{\infty}\right)$,

where $i=0$ *when* $\{u_t\}$ *is* $I(0)$, $i=1$ *when* $\{u_t\}$ *is* $I(1)$, *and* UR_{∞} *denotes* J_{∞}^{*} *or* BG_{∞}^{*}.

Theorem 1 demonstrates that $MeanW(c_{mean}^{UR})$ and $SupW(c_{sup}^{UR})$ are asymptotically free of nuisance parameters under the null hypothesis of no break when the errors are modeled as $I(0)$, and they only depend on the local unit root parameter $\overline{\alpha}$ when the errors are $I(1)$. Note that the distributions are affected by the scaling factors only when errors are $I(1)$. The dependence of asymptotic distributions on the kernel and the bandwidth is approximated by the random variable $\Phi_i(b,k,\lambda)$. Given a test statistic, kernel, bandwidth, and scaling factor, the scaling constants, c_{mean}^{UR} and $c_{\sup}^{UR}$, can be computed for a particular percentage point such that the test statistic remains at least asymptotically size correct, for both $I(0)$ errors and $I(1)$ errors for a wide range of values of $\overline{\alpha}$.

The asymptotic distributions given by Theorem 1 are nonstandard. However, the critical values, in addition to the values of the scaling constants, can be computed by Monte Carlo simulation methods. When computing critical values, independent and identically distributed (i.i.d.) standard normal random deviates were used to approximate the Brownian motions in the asymptotic distributions. Integrals were approximated by normalized partial sums of 1,000 steps using 10,000 replications. All simulations were carried out using $\lambda^* = 0.1$.

The simulations were carried out separately for the four models (1)–(4). For each model, asymptotic null critical values were simulated for $MeanW(c_{mean}^{UR})$ and $SupW(c_{sup}^{UR})$ for the grid of bandwidths given by $b = 0.02, 0.04, \ldots, 1.0$ and using five popular kernels: Bartlett, Parzen, Bohman, Daniell, and quadratic spectral. In each case the scaling constants c_{mean}^{UR} and $c_{\sup}^{UR}$ were computed for both J^* and BG^* as follows. Given a significance level, η, let $cv_{\eta}^{0}(b,k)$ denote the asymptotic critical value of a given statistic in the $I(0)$ case using bandwidth b and kernel $k(x)$. Let $\xi(\overline{\alpha},b,k,c_{stat}^{UR})$ denote the asymptotic limiting random variable of a given statistic in the $I(1)$ case with local parameter $\overline{\alpha}$ using bandwidth b, kernel $k(x)$, and scaling constant c_{stat}^{UR}. Consider the set $\Xi = \{0, 5, 10, \ldots, 50\}$ of possible values for $\overline{\alpha}$. The scaling constants are computed such that, for $stat = mean$, sup,

$$\sup_{\overline{\alpha}\in\Xi} P\left[\xi(\overline{\alpha},b,k,c_{stat}^{UR}) > cv_{\eta}^{0}(b,k)\right] = \eta. \tag{20}$$

A finer grid for $\overline{\alpha}$ was not considered because of the intense computational nature of the calculation even given modern computing power. Note that the value of

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

c_{stat}^{UR} that satisfies (20) is a function of b, $k(x)$, Ξ, and η. For the remainder of the paper, the more compact notation of $MeanW_{UR}$ and $SupW_{UR}$ is used to denote $MeanW(c_{mean}^{UR})$ and $SupW(c_{sup}^{UR})$ using c_{mean}^{UR} and c_{sup}^{UR} that satisfy (20).

In practice asymptotic critical values and scaling constants can be computed using information reported in the longer working paper Sayginsoy and Vogelsang (2008).[6] Because of power considerations discussed in Section 4, Sayginsoy and Vogelsary (2008) only report asymptotic critical values and scaling constants for the Daniell kernel. For commonly used significance levels, Tables $1.x.1-1.x.3$ in Sayginsoy and Vogelsary (2008), where $x = 1, 2, 3, 4$ indexes models (1)–(4), provide coefficients of polynomials in b that can be used to compute critical values and the scaling constants. For scaling constants, we found that a polynomial in b fit by OLS to simulated c_{mean}^{UR} and c_{sup}^{UR} values provided adequate fits. For the critical values, we found that a polynomial in b estimated by OLS provided a good fit to the natural logarithm of the critical values, $\ln(cv_\eta)$, rather than the levels. Given a value of b, the critical value is computed by evaluating the polynomial in b and then taking the exponential of the fitted value. We used a 12th-order polynomial in all cases, and the fits as measured by R^2 are excellent.

To illustrate how the scaling factors control the overrejection problem, we tabulate in Sayginsoy and Vogelsarg (2008) (see Tables $2.x$ [$x = 1$, 2, 3, 4.1, 4.2]) asymptotic null rejection probabilities for the optimal bandwidth statistics for a grid of $\overline{\alpha}$ ranging from 0 to 50. The nominal level is set to 5%. By construction, the rejection probabilities do not exceed the significance level, and the values of $\overline{\alpha}$ for which rejection probabilities are exactly 0.05 vary across models and statistics. In some cases, the rejection probabilities can be well below 0.05, indicating the conservative nature of the tests.

4. ASYMPTOTIC POWER ANALYSIS: OPTIMAL KERNELS AND BANDWIDTHS

This section presents a comprehensive analysis of local asymptotic power of the $MeanW_{UR}$ and $SupW_{UR}$ statistics. The analysis is carried out separately for each of the four models in (1)–(4). To compute asymptotic power functions, we require the asymptotic distribution of the statistics under local alternatives, that is, break magnitudes that are local to zero. Consider the alternative given by

$$H_1 : \beta_2 = w(T)\tau_{2T}\mathbf{D},$$

where

$$w(T) = \begin{cases} T^{-1/2}, & \text{if } \{u_t\} \text{ is } I(0) \\ T^{1/2}, & \text{if } \{u_t\} \text{ is } I(1) \end{cases};$$

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

τ_{2T} is a normalization matrix as defined in Assumption 2:

$$\tau_{2T} = \begin{cases} 1, & \text{for models (1) and (2)} \\ T^{-1}, & \text{for model (3)} \\ \begin{bmatrix} 1 & 0 \\ 0 & T^{-1} \end{bmatrix}, & \text{for model (4)} \end{cases},$$

and, for scalar constants D_1 and D_2,

$$\mathbf{D} = \begin{cases} D_1, & \text{for models (1) and (2)} \\ D_2, & \text{for model (3)} \\ [D_1 \quad D_2]' & \text{for model (4)} \end{cases}.$$

Although it is well known that $\widehat{\sigma}^2$, J, and BG are exactly invariant to the value of β_2 when the true break date, T_b^0, is used to estimate the models, this exact invariance does not hold when a break date different from T_b^0 is used. Therefore, the asymptotic distributions of $\widehat{\sigma}^2$, J, and BG depend on $\mathbf{D}$ and λ_0, and this complicates the asymptotic representations of $MeanW_{UR}$ and $SupW_{UR}$ under the alternatives. Definition 3 simplifies notation for these asymptotic results.

DEFINITION 3.

$$\boldsymbol{\delta} = \begin{cases} \boldsymbol{D}/\sigma, & \text{if } \{u_t\} \text{ is } I(0) \\ \boldsymbol{D}/d(1), & \text{if } \{u_t\} \text{ is } I(1) \end{cases},$$

$$\boldsymbol{G}(r) = \left[1 \;\; r \;\; r^2 \;\; r^3 \;\; \ldots \; r^9\right]',$$

$$\breve{\boldsymbol{F}}_2(r,\lambda) = \boldsymbol{F}_2(r,\lambda) - \left[\int_0^1 \boldsymbol{F}_2(r,\lambda)\boldsymbol{G}(r)'dr\right]\left[\int_0^1 \boldsymbol{G}(r)\boldsymbol{G}(r)'dr\right]^{-1}\boldsymbol{G}(r),$$

$$\boldsymbol{\Psi}(\lambda,\lambda_0) = \left[\int_0^1 \widetilde{\boldsymbol{F}}_2(r,\lambda)\widetilde{\boldsymbol{F}}_2(r,\lambda)'dr\right]^{-1}\left[\int_0^1 \widetilde{\boldsymbol{F}}_2(r,\lambda)\widetilde{\boldsymbol{F}}_2(r,\lambda_0)'dr\right],$$

$$\breve{\boldsymbol{\Psi}}(\lambda,\lambda_0) = \left[\int_0^1 \breve{\boldsymbol{F}}_2(r,\lambda)\breve{\boldsymbol{F}}_2(r,\lambda)'dr\right]^{-1}\left[\int_0^1 \breve{\boldsymbol{F}}_2(r,\lambda)\breve{\boldsymbol{F}}_2(r,\lambda_0)'dr\right],$$

$$\Delta(r,\lambda,\lambda_0,\boldsymbol{\delta}) = \left\{\left[\int_0^r \widetilde{\boldsymbol{F}}_2(s,\lambda)'ds\right]\boldsymbol{\Psi}(\lambda,\lambda_0) - \left[\int_0^r \widetilde{\boldsymbol{F}}_2(s,\lambda_0)'ds\right]\right\}\boldsymbol{\delta},$$

$$Q_{\delta 0}(r,\lambda,\lambda_0,\boldsymbol{\delta}) = \sigma\left[Q_0(r,\lambda) - \Delta(r,\lambda,\lambda_0,\boldsymbol{\delta})\right],$$

$$Q_{\delta 1}(r,\lambda,\lambda_0,\boldsymbol{\delta}) = d(1)\left[Q_1(r,\lambda) - \Delta(r,\lambda,\lambda_0,\boldsymbol{\delta})\right],$$

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

$$\Phi_{\delta i}(b,k,\lambda,\lambda_0,\boldsymbol{\delta}) =$$

$$\begin{cases} \int_0^1\int_0^1 -k^{*\prime\prime}(r-s)\,Q_{\delta i}(r,\lambda,\lambda_0,\boldsymbol{\delta})\,Q_{\delta i}(s,\lambda,\lambda_0,\boldsymbol{\delta})\,dr\,ds, & \text{if } k(\cdot) \text{ is Type 1} \\ \iint_{|r-s|<b} -k^{*\prime\prime}(r-s)\,Q_{\delta i}(r,\lambda,\lambda_0,\boldsymbol{\delta})\,Q_{\delta i}(s,\lambda,\lambda_0,\boldsymbol{\delta})\,dr\,ds & \\ \quad + 2k_-^{*\prime}(b)\int_0^{1-b} Q_{\delta i}(r+b,\lambda,\lambda_0,\boldsymbol{\delta}) & \\ \quad\times Q_{\delta i}(r,\lambda,\lambda_0,\boldsymbol{\delta})\,dr, & \text{if } k(\cdot) \text{ is Type 2.} \\ \frac{2}{b}\left\{\int_0^1 [Q_{\delta i}(r,\lambda,\lambda_0,\boldsymbol{\delta})]^2\,dr\right. & \\ \quad \left. -\int_0^{1-b} Q_{\delta i}(r+b,\lambda,\lambda_0,\boldsymbol{\delta})\,Q_{\delta i}(r,\lambda,\lambda_0,\boldsymbol{\delta})\,dr\right\}, & \text{if } k(\cdot) \text{ is Bartlett} \end{cases}$$

LEMMA 2. *Suppose that the regressions in (8) are estimated by using $T_b \in \Lambda^*$ and T_b^0 is the true break date. Further, suppose that Assumptions 1 and 2 hold. Then, under the local alternative, H_1, as $T \to \infty$, if $\{u_t\}$ is $I(0)$,*

$$J(T_b) \Rightarrow 0 \quad \textit{and} \quad BG(T_b) \Rightarrow 0,$$

for all $T_b \in \Lambda^*$, and, if $\{u_t\}$ is $I(1)$,

$$J(T_b) \Rightarrow J_\infty^\delta(\lambda,\lambda_0) \equiv \frac{\int_0^1 \left[\widehat{V}_{\overline{\alpha}}(r,\lambda) + \widetilde{\boldsymbol{F}}_2(r,\lambda)'[\boldsymbol{I} - \boldsymbol{\Psi}(\lambda,\lambda_0)]\boldsymbol{\delta}\right]^2 dr}{\int_0^1 \left[\breve{V}_{\overline{\alpha}}(r,\lambda) + \breve{\boldsymbol{F}}_2(r,\lambda)'\left[\boldsymbol{I} - \breve{\boldsymbol{\Psi}}(\lambda,\lambda_0)\right]\boldsymbol{\delta}\right]^2 dr} - 1,$$

$$BG(T_b) \Rightarrow BG_\infty^\delta(\lambda,\lambda_0) \equiv \frac{\int_0^1 [Q_{\delta 1}(r,\lambda,\lambda_0,\boldsymbol{\delta})]^2\,dr}{\int_0^1 \left[\widehat{V}_{\overline{\alpha}}(r,\lambda) + \widetilde{\boldsymbol{F}}_2(r,\lambda)'[\boldsymbol{I} - \boldsymbol{\Psi}(\lambda,\lambda_0)]\boldsymbol{\delta}\right]^2 dr}.$$

It directly follows from Lemma 2 by the CMT that, as $T \to \infty$,

$$J^* = \inf_{T_b \in \Lambda^*} J(T_b) \Rightarrow J_\infty^{\delta *} \equiv \begin{cases} 0, & \text{if } \{u_t\} \text{ is } I(0) \\ \inf_{\lambda \in [\lambda^*, 1-\lambda^*]} J_\infty^\delta(\lambda,\lambda_0), & \text{if } \{u_t\} \text{ is } I(1) \end{cases},$$

$$BG^* = \inf_{T_b \in \Lambda^*} BG(T_b) \Rightarrow BG_\infty^{\delta *} \equiv \begin{cases} 0, & \text{if } \{u_t\} \text{ is } I(0) \\ \inf_{\lambda \in [\lambda^*, 1-\lambda^*]} BG_\infty^\delta(\lambda,\lambda_0), & \text{if } \{u_t\} \text{ is } I(1) \end{cases}.$$

THEOREM 2. *Suppose that the regressions in* (8) *are estimated by using $T_b \in \Lambda^*$ and T_b^0 is the true break date. Further, suppose that Assumptions 1 and 2 hold. Let the bandwidth parameter be a fixed proportion of the sample size, $M = bT$, $b \in (0,1]$. Then, under the local alternative, H_1, as $T \to \infty$,*

(i) $\widehat{\sigma}^2 \Rightarrow \sigma^2 \Phi_{\delta 0}(b,k,\lambda,\lambda_0,\boldsymbol{\delta})$, *if* $\{u_t\}$ *is* $I(0)$; $T^{-2}\widehat{\sigma}^2 \Rightarrow d(1)^2\,\Phi_{\delta 1}\big(b,k,\lambda,\lambda_0,\boldsymbol{\delta}\big)$, *if* $\{u_t\}$ *is* $I(1)$,

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

$$(ii)\ W(T_b) \Rightarrow \frac{[\boldsymbol{P}_i(\lambda)+\boldsymbol{\Psi}(\lambda,\lambda_0)\boldsymbol{\delta}]'\left[\int_0^1 \widetilde{\boldsymbol{F}}_2(r,\lambda)\widetilde{\boldsymbol{F}}_2(r,\lambda)'dr\right][\boldsymbol{P}_i(\lambda)+\boldsymbol{\Psi}(\lambda,\lambda_0)\boldsymbol{\delta}]}{\Phi_{\delta i}(b,k,\lambda,\lambda_0,\boldsymbol{\delta})} \equiv W_{\delta i}^{\infty}(\lambda, \lambda_0, b, k, \boldsymbol{\delta}),$$

$$(iii)\ MeanW_{UR} \Rightarrow \left[\int_{\lambda^*}^{1-\lambda^*} W_{\delta i}^{\infty}(\lambda,\lambda_0,b,k,\boldsymbol{\delta})\,d\lambda\right] \cdot \exp\left(-c_{\text{mean}}^{UR}\cdot UR_{\infty}^{\delta}\right),$$

$$SupW_{UR} \Rightarrow \left[\sup_{\lambda\in[\lambda^*,1-\lambda^*]} W_{\delta i}^{\infty}(\lambda,\lambda_0,b,k,\boldsymbol{\delta})\right]\cdot\exp\left(-c_{\sup}^{UR}\cdot UR_{\infty}^{\delta}\right),$$

where $i = 0$ *when* $\{u_t\}$ *is* $I(0)$*,* $i = 1$ *when* $\{u_t\}$ *is* $I(1)$*, and* UR_{∞}^{δ} *generically denotes* $J_{\infty}^{\delta *}$ *and* $BG_{\infty}^{\delta *}$*.*

By Monte Carlo simulation methods similar to those used to generate asymptotic null distributions, the asymptotic distributions of $MeanW_{UR}$ and $SupW_{UR}$ given by Theorem 2 were simulated and used to compute power for various values of $\boldsymbol{\delta}$. In the case of model (4), we restrict attention to the cases where $D_2 = 0$ (in which case there is only an intercept shift) and $D_1 = 0$ (in which case there is only a slope shift). All simulations were carried out using $\lambda^* = 0.1$ and $\lambda_0 = 0.5$; other values for these parameters were not considered because of the intense computational nature of the simulations. Using 5% asymptotic null critical values, power was computed for $I(0)$ errors and for $I(1)$ errors for the grid $\overline{\alpha} = 0, 2, 4, \ldots, 148, 150$.

Under $I(0)$ errors, the scaling factors converge to unity and do not affect asymptotic power of tests. When the errors are $I(1)$, however, the scaling factors take on nondegenerate limiting distributions. Consequently, for a given value of $\overline{\alpha}$ power depends on whether J^* or BG^* is used in constructing the scaling factors.

Suppose the errors are $I(0)$. Given a value of $\boldsymbol{\delta}$**,** asymptotic power can be computed across statistics ($MeanW$ and $SupW$), across bandwidths ($b = 0.02, 0.04, \ldots, 0.98, 1.0$), and across the five kernels, and maximal power can be determined. Performing this calculation for a grid of δ traces out an asymptotic power envelope within this finite set of statistics. Similarly, when the errors are $I(1)$, power envelopes can be computed for each value of $\overline{\alpha}$. The difference in the $I(1)$ case is that the choice of unit root statistic for the scaling factor affects power, thus increasing the set of statistics. The relationship between the power envelopes and the choice of kernel, bandwidth, and scaling factor was investigated in a preliminary analysis by Sayginsoy (2004). It was found that no single statistic, $MeanW$ or $SupW$, and no single scaling factor, J^* or BG^*, attains the power envelopes across all values of $\overline{\alpha}$. Therefore, we proceed to focus on the choice of kernel and bandwidth as a function of the persistence of the errors for each of the statistics $MeanW_J$, $MeanW_{BG}$, $SupW_J$, and $SupW_{BG}$. For all four statistics, Sayginsoy (2004) found that the Daniell kernel either achieves or nearly achieves the power envelope both when the errors are $I(0)$ and when the errors are $I(1)$. Therefore, in what follows we focus attention on the Daniell kernel only. Sayginsoy (2004) found that the bandwidths that attain the power envelope depend on whether the errors are $I(0)$ or $I(1)$. In the stationary case, $b = 0.02$ effectively attains the

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

envelope.[7] When the errors are $I(1)$, it was found that the bandwidths that attain the envelopes vary across $\overline{\alpha}$. More problematic is the finding that, for a given value of $\overline{\alpha}$, the power envelope is attained with relatively small bandwidths when δ is small but with relatively large bandwidths when δ is large. This leads to power curves using small bandwidths, crossing power curves using large bandwidths, making it difficult to rank power across bandwidths. This indeterminacy in the relationship between bandwidth and power envelopes can be avoided by focusing on the area under a power curve, that is, integrated power.

Given a bandwidth b and values of $\overline{\alpha}$ and δ, let $\xi_\delta(\overline{\alpha}, b, k, c^{UR}_{stat}(b, k, \Xi, \eta))$ denote the local asymptotic limit of the *stat* statistic using the *UR* unit root test for the scaling factor and $c^{UR}_{stat}(b, \Xi, \eta)$ denote the scaling constant appropriate for a test with significance level η. Local asymptotic power is given by

$$P[\xi_\delta(\overline{\alpha}, b, k, c^{UR}_{stat}(b, \Xi, \eta))] > cv^0_\eta(b, k)],$$

where $cv^0_\eta(b, k)$ is the asymptotic critical value for the case of $I(0)$ errors. Integrated power is defined as

$$\int_0^\infty P[\xi_\delta(\overline{\alpha}, b, k, c^{UR}_{stat}(b, k, \Xi, \eta))] > cv^0_\eta(b, k)]d\delta.$$

Because this integral does not exist for statistics that have power going to one as δ increases, this is not a well-defined objective function. Instead we can use the function

$$-\int_0^\infty \left(1 - P[\xi_\delta(\overline{\alpha}, b, k, c^{UR}_{stat}(b, k, \Xi, \eta))] > cv^0_\eta(b, k)\right) d\delta,$$

which is the negative of the area between the power function and one. This objective function is well defined, is conceptually equivalent to integrated power, and will lead to the same power rankings of the statistics. Integrated power was used by Bunzel and Vogelsang (2005) and is similar in nature to other average power criteria used in the literature (see, e.g., Andrews and Ploberger, 1994). The advantage of using integrated power as a criterion is that for a given value of $\overline{\alpha}$ there must exist a statistic with maximal integrated power with respect to b. In other words, for a given value of $\overline{\alpha}$, an optimal bandwidth can be determined using integrated power as a criterion. In the few cases where power curves do not cross and there exists a value for b that attains the power envelope, that same bandwidth will maximize integrated power by construction.

Given a model, for each of the four statistics and for $\overline{\alpha} = 0, 2, 4, \ldots, 148, 150$, integrated power was computed for bandwidths $b = 0.02, 0.04, \ldots, 0.98, 1.0$ using simulation methods. This calculation was carried out for tests at the 5% significance level, that is, $\eta = 0.05$. The range of integration was truncated once δ was large enough so that $1 - P[\xi_\delta(\overline{\alpha}, b, k, c^{UR}_{stat}(b, \Xi, 0.05))]$ was very small. For each value of $\overline{\alpha}$, the bandwidth that maximizes integrated power was computed and is denoted by $b_{opt}(\overline{\alpha})$.[8] Because the computations were carried out for a relatively coarse grid over $\overline{\alpha}$, the computed $b_{opt}(\overline{\alpha})$ functions are effectively step functions.[9]

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

The $b_{opt}(\overline{\alpha})$ step functions for all the models are given in Table 3 of Sayginsoy and Vogelsang (2008) and are depicted in plots. The shape of the step functions can be summarized as follows. In all cases $b_{opt}(\overline{\alpha})$ is relatively large when $\overline{\alpha}$ is close to zero, and it decreases as $\overline{\alpha}$ increases and the serial correlation in the errors becomes less persistent. For big enough $\overline{\alpha}$, $b_{opt}(\overline{\alpha})$ decreases to 0.02, which is the power optimal bandwidth in the case of stationary errors. For the remainder of the paper, $MeanW_J$, $SupW_J$, $MeanW_{BG}$, and $SupW_{BG}$ are always implemented using optimal bandwidths (and the Daniell kernel).

To compare the power of the optimal bandwidth statistics, we first examine integrated power of the tests. Figures 1a–e plot the ratios of integrated power relative to the integrated power of $SupW_{BG}$ for a range of $\overline{\alpha}$. Obviously, the integrated power of $SupW_{BG}$ relative to itself is always equal to one. Because the tests are based on OLS estimates, they are most appropriate for $I(0)$ errors, and it is instructive to compare integrated power in this case. The case of $\overline{\alpha} \geq 50$ corresponds to errors that are quite stationary. The advantage of using the $\overline{\alpha} \geq 50$ rankings instead of rankings based on $I(0)$ asymptotics is that the $\overline{\alpha} \geq 50$ results depend on the scaling factors, whereas the scaling factors do not appear in the asymptotic results in the $I(0)$ case. As the figures show, in model (1), model (3), and case 2 (slope shift) of model (4), $MeanW_J$ has the highest integrated power whereas for model (2) and model (4), case 1 (intercept slope), $SupW_J$ has the highest integrated power.

When we examine other values of $\overline{\alpha}$, the rankings change except in model (2), where $SupW_J$ has the highest integrated power for all values of $\overline{\alpha}$. For model (1) $MeanW_J$ has the highest integrated power when $\overline{\alpha} > 24$, whereas $SupW_J$ has the highest integrated power for $\overline{\alpha} \leq 24$. In model (3), $MeanW_J$ has the highest integrated power for $\overline{\alpha} > 48$, whereas for $0 \leq \overline{\alpha} \leq 30$, $SupW_{BG}$ exhibits the highest integrated power and for $32 \leq \overline{\alpha} \leq 48$ $MeanW_{BG}$ dominates. For model (4), case 1, $SupW_J$ and $SupW_{BG}$ exhibit substantially higher integrated power than $MeanW_J$ and $MeanW_{BG}$ for all values of $\overline{\alpha}$. The statistic $SupW_{BG}$ is slightly better than $SupW_J$ for $\overline{\alpha} \leq 38$, and the reverse is true for $\overline{\alpha} > 38$. Finally, for model (4), case 2, $SupW_J$ and $SupW_{BG}$ dominate for $\overline{\alpha} \leq 26$, all four tests provide comparable integrated power for $26 \leq \overline{\alpha} \leq 44$, and $MeanW_J$ dominates for $\overline{\alpha} > 44$.

If a practitioner is looking for tests that are powerful in the case of $I(0)$ errors but are robust to size distortions caused by strong serial correlation, the integrated power results suggest that $MeanW_J$ be used in model (1), model (3), and model (4), case 2, and $SupW_J$ be used in model (2) and model (4), case 1. Figure 2a plots power of $MeanW$ and $SupW$ for the case of $I(0)$ errors along with the power envelopes in model (2). Plots for the other models are reported in Sayginsoy and Vogelsang (2008). We clearly see in the figures that the tests recommended on the basis of integrated power have power close (sometimes identical) to the envelope. Thus, the recommended Daniell-based tests with integrated power optimal bandwidths are approximately optimal in terms of power.

Figure 2b plots power of the tests in model (2) when $\overline{\alpha} = 50$. This value of $\overline{\alpha}$ would be considered in the more stationary region of the parameter space, but

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

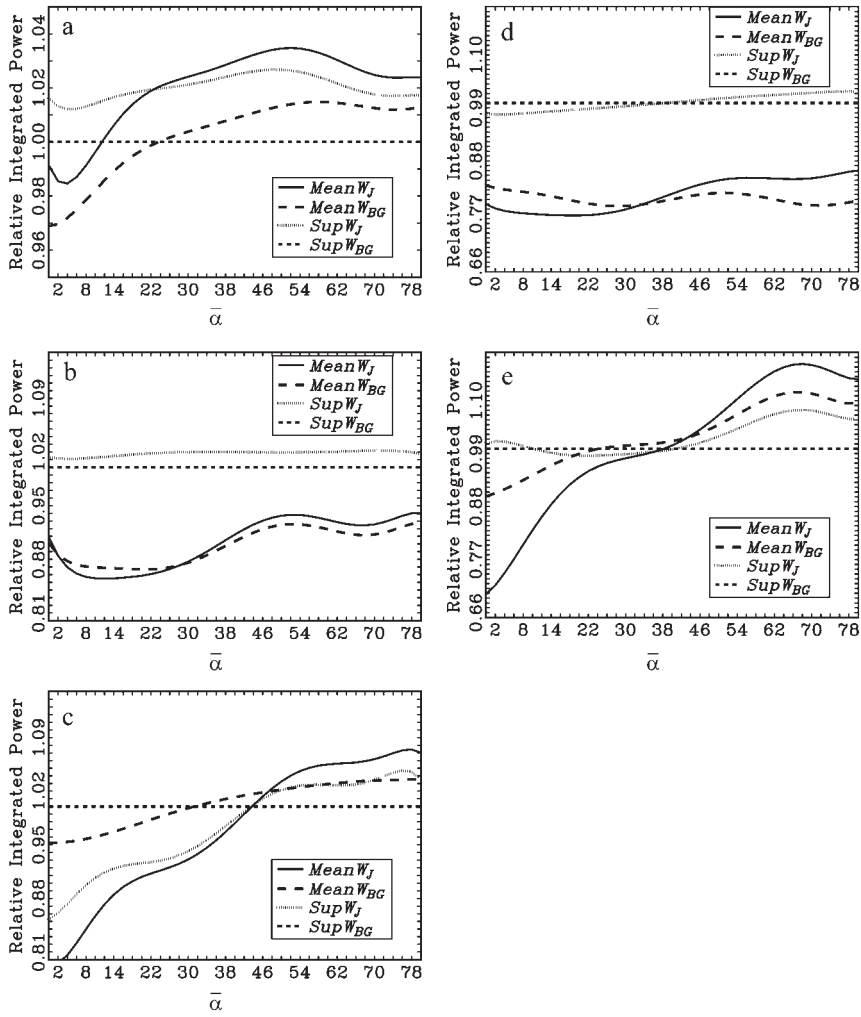


FIGURE 1. Ratios of integrated asymptotic power relative to $SupW_{BG}$: Panel a: model (1); panel b: model (2); panel c: model (3); panel d: model (4), case 1; panel e: model (4), case 2.

because the power calculations are done using the nearly integrated asymptotics, the effects of the scaling factors on power can be seen (in contrast with the $I(0)$ asymptotic power calculations where the scaling factor plays no role). Power plotted in Figure 2b is more likely to be similar to finite-sample power when there is modest serial correlation in the data than power depicted in Figure 2a. Note that the test with highest integrated power, $SupW_J$, has highest power over the full range of break magnitudes.[10]

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

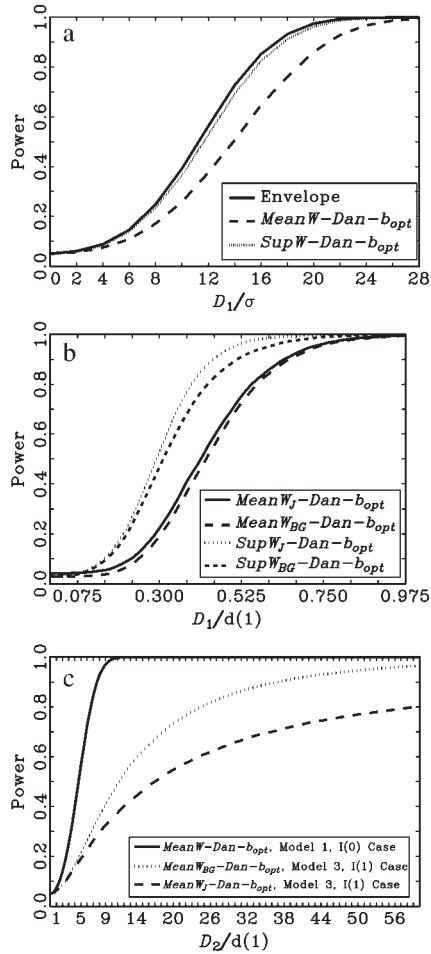


FIGURE 2. Local asymptotic power, 5% level, $\lambda_0 = 0{:}5$, $\lambda^* = 0{:}1$. Panel a: model (2), $I(0)$ errors; panel b: model (2), $I(1)$ errors, $\bar{\alpha} = 50$; panel c: model (3), $I(1)$ errors, $\bar{\alpha} = 0$ and model (1), $I(0)$ errors.

When the errors are $I(1)$ all the tests in this paper will have power that is dominated by tests based on GLS under the assumption that $\overline{\alpha}$ is known. This is because the asymptotic equality between OLS and GLS no longer holds when the errors are $I(1)$. To illustrate this, we report in Figure 2c the power of $MeanW_J$ and $MeanW_{BG}$ in model (3) when $\overline{\alpha} = 0$. The $MeanW$ test has the highest power in model (3) when the errors are $I(0)$. Also included in the figure is the power of the $MeanW$ test obtained from first differencing the data. When $\overline{\alpha} = 0$, first differencing transforms model (3) to model (1). Because the errors are $I(0)$ after first differencing, OLS estimation of the first difference specification delivers

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

the GLS estimates of model (3). If we compute the *MeanW* test using the first difference model, its power is equivalent to the power of *MeanW* in model (1) when the errors are $I(0)$. We see in Figure 2c that the power of *MeanW* applied to the first differences provides a test with much higher power than the OLS-based test applied to the levels of the data. When the strength of the serial correlation in the errors is not known, a practitioner does not know whether it would be better to work with the levels or the first differences of the data. HLT investigated the use of a linear combination of levels and first differences–based tests using traditional asymptotics and ad hoc bandwidth rules for the HAC standard errors. It would be interesting to see if the fixed-b asymptotic approximation could be used to improve the asymptotic approximation of the tests proposed by HLT, and it might be possible to compute integrated power optimal bandwidth rules for their tests.

5. FINITE-SAMPLE EVIDENCE

5.1. A Feasible Data Dependent Bandwidth Rule

To implement the optimal bandwidth tests in practice, the $b_{opt}(\overline{\alpha})$ step functions need to be made operational. Because $\overline{\alpha}$ is an unknown parameter that cannot be consistently estimated, we replace $\overline{\alpha}$ with a proxy. The effect of using a proxy for $\overline{\alpha}$ on the finite-sample performance of the tests is assessed by simulations. Consider the estimator

$$\widehat{\overline{\alpha}} = T\left(1-\widehat{\alpha}\right), \tag{21}$$

where

$$\widehat{\alpha} = \frac{\sum_{t=2}^{T}\widehat{u}_t\widehat{u}_{t-1}}{\sum_{t=2}^{T}\widehat{u}_{t-1}^2}$$

and $\widehat{u}_t$ are the OLS estimates using the estimated break date, $\widehat{T}_b$, where $\widehat{T}_b$ is obtained by minimizing the sum of squared residuals across break dates ($\widehat{T}_b$ is the least squares estimate of the break date). The estimator $\widehat{\overline{\alpha}}$ is not a consistent estimator of $\overline{\alpha}$, and "proxy" is a more appropriate label. A data dependent integrated power optimal bandwidth rule can be defined as

$$\widehat{b}_{opt} = b_{opt}\left(\widehat{\overline{\alpha}}\right).$$

When implementing test statistics, $\widehat{b}_{opt}$ is used to compute (11) by using

$$\widehat{M} = \max\left(T\cdot\widehat{b}_{opt},\ 2\right) \tag{22}$$

in place of M. The lower bound of 2 in (22) is necessary to ensure that $\widehat{M}$ is not too small when both T and $\widehat{b}_{opt}$ are small. Furthermore, the critical value and

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

scaling constants required to compute the statistics and carry out the tests can be computed as $c^{UR}(\widehat{b}_{opt})$ and $cv^0_{0.05}(\widehat{b}_{opt})$, respectively. The analysis of the impact of using $\widehat{b}_{opt}$ on the asymptotic distributions of the test statistics is well beyond the scope of this paper, and so we treat $\widehat{b}_{opt}$ as if it were not a function of the data. Finite-sample simulation results in the next section indicate that treating $\widehat{b}_{opt}$ as a constant has minor effects on the performance of the tests in most cases.

5.2. Finite-Sample Analysis of the Feasible Tests

In this section, Monte Carlo simulation methods are used to analyze finite-sample size and power performance of the Daniell kernel–based tests using the feasible optimal bandwidth rule described in Section 5.1. The data generating process (DGP) for each model is generically given by (8),

$$y_t = \mathbf{f}_1(t)'\boldsymbol{\beta}_1 + \mathbf{f}_2(t,\lambda)'\boldsymbol{\beta}_2 + u_t,$$

with ARMA(1, 1) errors,

$$u_t = \alpha u_{t-1} + e_t + \theta e_{t-1},$$

where $\{e_t\}$ is i.i.d. $\mathbb{N}(0,1)$, $u_0 = 0$, and $e_0 = 0$. Null rejection probabilities were calculated for various values of α and θ by using the appropriate 5% asymptotic critical values given by $cv^0_{0.05}(\widehat{b}_{opt})$. Results are reported for sample sizes equal to 50 and 100, and 10,000 replications were performed in each case. All computations were carried out using $\lambda_0 = 0.5$ and 10% trimming (i.e., $\lambda^* = 0.1$).

Table 1 presents empirical null hypothesis rejection probabilities for model (3). Tables for the other models appear in Sayginsoy and Vogelsang (2008). For comparison, we report rejection probabilities for the trend break test recently proposed by HLT. The HLT test involves the weighted average of the two supremum t-statistics for testing $\gamma = 0$. The first t-statistic is computed using model (3) analogously to the Wald statistic given by (10) using the Bartlett kernel. Let t_0^* denote the supremum of the absolute value of this t-statistic across breaks in Λ^*. Let $\widehat{\lambda}$ denote the break point corresponding to the supremum. The second t-statistic is the Bartlett kernel t-statistic for testing $\varphi = 0$ in model (1) applied to the first differences of the data, Δy_t. Let t_1^* denote the supremum of the absolute value of this t-statistic across breaks in Λ^* and let $\widetilde{\lambda}$ denote the break point corresponding to the supremum.

The HLT statistic is then defined as

$$HLT = \exp\left(-[500KPSS_0(\widehat{\lambda})\cdot KPSS_1(\widetilde{\lambda})]^2\right)t_0^* + m_\xi\left(1-\exp\left(-[500KPSS_0(\widehat{\lambda})*KPSS_1(\widetilde{\lambda})]^2\right)\right)t_1^*,$$

where $KPSS_0(\widehat{\lambda})$ is the KPSS statistic computed using the OLS residuals from model (3) with break date $\widehat{\lambda}T$ and $KPSS_1(\widetilde{\lambda})$ is the KPSS statistic computed

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

using the OLS residuals from model (1) applied to Δy_t using the break date $\widetilde{\lambda}T$. Both KPSS statistics are implemented using the Bartlett kernel. The constant, m_ξ, depends on the significance level of the test and is configured so that the asymptotic null distribution of the HLT statistic is the same for $I(0)$ and $I(1, \overline{\alpha} = 0)$ errors. For the case of 10% trimming the asymptotic 5% critical value of the HLT statistic is 2.563 and $m_\xi = 0.853$. We report results using two bandwidth choices for the two Bartlett kernel HAC estimators. The first follows HLT and uses the deterministic rule $M(T) = [4(T/100)^{1/4}]$ so that the same bandwidth is used for t_0^*, t_1^*, and the KPSS statistics. The second uses the data dependent $AR(1)$ plug-in method following Andrews (1991), in which case $t_0^*, KPSS_0$ use a different bandwidth than $t_1^*, KPSS_1$.

In Table 1 the first six columns give rejections for $T = 50$, whereas the next six columns give rejections for $T = 100$. The first thing to note in Table 1 is that the OLS Daniell kernel statistics have rejections that are usually close to or below the nominal level of 0.05. Exceptions occur when the MA(1) parameter is negative ($\theta = -0.4$) and the $AR(1)$ parameter is close to or equal to one, in which case overrejections occur. Without the scaling factors the Daniell kernel statistics would show substantial over-rejection problems as α approaches one regardless of the value of θ (see Sayginsoy and Vogelsang 2008). The rejections of the HLT statistic are less robust, with overrejection problems occurring when α is close to or equal to one. When the deterministic bandwidth rule is used, increasing T from 50 to 100 causes most of the overrejections to decrease in magnitude. Indeed, simulations reported by HLT for the same DGP with $T = 150$ and $T = 300$ indicate that overrejections reduce further as T increases. Although the HLT statistic has correct size asymptotically, it can take relatively large sample sizes for the asymptotic approximation to be reasonable when there is a unit root in the errors. When data dependent bandwidths are used for the HLT statistic, the overrejection problem becomes slightly worse.

We now focus on finite-sample power results. We first give results for model (2) to show that the predictions made by the local asymptotic power calculations are a good predictor of finite-sample behavior. We focus on $AR(1)$ errors with $\alpha = 0.0, 0.5, 0.8, 1.0$. Figures 3a–d plot the power of the Daniell kernel tests for $T = 100$. Plots for $T = 50$ can be found in Sayginsoy and Vogelsang (2008). The test recommended for model (2), $SupW_J$, has the highest power in all cases as predicted by the asymptotic power analysis. These results show that the predictions of the asymptotic power analysis are reflected in the finite-sample behavior of the tests at least for $AR(1)$ models.

In Table 2 we report finite-sample power with $T = 100$ for model (3) for the case of the $AR(1)$ errors. In addition to the OLS Daniell kernel tests and the HLT test, we report power for the exponential average GLS test proposed by PY. The practical implementation of the PY statistic is complicated and we do not review the details here. Instead we refer the reader to Sections 2 and 3 of PY, which provide the details on practical implementation for the $AR(1)$ error case. The $AR(1)$ GLS transformation used by the PY statistic is based on a bias corrected estimate

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

TABLE 1. Empirical null hypothesis rejection probabilities for model (3),[a] 5% nominal level, 10% trimming ($\lambda^* = 0.10$)

		$T = 50$						$T = 100$					
		Proposed tests[b]						Proposed Tests[b]					
		Daniell kernel, $b = \widehat{b}_{opt}$				HLT test[c]		Daniell kernel, $b = \widehat{b}_{opt}$				HLT test[c]	
		Mean	*Sup*	*Mean*	*Sup*	Bandwidth		*Mean*	*Sup*	*Mean*	*Sup*	Bandwidth	
θ	α	W_J	W_J	W_{BG}	W_{BG}	$M(T)$	$\widehat{M}$	W_J	W_J	W_{BG}	W_{BG}	$M(T)$	$\widehat{M}$
−0.4	.0	.010	.009	.003	.002	.003	.009	.005	.003	.003	.002	.002	.003
	.3	.020	.016	.004	.002	.007	.001	.012	.007	.006	.002	.006	.007
	.5	.027	.020	.006	.003	.011	.018	.026	.017	.012	.006	.009	.017
	.7	.038	.032	.016	.007	.018	.038	.045	.036	.021	.013	.017	.040
	.9	.061	.061	.041	.031	.049	.092	.050	.049	.038	.030	.033	.085
	.95	.078	.081	.064	.053	.085	.132	.071	.066	.055	.052	.051	.127
	1.00	.104	.108	.091	.078	.124	.177	.120	.115	.100	.096	.120	.212
0.0	.0	.024	.018	.005	.002	.009	.014	.018	.012	.009	.004	.007	.011
	.3	.030	.021	.006	.002	.016	.030	.022	.014	.008	.004	.001	.030
	.5	.030	.024	.008	.003	.020	.045	.021	.018	.008	.005	.014	.059
	.7	.025	.024	.010	.006	.028	.069	.029	.022	.008	.007	.020	.090
	.9	.032	.032	.020	.018	.094	.123	.028	.027	.012	.010	.035	.110
	.95	.040	.038	.030	.028	.160	.179	.035	.037	.023	.021	.068	.147
	1.00	.050	.048	.046	.045	.225	.235	.058	.060	.052	.053	.172	.255

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

0.4	.0	.033	.024	.006	.002	.013	.034	.016	.010	.005	.002	.008	.039
	.3	.028	.025	.007	.003	.017	.059	.022	.018	.007	.003	.011	.085
	.5	.022	.019	.007	.002	.019	.064	.034	.027	.008	.005	.015	.094
	.7	.018	.015	.006	.004	.033	.058	.031	.029	.007	.005	.021	.084
	.9	.024	.022	.016	.014	.112	.108	.024	.023	.010	.009	.037	.073
	.95	.032	.026	.025	.022	.185	.171	.030	.027	.023	.018	.076	.108
	1.00	.039	.032	.038	.035	.255	.231	.050	.047	.049	.045	.193	.220
0.8	.0	.032	.029	.008	.003	.013	.056	.020	.016	.004	.002	.009	.078
	.3	.023	.022	.007	.002	.016	.059	.032	.024	.008	.004	.011	.089
	.5	.020	.017	.006	.002	.018	.053	.036	.033	.009	.006	.015	.077
	.7	.016	.013	.007	.004	.033	.044	.030	.027	.069	.005	.021	.059
	.9	.021	.019	.017	.012	.114	.098	.023	.020	.012	.008	.037	.055
	.95	.029	.023	.026	.020	.191	.163	.029	.025	.024	.017	.077	.091
	1.00	.035	.026	.039	.032	.258	.231	.048	.043	.049	.043	.197	.204

[a]DGP: $y_t = \mu + \theta t + \varphi DU_t + u_t,\ u_t = \alpha u_{t-1} + e_t + \rho e_{t-1}$ with $H_0 : \varphi = 0$, where $\{e_t\}$ are i.i.d. $\mathbb{N}(0,1)$, $u_0 = 0$, $e_0 = 0$; 10,000 replications were performed for each ARMA(1, 1) specification. The rejection probabilities were computed by using the asymptotic 5% critical value appropriate for each test as explained in the text.

[b]For each replication of the DGP, $\overline{\alpha}$ was estimated as $\widehat{\overline{\alpha}} = T\left(1-\widehat{\alpha}\right)$, where $\widehat{\alpha} = \sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1} / \sum_{t=2}^{T} \widehat{u}_{t-1}^2$ and $\{\widehat{u}_t\}$ are the OLS residuals using the estimated break date, $\widehat{T}_b$, where $\widehat{T}_b$ is obtained by minimizing the sum of squared residuals across break dates. When implementing test statistics, $\widehat{b}_{opt} = b_{opt}(\widehat{\overline{\alpha}})$ is used to compute the long-run variance by using $\widehat{M} = \max\left(T \cdot \widehat{b}_{opt},\ 2\right)$. Then $\widehat{b}_{opt}$ was used to determine the corresponding asymptotic fixed-b critical value.

[c]The HLT tests use the Bartlett kernel with bandwidths given by $M(T) = [4(T/100)^{1/4}]$, and $\widehat{M}$ is the $AR(1)$ plug-in data dependent bandwidth of Andrews (1991).

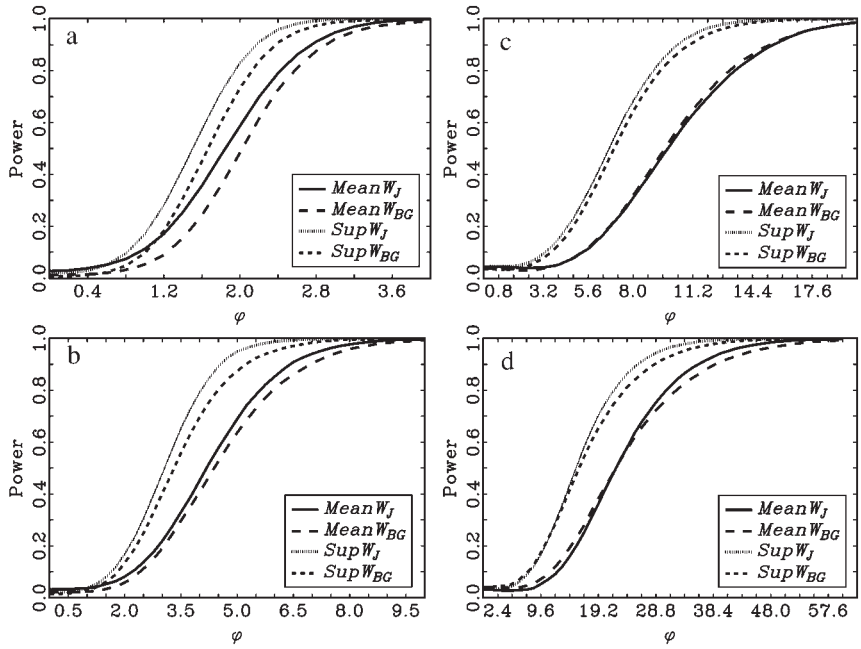


FIGURE 3. Finite sample power in model (2), $AR(1)$ errors, $T = 100$, Daniell kernel, $b = \hat{b}_{opt}$, 5% level, $\lambda_0 = 0{:}5$, $\lambda^* = 0{:}1$. panel a: $\alpha = 0$; panel b: $\alpha = 0{:}5$; panel c: $\alpha = 0{:}8$; panel d: $\alpha = 1{:}0$.

of α due to Roy, Falk, and Fuller (2004) denoted by $\widehat{\alpha}_M$ (see PY, eqn. (7)). Given $\widehat{\alpha}_M$, PY construct the estimator

$$\widehat{\alpha}_{MS} = \begin{cases} \widehat{\alpha}_M & \text{if } \left|\widehat{\alpha}_M - 1\right| > T^{-1/2} \\ 1 & \text{if } \left|\widehat{\alpha}_M - 1\right| \leq T^{-1/2} \end{cases}$$

and use it for the GLS transformation.

This hybrid estimator of α can be given a pretest interpretation. Note that we can define $\widehat{\alpha}_{MS}$ equivalently as

$$\widehat{\alpha}_{MS} = \begin{cases} \widehat{\alpha}_M & \text{if } T\left|\widehat{\alpha}_M - 1\right| > T^{1/2} \\ 1 & \text{if } T\left|\widehat{\alpha}_M - 1\right| \leq T^{1/2} \end{cases},$$

where $T\left|\widehat{\alpha}_M - 1\right|$ can be viewed as a unit root statistic with a critical value that increases with the sample size at rate $T^{1/2}$. When $T\left|\widehat{\alpha}_M - 1\right| \leq T^{1/2}$, we do not reject a unit root in the error, and the unit root is taken as given. Then GLS is equivalent to taking first differences of the data, and model (1) is used. When $T\left|\widehat{\alpha}_M - 1\right| > T^{1/2}$, the unit root is rejected, and feasible GLS proceeds with the estimated value of α.

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

TABLE 2. Finite-sample power for model (3), 5% nominal level, 10% trimming ($\lambda^* = 0.10$), $T = 100$

		Proposed tests, Daniell kernel, $b = \widehat{b}_{opt}$						HLT test		PY
		Mean	*Sup*	*Mean*	*Sup*	Pretests[a]		Bandwidth		
α	δ	W_J	W_J	W_{BG}	W_{BG}	$\widehat{\alpha}$	$\widetilde{\alpha}_M$	$M(T)$	$\widehat{M}$	GLS
0	0	.018	.012	.009	.004	.018	.018	.007	.011	.064
	.012	.052	.031	.029	.013	.052	.052	.020	.027	.144
	.024	.197	.120	.132	.066	.197	.197	.068	.098	.380
	.036	.462	.334	.360	.218	.461	.461	.162	.251	.681
	.048	.736	.603	.641	.473	.736	.736	.301	.444	.895
	.060	.904	.829	.847	.733	.904	.904	.424	.607	.980
	.072	.975	.946	.946	.894	.975	.975	.520	.716	.997
	.084	.993	.984	.985	.968	.993	.993	.589	.781	1.00
	.096	.998	.995	.996	.989	.998	.998	.645	.819	1.00
0.5	0	.021	.018	.008	.005	.020	.024	.014	.059	.063
	.03	.075	.056	.038	.021	.074	.082	.049	.156	.164
	.06	.265	.210	.186	.116	.267	.271	.166	.428	.449
	.09	.526	.429	.421	.323	.527	.509	.345	.704	.740
	.12	.731	.642	.651	.553	.735	.696	.494	.851	.902
	.15	.842	.797	.800	.732	.847	.791	.601	.915	.959
	.18	.902	.882	.885	.840	.907	.849	.679	.945	.979
	.21	.934	.929	.932	.907	.939	.885	.744	.960	.988
	.24	.953	.955	.958	.942	.956	.910	.803	.969	.991
0.8	0	.029	.025	.008	.007	.028	.020	.024	.096	.030
	.2	.282	.281	.248	.209	.287	.147	.318	.711	.184
	.4	.532	.539	.610	.611	.539	.343	.768	.931	.316
	.6	.679	.691	.788	.821	.686	.894	.993	.995	.902
	.8	.766	.783	.875	.913	.775	.991	1.00	1.00	1.00
	1.0	.827	.842	.919	.956	.834	.995	1.00	1.00	1.00
	1.2	.851	.868	.937	.971	.871	.998	1.00	1.00	1.00
	1.4	.891	.908	.963	.988	.897	.999	1.00	1.00	1.00
	1.6	.910	.928	.972	.994	.918	.999	1.00	1.00	1.00
1.0	0	.058	.060	.052	.053	.066	.061	.172	.255	.088
	.2	.085	.087	.093	.091	.102	.148	.282	.359	.176
	.4	.149	.151	.184	.192	.203	.420	.541	.612	.454
	.6	.216	.224	.295	.313	.324	.734	.807	.841	.768
	.8	.281	.292	.396	.438	.440	.929	.955	.962	.947
	1.0	.334	.350	.482	.543	.509	.988	.993	.994	.994
	1.2	.384	.403	.553	.631	.561	.998	1.00	1.00	1.00
	1.4	.427	.452	.613	.703	.601	.999	1.00	1.00	1.00
	1.6	.467	.493	.659	.761	.638	.999	1.00	1.00	1.00

Note: See the notes to Table 1.

[a] The pretests use $MeanW_J$ applied to either model (3) or model (1). Model (3) is used if $T\left|\widehat{\alpha}-1\right| > T^{1/2}$, $T\left|\widetilde{\alpha}_M-1\right| > T^{1/2}$. Model (1) is used for the first differences of the data if $T\left|\widehat{\alpha}-1\right| \le T^{1/2}$, $T\left|\widetilde{\alpha}_M-1\right| \le T^{1/2}$.

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

Using the decision rule that defines $\widehat{\alpha}_{MS}$ it is possible to configure a pretest version of the OLS Daniell kernel tests. Recall that for model (3) the $MeanW_J$ statistic has the highest integrated power when the errors are $I(0)$. When the errors are $I(1)$, first differencing the data results in model (1) with $I(0)$ errors, in which case $MeanW_J$ continues to have the highest integrated power. The pretest procedure is implemented as follows. If $T\left|\widehat{\alpha}_M - 1\right| > T^{1/2}$, indicating that the errors are $I(0)$, $MeanW_J$ is computed using the levels of the data and model (3). If $T\left|\widehat{\alpha}_M - 1\right| \leq T^{1/2}$, indicating a unit root in the errors, the data are first differenced, and $MeanW_J$ is computed using model (1). We report in Table 2 finite-sample power of this pretest version of $MeanW_J$. We also implemented this pretest using $\widehat{\alpha}$ in place of $\widehat{\alpha}_M$.

The power results reported in Table 2 are organized as follows. The first four columns give results for the four OLS Daniell kernel tests. Columns five and six give results for the pretest version of $MeanW_J$. Columns seven and eight give results for the HLT test for the two bandwidth rules. The last column gives the power of the PY statistic. Overall, no test dominates the others in terms of power, especially when null rejection probabilities are taken into account. When the errors have no serial correlation ($\alpha = 0$), the PY test has the highest power. Notice that the PY test slightly overrejects, whereas the other tests substantially underreject. The superior power of the PY test is likely due to the differences in the null rejection probabilities. Similar patterns hold for $\alpha = 0.5$, although the HLT test using the data dependent bandwidth has power comparable to the PY test. When $\alpha = 0.8$ several of the tests have similar null rejection probabilities. Ignoring the HLT test based on the data dependent bandwidth because it is nontrivially oversized, we see that the HLT test based on the deterministic bandwidth has the highest power. For small values of δ the $MeanW_J$, the $SupW_J$, and the pretest based on $\widehat{\alpha}$ have more power than the PY test, whereas for medium to large values of δ the PY test has higher power. Not surprisingly, the pretest based on $\widehat{\alpha}_M$ has power very similar to the PY test. When $\alpha = 1.0$, the HLT and PY tests overreject, whereas the Daniell kernel tests have null rejections closer to 0.05. Power of the OLS Daniell kernel tests is generally low because of the unit root in the errors. Using the pretest with $\widehat{\alpha}_M$ substantially boosts the power of the OLS Daniell kernel tests without causing overrejection problems and gives power very similar to the PY test.

6. CONCLUSION

This paper analyzes tests for a shift in trend at an unknown date based on OLS estimates and nonparametric kernel HAC robust standard errors. It is shown that the fixed-b asymptotic approximation of Kiefer and Vogelsang (2005) used in conjunction with the scaling approach of Vogelsang (1998a) makes it possible to configure the $SupW$ test of Andrews (1993) and the $MeanW$ test of Andrews and Ploberger (1994) so that the tests do not suffer from the well-known overrejection problems caused by strong serial correlation in the errors. Concrete

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

recommendations are provided for the kernel and bandwidth choices required to implement the tests. Using local asymptotic power calculations, the Daniell kernel is shown to deliver tests with approximately the highest integrated (average) power within a class of popular kernels. For 5% level tests a bandwidth rule is provided that maximizes integrated power of the Daniell kernel–based tests, and a finite-sample simulation experiment shows that a feasible data dependent version of the bandwidth rule works well in practice. For four different trend break specifications our analysis provides practitioners with specific recommendations regarding the form of the statistic (*Sup* vs. *Mean*), kernel, and bandwidth rule. The tests are not prone to overrejection caused by strong serial correlation, and the recommended tests maximize integrated power when the data are trend stationary. Power of the tests can be increased without inducing overrejection problems through the use of a unit root pretest.

NOTES

1. For discussions on U.S. regional convergence, see, for example, Bernard and Durlauf (1995, 1996), Carlino and Mills (1993), Loewy and Papell (1996), Sayginsoy (2006), and Tomljanovic and Vogelsang (2002).

2. Throughout the paper, whenever errors are said to be $I(1)$, it is intended to mean that errors exhibit a local-to-unity process, unless otherwise stated.

3. The constant λ^* is sometimes referred to as the "trimming."

4. Other unit root statistics could be used to construct scaling factors, and the choice of unit root statistic would affect the power of the scaled statistics in a complicated manner when the errors are $I(1)$. Obtaining analytical results on the choice of unit root statistic would be very difficult if even feasible. We do not consider unit root statistics that have tuning parameters that are not captured by the asymptotic theory so that we can focus attention on the choice of tuning parameters (kernel, bandwidth) needed to implement the HAC trend statistics.

5. The J^* and BG^* statistics can also be used to test for a unit root while allowing for a break in trend of unknown timing. The infimum is used because both unit root tests are left tailed.

6. The longer working paper is available on Tim Vogelsang's MSU Web page: https://www.msu.edu/~tjv/working.html.

7. Technically asymptotic power is highest in the $I(0)$ case when $b = 0$, but we cannot recommend using $b = 0$ in practice because this would make $\widehat{\sigma}^2$ an estimate of the variance of u_t rather than an estimate of the long-run variance of u_t.

8. Extending the range of Ξ to include larger values of $\overline{\alpha}$ would have no effect on the power results because the supremum in (20) occurs for $\overline{\alpha} = 0, 5$, or 10 (depending on the statistic), which implies that the scaling constants would not change by the range extension. This can be seen in Tables 2.1–2.4.2 of Sayginsoy and Vogelsang (2008), where asymptotic null rejections for the optimal bandwidth tests are equal to 0.05 for $\overline{\alpha} = 0, 5$, or 10.

9. It would be possible in principle to obtain smoother functions for $b_{opt}(\overline{\alpha})$ by using a very fine grid over $\overline{\alpha}$, but the computational cost would be prohibitive.

10. Plots for the other models are reported in Sayginsoy and Vogelsang (2008) and can be summarized as follows. In model (1), model (3), and model (4), case 2, the test that has highest integrated power has highest power over the full range of the break magnitude. But, in model (4), case 1, the test with highest integrated power, $SupW_J$, has the highest power for small to medium intercept shifts, whereas for larger intercept shifts, $SupW_{BG}$ has highest power. Thus the power curves intersect and no single test dominates in terms of power, whereas in terms of integrated power $SupW_J$ dominates $SupW_{BG}$.

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–854.

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.

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

Bai, J. (1994) Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis* 15, 453–470.

Bai, J., R.L. Lumsdaine, & J.H. Stock (1998) Testing for and dating breaks in integrated and cointegrated time series. *Review of Economic Studies* 65, 395–432.

Bernard, A.B. & S.N. Durlauf (1995) Convergence in international output. *Journal of Applied Econometrics* 10, 161–173.

Bernard, A.B. & S.N. Durlauf (1996) Interpreting tests for the convergence hypothesis. *Journal of Econometrics* 91, 161–173.

Breitung, J. (2002) Nonparametric tests for unit roots and cointegration. *Journal of Econometrics* 108, 343–363.

Bunzel, H. & T.J. Vogelsang (2005) Powerful trend function tests that are robust to strong serial correlation with an application to the Prebish Singer hypothesis. *Journal of Business & Economic Statistics* 23, 381–394.

Carlino, G.A. & L.O. Mills (1993) Are U.S. regional incomes converging? *Journal of Monetary Economics* 32, 335–346.

Chen, C.L. & Tiao, G.C. (1990) Random level-shift time series models, ARIMA approximations, and level-shift detection. *Journal of Business & Economic Statistics* 8, 83–97.

Chu, C.S.J. & H. White (1992) A direct test for changing trend. *Journal of Business & Economic Statistics* 10, 289–300.

Frisch, R. & F.V. Waugh (1933) Partial time regressions as compared with individual trends. *Econometrica* 1, 387–401.

Gregory, A. & B.E. Hansen (1996) Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics* 70, 99–126.

Grenander, U. & M. Rosenblatt (1957) *Statistical Analysis of Stationary Time Series*. Wiley.

Harvey, D., S.J. Leybourne, & A.M.R. Taylor (2009) Simple, robust and powerful tests of the breaking trend hypothesis. *Econometric Theory* 25, 995–1029.

Kiefer, N.M. & T.J. Vogelsang (2005) A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130–1164.

Kramer, W., W. Ploberger, & R. Alt (1988) Testing for structural change in dynamic models. *Econometrica* 56, 1355–1369.

Kwiatkowski, D., P. Phillips, P. Schmidt & Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root. *Journal of Econometrics* 54, 154–178.

Loewy, M.B. & Papell, D.H. (1996) Are U.S. regional incomes converging? Some further evidence, *Journal of Monetary Economics* 38, 587–598.

Park, J.Y. (1990) Testing for unit roots and cointegration by variable addition. In T. Fomby & R. Rhodes (eds.), *Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots*, pp. 107–134. Jai Press.

Park, J.Y. & B. Choi (1988) A New Approach to Testing for a Unit Root. CAE Working paper 88–23, Cornell University.

Perron, P. (1989) The Great Crash, the oil price shock and the unit root hypothesis. *Econometrica* 57, 1361–1401.

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

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

Perron, P. & T. Yabu (2009) Testing for shifts in trend with an integrated or stationary noise component. *Journal of Business & Economic Statistics* 27, 369–396.

Perron, P. & X. Zhu (2005) Structural breaks with stochastic and deterministic trends. *Journal of Econometrics* 129, 65–119.

Phillips, P.C.B. (1987) Time series regression with unit roots. *Econometrica* 55, 277–302.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Roy, A., B. Falk, & W. Fuller (2004) Testing for trend in the presence of autoregressive error. *Journal of the American Statistical Association* 99, 1082–1092.

Sayginsoy, O. (2004) Essays on nonstandard testing of trend functions in time series models. *Dissertation Abstracts International* 65, 3938–4167A. Ph.D. dissertation, Cornell University.

Sayginsoy, O. (2006) Joint Inequality Tests of Trend Function Parameters. Working paper, Department of Economics, SUNY–University at Albany.

Sayginsoy, O. and Vogelsang, T.J. (2008) Testing for a Shift in Trend at an Unknown Date: A Fixed-*b* Analysis of Heteroskedasticity Autocorrelation Robust OLS Based Tests. Working paper, Department of Economics, Michigan State University.

Shin, Y. & P. Schmidt (1992) The KPSS stationarity test as a unit root test. *Economics Letters* 38, 387–392.

Tanaka, K. (1990) Testing for a moving average root. *Econometric Theory* 6, 433–444.

Tomljanovic, M. & T.J. Vogelsang (2002) Are U.S. regions converging? Using new econometric methods to examine old issues. *Empirical Economics* 27, 49–62.

Vogelsang, T.J. (1997) Wald-type tests for detecting shifts in the trend function of a dynamic time series. *Econometric Theory* 13, 818–849.

Vogelsang, T.J. (1998a) Testing for a shift in mean without having to estimate serial correlation parameters. *Journal of Business & Economic Statistics* 16, 73–80.

Vogelsang, T.J. (1998b) Trend function hypothesis testing in the presence of serial correlation. *Econometrica* 65, 123–148.

APPENDIX: Supplementary Results and Proofs

The following lemmas are used to prove Lemma 2 and Theorems 1 and 2. The proofs of these lemmas are straightfoward, and details are given in Sayginsoy and Vogelsang (2008).

LEMMA A.1. *Suppose that the regressions in (8) are estimated by using* $T_b \in \Lambda^*$. *Further, suppose that Assumptions 1 and 2 hold. Then, under the null hypothesis as* $T \to \infty$, *if* $\{u_t\}$ *is* $I(0)$,

$$T^{-1/2}\widehat{S}_{[rT]} \Rightarrow \sigma\left[W(r) - \int_0^r \widetilde{\boldsymbol{F}}_2(s,\lambda)'\boldsymbol{P}_0(\lambda)\,ds\right] \equiv \sigma\, Q_0(r,\lambda),$$

and, if $\{u_t\}$ *is* $I(1)$,

$$T^{-3/2}\widehat{S}_{[rT]} \Rightarrow d(1)\left[\int_0^r V_{\overline{\alpha}}(s)\,ds - \int_0^r \widetilde{\boldsymbol{F}}_2(s,\lambda)'\boldsymbol{P}_1(\lambda)\,ds\right] \equiv d(1)\,Q_1(r,\lambda).$$

LEMMA A.2. *Suppose that the regressions in (8) are estimated by using* $T_b \in \Lambda^*$ *and* T_b^0 *is the true break date. Further, suppose that Assumptions 1 and 2 hold. Then, under the local alternative,* H_1, *as* $T \to \infty$, *if* $\{u_t\}$ *is* $I(0)$,

$$T^{1/2}\boldsymbol{\tau}_{2T}^{-1}\widehat{\boldsymbol{\beta}}_2 \Rightarrow \sigma\boldsymbol{P}_0(\lambda) + \boldsymbol{\Psi}(\lambda,\lambda_0)\boldsymbol{D},$$

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

and, if $\{u_t\}$ *is* $I(1)$,

$$T^{-1/2}\boldsymbol{\tau}_{2T}^{-1}\widehat{\boldsymbol{\beta}}_2 \Rightarrow d(1)\boldsymbol{P}_1(\lambda)+\boldsymbol{\Psi}(\lambda,\lambda_0)\boldsymbol{D}.$$

LEMMA A.3. *Suppose that the regressions in (8) are estimated by using* $T_b \in \Lambda^*$ *and* T_b^0 *is the true break date. Further, suppose that Assumptions 1 and 2 hold. Then, under the local alternative,* H_1, *as* $T \to \infty$, *if* $\{u_t\}$ *is* $I(0)$,

$$T^{-1/2}\widehat{S}_{[rT]} \Rightarrow \sigma\left[Q_0(r,\lambda)-\left\{\left[\int_0^r \widetilde{\boldsymbol{F}}_2(s,\lambda)'\,ds\right]\boldsymbol{\Psi}(\lambda,\lambda_0)-\left[\int_0^r \widetilde{\boldsymbol{F}}_2(s,\lambda_0)'\,ds\right]\right\}\boldsymbol{\delta}\right],$$

$$\equiv \sigma\, Q_{\delta 0}(r,\lambda,\lambda_0,\boldsymbol{\delta}),$$

and, if $\{u_t\}$ *is* $I(1)$,

$$T^{-3/2}\widehat{S}_{[rT]} \Rightarrow d(1)\left[Q_1(r,\lambda)-\left\{\left[\int_0^r \widetilde{\boldsymbol{F}}_2(s,\lambda)'\,ds\right]\boldsymbol{\Psi}(\lambda,\lambda_0)-\left[\int_0^r \widetilde{\boldsymbol{F}}_2(s,\lambda_0)'\,ds\right]\right\}\boldsymbol{\delta}\right],$$

$$\equiv d(1)\,Q_{\delta 1}(r,\lambda,\lambda_0,\boldsymbol{\delta}).$$

Proof of Lemma 2. The proof of the first part of the lemma (i.e., when $\{u_t\}$ is $I(0)$) can be obtained by trivial modifications to the proof of the second part of the lemma (i.e., when $\{u_t\}$ is $I(1)$). The proof of the second part of the lemma is given in what follows. First, the asymptotic distribution of $BG(T_b)$ will be obtained. Suppose that $T_b \in \Lambda^*$ is used as the break date. The residuals are

$$\widehat{u}_t = \widetilde{y}_t - \widetilde{\mathbf{f}}_2(t,\lambda)'\widehat{\boldsymbol{\beta}}_2 = \widetilde{\mathbf{f}}_2(t,\lambda_0)'\boldsymbol{\beta}_2 + u_t - \widetilde{\mathbf{f}}_2(t,\lambda)'\widehat{\boldsymbol{\beta}}_2.$$

Under the local alternative, H_1, if $\{u_t\}$ is $I(1)$,

$$\widehat{u}_t = \widetilde{\mathbf{f}}_2(t,\lambda_0)'T^{1/2}\boldsymbol{\tau}_{2T}\mathbf{D} + u_t - \widetilde{\mathbf{f}}_2(t,\lambda)'\widehat{\boldsymbol{\beta}}_2. \tag{A.1}$$

Rewriting (A.1),

$$T^{-1/2}\widehat{u}_t = \widetilde{\mathbf{f}}_2(t,\lambda_0)'\boldsymbol{\tau}_{2T}\mathbf{D} + T^{-1/2}u_t - \widetilde{\mathbf{f}}_2(t,\lambda)'\boldsymbol{\tau}_{2T}T^{-1/2}\boldsymbol{\tau}_{2T}^{-1}\widehat{\boldsymbol{\beta}}_2. \tag{A.2}$$

Noting that $\widetilde{\mathbf{f}}_2([rT],\lambda_0)'\boldsymbol{\tau}_{2T} \to \widetilde{\mathbf{F}}_2(r,\lambda_0)'$, using (6) and Lemma A.2, it follows by the CMT that, as $T \to \infty$,

$$\begin{aligned}
T^{-1/2}\widehat{u}_{[rT]} &\Rightarrow \widetilde{\mathbf{F}}_2(r,\lambda_0)'\mathbf{D} + d(1)\,V_{\overline{\alpha}}(r) - \widetilde{\mathbf{F}}_2(r,\lambda)'\left[d(1)\mathbf{P}_1(\lambda)+\boldsymbol{\Psi}(\lambda,\lambda_0)\mathbf{D}\right] \\
&= d(1)\left[V_{\overline{\alpha}}(r) - \widetilde{\mathbf{F}}_2(r,\lambda)'\mathbf{P}_1(\lambda)\right] + \widetilde{\mathbf{F}}_2(r,\lambda)'\left[\mathbf{I}-\boldsymbol{\Psi}(\lambda,\lambda_0)\right]\mathbf{D} \\
&= d(1)\,\widehat{V}_{\overline{\alpha}}(r,\lambda) + \widetilde{\mathbf{F}}_2(r,\lambda)'\left[\mathbf{I}-\boldsymbol{\Psi}(\lambda,\lambda_0)\right]\mathbf{D} \\
&= d(1)\left[\widehat{V}_{\overline{\alpha}}(r,\lambda) + \widetilde{\mathbf{F}}_2(r,\lambda)'\left[\mathbf{I}-\boldsymbol{\Psi}(\lambda,\lambda_0)\right]\boldsymbol{\delta}\right].
\end{aligned} \tag{A.3}$$

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

Then, under the local alternative, H_1, if $\{u_t\}$ is $I(1)$, it follows by the CMT and (A.3) that, as $T \to \infty$,

$$T^{-2}\sum_{t=1}^{T}\widehat{u}_t^2 = T^{-1}\sum_{t=1}^{T}\left[T^{-1/2}\widehat{u}_t\right]^2$$

$$\Rightarrow d(1)^2\int_0^1\left[\widehat{V}_{\overline{\alpha}}(r,\lambda)+\widetilde{\mathbf{F}}_2(r,\lambda)'\left[\mathbf{I}-\boldsymbol{\Psi}(\lambda,\lambda_0)\right]\boldsymbol{\delta}\right]^2 dr. \tag{A.4}$$

Similarly, under the alternative, H_1, if $\{u_t\}$ is $I(1)$, it follows by the CMT and Lemma A.3 that, as $T \to \infty$,

$$T^{-4}\sum_{t=1}^{T}\widehat{S}_t^2 = T^{-1}\sum_{t=1}^{T}\left[T^{-3/2}\widehat{S}_t\right]^2 \Rightarrow d(1)^2\int_0^1\left[Q_{\delta 1}(r,\lambda,\lambda_0,\boldsymbol{\delta})\right]^2 dr. \tag{A.5}$$

Then, under the alternative, H_1, if $\{u_t\}$ is $I(1)$, it follows by (A.4), (A.5), and the CMT that, as $T \to \infty$,

$$BG(T_b) = \frac{T^{-2}\sum_{t=1}^{T}\widehat{S}_t^2}{\sum_{t=1}^{T}\widehat{u}_t^2} = \frac{T^{-4}\sum_{t=1}^{T}\widehat{S}_t^2}{T^{-2}\sum_{t=1}^{T}\widehat{u}_t^2}$$

$$\Rightarrow \frac{\int_0^1\left[Q_{\delta 1}(r,\lambda,\lambda_0,\boldsymbol{\delta})\right]^2 dr}{\int_0^1\left[\widehat{V}_{\overline{\alpha}}(r,\lambda)+\widetilde{\mathbf{F}}_2(r,\lambda)'\left[\mathbf{I}-\boldsymbol{\Psi}(\lambda,\lambda_0)\right]\boldsymbol{\delta}\right]^2 dr}.$$

The asymptotic distribution of $J(T_b)$ is now obtained. The asymptotic distribution of RSS_R has already been established in (A.4). To establish the asymptotic distribution of RSS_U, let $\mathbf{g}(t) = \left[1\ t\ t^2\ t^3\ \ldots\ t^9\right]'$. Define $\left\{\breve{\mathbf{f}}_2(t,\lambda)\right\}$ and $\{\breve{y}_t\}$ to be the residuals from the OLS regressions of $\{\mathbf{f}_2(t,\lambda)\}$ and $\{y_t\}$ on $\{\mathbf{g}(t)\}$, respectively. Let $\breve{\boldsymbol{\tau}}_{1T}$ be the appropriate normalization matrix such that

$$\breve{\boldsymbol{\tau}}_{1T}\mathbf{g}(t) = \mathbf{g}(t/T)+o(1) \longrightarrow \mathbf{G}(r) = \left[1\ r\ r^2\ r^3\ \ldots\ r^9\right]'.$$

Then note that

$$\boldsymbol{\tau}_{2T}\breve{\mathbf{f}}_2(t,\lambda) = \boldsymbol{\tau}_{2T}\mathbf{f}_2(t,\lambda) - \left[\sum_{t=1}^{T}\boldsymbol{\tau}_{2T}\mathbf{f}_2(t,\lambda)\mathbf{g}(t)'\breve{\boldsymbol{\tau}}_{1T}\right]\left[\sum_{t=1}^{T}\breve{\boldsymbol{\tau}}_{1T}\mathbf{g}(t)\mathbf{g}(t)'\breve{\boldsymbol{\tau}}_{1T}\right]^{-1}$$

$$\times\breve{\boldsymbol{\tau}}_{1T}\mathbf{g}(t) \to \mathbf{F}_2(r,\lambda) - \left[\int_0^1\mathbf{F}_2(r,\lambda)\mathbf{G}(r)'dr\right]\left[\int_0^1\mathbf{G}(r)\mathbf{G}(r)'dr\right]^{-1}$$

$$\times\mathbf{G}(r) \equiv \breve{\mathbf{F}}_2(r,\lambda). \tag{A.6}$$

Let $\breve{\boldsymbol{\beta}}_2$ be the OLS estimate of $\boldsymbol{\beta}_2$ in (14). Then, by the Frisch and Waugh (1933) theorem,

$$\breve{\boldsymbol{\beta}}_2 = \left[\sum_{t=1}^{T}\breve{\mathbf{f}}_2(t,\lambda)\breve{\mathbf{f}}_2(t,\lambda)'\right]^{-1}\sum_{t=1}^{T}\breve{\mathbf{f}}_2(t,\lambda)\breve{y}_t$$

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

$$= \left[\sum_{t=1}^{T} \breve{\mathbf{f}}_2(t,\lambda)\breve{\mathbf{f}}_2(t,\lambda)'\right]^{-1} \sum_{t=1}^{T} \breve{\mathbf{f}}_2(t,\lambda)\left(\breve{\mathbf{f}}_2(t,\lambda_0)'\boldsymbol{\beta}_2 + u_t\right)$$

$$= \left[\sum_{t=1}^{T} \breve{\mathbf{f}}_2(t,\lambda)\breve{\mathbf{f}}_2(t,\lambda)'\right]^{-1} \left[\sum_{t=1}^{T} \breve{\mathbf{f}}_2(t,\lambda)\breve{\mathbf{f}}_2(t,\lambda_0)'\boldsymbol{\beta}_2 + \sum_{t=1}^{T} \breve{\mathbf{f}}_2(t,\lambda)u_t\right].$$

Under the alternative, H_1, if $\{u_t\}$ is $I(1)$, as $T \to \infty$,

$$T^{-1/2}\boldsymbol{\tau}_{2T}^{-1}\breve{\boldsymbol{\beta}}_2 = \left[T^{-1}\sum_{t=1}^{T} \boldsymbol{\tau}_{2T}\breve{\mathbf{f}}_2(t,\lambda)\breve{\mathbf{f}}_2(t,\lambda)'\boldsymbol{\tau}_{2T}\right]^{-1}$$

$$\times \left[T^{-1}\sum_{t=1}^{T} \boldsymbol{\tau}_{2T}\breve{\mathbf{f}}_2(t,\lambda)\breve{\mathbf{f}}_2(t,\lambda_0)'\boldsymbol{\tau}_{2T}\mathbf{D} + T^{-1}\sum_{t=1}^{T} \boldsymbol{\tau}_{2T}\breve{\mathbf{f}}_2(t,\lambda)T^{-1/2}u_t\right]$$

$$\Rightarrow d(1)\breve{\mathbf{P}}_1(\lambda) + \breve{\boldsymbol{\Psi}}(\lambda,\lambda_0)\mathbf{D}, \tag{A.7}$$

where

$$\breve{\mathbf{P}}_1(\lambda) = \left[\int_0^1 \breve{\mathbf{F}}_2(r,\lambda)\breve{\mathbf{F}}_2(r,\lambda)'dr\right]^{-1}\left[\int_0^1 \breve{\mathbf{F}}_2(r,\lambda)V_{\overline{\alpha}}(r)dr\right],$$

and $\breve{\mathfrak{g}}(\lambda,\lambda_0)$ is provided in Definition 3.

The residuals from the OLS estimation of (14) are

$$\breve{u}_t = \breve{y}_t - \breve{\mathbf{f}}_2(t,\lambda)'\breve{\boldsymbol{\beta}}_2 = \breve{\mathbf{f}}_2(t,\lambda_0)'\boldsymbol{\beta}_2 + u_t - \breve{\mathbf{f}}_2(t,\lambda)'\breve{\boldsymbol{\beta}}_2.$$

Under the local alternative, H_1, if $\{u_t\}$ is $I(1)$,

$$\breve{u}_t = \breve{\mathbf{f}}_2(t,\lambda_0)'T^{1/2}\boldsymbol{\tau}_{2T}\mathbf{D} + u_t - \breve{\mathbf{f}}_2(t,\lambda)'\breve{\boldsymbol{\beta}}_2. \tag{A.8}$$

Rewriting (A.8),

$$T^{-1/2}\breve{u}_t = \breve{\mathbf{f}}_2(t,\lambda_0)'\boldsymbol{\tau}_{2T}\mathbf{D} + T^{-1/2}u_t - \breve{\mathbf{f}}_2(t,\lambda)'\boldsymbol{\tau}_{2T}T^{-1/2}\boldsymbol{\tau}_{2T}^{-1}\breve{\boldsymbol{\beta}}_2.$$

Using (6), (A.6) and (A.7), it follows by the CMT that, as $T \to \infty$,

$$T^{-1/2}\breve{u}_{[rT]} \Rightarrow \breve{\mathbf{F}}_2(r,\lambda_0)'\mathbf{D} + d(1)V_{\overline{\alpha}}(r) - \breve{\mathbf{F}}_2(r,\lambda)'\left[d(1)\breve{\mathbf{P}}_1(\lambda) + \breve{\boldsymbol{\Psi}}(\lambda,\lambda_0)\mathbf{D}\right]$$

$$= d(1)\left[V_{\overline{\alpha}}(r) - \breve{\mathbf{F}}_2(r,\lambda)'\breve{\mathbf{P}}_1(\lambda)\right] + \breve{\mathbf{F}}_2(r,\lambda)'\left[\mathbf{I} - \breve{\boldsymbol{\Psi}}(\lambda,\lambda_0)\right]\mathbf{D}$$

$$= d(1)\breve{V}_{\overline{\alpha}}(r,\lambda) + \breve{\mathbf{F}}_2(r,\lambda)'\left[\mathbf{I} - \breve{\boldsymbol{\Psi}}(\lambda,\lambda_0)\right]\mathbf{D}$$

$$= d(1)\left[\breve{V}_{\overline{\alpha}}(r,\lambda) + \breve{\mathbf{F}}_2(r,\lambda)'\left[\mathbf{I} - \breve{\boldsymbol{\Psi}}(\lambda,\lambda_0)\right]\boldsymbol{\delta}\right]. \tag{A.9}$$

Then, under the alternative, H_1, if $\{u_t\}$ is $I(1)$, it follows by the CMT and (A.9) that, as $T \to \infty$,

$$T^{-2}\sum_{t=1}^{T} \breve{u}_t^2 = T^{-1}\sum_{t=1}^{T}\left[T^{-1/2}\breve{u}_t\right]^2$$

$$\Rightarrow d(1)^2\int_0^1 \left[\breve{V}_{\overline{\alpha}}(r,\lambda) + \breve{\mathbf{F}}_2(r,\lambda)'\left[\mathbf{I} - \breve{\boldsymbol{\Psi}}(\lambda,\lambda_0)\right]\boldsymbol{\delta}\right]^2 dr. \tag{A.10}$$

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press

Finally, under the alternative, H_1, if $\{u_t\}$ is $I(1)$, it follows by (A.4), (A.10), and the CMT that, as $T \to \infty$,

$$J(T_b) = \frac{T^{-2}RSS_R - T^{-2}RSS_U}{T^{-2}RSS_U} = \frac{T^{-2}\sum_{t=1}^{T}\widehat{u}_t^2}{T^{-2}\sum_{t=1}^{T}\breve{u}_t^2} - 1$$

$$\Rightarrow \frac{\int_0^1 \left[\widehat{V}_{\overline{\alpha}}(r,\lambda) + \widetilde{\mathbf{F}}_2(r,\lambda)'\left[\mathbf{I} - \boldsymbol{\Psi}(\lambda,\lambda_0)\right]\boldsymbol{\delta}\right]^2 dr}{\int_0^1 \left[\breve{V}_{\overline{\alpha}}(r,\lambda) + \breve{\mathbf{F}}_2(r,\lambda)'\left[\mathbf{I} - \breve{\boldsymbol{\Psi}}(\lambda,\lambda_0)\right]\boldsymbol{\delta}\right]^2 dr} - 1.$$

■

Proofs of Theorems 1 and 2. Theorem 1 follows from Theorem 2 when $\delta = 0$, and so only a proof of Theorem 2 is needed. The result in part (i) of Theorem 2 follows directly by the CMT using Theorem 1 of Bunzel and Vogelsang (2005) and Lemma A.3. To show part (ii), $W(T_b)$ is scaled depending on whether $\{u_t\}$ is $I(0)$ or $I(1)$. When $\{u_t\}$ is $I(0)$,

$$W(T_b) = \frac{\left(T^{1/2}\boldsymbol{\tau}_{2T}^{-1}\widehat{\boldsymbol{\beta}}_2\right)'\left[T^{-1}\sum_{t=1}^{T}\boldsymbol{\tau}_{2T}\widetilde{\mathbf{f}}_2(t,\lambda)\widetilde{\mathbf{f}}_2(t,\lambda)'\boldsymbol{\tau}_{2T}\right]\left(T^{1/2}\boldsymbol{\tau}_{2T}^{-1}\widehat{\boldsymbol{\beta}}_2\right)}{\widehat{\sigma}^2}, \quad \textbf{(A.11)}$$

and when $\{u_t\}$ is $I(1)$,

$$W(T_b) = \frac{\left(T^{-1/2}\boldsymbol{\tau}_{2T}^{-1}\widehat{\boldsymbol{\beta}}_2\right)'\left[T^{-1}\sum_{t=1}^{T}\boldsymbol{\tau}_{2T}\widetilde{\mathbf{f}}_2(t,\lambda)\widetilde{\mathbf{f}}_2(t,\lambda)'\boldsymbol{\tau}_{2T}\right]\left(T^{-1/2}\boldsymbol{\tau}_{2T}^{-1}\widehat{\boldsymbol{\beta}}_2\right)}{T^{-2}\widehat{\sigma}^2}. \quad \textbf{(A.12)}$$

Simple calculations give

$$T^{-1}\sum_{t=1}^{T}\boldsymbol{\tau}_{2T}\widetilde{\mathbf{f}}_2(t,\lambda_0)\widetilde{\mathbf{f}}_2(t,\lambda_0)'\boldsymbol{\tau}_{2T} \longrightarrow \int_0^1 \widetilde{\mathbf{F}}_2(r,\lambda_0)\widetilde{\mathbf{F}}_2(r,\lambda_0)'dr. \quad \textbf{(A.13)}$$

Using Lemma A.2, part (i) of Theorem 2, and (A.13), it follows by the CMT from (A.11) and (A.12) that, if $\{u_t\}$ is $I(0)$,

$$W(T_b) \Rightarrow \frac{\left[\mathbf{P}_0(\lambda) + \boldsymbol{\Psi}(\lambda,\lambda_0)\boldsymbol{\delta}\right]'\left[\int_0^1 \widetilde{\mathbf{F}}_2(r,\lambda)\widetilde{\mathbf{F}}_2(r,\lambda)'dr\right]\left[\mathbf{P}_0(\lambda) + \boldsymbol{\Psi}(\lambda,\lambda_0)\boldsymbol{\delta}\right]}{\Phi_{\delta 0}(b,k,\lambda,\lambda_0,\boldsymbol{\delta})} \quad \textbf{(A.14)}$$

and, if $\{u_t\}$ is $I(1)$,

$$W(T_b) \Rightarrow \frac{\left[\mathbf{P}_1(\lambda) + \boldsymbol{\Psi}(\lambda,\lambda_0)\boldsymbol{\delta}\right]'\left[\int_0^1 \widetilde{\mathbf{F}}_2(r,\lambda)\widetilde{\mathbf{F}}_2(r,\lambda)'dr\right]\left[\mathbf{P}_1(\lambda) + \boldsymbol{\Psi}(\lambda,\lambda_0)\boldsymbol{\delta}\right]}{\Phi_{\delta 1}(b,k,\lambda,\lambda_0,\boldsymbol{\delta})}. \quad \textbf{(A.15)}$$

Because under fixed-b asymptotics we write the statistics as continuous functions of partial sums of the data, the approach of Gregory and Hansen (1996) can be used to show that the convergence result in (A.15) holds uniformly in λ (see Gregory and Hansen, 1996, p. 118). Therefore the CMT can be used to establish part (iii) of Theorem 2 given (A.15). ■

https://doi.org/10.1017/S0266466610000617 Published online by Cambridge University Press