

Article

Fixed- b Inference for Testing Structural Change in a Time Series Regression

Cheol-Keun Cho ^{1,*} and Timothy J. Vogelsang ²¹ Korea Energy Economics Institute, Ulsan 44543, Korea² Michigan State University, East Lansing, MI 48824, USA; tjv@msu.edu* Correspondence: chkcho@gmail.com; Tel.: +82-52-714-2070

Academic Editor: Pierre Perron

Received: 19 August 2016; Accepted: 14 December 2016; Published: 30 December 2016

Abstract: This paper addresses tests for structural change in a weakly dependent time series regression. The cases of full structural change and partial structural change are considered. Heteroskedasticity-autocorrelation (HAC) robust Wald tests based on nonparametric covariance matrix estimators are explored. Fixed- b theory is developed for the HAC estimators which allows fixed- b approximations for the test statistics. For the case of the break date being known, the fixed- b limits of the statistics depend on the break fraction and the bandwidth tuning parameter as well as on the kernel. When the break date is unknown, supremum, mean and exponential Wald statistics are commonly used for testing the presence of the structural break. Fixed- b limits of these statistics are obtained and critical values are tabulated. A simulation study compares the finite sample properties of existing tests and proposed tests.

Keywords: HAC estimator; kernel; bandwidth; partial structural change; break point

JEL Classification: C10; C22

1. Introduction

This paper focuses on fixed- b inference of heteroskedasticity and autocorrelation (HAC) robust Wald statistics for testing for a structural break in a time series regression. We focus on kernel-based nonparametric HAC estimators which are commonly used to estimate the asymptotic variance. HAC estimators allow for arbitrary structure of the serial correlation and heteroskedasticity of weakly dependent time series and are consistent estimators of the long run variance under the assumption that the bandwidth (M) is growing at a certain rate slower than the sample size (T). Under consistency assumptions, the Wald statistics converge to the usual chi-square distributions. However, because the critical values from the chi-square distribution are based on a consistency approximation for the HAC estimator, the chi-square limit does not reflect the often substantial finite sample randomness of the HAC estimator. Furthermore, the chi-square approximation does not capture the impact of the choice of the kernel or the bandwidth on the Wald statistics. The sensitivity of the statistics to the finite sample bias and variability of the HAC estimator is well known in the literature; Kiefer and Vogelsang (2005) [1] among others have illustrated by simulation that the traditional inference with a HAC estimator can have poor finite sample properties.

Departing from the traditional approach, Kiefer and Vogelsang [1–3] obtain an alternative asymptotic approximation by assuming that the ratio of the bandwidth to the sample size, $b = M/T$, is held constant as the sample size increases. Under this alternative nesting of the bandwidth, they obtain pivotal asymptotic distributions for the test statistics which depend on the choice of kernel and bandwidth tuning parameter. Simulation results indicate that the resulting fixed- b approximation has

less size distortions in finite samples than the traditional approach, especially when the bandwidth is not small.

Theoretical explanations for the finite sample properties of the fixed- b approach include the studies by Hashimzade and Vogelsang (2008) [4], Jansson (2004) [5], Sun, Phillips and Jin (2008, hereafter SPJ) [6], Gonçalves and Vogelsang (2011) [7] and Sun (2013) [8]. Hashimzade and Vogelsang (2008) [4] provides an explanation for the better performance of the fixed- b asymptotics by analyzing the bias and variance of the HAC estimator. Gonçalves and Vogelsang (2011) [7] provides a theoretical treatment of the asymptotic equivalence between the naive bootstrap distribution and the fixed- b limit. Higher order theory is used by Jansson (2004) [5], SPJ (2008) [6] and Sun (2013) [8] to show that the error in rejection probability using the fixed- b approximation is more accurate than the traditional approximation. In a Gaussian location model, Jansson (2004) [5] proves that for the Bartlett kernel with bandwidth equal to sample size (i.e., $b = 1$), the error in rejection probability of fixed- b inference is $O(T^{-1} \log T)$ which is smaller than the usual rate of $O(T^{-1/2})$. The results in SPJ (2008) [6] complement Jansson's result by extending the analysis for a larger class of kernels and focusing on smaller values of bandwidth ratio b . In particular, they find that the error in rejection probability of the fixed- b approximation is $O(T^{-1})$ around $b = 0$. They also show that for positively autocorrelated series, which is typical for economic time series, the fixed- b approximation has smaller error than the chi-square or standard normal approximation, even when b is assumed to decrease to zero although the stochastic orders are same.

In this paper, fixed- b asymptotics is applied to testing for structural change in a weakly dependent time series regression. The structural change literature is now enormous and no attempt will be made here to summarize the relevant literature. Some key references include Andrews (1993) [9], Andrews and Ploberger (1994) [10], and Bai and Perron (1998) [11]. Andrews (1993) [9] treats the issue of testing for a structural break in the generalized method of moments framework when the one-time break date is unknown and Andrews and Ploberger (1994) [10] derive asymptotically optimal tests. Bai and Perron (1998) [11] considers multiple structural change occurring at unknown dates and covers the issues of estimation of break dates, testing for the presence of structural change and testing for the number of breaks. For a comprehensive survey of the recent structural break literature see Perron (2006) [12], Banerjee and Urga (2005) [13], and Aue and Horváth (2013) [14]. The fixed- b analysis can be extended to the case of multiple breaks but the simulation of critical values will be computationally intensive. Therefore, we leave the case of multiple breaks for future research and we consider the case of a single break in this paper.

For testing the presence of break, the robust version of the Wald statistic is considered in this paper and a HAC estimator is used to construct the test statistic. The ways of constructing HAC estimators in the context of structural change tests are well described in Bai and Perron (2003) [15] and Bai and Perron (1998) [11]. We focus mainly on the HAC estimator documented in Bai and Perron (2003) (Section 4.1, [15]) in which the usual "Newey-West-Andrews" approach is applied directly to the regression with regime dummies. Under the assumption of a fixed bandwidth ratio (fixed- b assumption), the asymptotic limit of the test statistic is a nonstandard distribution but it is pivotal. As in standard fixed- b theory, the impact of choice of bandwidth on the limiting distribution is substantial. In particular, the bandwidth interplays with the hypothesized break fraction so that the limit of the test statistic depends on both of them. For the unknown break date case, three existing test statistics (Sup-, Mean-, Exp-Wald) are considered and their fixed- b critical values are tabulated. The finite sample performance is examined by simulation experiments with comparisons made to existing tests. For practitioners, we include results using a data-dependent bandwidth rule based on Andrews (1991) [16]. This data-dependent bandwidth is calculated from the regression using the break fraction that yields the minimum sum of squared residuals (Bai and Perron, 1998 [11]). One can calculate a bandwidth ratio ($b^* = \frac{M^*}{T}$) with this data-dependent bandwidth (M^*) and proceed to apply the fixed- b critical values corresponding to this specific value of b^* .

The remainder of this paper is organized as follows. In Section 2, the basic setup of the full/partial structural-change model is presented and preliminary results are provided. Section 3 derives the fixed- b limit of the Wald statistic and the fixed- b critical values, for the case of unknown break dates, are tabulated in Section 4. Section 5 compares empirical null rejection probabilities and provides the size-adjusted power for tests based on the b^* data-dependent bandwidth ratio. Section 6 concludes. Proofs and definitions are collected in Appendix A.

2. Setup and Preliminary Results

Consider a weakly dependent time series regression model with a structural break given by

$$\begin{aligned} y_t &= w_t' \beta + u_t, \\ w_t' &= (x_{1t}', x_{2t}'), \quad \beta' = (\beta_1', \beta_2'), \\ x_{1t} &= x_t \cdot \mathbf{1}(t \leq [\lambda T]), \quad x_{2t} = x_t \cdot \mathbf{1}(t \geq [\lambda T] + 1), \end{aligned} \quad (1)$$

where x_t is $p \times 1$ regressor vector, $\lambda \in (0, 1)$ is a break point, and $\mathbf{1}(\cdot)$ is the indicator function. Define $v_t = x_t u_t$ and $w_t = w_t' u_t$. Recalling that $[x]$ denotes the integer part of a real number, x , notice that $x_{2t} = \mathbf{0}$ for $t = 1, 2, \dots, [\lambda T]$ and $x_{1t} = \mathbf{0}$ for $t = [\lambda T] + 1, \dots, T$. For the time being, the potential break point (fraction) λ is assumed to be known in order to develop the asymptotic theory for a test statistic and characterize its asymptotic limit. We will relax this assumption to deal with the empirically relevant case of an unknown break date. The regression model (1) implies that coefficients of all explanatory variables are subject to potential structural change and this model is labeled the ‘full’ structural change model.

We are interested in testing the presence of a structural change in the regression parameters. Consider the null hypothesis of the form

$$H_0 : \mathbf{R}\beta = \mathbf{0}, \quad (2)$$

where

$$\mathbf{R}_{(l \times 2p)} = (\mathbf{R}_1, -\mathbf{R}_1),$$

and \mathbf{R}_1 is an $l \times p$ matrix with $l \leq p$. Under the null hypothesis, we are testing that one or more linear relationships on the regression parameter(s) do not experience structural change before and after the break point. Tests of the null hypothesis of no structural change about a subset of the slope parameters are special cases. For example, we can test the null hypothesis that the slope parameter on the first regressor did not change by setting $\mathbf{R}_1 = (1, 0, \dots, 0)$. We can test the null hypothesis that none of the regression parameters have structural change by setting $\mathbf{R}_1 = I_p$. We focus on the OLS estimator of β given by $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2')' = \left(\sum_{t=1}^T w_t w_t' \right)^{-1} \left(\sum_{t=1}^T w_t y_t \right)$.

In order to establish the asymptotic limits of the HAC estimators and the Wald statistics, two assumptions are sufficient. These assumptions imply that there is no heterogeneity in the regressors across the segments and the covariance structure of the errors is assumed to be the same across segments as well.

Assumption 1. $T^{-1} \sum_{t=1}^{[rT]} x_t x_t' \xrightarrow{p} rQ$, uniformly in $r \in [0, 1]$, and Q^{-1} exists.

Assumption 2. $T^{-1/2} \sum_{t=1}^{[rT]} x_t u_t = T^{-1/2} \sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda W_p(r)$, $r \in [0, 1]$, where $\Lambda \Lambda' = \Sigma$, $W_p(r)$ is a $p \times 1$ standard Wiener process, and \Rightarrow denotes weak convergence.

For later use, we define a $l \times l$ nonsingular matrix A such that $\mathbf{R}_1 Q^{-1} \Lambda \Lambda' Q^{-1} \mathbf{R}_1' = A A'$ and $\mathbf{R}_1 Q^{-1} \Lambda W_p(r) \stackrel{d}{=} A W_l(r)$, where $W_l(r)$ is $l \times 1$ standard Wiener process. For a more detailed discussion about the regularity conditions under which Assumptions 1 and 2 hold, refer to Kiefer and Vogelsang (2002) [3] and see Davidson (1994) [17], Phillips and Durlauf (1986) [18], Phillips and Solo (1992) [19], and Wooldridge and White (1988) [20] for more details.

The matrix Q is the second moment matrix of x_t and is typically estimated using the quantity $\hat{Q} = \frac{1}{T} \sum_{t=1}^T x_t x_t'$. The matrix $\Sigma \equiv \Lambda \Lambda'$ is the asymptotic variance of $T^{-1/2} \sum_{t=1}^T v_t$, which is, for a covariance stationary series, given by

$$\Sigma = \Gamma_0 + \sum_{j=1}^{\infty} (\Gamma_j + \Gamma_j') \text{ with } \Gamma_j = E(v_t' v_{t-j}).$$

Consider the non-structural change regression equation where $\beta_1 = \beta_2$ and this coefficient parameter is estimated by OLS ($\hat{\beta}$). In this particular setup, the long run variance, Σ , is commonly estimated by the kernel-based nonparametric HAC estimator given by

$$\hat{\Sigma} = T^{-1} \sum_{t=1}^T \sum_{s=1}^T K\left(\frac{|t-s|}{M}\right) \hat{v}_t \hat{v}_s' = \hat{\Gamma}_0 + \sum_{j=1}^{T-1} K\left(\frac{j}{M}\right) (\hat{\Gamma}_j + \hat{\Gamma}_j'),$$

where $\hat{\Gamma}_j = T^{-1} \sum_{t=j+1}^T \hat{v}_t \hat{v}_{t-j}'$, $\hat{v}_t = x_t \hat{u}_t = x_t (y_t - x_t' \hat{\beta})$, M is a bandwidth, and $K(\cdot)$ is a kernel weighting function.

Under some regularity conditions (see Andrews (1991) [16], DeJong and Davidson (2000) [21], Hansen (1992) [22], Jansson (2002) [23] or Newey and West (1987) [24]), $\hat{\Sigma}$ is a consistent estimator of Σ , i.e., $\hat{\Sigma} \xrightarrow{p} \Sigma$. These regularity conditions include the necessary condition that $M/T \rightarrow 0$ as $M, T \rightarrow \infty$. This asymptotics is called “traditional” asymptotics throughout this paper.

In contrast to the traditional approach, fixed- b asymptotics assumes $M = bT$ where b is held constant as T increases. Assumptions 1 and 2 are the only regularity conditions required to obtain a fixed- b limit for $\hat{\Sigma}$. Under the fixed- b approach, for $b \in (0, 1]$, Kiefer and Vogelsang (2005) [1] show that

$$\hat{\Sigma} \Rightarrow \Lambda \mathbf{P}(b, \tilde{W}_p) \Lambda', \quad (3)$$

where $\tilde{W}_p(r) = W_p(r) - r W_p(1)$ is a p -vector of standard Brownian bridges and the form of the random matrix $\mathbf{P}(b, \tilde{W}_p)$ depends on the kernel. Following Kiefer and Vogelsang (2005) [1], we consider three classes of kernels which give three forms of \mathbf{P} . Let $H_p(r)$ denote a generic vector of stochastic processes. $H_p(r)'$ denotes its transpose. $\mathbf{P}(b, H_p)$ is defined in Appendix A.

Getting back to our structural change regression model, fixed- b results depend on the limiting behavior of the following partial sum process given by

$$\begin{aligned} \hat{S}_t &= \sum_{j=1}^t w_j \hat{u}_j = \sum_{j=1}^t w_j (y_j - x_{1j}' \hat{\beta}_1 - x_{2j}' \hat{\beta}_2) \\ &= \sum_{j=1}^t w_j (u_j - x_{1j}' (\hat{\beta}_1 - \beta_1) - x_{2j}' (\hat{\beta}_2 - \beta_2)). \end{aligned} \quad (4)$$

Under Assumptions 1 and 2, the limiting behavior of $\hat{\beta}$ and the partial sum process \hat{S}_t are given as follows.

Proposition 1. Let $\lambda \in (0, 1)$ be given. Suppose the data generation process is given by (1) and let $[rT]$ denote the integer part of rT where $r \in [0, 1]$. Then, under Assumptions 1 and 2 as $T \rightarrow \infty$,

$$\sqrt{T}(\hat{\beta} - \beta) = \begin{pmatrix} \sqrt{T}(\hat{\beta}_1 - \beta_1) \\ \sqrt{T}(\hat{\beta}_2 - \beta_2) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} (\lambda Q)^{-1} \Lambda W_p(\lambda) \\ ((1 - \lambda)Q)^{-1} \Lambda (W_p(1) - W_p(\lambda)) \end{pmatrix},$$

and

$$T^{-1/2} \hat{S}_{[rT]} \Rightarrow \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \Lambda \end{pmatrix} F_p(r, \lambda) \equiv \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \Lambda \end{pmatrix} \begin{pmatrix} F_p^{(1)}(r, \lambda) \\ F_p^{(2)}(r, \lambda) \end{pmatrix},$$

where

$$F_p^{(1)}(r, \lambda) = \left(W_p(r) - \frac{r}{\lambda} W_p(\lambda) \right) \cdot \mathbf{1}(r \leq \lambda),$$

and

$$F_p^{(2)}(r, \lambda) = \left(W_p(r) - W_p(\lambda) - \frac{r - \lambda}{1 - \lambda} (W_p(1) - W_p(\lambda)) \right) \cdot \mathbf{1}(r > \lambda).$$

See Appendix A for the proof.

It is easily seen that the asymptotic distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$ are Gaussian and are independent of each other. Hence the asymptotic covariance of $\hat{\beta}_1$ and $\hat{\beta}_2$ is zero. The asymptotic variance of $\sqrt{T}(\hat{\beta} - \beta)$ is given by $\mathbf{Q}_\lambda^{-1} \Omega \mathbf{Q}_\lambda^{-1}$, where

$$\mathbf{Q}_\lambda \equiv \begin{pmatrix} \lambda Q & \mathbf{0} \\ \mathbf{0} & (1 - \lambda)Q \end{pmatrix} \text{ and } \Omega \equiv \begin{pmatrix} \lambda \Sigma & \mathbf{0} \\ \mathbf{0} & (1 - \lambda)\Sigma \end{pmatrix}. \quad (5)$$

In order to test the null hypothesis (2), HAC robust Wald statistics are considered. These statistics are robust to heteroskedasticity and autocorrelation in the vector process, $v_t = x_t u_t$. The generic form of the robust Wald statistic is given by

$$Wald = T \left(\mathbf{R} \hat{\beta} \right)' \left(\mathbf{R} \hat{\mathbf{Q}}_\lambda^{-1} \hat{\Omega} \hat{\mathbf{Q}}_\lambda^{-1} \mathbf{R}' \right)^{-1} \left(\mathbf{R} \hat{\beta} \right), \quad (6)$$

where

$$\hat{\mathbf{Q}}_\lambda = \begin{pmatrix} T^{-1} \sum_{t=1}^{[\lambda T]} x_t x_t' & \mathbf{0} \\ \mathbf{0} & T^{-1} \sum_{t=[\lambda T]+1}^T x_t x_t' \end{pmatrix},$$

and $\hat{\Omega}$ is a HAC robust estimator of Ω .

We consider a particular way of constructing the HAC estimator. This estimator is the same one as in Bai and Perron (2003) [15]. Denoted by $\hat{\Omega}^{(F)}$, it is constructed using the residuals directly from the dummy regression (1):

$$\hat{\Omega}^{(F)} = T^{-1} \sum_{t=1}^T \sum_{s=1}^T K \left(\frac{|t-s|}{M} \right) \hat{v}_t \hat{v}_s', \quad (7)$$

where $\hat{v}_t = w_t \hat{u}_t = \begin{pmatrix} x_{1t}' \hat{u}_t & x_{2t}' \hat{u}_t \end{pmatrix}'_{2p \times 1}$. We denote the components of \hat{v}_t as $\hat{v}_t^{(1)} = x_{1t}' \hat{u}_t = x_t \hat{u}_t \mathbf{1}(t \leq [\lambda T])$ and $\hat{v}_t^{(2)} = x_{2t}' \hat{u}_t = x_t \hat{u}_t \mathbf{1}(t \geq [\lambda T] + 1)$. Notice that $\hat{\Omega}^{(F)}$ is the variance estimator one would be using if the usual “Newey-West-Andrews” approach is applied directly to the dummy regression (1).

Using $\hat{v}_t' = (\hat{v}_t^{(1)'}, \hat{v}_t^{(2)'})'$ we can write $\hat{\Omega}^{(F)}$ as

$$\begin{aligned}\hat{\Omega}^{(F)} &= \begin{pmatrix} \hat{\Omega}_{11}^{(F)} & \hat{\Omega}_{12}^{(F)} \\ \hat{\Omega}_{21}^{(F)} & \hat{\Omega}_{22}^{(F)} \end{pmatrix} \\ &= \begin{pmatrix} T^{-1} \sum_{t=1}^T \sum_{s=1}^T K\left(\frac{|t-s|}{M}\right) \hat{v}_t^{(1)} \hat{v}_s^{(1)'} & T^{-1} \sum_{t=1}^T \sum_{s=1}^T K\left(\frac{|t-s|}{M}\right) \hat{v}_t^{(1)} \hat{v}_s^{(2)'} \\ T^{-1} \sum_{t=1}^T \sum_{s=1}^T K\left(\frac{|t-s|}{M}\right) \hat{v}_t^{(2)} \hat{v}_s^{(1)'} & T^{-1} \sum_{t=1}^T \sum_{s=1}^T K\left(\frac{|t-s|}{M}\right) \hat{v}_t^{(2)} \hat{v}_s^{(2)'} \end{pmatrix} \\ &= \begin{pmatrix} T^{-1} \sum_{t=1}^{\lfloor \lambda T \rfloor} \sum_{s=1}^{\lfloor \lambda T \rfloor} K\left(\frac{|t-s|}{M}\right) \hat{v}_t^{(1)} \hat{v}_s^{(1)'} & T^{-1} \sum_{t=1}^{\lfloor \lambda T \rfloor} \sum_{s=\lfloor \lambda T \rfloor+1}^T K\left(\frac{|t-s|}{M}\right) \hat{v}_t^{(1)} \hat{v}_s^{(2)'} \\ T^{-1} \sum_{t=\lfloor \lambda T \rfloor+1}^T \sum_{s=1}^{\lfloor \lambda T \rfloor} K\left(\frac{|t-s|}{M}\right) \hat{v}_t^{(2)} \hat{v}_s^{(1)'} & T^{-1} \sum_{t=\lfloor \lambda T \rfloor+1}^T \sum_{s=\lfloor \lambda T \rfloor+1}^T K\left(\frac{|t-s|}{M}\right) \hat{v}_t^{(2)} \hat{v}_s^{(2)'} \end{pmatrix} \quad (8) \\ &= \begin{pmatrix} T^{-1} \lfloor \lambda T \rfloor \hat{\Sigma}^{(1)} & T^{-1} \sum_{t=1}^{\lfloor \lambda T \rfloor} \sum_{s=\lfloor \lambda T \rfloor+1}^T K\left(\frac{|t-s|}{M}\right) \hat{v}_t^{(1)} \hat{v}_s^{(2)'} \\ T^{-1} \sum_{t=\lfloor \lambda T \rfloor+1}^T \sum_{s=1}^{\lfloor \lambda T \rfloor} K\left(\frac{|t-s|}{M}\right) \hat{v}_t^{(2)} \hat{v}_s^{(1)'} & T^{-1} (T - \lfloor \lambda T \rfloor) \hat{\Sigma}^{(2)} \end{pmatrix}\end{aligned}$$

Three important observations are in order. First, the main component of the two diagonal blocks are within regime HAC estimators of Σ , the long run variance of $\{v_t\}$. However, one should see that the “effective” bandwidth ratio being applied to $\hat{\Sigma}^{(1)}$ is not b ($= \frac{M}{T}$) but $\frac{M}{\lambda T} = \frac{bT}{\lambda T} = \frac{b}{\lambda}$, which is bigger than b since $0 < \lambda < 1$. Similarly, the effective bandwidth ratio for $\hat{\Sigma}^{(2)}$ is $\frac{M}{(1-\lambda)T} = \frac{b}{1-\lambda}$. As documented in fixed- b literature (e.g., Kiefer and Vogelsang (2005) [1]), the bias in HAC estimators not accounted by traditional inference increases as the bandwidth ratio gets bigger. So, when the HAC estimator is constructed as in (8), traditional inference might be often exposed to size distortion—more than expected—due to this mechanism of determining effective bandwidths. The second issue is that the above estimator has non-zero off-diagonal blocks. So, the methodology based on partial samples such as in Andrews (1993) [9] does not exactly cover this case because the off-diagonal blocks in Andrews (1993) [9] are assumed to be zero, matching the zero asymptotic covariance of the OLS estimators of the slope coefficients between pre- and post-regimes. It is presumable that the influence of having non-zero off diagonal terms might be small since the off-diagonal blocks converge to zero under the traditional assumption $\frac{M}{T} \rightarrow 0$ as sample size grows (see a proof in Cho (2014) [25] for the Bartlett kernel) but it might still negatively affect the performance of tests in finite samples and we need to develop an alternative asymptotic theory to explicitly reflect the presence of these components. Third, there is another issue when a researcher uses a data-dependent bandwidth as in Andrews (1991) [16]. For a given hypothesized break fraction, a data-dependent bandwidth can be calculated based on the pooled series of $\{\hat{v}_t^{(1)}\}_{t=1}^{\lfloor \lambda T \rfloor}$ and $\{\hat{v}_t^{(2)}\}_{t=\lfloor \lambda T \rfloor+1}^T$. This method would result in an optimal bandwidth which minimizes the MSE in estimating Σ but the presence of non-zero off-diagonal terms are not taken into account in this procedure. Moreover, when the break date is treated as unknown, a sequence of data-dependent bandwidth across potential break dates will be generated. In this case, the fixed- b limits are not useful approximations because the sequence of the data-dependent bandwidth is random by nature so the limiting distributions of corresponding test statistics cannot be characterized by a single particular value of b .

Denote by $Wald^{(F)}(T_b)$, the Wald statistic given by (6) using the break date T_b with $\hat{\Omega}^{(F)}$ used for $\hat{\Omega}$. Tests for a potential structural break with an unknown break date are well studied in Andrews (1993) [9], Andrews and Ploberger (1994) [10], and Bai and Perron (1998) [11]. Andrews (1993) [9] considers several tests based on the supremum across breakpoints of Wald and Lagrange multiplier statistics and shows that they are asymptotically equivalent. Andrews and Ploberger (1994) [10] derives tests that maximize average power across potential breakpoints.

As argued by Andrews (1993) [9] and Andrews and Ploberger (1994) [10], break dates close to the end points of the sample cannot be used and so some trimming is needed. To that end, define $\Xi^* = [\epsilon T, T - \epsilon T]$ with $0 < \epsilon < 1$ to be the set of admissible break dates. The tuning parameter, ϵ ,

denotes the amount of trimming of potential break dates. We consider the three statistics following Andrews (1993) [9]¹ and Andrews and Ploberger (1994) [10]² defined as

$$SupW^{(F)} \equiv \sup_{T_b \in \Xi^*} Wald^{(F)}(T_b), \quad (9)$$

$$MeanW^{(F)} \equiv \frac{1}{T} \sum_{T_b \in \Xi^*} Wald^{(F)}(T_b), \quad (10)$$

$$ExpW^{(F)} \equiv \log \left(\frac{1}{T} \sum_{T_b \in \Xi^*} \exp \left[\frac{1}{2} Wald^{(F)}(T_b) \right] \right). \quad (11)$$

The next section provides asymptotic results for the robust Wald statistics under the fixed- b asymptotics.

3. Asymptotic Results

3.1. Asymptotic Results under the Fixed- b Approach

We now provide fixed- b limits for the HAC estimators and the test statistics in the full structural change model (1). The fixed- b limits presented in the next Lemma and Corollary approximate the diagonal blocks of $\widehat{\Omega}^{(F)}$ by random matrices. Also, it is shown that the fixed- b approach gives a non-zero limit for the off-diagonal blocks, which further distinguishes fixed- b asymptotics from traditional asymptotics.

Lemma 1. Let $b \in (0, 1]$ be given and suppose $M = bT$. Then under Assumptions 1 and 2, as $T \rightarrow \infty$,

$$\widehat{\Omega}^{(F)} \Rightarrow \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \Lambda \end{pmatrix} \times \mathbf{P}(b, F_p(r, \lambda)) \times \begin{pmatrix} \Lambda' & \mathbf{0} \\ \mathbf{0} & \Lambda' \end{pmatrix}, \quad (12)$$

where

$$F_p(r, \lambda) = \begin{pmatrix} F_p^{(1)}(r, \lambda) \\ F_p^{(2)}(r, \lambda) \end{pmatrix}, \quad (13)$$

$$F_p^{(1)}(r, \lambda) = \left(W_p(r) - \frac{r}{\lambda} W_p(\lambda) \right) \mathbf{1}(0 \leq r \leq \lambda), \quad (14)$$

$$F_p^{(2)}(r, \lambda) = \left(W_p(r) - W_p(\lambda) - \frac{r - \lambda}{1 - \lambda} (W_p(1) - W_p(\lambda)) \right) \mathbf{1}(\lambda < r \leq 1), \quad (15)$$

and $\mathbf{P}(b, F_p(r, \lambda))$ is defined by (A1)–(A3) with $H_p(r) = F_p(r, \lambda)$.

See Appendix A for the proof.

Next, Corollary presents alternative representations for $\mathbf{P}(b, F_p(r, \lambda))$ for three classes of kernels. The definitions of these classes of kernels (Classes 1, 2 and 3) are given in Appendix A. Three popular kernels—the Quadratic Spectral, Bartlett and Parzen kernels—belong to Classes 1, 2 and 3, respectively. See Cho (2014) [25] for the proof of this Corollary.

¹ We used the critical values provided in Andrews (2003) [26] for traditional inference.

² The definitions for the mean and exponential statistics are slightly different in the divisor of the summation. For traditional inference, we adjusted the critical values in Andrews and Ploberger (1994) [10] to our definitions of the statistics.

Corollary 1.

$$\mathbf{P}(b, F_p(r, \lambda)) = \begin{pmatrix} \mathbf{P}(b, F_p^{(1)}(r, \lambda)) & \mathbf{C}(b, F_p^{(1)}(r, \lambda), F_p^{(2)}(r, \lambda)) \\ \mathbf{C}(b, F_p^{(1)}(r, \lambda), F_p^{(2)}(r, \lambda))' & \mathbf{P}(b, F_p^{(2)}(r, \lambda)) \end{pmatrix}, \quad (16)$$

where

$$\mathbf{C}(b, F_p^{(1)}(r, \lambda), F_p^{(2)}(r, \lambda)) = \begin{cases} - \int_0^1 \int_0^1 \frac{1}{b^2} K''\left(\frac{|r-s|}{b}\right) F_p^{(1)}(r, \lambda) F_p^{(2)}(s, \lambda)' dr ds, \\ \frac{1}{b} \int_0^{1-b} F_p^{(1)}(r, \lambda) F_p^{(2)}(r+b, \lambda)' dr, \\ - \int \int_{|r-s| < b} \frac{1}{b^2} K''\left(\frac{|r-s|}{b}\right) F_p^{(1)}(r, \lambda) F_p^{(2)}(s, \lambda)' dr ds + \frac{K'(1)}{b} \int_0^{1-b} F_p^{(1)}(r, \lambda) F_p^{(2)}(r+b, \lambda)' dr, \end{cases}$$

for Classes 1, 2 and 3 kernels respectively.

The expression for $\mathbf{P}(b, F_p(r, \lambda))$ in this Corollary makes it easier to compare the fixed- b limit of $\hat{\Omega}^{(F)}$ with the standard fixed- b limit (see (3)) appearing in a non-structural change setting. Since each diagonal block of $\hat{\Omega}^{(F)}$ is basically a HAC estimator (up to a scale factor; see (8)) based on one of the pre- or post- break data, its limit should take the same form as (3), which is verified in this Corollary. So, each diagonal component of $\mathbf{P}(b, F_p(r, \lambda))$ serves to reflect the randomness and bandwidth/kernel-dependence of the associated HAC estimator. Second, unlike the traditional approach, the fixed- b limit of the off-diagonal component is non-zero. This implies that the fixed- b approach is able to take account of the covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$ which is generally non-zero in finite samples. The limits of the Wald statistics can be derived by using Lemma 1 and the result is presented in the next Theorem.

Theorem 1. Let $b \in (0, 1]$ be given. Suppose $M = bT$. Then under Assumptions 1 and 2, as $T \rightarrow \infty$,

$$\text{Wald}^{(F)} \Rightarrow \left(\frac{1}{\lambda} W_l(\lambda) - \frac{1}{1-\lambda} (W_l(1) - W_l(\lambda)) \right)' \times \left(\mathbf{P} \left(b, \frac{1}{\lambda} F_l^{(1)}(r, \lambda) - \frac{1}{1-\lambda} F_l^{(2)}(r, \lambda) \right) \right)^{-1} \times \left(\frac{1}{\lambda} W_l(\lambda) - \frac{1}{1-\lambda} (W_l(1) - W_l(\lambda)) \right) \quad (17)$$

See Appendix A for the proof.

The next Corollary provides an alternative representation for the limit given in (17). The proof for this Corollary is given in Cho (2014) [25].

Corollary 2. For a given value of $\lambda \in (0, 1)$, the fixed- b limit of $\text{Wald}^{(F)}$ has the same distribution as

$$\frac{1}{\lambda(1-\lambda)} W_l(1)' \left(\frac{1}{\lambda} \mathbf{P} \left(\frac{b}{\lambda}, \tilde{W}_l(r) \right) + \frac{1}{1-\lambda} \mathbf{P} \left(\frac{b}{1-\lambda}, \tilde{W}_l^*(r) \right) + \mathbf{CP}(\lambda, b) + \mathbf{CP}(\lambda, b)' \right)^{-1} W_l(1), \quad (18)$$

where

$$\mathbf{CP}(\lambda, b) = \begin{cases} \frac{\sqrt{\lambda}\sqrt{1-\lambda} \int_0^1 \int_0^1 K''\left(\frac{|\lambda t - (1-\lambda)s - \lambda|}{b}\right) \tilde{W}_l(t) \tilde{W}_l^*(s)' dt ds}{b^2} & \text{for Class-1 kernels,} \\ \frac{\int_0^{1-b} \tilde{W}_l\left(\frac{r}{\lambda}\right) \tilde{W}_l^*\left(\frac{r+b-\lambda}{1-\lambda}\right)' 1(\lambda-b < r \leq \lambda) dr}{b\sqrt{\lambda}\sqrt{1-\lambda}} & \text{for Class-2 kernels,} \\ \frac{\sqrt{\lambda}\sqrt{1-\lambda} \int_0^1 \int_0^1 K''\left(\frac{|\lambda t - (1-\lambda)s - \lambda|}{b}\right) \tilde{W}_l(t) \tilde{W}_l^*(s)' 1(|\lambda t - (1-\lambda)s - \lambda| < b) dt ds}{b^2} \\ - \frac{\int_0^{1-b} K'(1) \tilde{W}_l\left(\frac{r}{\lambda}\right) \tilde{W}_l^*\left(\frac{r+b-\lambda}{1-\lambda}\right)' 1(\lambda-b < r \leq \lambda) dr}{b\sqrt{\lambda}\sqrt{1-\lambda}} & \text{for Class-3 kernels,} \end{cases}$$

and $\tilde{W}_l(r)$ and $\tilde{W}_l^*(r)$ are $l \times 1$ Brownian Bridge processes which are independent of each other and of $W_l(1)$.

The limit in (18) shows how the components of $\hat{\Omega}^{(F)}$ affect the distribution of $Wald^{(F)}$. As mentioned earlier, the random matrix $\mathbf{P}\left(\frac{b}{\lambda}, \tilde{W}_l(r)\right)$ reflects the random nature of $\hat{\Omega}_{11}^{(F)}$ which is part of the estimator of the asymptotic variance of $\hat{\beta}_1$. Notice that the effective bandwidth for $\hat{\Omega}_{11}^{(F)}$ turns out to be $\frac{b}{\lambda}$ not b . Thus, we implicitly use the bandwidth ratio $\frac{b}{\lambda}$ for $\hat{\Omega}_{11}^{(F)}$ when we use a full sample bandwidth ratio b for constructing $\hat{\Omega}^{(F)}$. The second component, $\mathbf{P}\left(\frac{b}{1-\lambda}, \tilde{W}_l^*(r)\right)$, is related to $\hat{\Omega}_{22}^{(F)}$ (and $\hat{\beta}_2$) in exactly the same fashion. Finally, the third component, $\mathbf{CP}(\lambda, b)$, captures the impact of finite sample covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$ on structural change inference.

Now consider the unknown break date case and let $Wald_{\infty}^{(F)}(\lambda)$ denote the limit of $Wald^{(F)}(T_b)$, where the form of $Wald_{\infty}^{(F)}(\lambda)$ depends on whether traditional or fixed- b asymptotic theory is being used. In the case of fixed- b theory, $Wald_{\infty}^{(F)}(\lambda)$ is given in (17). Under the traditional assumption that the bandwidth ratio goes to zero as T grows,

$$Wald^{(F)} \Rightarrow \lambda(1-\lambda) \left(\frac{1}{\lambda} W_l(\lambda) - \frac{1}{1-\lambda} (W_l(1) - W_l(\lambda)) \right)' \\ \times \left(\frac{1}{\lambda} W_l(\lambda) - \frac{1}{1-\lambda} (W_l(1) - W_l(\lambda)) \right)$$

The asymptotic limits of Sup-, Mean-, and Exp-Wald statistics immediately follow from the continuous mapping theorem given by

$$SupW^{(F)} \xrightarrow{d} \sup_{\lambda \in (\epsilon, 1-\epsilon)} Wald_{\infty}^{(F)}(\lambda), \\ MeanW^{(F)} \xrightarrow{d} \int_{\epsilon}^{1-\epsilon} Wald_{\infty}^{(F)}(\lambda) d\lambda, \\ ExpW^{(F)} \xrightarrow{d} \log \left(\int_{\epsilon}^{1-\epsilon} \exp \left[\frac{1}{2} Wald_{\infty}^{(F)}(\lambda) \right] d\lambda \right).$$

3.2. Extension to the Partial Structural Change Model

This section derives the fixed- b limit of $Wald^{(F)}$ in the partial structural change model. The main result of this section is that the limit is the same as the limit for the full structural change model. The regression model with partial structural change is given by

$$y_t = z_t' \alpha + x_{1t}' \beta_1 + x_{2t}' \beta_2 + u_t \\ = z_t' \alpha + X_t' \beta + u_t, \quad (19)$$

where x_t is $p \times 1$ and z_t is $q \times 1$ vector and

$$x_{1t} = x_t 1(t \leq [\lambda T]), \quad x_{2t} = x_t 1(t \geq [\lambda T] + 1), \\ X_t' = (x_{1t}' \ x_{2t}'), \text{ and } \beta' = (\beta_1' \ \beta_2').$$

The coefficients on the x_t regressors are unrestricted in terms of a structural change whereas the coefficients on the z_t regressors are assumed to not have structural change. Denote

$$y = (y_1, y_2, \dots, y_T)', \quad X = (X_1, X_2, \dots, X_T)', \\ Z = (z_1, z_2, \dots, z_T)', \quad u = (u_1, u_2, \dots, u_T)'.$$

The parameters (α, β) are estimated by OLS and the OLS residual vector can be written as

$$\hat{u} = \tilde{y} - \tilde{X} \hat{\beta} = u - \tilde{X} (\hat{\beta} - \beta) - P_Z u,$$

where

$$\tilde{y} = (I - P_Z) y, \quad \tilde{X} = (I - P_Z) X, \quad \text{and} \quad P_Z = Z(Z'Z)^{-1}Z'.$$

The residual for an individual observation is given by

$$\hat{u}_t = u_t - \tilde{X}_t'(\hat{\beta} - \beta) - z_t'(Z'Z)^{-1}Z'u. \quad (20)$$

Also, note that

$$\tilde{X}_t = X_t - X'Z(Z'Z)^{-1}z_t = \begin{pmatrix} \tilde{X}_t^{(1)} \\ p \times 1 \\ \tilde{X}_t^{(2)} \\ p \times 1 \end{pmatrix}.$$

The following assumptions replace Assumptions 1 and 2:

Assumption 3. $T^{-1/2} \sum_{t=1}^{[rT]} \begin{pmatrix} x_t u_t \\ z_t u_t \end{pmatrix} \Rightarrow \Lambda W_{p+q}(r) \equiv \begin{pmatrix} \Lambda_1 \\ \Lambda_2 \end{pmatrix} W_{p+q}(r)$, where Λ_1 is a $p \times (p+q)$ matrix, Λ_2 is a $q \times (p+q)$ matrix, and $W_{p+q}(r)$ is a $(p+q) \times 1$ vector of independent Wiener process.

Assumption 4. $p \lim \frac{1}{T} \sum_{t=1}^{[rT]} z_t z_t' = r Q_{ZZ}$, $p \lim \frac{1}{T} \sum_{t=1}^{[rT]} x_t x_t' = r Q_{xx}$, and $p \lim \frac{1}{T} \sum_{t=1}^{[rT]} x_t z_t' = r Q_{xz}$ uniformly in $r \in [0, 1]$, and Q_{ZZ}^{-1} and Q_{xx}^{-1} exist.

We continue to focus on tests of the null hypothesis of no structural change in the x_t slope parameters of the form

$$H_0 : R\beta = r$$

with

$$R = \begin{pmatrix} R_1 & -R_1 \\ l \times p & l \times p \end{pmatrix} \quad \text{and} \quad r = 0. \quad (21)$$

Recall that the OLS estimator, $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2')'$ can be rewritten as

$$\hat{\beta} = \left(\sum_{t=1}^T \tilde{X}_t \tilde{X}_t' \right)^{-1} \left(\sum_{t=1}^T \tilde{X}_t \tilde{y}_t \right). \quad (22)$$

Proposition 2. Under Assumptions 3 and 4, as $T \rightarrow \infty$

$$T^{1/2} (\hat{\beta} - \beta) \xrightarrow{d} Q_{\tilde{X}\tilde{X}}^{-1} \begin{pmatrix} \Lambda_1 W_{p+q}(\lambda) - \lambda Q_{xz} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1) \\ \Lambda_1 (W_{p+q}(1) - W_{p+q}(\lambda)) - (1 - \lambda) Q_{xz} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1) \end{pmatrix},$$

and

$$\sqrt{T} (R\hat{\beta} - r) \Rightarrow R_1 Q_{xx}^{-1} \Lambda_1 \left(\frac{1}{\lambda} W_{p+q}(\lambda) - \frac{1}{1-\lambda} (W_{p+q}(1) - W_{p+q}(\lambda)) \right), \quad (23)$$

where $Q_{\tilde{X}\tilde{X}} = p \lim \left(T^{-1} \sum_{t=1}^T \tilde{X}_t \tilde{X}_t' \right)$.

See Appendix A for the proof.

As seen from the above proposition, $\hat{\beta}_1$ and $\hat{\beta}_2$ are not asymptotically independent in the partial structural change regression model. This is true because we are projecting out the variation of explanatory variables z_t so that $\hat{\beta}_1$ and $\hat{\beta}_2$ depend on the entire series of x_t and z_t . The dichotomy that $\hat{\beta}_1$ is dependent only on the pre-break data and that $\hat{\beta}_2$ depends only on the post-break data no longer holds in the partial structural change model. The dependence manifests in the common term, $Q_{xz} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1)$, in Proposition 2. However, this term cancels out in (23) when the restriction

matrix takes the form of (21). As a result, and also as suggested by Equation (23), in principle we need to estimate only $\Lambda_1 \Lambda_1'$ for testing for partial structural change. Because $\hat{\Omega}^{(F)}$, extended for the case of partial structural change, does not impose any restrictions on the asymptotic correlation between $\hat{\beta}_1$ and $\hat{\beta}_2$, $Wald^{(F)}$ continues to allow asymptotically pivotal fixed- b tests for partial structural change. While not obvious at first glance, $Wald^{(F)}$ has the same fixed- b limit in the partial structural change case as it does in the full structural change case.

The Wald statistic for testing for partial structural change is given by

$$Wald = T \left(\mathbf{R} \hat{\beta} \right)' \left(\mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} \hat{\Omega} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} \mathbf{R}' \right)^{-1} \left(\mathbf{R} \hat{\beta} \right), \quad (24)$$

where $\hat{Q}_{\tilde{X}\tilde{X}} = T^{-1} \sum_{t=1}^T \tilde{X}_t \tilde{X}_t'$. For constructing $Wald^{(F)}$, we use the HAC estimator $\hat{\Omega}^{(F)}$ which is computed using $\left\{ \tilde{X}_t \hat{u}_t \right\}_{t=1}^T$:

$$\hat{\Omega}^{(F)} = T^{-1} \sum_{t=1}^T \sum_{s=1}^T K \left(\frac{|t-s|}{M} \right) \hat{\xi}_t \hat{\xi}_s', \quad (25)$$

where $\hat{\xi}_t = \tilde{X}_t \hat{u}_t$. By the Frisch-Waugh-Lovell Theorem, this is the straightforward extension of $Wald^{(F)}$ to the case of partial structural change.

The next Lemma provides the limit of the scaled partial sum process of $\hat{\xi}_t$ premultiplied by an appropriate term.

Lemma 2. Let $\hat{S}_t^{\tilde{\xi}} = \sum_{j=1}^t \hat{\xi}_j$. Under Assumptions 3 and 4, as $T \rightarrow \infty$,

$$\mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} T^{-1/2} \hat{S}_{[rT]}^{\tilde{\xi}} \Rightarrow \mathbf{R}_1 Q_{xx}^{-1} \Lambda_1 \left(\frac{1}{\lambda} F_{p+q}^{(1)}(r, \lambda) - \frac{1}{1-\lambda} F_{p+q}^{(2)}(r, \lambda) \right),$$

where

$$F_{p+q}^{(1)}(r, \lambda) = \left(W_{p+q}(r) - \frac{r}{\lambda} W_{p+q}(\lambda) \right) \mathbf{1}(0 \leq r \leq \lambda),$$

$$F_{p+q}^{(2)}(r, \lambda) = \left(W_{p+q}(r) - W_{p+q}(\lambda) - \frac{r-\lambda}{1-\lambda} (W_{p+q}(1) - W_{p+q}(\lambda)) \right) \mathbf{1}(\lambda < r \leq 1).$$

See Appendix A for the proof.

As Lemma 2 shows, the partial sums of the inputs to $\hat{\Omega}^{(F)}$ are asymptotically proportional to the same nuisance parameters as $\sqrt{T}(\mathbf{R}\hat{\beta} - \mathbf{r})$. This is the key condition for a pivotal fixed- b limit. The next Theorem provides the fixed- b limit of $Wald^{(F)}$.

Theorem 2. Let $b \in (0, 1]$ be given. Suppose $M = bT$. Then, under Assumptions 3 and 4, $Wald^{(F)}$ weakly converges to the same limit in (17), i.e., as $T \rightarrow \infty$,

$$Wald^{(F)} \Rightarrow \left(\frac{1}{\lambda} W_l(\lambda) - \frac{1}{1-\lambda} (W_l(1) - W_l(\lambda)) \right)'$$

$$\times \left(\mathbf{P} \left(b, \frac{1}{\lambda} F_l^{(1)}(r, \lambda) - \frac{1}{1-\lambda} F_l^{(2)}(r, \lambda) \right) \right)^{-1} \times \left(\frac{1}{\lambda} W_l(\lambda) - \frac{1}{1-\lambda} (W_l(1) - W_l(\lambda)) \right).$$

See Appendix A for the proof.

According to Theorem 2, the limit of $Wald^{(F)}$ in the partial structural change model is the same as in the full structural change model.

4. Critical Values

While the fixed- b limiting distributions are nonstandard, asymptotic critical values are easily obtained via simulations. We approximate the Wiener processes in the limiting distributions using scaled partial sums of 1000 i.i.d. $N(0, 1)$ random variables. Critical values are tabulated based on 50,000 replications³.

In Table 1, fixed- b critical values for $SupW^{(F)}$, $MeanW^{(F)}$, and $ExpW^{(F)}$ are provided for $l = 2$, $\epsilon = 0.05, 0.1, 0.2$ and for $b \in \{0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.3, \dots, 0.9, 1\}$. Critical values over the entire grid of 0.02-increment of b are available upon request.

Table 1. (a) Fixed- b 95% Critical Values of $Sup-/Mean-/Exp-W^{(F)}$, Bartlett kernel, $l = 2$; (b) Fixed- b 95% Critical Values of $Sup-/Mean-/Exp-W^{(F)}$, QS kernel, $l = 2$.

(a)									
b	$\epsilon = 0.05$			$\epsilon = 0.1$			$\epsilon = 0.2$		
	$SupW$	$MeanW$	$ExpW$	$SupW$	$MeanW$	$ExpW$	$SupW$	$MeanW$	$ExpW$
0.02	30.293	4.861	9.588	18.230	4.235	5.051	13.542	3.263	3.539
0.04	48.447	5.9489	18.194	26.034	4.974	8.173	16.313	3.688	4.654
0.06	61.976	7.0183	24.816	33.172	5.729	11.483	19.496	4.162	5.967
0.08	73.862	8.001	30.656	39.957	6.496	14.695	22.812	4.617	7.364
0.1	84.848	8.973	36.109	46.263	7.278	17.653	26.323	5.146	8.998
0.2	138.92	14.018	63.068	76.971	11.323	32.706	46.122	8.052	18.156
0.3	193.94	19.113	90.408	109.11	15.596	48.657	67.262	11.216	28.446
0.4	254.14	24.443	120.71	142.31	20.009	65.120	89.241	14.464	39.161
0.5	313.06	29.999	149.85	176.51	24.565	82.037	111.18	17.912	49.818
0.6	374.36	35.304	180.46	212.05	29.202	99.596	134.00	21.386	61.205
0.7	433.71	40.902	210.22	245.66	33.625	116.32	153.93	24.666	70.991
0.8	491.83	46.205	239.08	279.65	38.016	133.32	173.96	27.702	81.134
0.9	549.63	51.450	268.05	311.37	42.238	149.22	192.52	30.670	90.145
1	608.99	57.142	297.78	344.26	46.623	165.51	212.76	33.936	100.36

(b)									
b	$\epsilon = 0.05$			$\epsilon = 0.1$			$\epsilon = 0.2$		
	$SupW$	$MeanW$	$ExpW$	$SupW$	$MeanW$	$ExpW$	$SupW$	$MeanW$	$ExpW$
0.02	64.848	5.678	26.200	24.831	4.641	7.548	15.051	3.458	4.111
0.04	122.00	8.102	54.483	46.350	6.059	17.433	20.670	4.205	6.401
0.06	161.74	10.617	74.329	68.158	7.630	28.148	28.305	5.060	9.666
0.08	207.65	13.202	97.163	91.258	9.461	39.595	38.905	6.143	14.409
0.1	257.31	16.139	122.02	118.67	11.671	53.066	52.759	7.491	20.987
0.2	832.93	40.501	409.56	452.33	30.155	219.29	240.65	19.924	113.55
0.3	3339.8	99.975	1663.0	2055.3	77.012	1020.8	1144.7	51.677	565.45
0.4	13,932	239.82	6959.4	8975.9	185.18	4481.1	4771.4	124.22	2378.8
0.5	47,253	537.89	23,620	31,752	411.53	15,869	16,684	276.98	8334.9
0.6	136,211	1115.4	68,099	91,828	850.69	45,907	49,492	580.43	24,740
0.7	328,737	2170.5	164,361	224,463	1674.7	112,225	128,234	1140.0	64,110
0.8	719,812	3982.4	359,899	488,008	3100.4	243,997	283,267	2099.3	141,627
0.9	1,444,833	7015.5	722,409	970,172	5395.5	485,079	565,285	3626.6	282,635
1	2,647,520	11566	1,323,754	1,829,406	9072.3	914,696	1,062,685	5951.4	531,336

³ For the case of a known break date, the 95% critical values for $l = 2$ are available for selected values of b and λ in Cho and Vogelsang (2014) [27]. The critical values display two main patterns. First, for each given λ the critical values increase as the bandwidth gets bigger. This can be expected given the well known downward bias induced into HAC estimators from estimation error. The fixed- b approximation captures this downward bias and reflects it through larger critical values. Second, for a given value of the bandwidth, the critical values display a V-shaped pattern as a function of λ . As the break point moves closer to zero or one, the critical values increase and the minimum critical values occur at $\lambda = 0.5$.

5. Finite Sample Properties

In this section, we report the results of a finite sample simulation study that illustrates the performance of fixed- b critical values relative to traditional critical values. The data generating process (DGP) is given by (1) with $x'_t = [1, q_t]$ where q_t is a scalar time series, $\beta'_1 = [\beta_1^c, \beta_1^s]$, and $\beta'_2 = [\beta_2^c, \beta_2^s]$. We use the break point $\lambda = 0.4$. The regressor q_t and the regression error u_t are generated as $q_t = \theta q_{t-1} + \epsilon_t$ and $u_t = \rho u_{t-1} + \eta_t + \varphi \eta_{t-1}$, where ϵ_t and η_t are independent of each other with $\epsilon_t, \eta_t \sim \text{i.i.d. } N(0, 1)$. We use the parameter values: $\theta \in \{0.5, 0.8, 0.9\}$, and $(\rho, \varphi) \in \{(0, 0), (0.5, 0.5), (0.9, 0.9)\}$ (see Table 2):

Table 2. Parameter values for simulations

DGP	θ	ρ	φ	q_t	u_t	$v_{1t} = q_t u_t$
A	0.5	0	0	AR(1)	IID	White Noise
B	0.8	0.5	0.5	AR(1)	ARMA(1,1)	Serially Correlated
C	0.9	0.9	0.9	AR(1)	ARMA(1,1)	Serially Correlated

The value of θ measures the persistence of the time varying regressor q_t . The parameters ρ and φ jointly determine the serial correlation structure of the error term u_t . Bigger values of these three parameters lead to higher persistence of the series $v_{1t} \equiv q_t u_t$ except for specification A where bigger values of θ would not increase persistence in v_{1t} . We set $\beta_1^c = 0$, $\beta_1^s = 0$ and $\beta_2^c = \delta$, $\beta_2^s = \delta$. Under the null hypothesis of no structural change, $\delta = 0$, whereas for $\delta \neq 0$ there is structural change in both the intercept and slope parameters. We report results for sample sizes $T = 100, 200, 500$, and 1000 and the number of replications is 2500. The nominal level of all tests is 5%. We compute the *Sup/ Mean/ Exp- $W^{(F)}$* statistics for testing the joint null hypothesis of no structural change in both the intercept and slope parameters. The frequency of rejections for the case of $\delta = 0$ measures the empirical type-I error.⁴

We report empirical rejection frequencies for traditional inference and for fixed- b inference. In traditional inference, we select the bandwidth following Andrews (1991) [16] for each hypothesized break date using the AR(1) plug-in formula. For fixed- b inference, we report results for different values of b to show how the null rejection probability varies with the choice of b . We also give results for another test in which a single data-dependent bandwidth ratio, denoted by b^* , is used across all hypothetical break dates and a fixed- b critical value is applied. The data-dependent bandwidth ratio, b^* , is computed as follows. We find the break date which minimizes the sum of squared residuals; we use that break date to select Andrews (1991) [16] data-dependent bandwidth (M^*) with the AR(1) plug-in formula and calculate the implied bandwidth ratio ($b^* = M^*/T$); we implement the test using the fixed- b critical values for b^* .

The rationale behind b^* is as follows. If a different bandwidth is used for each potential break point within the trimming range, then the fixed- b limits of the sup/mean/exp statistics will be functions of those bandwidth ratios and tabulation of fixed- b critical values will be computationally prohibitive. To provide practitioners with a data-dependent bandwidth approach that can be implemented with fixed- b critical values, we need a single data-dependent bandwidth to be used for all potential break points in which case the tabulated critical values can be used. Given the nice properties of the least squares estimator of the break point under the alternative of structural change (see Bai and Perron (1998) [11]), it is natural to use the least squares estimator of the break point to generate residuals needed to implement the Andrews (1991) [16] plug-in formula. Under the null of no structural change, any break point, including the least squares break point, will generate useful residuals for the Andrews (1991) [16] plug-in formula. Crainiceanu and Vogelsang (2007) [28] also considered using the least squares estimator of the break point to deal with the nonmonotonic power of the CUSUM test.

⁴ Cho and Vogelsang (2014) [27] also contains results for the known break date case along with a local power analysis.

Table 3 provides empirical null rejection frequencies for the traditional tests. For each hypothetical break date, the HAC estimator is constructed using the data-dependent bandwidth. For DGP A with zero persistence, all tests using $\epsilon = 0.05$ are subject to severe size distortions when the sample size is 100. Having more data or using more trimming helps reduce the size distortions. The null rejections decrease towards the 5% nominal level for all statistics when T is 500 and $\epsilon = 0.2$. Under the DGP B, as the sample size increases from 100 to 500, the null rejection probabilities drop to 0.194 from 0.594 for the supremum test with $\epsilon = 0.2$ and the QS kernel being used. The $T = 500$ rejection rate is still far from the nominal level. Size distortions get worse under more persistent data (DGP C). The mean test, which has the least size distortion of the three statistics, only attains a null rejection of 0.368 with the larger trimming value and $T = 500$. While traditional inference provides tests with reasonable size under DGPs with zero or mild persistence, as the DGP becomes more persistent, over-rejections can be substantial.

Table 3. Empirical Null Rejection Probabilities, traditional $Sup/Mean/Exp-W^{(F)}$ tests with 5% Nominal Size, H_0 : No Structural Change ($\delta = 0$).

DGP	T	$SupW^{(F)}$				$MeanW^{(F)}$				$ExpW^{(F)}$			
		$\epsilon = 0.05$		$\epsilon = 0.2$		$\epsilon = 0.05$		$\epsilon = 0.2$		$\epsilon = 0.05$		$\epsilon = 0.2$	
		Bartlett	QS	Bartlett	QS	Bartlett	QS	Bartlett	QS	Bartlett	QS	Bartlett	QS
A	100	0.699	0.742	0.164	0.171	0.278	0.306	0.115	0.121	0.676	0.728	0.176	0.186
	200	0.368	0.408	0.090	0.095	0.111	0.124	0.069	0.072	0.322	0.356	0.094	0.102
	500	0.165	0.177	0.066	0.068	0.070	0.072	0.060	0.060	0.132	0.146	0.070	0.070
B	100	0.967	0.980	0.588	0.594	0.855	0.898	0.440	0.428	0.972	0.981	0.604	0.609
	200	0.918	0.940	0.392	0.371	0.622	0.653	0.261	0.238	0.906	0.930	0.400	0.371
	500	0.745	0.750	0.218	0.194	0.315	0.297	0.152	0.134	0.688	0.699	0.217	0.196
C	100	0.992	0.993	0.910	0.918	0.982	0.984	0.853	0.866	0.995	0.995	0.924	0.930
	200	0.980	0.984	0.800	0.804	0.946	0.952	0.679	0.672	0.987	0.988	0.819	0.814
	500	0.949	0.955	0.540	0.509	0.784	0.780	0.405	0.368	0.949	0.952	0.548	0.514

Tables 4–6 present simulation results for fixed- b inference. A single bandwidth ratio, b , is applied across all hypothetical break dates in constructing HAC estimators. We report results for $b = 0.02, 0.1, 0.5$, and 1. These tables also contain the null rejection probability when the traditional critical values in Andrews (1993) [9] or Andrews and Ploberger (1994) [10] are used. The traditional critical values are not designed to work well with relatively large bandwidths and this can be clearly seen in the tables. In general, as the bandwidth ratio gets bigger, the tendency to over-reject becomes more and more pronounced because using more lags generates a systematic downward bias in the HAC estimator and pushes up the value of test statistic. The traditional critical values do not take this impact of lag-choice into account. Because the effective bandwidths play important roles for the behavior of the HAC estimator (8), the impact of using large values of b is greater than for HAC estimators in non-structural change settings.

For fixed- b inference, several patterns stand out in Table 4 for the supremum test. Rejections using fixed- b critical values are similar to the rejections in traditional inference when a small bandwidth ratio is used. However, as the bandwidth increases, rejections using fixed- b critical values systematically decrease towards the nominal level of 0.05. Under DGP B, the null rejections decrease as $0.131 \rightarrow 0.096 \rightarrow 0.083 \rightarrow 0.086$ over the range of b with $T = 500$ and the Bartlett kernel and $\epsilon = 0.2$ being used. Even under DGP C, the null rejections approach the nominal level as b increases for all sample sizes when the QS kernel and the trimming value of 0.2 are used.

Table 7 gives null rejection probabilities when using the data-dependent bandwidth ratio b^* . Columns on the left give rejections using fixed- b critical values whereas columns on the right give rejections using traditional critical values. Patterns in Table 7 are similar to patterns in Tables 4–6. Over-rejections are often large when traditional critical values are used. Over-rejections are systematically smaller when fixed- b critical values are used and b^* works reasonably well if the sample size is large enough relative to the strength of the persistence in the data. This is particularly

true when the QS kernel is used with 0.2 trimming for the mean statistic and 0.05 trimming for the supremum and exponential statistics.

Table 4. Empirical Null Rejection Probabilities, $SupW^{(F)}$ test with 5% nominal size, $M = bT$, H_0 : No Structural Change ($\delta = 0$), $T = 100, 200, 500$.

DGP	ϵ	c.v.	kernel	$b = 0.02$			$b = 0.1$			$b = 0.5$			$b = 1$		
				T = 100	200	500	T = 100	200	500	T = 100	200	500	T = 100	200	500
A	0.05	fixed- b	Bartlett	0.331	0.146	0.093	0.277	0.132	0.084	0.250	0.131	0.083	0.253	0.131	0.081
			QS	0.184	0.094	0.084	0.212	0.118	0.077	0.036	0.028	0.046	0.012	0.016	0.026
		A93	Bartlett	0.721	0.555	0.472	0.954	0.930	0.904	1.00	1.00	1.00	1.00	1.00	1.00
			QS	0.810	0.735	0.696	0.995	0.993	0.992	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	fixed- b	Bartlett	0.104	0.072	0.062	0.094	0.064	0.062	0.081	0.057	0.052	0.080	0.056	0.055
			QS	0.099	0.072	0.062	0.071	0.051	0.056	0.019	0.024	0.042	0.009	0.015	0.028
		A93	Bartlett	0.163	0.124	0.111	0.447	0.397	0.381	0.923	0.912	0.908	0.994	0.990	0.992
			QS	0.201	0.161	0.146	0.649	0.610	0.608	0.999	0.998	0.999	1.00	1.00	1.00
B	0.05	fixed- b	Bartlett	0.665	0.420	0.247	0.351	0.268	0.170	0.302	0.243	0.144	0.309	0.238	0.149
			QS	0.308	0.184	0.109	0.161	0.137	0.089	0.034	0.039	0.037	0.020	0.018	0.032
		A93	Bartlett	0.947	0.855	0.697	0.985	0.975	0.956	1.00	1.00	1.00	1.00	1.00	1.00
			QS	0.947	0.874	0.774	0.994	0.993	0.992	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	fixed- b	Bartlett	0.493	0.274	0.131	0.255	0.160	0.096	0.213	0.132	0.083	0.209	0.131	0.087
			QS	0.376	0.189	0.090	0.128	0.081	0.070	0.034	0.036	0.041	0.018	0.020	0.030
		A93	Bartlett	0.604	0.382	0.216	0.674	0.543	0.447	0.966	0.949	0.926	0.999	0.997	0.994
			QS	0.539	0.333	0.205	0.757	0.686	0.620	0.999	1.00	0.999	1.00	1.00	1.00
C	0.05	fixed- b	Bartlett	0.934	0.824	0.591	0.346	0.257	0.212	0.307	0.216	0.176	0.296	0.209	0.174
			QS	0.586	0.365	0.195	0.092	0.064	0.059	0.036	0.030	0.041	0.026	0.024	0.038
		A93	Bartlett	0.998	0.991	0.945	0.990	0.980	0.971	1.00	1.00	1.00	1.00	1.00	1.00
			QS	0.996	0.986	0.926	0.988	0.987	0.980	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	fixed- b	Bartlett	0.942	0.843	0.512	0.596	0.413	0.204	0.488	0.340	0.180	0.494	0.335	0.174
			QS	0.886	0.731	0.354	0.304	0.176	0.098	0.064	0.050	0.050	0.032	0.036	0.044
		A93	Bartlett	0.967	0.902	0.632	0.900	0.809	0.630	0.994	0.982	0.956	1.00	0.999	0.997
			QS	0.947	0.846	0.532	0.904	0.835	0.698	1.00	1.00	0.998	1.00	1.00	1.00

Note: A93 are critical values from Andrews (2003) [26].

Table 5. Empirical Null Rejection Probabilities, $MeanW^{(F)}$ test with 5% nominal size, $M = bT$, H_0 : No Structural Change ($\delta = 0$), $T = 100, 200, 500$.

DGP	ϵ	c.v.	Kernel	$b = 0.02$			$b = 0.1$			$b = 0.5$			$b = 1$		
				T = 100	200	500	T = 100	200	500	T = 100	200	500	T = 100	200	500
A	0.05	fixed- b	Bartlett	0.162	0.077	0.064	0.190	0.090	0.064	0.216	0.100	0.065	0.217	0.100	0.067
			QS	0.148	0.085	0.066	0.226	0.108	0.070	0.120	0.075	0.057	0.097	0.062	0.055
		AP94	Bartlett	0.290	0.174	0.134	0.759	0.623	0.570	0.999	0.998	0.995	1.00	1.00	1.00
			QS	0.376	0.265	0.216	0.952	0.908	0.877	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	fixed- b	Bartlett	0.084	0.057	0.056	0.086	0.055	0.061	0.089	0.059	0.060	0.087	0.060	0.060
			QS	0.082	0.056	0.060	0.081	0.050	0.055	0.065	0.053	0.051	0.055	0.050	0.050
		AP94	Bartlett	0.120	0.087	0.084	0.291	0.248	0.231	0.850	0.828	0.818	0.982	0.974	0.971
			QS	0.132	0.106	0.100	0.449	0.403	0.390	0.998	0.996	0.998	1.00	1.00	1.00
B	0.05	fixed- b	Bartlett	0.664	0.358	0.160	0.445	0.270	0.128	0.462	0.275	0.148	0.455	0.276	0.141
			QS	0.530	0.242	0.110	0.291	0.178	0.103	0.155	0.112	0.078	0.121	0.088	0.073
		AP94	Bartlett	0.806	0.534	0.291	0.926	0.834	0.694	1.00	1.00	0.999	1.00	1.00	1.00
			QS	0.783	0.512	0.318	0.973	0.946	0.908	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	fixed- b	Bartlett	0.420	0.207	0.109	0.216	0.132	0.081	0.224	0.135	0.082	0.229	0.137	0.086
			QS	0.324	0.148	0.085	0.154	0.102	0.070	0.114	0.084	0.065	0.089	0.072	0.065
		AP94	Bartlett	0.488	0.270	0.152	0.521	0.386	0.295	0.938	0.889	0.848	0.996	0.992	0.982
			QS	0.428	0.219	0.137	0.603	0.483	0.417	0.998	0.998	0.998	1.00	1.00	1.00
C	0.05	fixed- b	Bartlett	0.985	0.927	0.653	0.723	0.546	0.324	0.668	0.494	0.316	0.664	0.492	0.304
			QS	0.948	0.827	0.451	0.407	0.247	0.145	0.207	0.137	0.099	0.174	0.127	0.091
		AP94	Bartlett	0.996	0.966	0.788	0.979	0.955	0.868	1.00	1.00	1.00	1.00	1.00	1.00
			QS	0.992	0.949	0.734	0.981	0.968	0.928	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	fixed- b	Bartlett	0.916	0.776	0.429	0.586	0.396	0.206	0.552	0.384	0.204	0.547	0.375	0.204
			QS	0.854	0.661	0.318	0.409	0.257	0.124	0.223	0.144	0.105	0.171	0.114	0.084
		AP94	Bartlett	0.943	0.822	0.502	0.838	0.686	0.469	0.988	0.967	0.918	0.999	0.998	0.992
			QS	0.909	0.753	0.408	0.843	0.705	0.529	1.00	1.00	0.999	1.00	1.00	1.00

Note: AP94 are critical values from Andrews and Ploberger (1994) [10] with an adjustment.

Table 6. Empirical Null Rejection Probabilities, $ExpW^{(F)}$ test with 5% nominal size, $M = bT$, H_0 : No Structural Change ($\delta = 0$), $T = 100, 200, 500$.

DGP	ϵ	c.v.	Kernel	$b = 0.02$			$b = 0.1$			$b = 0.5$			$b = 1$		
				T = 100	200	500	T = 100	200	500	T = 100	200	500	T = 100	200	500
A	0.05	fixed- b	Bartlett	0.368	0.162	0.094	0.291	0.140	0.086	0.256	0.133	0.083	0.254	0.132	0.082
			QS	0.198	0.100	0.086	0.217	0.120	0.078	0.036	0.028	0.046	0.012	0.016	0.026
		AP94	Bartlett	0.712	0.504	0.404	0.956	0.920	0.888	1.00	1.00	1.00	1.00	1.00	1.00
			QS	0.802	0.693	0.640	0.996	0.992	0.987	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	fixed- b	Bartlett	0.108	0.068	0.062	0.104	0.068	0.060	0.086	0.060	0.055	0.082	0.058	0.055
			QS	0.104	0.072	0.064	0.080	0.055	0.055	0.019	0.024	0.042	0.009	0.015	0.028
		AP94	Bartlett	0.179	0.131	0.115	0.454	0.390	0.367	0.929	0.909	0.906	0.996	0.991	0.992
			QS	0.210	0.167	0.142	0.646	0.605	0.591	0.999	0.999	1.00	1.00	1.00	1.00
B	0.05	fixed- b	Bartlett	0.708	0.456	0.260	0.372	0.282	0.172	0.309	0.246	0.144	0.313	0.241	0.150
			QS	0.333	0.194	0.113	0.165	0.140	0.090	0.034	0.039	0.037	0.020	0.018	0.032
		AP94	Bartlett	0.954	0.836	0.647	0.986	0.974	0.946	1.00	1.00	1.00	1.00	1.00	1.00
			QS	0.952	0.855	0.734	0.996	0.992	0.993	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	fixed- b	Bartlett	0.518	0.281	0.132	0.275	0.169	0.096	0.219	0.136	0.085	0.214	0.133	0.087
			QS	0.398	0.195	0.094	0.139	0.088	0.072	0.034	0.036	0.041	0.018	0.020	0.030
		AP94	Bartlett	0.627	0.384	0.211	0.689	0.554	0.446	0.970	0.956	0.925	1.00	0.998	0.997
			QS	0.564	0.332	0.202	0.763	0.687	0.610	0.999	1.00	0.999	1.00	1.00	1.00
C	0.05	fixed- b	Bartlett	0.958	0.870	0.634	0.370	0.277	0.220	0.313	0.219	0.177	0.301	0.210	0.174
			QS	0.624	0.387	0.204	0.095	0.066	0.060	0.036	0.030	0.041	0.026	0.024	0.038
		AP94	Bartlett	0.999	0.992	0.942	0.993	0.988	0.970	1.00	1.00	1.00	1.00	1.00	1.00
			QS	0.998	0.986	0.924	0.993	0.989	0.985	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	fixed- b	Bartlett	0.951	0.853	0.517	0.617	0.437	0.218	0.502	0.348	0.183	0.498	0.342	0.177
			QS	0.902	0.751	0.372	0.327	0.190	0.102	0.064	0.050	0.050	0.032	0.036	0.044
		AP94	Bartlett	0.971	0.907	0.641	0.918	0.829	0.637	0.996	0.985	0.961	1.00	1.00	0.998
			QS	0.954	0.854	0.532	0.919	0.845	0.710	1.00	1.00	0.999	1.00	1.00	1.00

Note: AP94 are critical values from Andrews and Ploberger (1994) [10] with an adjustment.

We now examine the power of the tests when using b^* . We report size-adjusted power for $T = 200$ in Figures 1–6. Recall the break point under the alternative is $\lambda = 0.4$. Odd (even) numbered figures give results with 0.05 (0.2) trimming. Results are given for the three DGPs used for the tables. First note that more trimming leads to higher power in all cases as one would expect. Second, the mean statistic tends to have the highest power regardless of the DGP or kernel. This is not surprising given the power optimality properties of the mean statistic derived by Andrews and Ploberger (1994) [10] using traditional asymptotics. Third, for a given kernel, the supremum and exponential statistics have almost the same power across DGPs and trimming. This is somewhat surprising given that under traditional asymptotics, the exponential statistic is in the class of power optimal tests but the supremum statistic is not. This finding could be driven by values of b^* being far away from zero in which case the traditional asymptotics might not be accurately reflecting finite sample power. Finally, the Bartlett kernel tends to give tests with higher power than the QS kernel; a similar finding was made by Kiefer and Vogelsang (2005) [1] in models without structural change.

The size and power results for the statistics implemented with b^* point to the typical size-power tradeoff when using HAC variance estimators. Configurations that give the least size distortions also tend to have low power. As long as the data is not too persistent relative to the sample size, a reasonable approach for practice that balances size distortions and power is to use the mean statistic with 0.2 trimming implemented with the QS kernel with b^* and fixed- b critical values.

6. Summary and Conclusions

In this paper, fixed- b asymptotics is applied to the problem of testing for the presence of a structural break in a weakly dependent time series regression. The $Wald^{(F)}$ statistic is the Wald statistic that one obtains when structural change is expressed in terms of dummy variables interacted with regressors as in Bai and Perron (1998, 2003) [11,15]. We derived the fixed- b limit of the statistic. In both the full structural change and partial structural change model, the Wald statistic has the same pivotal fixed- b limit. We tabulated fixed- b critical values for $Sup/Mean/Exp-Wald^{(F)}$ statistics which

are commonly used for testing parameter instability when the break point is unknown. In a simulation study, we examined the finite sample properties of traditional and fixed- b inference. With persistent data, traditional inference suffers from substantial size distortions. Using fixed- b critical values markedly improves over-rejection problem. A reasonable approach for practice that balances size distortions and power is to use the mean statistic with 0.2 trimming implemented with the QS kernel, b^* and fixed- b critical values.

Table 7. Empirical Null Rejection Probabilities, $Sup/Mean/Exp-W^{(F)}$ test using bandwidth ratio b^* with 5% nominal size, H_0 : No Structural Change ($\delta = 0$), $T = 100, 200, 500, 1000$.

$SupW^{(F)}$		Fixed- b c.v.				Andrews (1993) [9] c.v.			
DGP	T	$\epsilon = 0.05$		$\epsilon = 0.2$		$\epsilon = 0.05$		$\epsilon = 0.2$	
		Bartlett	QS	Bartlett	QS	Bartlett	QS	Bartlett	QS
A	100	0.318	0.182	0.103	0.099	0.773	0.830	0.189	0.221
	200	0.147	0.094	0.073	0.072	0.558	0.735	0.126	0.161
	500	0.089	0.081	0.062	0.062	0.479	0.701	0.111	0.146
B	100	0.384	0.164	0.287	0.177	0.972	0.985	0.612	0.627
	200	0.302	0.135	0.187	0.134	0.931	0.962	0.416	0.415
	500	0.206	0.102	0.110	0.088	0.826	0.801	0.249	0.218
C	100	0.328	0.083	0.574	0.289	0.992	0.991	0.915	0.924
	200	0.278	0.070	0.428	0.219	0.982	0.985	0.806	0.814
	500	0.267	0.082	0.250	0.158	0.954	0.965	0.556	0.548
	1000	0.254	0.072	0.188	0.120	0.900	0.928	0.375	0.368
$MeanW^{(F)}$		Fixed- b c.v.				Andrews and Ploberger (1994) [10] c.v.			
DGP	T	$\epsilon = 0.05$		$\epsilon = 0.2$		$\epsilon = 0.05$		$\epsilon = 0.2$	
		Bartlett	QS	Bartlett	QS	Bartlett	QS	Bartlett	QS
A	100	0.172	0.154	0.086	0.082	0.348	0.417	0.132	0.142
	200	0.077	0.085	0.058	0.056	0.176	0.265	0.088	0.106
	500	0.064	0.066	0.056	0.060	0.134	0.216	0.084	0.100
B	100	0.465	0.312	0.237	0.168	0.869	0.912	0.454	0.456
	200	0.278	0.190	0.145	0.110	0.667	0.717	0.277	0.269
	500	0.142	0.108	0.090	0.085	0.408	0.349	0.171	0.143
C	100	0.701	0.382	0.566	0.401	0.982	0.983	0.853	0.875
	200	0.555	0.293	0.408	0.283	0.948	0.956	0.688	0.688
	500	0.374	0.203	0.232	0.171	0.804	0.812	0.415	0.391
	1000	0.258	0.133	0.155	0.104	0.574	0.619	0.257	0.238
$ExpW^{(F)}$		Fixed- b c.v.				Andrews and Ploberger (1994) [10] c.v.			
DGP	T	$\epsilon = 0.05$		$\epsilon = 0.2$		$\epsilon = 0.05$		$\epsilon = 0.2$	
		Bartlett	QS	Bartlett	QS	Bartlett	QS	Bartlett	QS
A	100	0.331	0.196	0.095	0.107	0.761	0.820	0.203	0.230
	200	0.161	0.100	0.068	0.072	0.506	0.593	0.132	0.167
	500	0.093	0.083	0.062	0.064	0.404	0.640	0.115	0.142
B	100	0.402	0.170	0.296	0.195	0.976	0.988	0.626	0.638
	200	0.278	0.143	0.167	0.142	0.923	0.956	0.427	0.423
	500	0.135	0.104	0.068	0.093	0.788	0.762	0.251	0.215
C	100	0.348	0.087	0.601	0.308	0.995	0.994	0.929	0.935
	200	0.298	0.072	0.449	0.241	0.987	0.990	0.826	0.829
	500	0.277	0.083	0.263	0.169	0.955	0.961	0.563	0.556
	1000	0.186	0.072	0.134	0.121	0.881	0.912	0.378	0.369

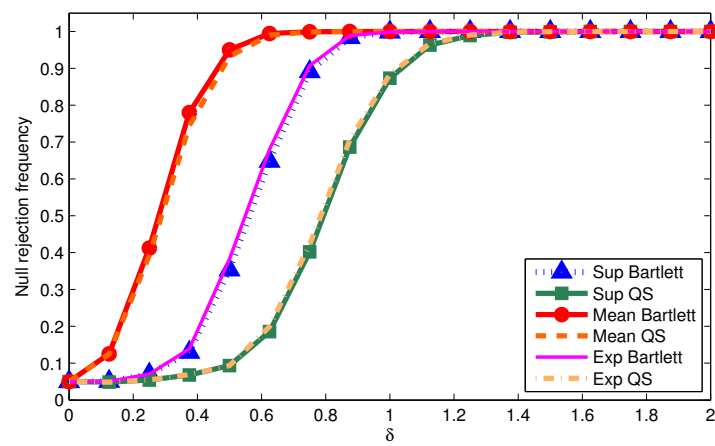


Figure 1. Size adjusted power, DGP A, $\epsilon = 0.05$, $T = 200$.

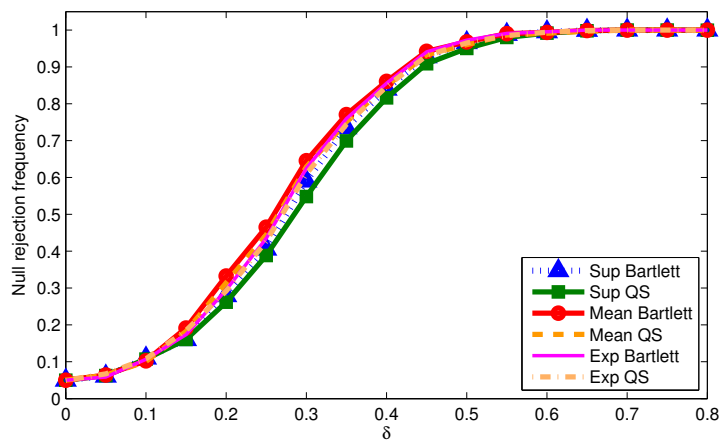


Figure 2. Size adjusted power, DGP A, $\epsilon = 0.2$, $T = 200$.

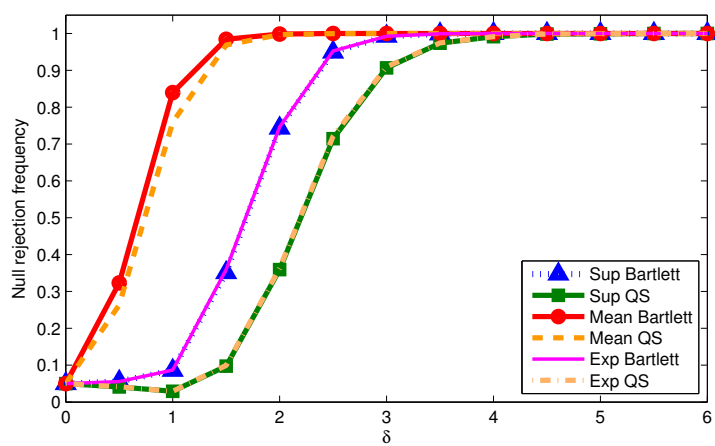


Figure 3. Size adjusted power, DGP B, $\epsilon = 0.05$, $T = 200$.

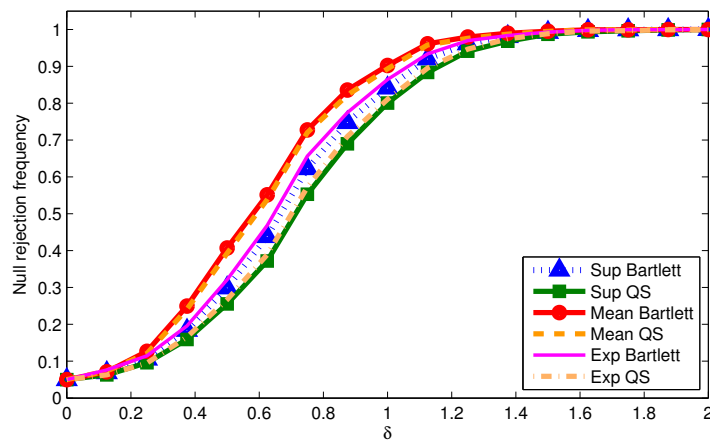


Figure 4. Size adjusted power, DGP B, $\epsilon = 0.2$, $T = 200$.

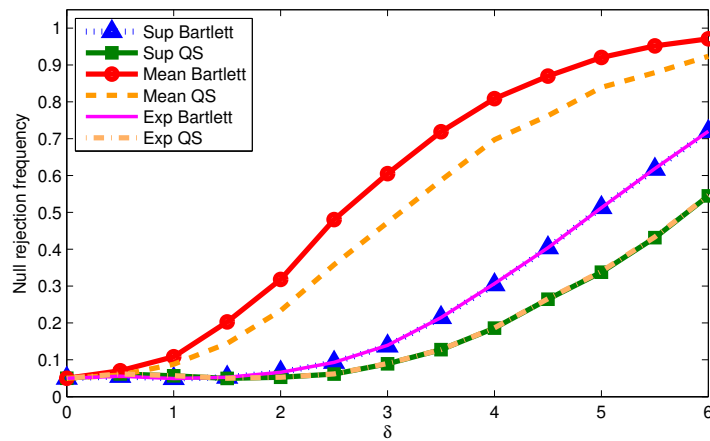


Figure 5. Size adjusted power, DGP C, $\epsilon = 0.05$, $T = 200$.

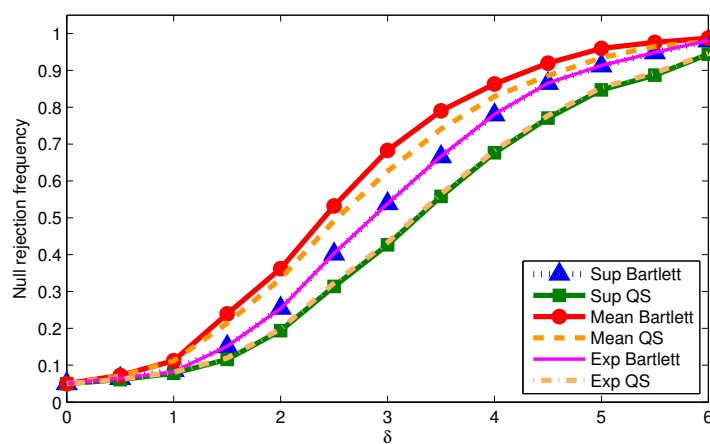


Figure 6. Size adjusted power, DGP C, $\epsilon = 0.2$, $T = 200$.

Acknowledgments: We thank the editor and three anonymous referees for helpful suggestions and comments.

Author Contributions: Both authors contributed equally to the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Definitions and Proofs

Definitions

Case 1. Suppose $K(x)$ is twice continuously differentiable everywhere (Class 1) such as the Quadratic Spectral kernel (QS), then

$$\mathbf{P}(b, H_p) \equiv - \int_0^1 \int_0^1 \frac{1}{b^2} K'' \left(\frac{r-s}{b} \right) H_p(r) H_p(s)' dr ds, \quad (\text{A1})$$

where $K''(\cdot)$ is the second derivative of the kernel $K(\cdot)$.

Case 2. Suppose $K(x)$ is the Bartlett kernel (Class 2), then

$$\mathbf{P}(b, H_p) \equiv \frac{2}{b} \int_0^1 H_p(r) H_p(r)' dr - \frac{1}{b} \int_0^{1-b} (H_p(r) H_p(r+b)' + H_p(r+b) H_p(r)') dr. \quad (\text{A2})$$

Case 3. Suppose $K(x)$ is continuous, $K(x) = 0$ for $|x| \geq 1$, and $K(x)$ is twice continuously differentiable everywhere except for $|x| = 1$ (Class 3) (e.g., Parzen kernel), then

$$\begin{aligned} \mathbf{P}(b, H_p) \equiv & - \int \int_{|r-s| < b} \frac{1}{b^2} K'' \left(\frac{|r-s|}{b} \right) H_p(r) H_p(s)' dr ds \\ & + \frac{K'_-(1)}{b} \int_0^{1-b} (H_p(r+b) H_p(r)' + H_p(r) H_p(r+b)') dr, \end{aligned} \quad (\text{A3})$$

where $K'_-(1) = \lim_{h \downarrow 0} [(K(1) - K(1-h)) / h]$, i.e., $K'_-(1)$ is the derivative of $K(x)$ from the left at $x = 1$.

The following expression is a general representation of the HAC estimators:

$$\hat{\Omega} = T^{-1} \sum_{t=1}^T \sum_{s=1}^T K \left(\frac{|t-s|}{M} \right) \hat{v}_t \hat{v}_s'.$$

This representation can be rewritten in terms of the partial sum processes $\hat{S}_t = \sum_{j=1}^t \hat{v}_j$ following Kiefer and Vogelsang (2005) [1] and Hashimzade and Vogelsang (2008) [4] as follows. Let $M = bT$. Then, for the kernels in Class 1, we have

$$\hat{\Omega} = T^{-2} \sum_{t=1}^{T-1} \sum_{s=1}^{T-1} T^{-1/2} \hat{S}_t \left(T^2 \Delta_{t,s}^2 \right) T^{-1/2} \hat{S}_s', \quad (\text{A4})$$

where

$$\Delta_{t,s}^2 \equiv (K_{t,s} - K_{t,s+1}) - (K_{t+1,s} - K_{t+1,s+1}) \text{ with } K_{t,s} = K \left(\frac{|t-s|}{bT} \right).$$

For the Class 2 kernel (Bartlett), we have

$$\hat{\Omega} = \frac{2}{bT} \sum_{t=1}^{T-1} \left(T^{-1} \hat{S}_t \hat{S}_t' \right) - \frac{1}{bT} \sum_{t=1}^{T-M-1} \left(T^{-1} \hat{S}_{t+bT} \hat{S}_t' + T^{-1} \hat{S}_t \hat{S}_{t+bT}' \right). \quad (\text{A5})$$

For the kernels in Class 3, we have

$$\begin{aligned} \hat{\Omega} = & T^{-2} \sum_{|t-s| < bT} \sum T^{-1} \hat{S}_t \left(T^2 \Delta_{t,s}^2 \right) \hat{S}_s' + \frac{1}{bT} \sum_{s=1}^{T-bT} T^{-1/2} \hat{S}_s T^{-1/2} \hat{S}_{s+bT}' \left(\frac{K(1) - K(1 - \frac{1}{bT})}{\frac{1}{bT}} \right) \\ & - \frac{1}{bT} \sum_{s=1}^{T-bT} T^{-1/2} \hat{S}_s T^{-1/2} \hat{S}_{s+bT}' \left(\frac{K(-1 + \frac{1}{bT}) - K(-1)}{\frac{1}{bT}} \right). \end{aligned} \quad (\text{A6})$$

Proof of Proposition 1. The limit of the $\hat{\mathbf{f}}_n$ follows immediately under Assumptions 1 and 2. Also, plugging the limits of $\hat{\beta}_1$ and $\hat{\beta}_2$ into Equation (4) yields, for $r \leq \lambda$,

$$T^{-1/2} \hat{S}_{[rT]} \Rightarrow \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \Lambda \end{pmatrix} \begin{pmatrix} (W_p(r) - \frac{r}{\lambda} W_p(\lambda)) \\ \mathbf{0} \end{pmatrix},$$

and for $r > \lambda$,

$$T^{-1/2} \hat{S}_{[rT]} \Rightarrow \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \Lambda \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ (W_p(r) - W_p(\lambda) - \frac{r-\lambda}{1-\lambda} (W_p(1) - W_p(\lambda))) \end{pmatrix}.$$

Thus, we can rewrite this result by using indicator functions as

$$T^{-1/2} \hat{S}_{[rT]} \Rightarrow \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \Lambda \end{pmatrix} F_p(r, \lambda) \equiv \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \Lambda \end{pmatrix} \begin{pmatrix} F_p^{(1)}(r, \lambda) \\ F_p^{(2)}(r, \lambda) \end{pmatrix},$$

where

$$F_p^{(1)}(r, \lambda) = (W_p(r) - \frac{r}{\lambda} W_p(\lambda)) \cdot \mathbf{1}(r \leq \lambda) \text{ and } F_p^{(2)}(r, \lambda) = (W_p(r) - W_p(\lambda) - \frac{r-\lambda}{1-\lambda} (W_p(1) - W_p(\lambda))) \cdot \mathbf{1}(r > \lambda).$$

□

Proof of Lemma 1. Plugging the limit of the partial sum process in Proposition 1 into the HAC estimators in (A4)–(A6), the desired result follows from direct application of the continuous mapping theorem to obtain the desired result in (12). □

Proof of Theorem 1. Recall that

$$Wald^{(F)} = T \left(\mathbf{R} \hat{\beta} \right)' \left(\mathbf{R} \hat{Q}_\lambda^{-1} \hat{\Omega}^{(F)} \hat{Q}_\lambda^{-1} \mathbf{R}' \right)^{-1} \left(\mathbf{R} \hat{\beta} \right).$$

Using $\mathbf{R} = (\mathbf{R}_1, -\mathbf{R}_1)$ it follows that

$$\begin{aligned} T^{1/2} \left(\mathbf{R} \hat{\beta} \right) &\stackrel{H_0}{\Rightarrow} \mathbf{R}_1 \left(T^{1/2} \left(\hat{\beta}_1 - \beta_1 \right) - T^{1/2} \left(\hat{\beta}_2 - \beta_2 \right) \right) \Rightarrow \\ &\mathbf{R}_1 Q^{-1} \Lambda \left(\frac{1}{\lambda} W_p(\lambda) - \frac{1}{1-\lambda} (W_p(1) - W_p(\lambda)) \right). \end{aligned}$$

Using Assumption 1 and Lemma 1,

$$\mathbf{R} \hat{Q}_\lambda^{-1} \hat{\Omega}^{(F)} \hat{Q}_\lambda^{-1} \mathbf{R}' \Rightarrow \left(\frac{1}{\lambda} \mathbf{R}_1 Q^{-1} \Lambda, \frac{-1}{1-\lambda} \mathbf{R}_1 Q^{-1} \Lambda \right) \times \mathbf{P}(b, F_p(r, \lambda)) \times \left(\frac{1}{\lambda} \mathbf{R}_1 Q^{-1} \Lambda, \frac{-1}{1-\lambda} \mathbf{R}_1 Q^{-1} \Lambda \right)'.$$

By writing $\mathbf{P}(b, F_p(r, \lambda))$ in the form (A1)–(A3) using $F_p(r, \lambda)' = (F_p^{(1)}(r, \lambda)', F_p^{(2)}(r, \lambda)')$, we obtain, after some algebra, the following expression for the above limit:

$$\mathbf{R}_1 Q^{-1} \Lambda \mathbf{P} \left(b, \frac{1}{\lambda} F_p^{(1)}(r, \lambda) - \frac{1}{1-\lambda} F_p^{(2)}(r, \lambda) \right) \Lambda' Q^{-1} \mathbf{R}_1'.$$

Now apply the transformation: $\mathbf{R}_1 Q^{-1} \Lambda W_p(r) \stackrel{d}{=} A W_l(r)$ with $\mathbf{R}_1 Q^{-1} \Lambda \Lambda' Q^{-1} \mathbf{R}_1' = A A'$, and conclude

$$\mathbf{R} \hat{Q}_\lambda^{-1} \hat{\Omega}^{(F)} \hat{Q}_\lambda^{-1} \mathbf{R}' \Rightarrow A \mathbf{P} \left(b, \frac{1}{\lambda} F_l^{(1)}(r, \lambda) - \frac{1}{1-\lambda} F_l^{(2)}(r, \lambda) \right) A',$$

yielding the desired result:

$$\begin{aligned} Wald^{(F)} &\Rightarrow \left(\frac{1}{\lambda} W_l(\lambda) - \frac{1}{1-\lambda} (W_l(1) - W_l(\lambda)) \right)' \\ &\times \left(\mathbf{P} \left(b, \frac{1}{\lambda} F_l^{(1)}(r, \lambda) - \frac{1}{1-\lambda} F_l^{(2)}(r, \lambda) \right) \right)^{-1} \times \left(\frac{1}{\lambda} W_l(\lambda) - \frac{1}{1-\lambda} (W_l(1) - W_l(\lambda)) \right). \end{aligned}$$

□

Proof of Proposition 2. Standard algebra gives

$$\begin{aligned} \hat{\beta} &= \left(\sum_{t=1}^T \tilde{X}_t \tilde{X}_t' \right)^{-1} \left(\sum_{t=1}^T \tilde{X}_t \tilde{y}_t \right) \\ &= \left(\sum_{t=1}^T \tilde{X}_t \tilde{X}_t' \right)^{-1} \left(\sum_{t=1}^T \tilde{X}_t \tilde{X}_t' \beta + \sum_{t=1}^T \tilde{X}_t u_t - \sum_{t=1}^T \tilde{X}_t z_t' (Z'Z)^{-1} Z'u \right) \\ &= \left(\sum_{t=1}^T \tilde{X}_t \tilde{X}_t' \right)^{-1} \left(\sum_{t=1}^T \tilde{X}_t \tilde{X}_t' \beta + \sum_{t=1}^T \tilde{X}_t u_t \right), \end{aligned}$$

and it immediately follows that

$$\begin{aligned} T^{1/2} (\hat{\beta} - \beta) &= \left(T^{-1} \sum_{t=1}^T \tilde{X}_t \tilde{X}_t' \right)^{-1} \left(T^{-1/2} \sum_{t=1}^T \tilde{X}_t u_t \right) \\ &= \left(T^{-1} \sum_{t=1}^T (X_t - X'Z(Z'Z)^{-1}z_t) (X_t' - z_t'(Z'Z)^{-1}Z'X) \right)^{-1} \\ &\quad \times \left(T^{-1/2} \sum_{t=1}^T (X_t - X'Z(Z'Z)^{-1}z_t) u_t \right). \end{aligned}$$

Under Assumptions 3 and 4, it follows in a straightforward manner that

$$\sqrt{T}(\hat{\beta} - \beta) \Rightarrow Q_{\tilde{X}\tilde{X}}^{-1} \begin{pmatrix} \Lambda_1 W_{p+q}(\lambda) - \lambda Q_{XZ} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1) \\ \Lambda_1 (W_{p+q}(1) - W_{p+q}(\lambda)) - (1-\lambda) Q_{XZ} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1) \end{pmatrix} \quad (A7)$$

In order to derive the limit of $\sqrt{T}(\mathbf{R}\hat{\beta} - \mathbf{r})$, the following standard results are useful:

$$\begin{aligned} Q_{XZ} &\equiv p \lim \left(T^{-1} \sum_{t=1}^T X_t z_t' \right) = \begin{pmatrix} \lambda Q_{XZ} \\ (1-\lambda) Q_{XZ} \end{pmatrix}_{2p \times q}, \\ Q_{XX} &\equiv p \lim \left(T^{-1} \sum_{t=1}^T X_t X_t' \right) = \begin{pmatrix} \lambda Q_{XX} & \mathbf{0} \\ \mathbf{0} & (1-\lambda) Q_{XX} \end{pmatrix}_{2p \times 2p}, \\ Q_{\tilde{X}\tilde{X}} &= p \lim \left(T^{-1} \sum_{t=1}^T \tilde{X}_t \tilde{X}_t' \right) = Q_{XX} - Q_{XZ} Q_{ZZ}^{-1} Q_{XZ}'. \end{aligned}$$

Also, well known matrix algebra properties (see e.g., Schott (1997) [29]), we can write

$$Q_{\tilde{X}\tilde{X}}^{-1} = Q_{XX}^{-1} + Q_{XX}^{-1} Q_{XZ} \left(Q_{ZZ} - Q_{XZ}' Q_{XX}^{-1} Q_{XZ} \right)^{-1} Q_{XZ}' Q_{XX}^{-1}, \quad (A8)$$

and using (A8), one can further show that

$$Q_{\tilde{X}\tilde{X}}^{-1} = \begin{pmatrix} \frac{1}{\lambda} Q_{XX}^{-1} + P & P \\ P & \frac{1}{1-\lambda} Q_{XX}^{-1} + P \end{pmatrix}, \quad (A9)$$

where

$$P = Q_{XX}^{-1} Q_{XZ} \left(Q_{ZZ} - Q_{XZ}' Q_{XX}^{-1} Q_{XZ} \right)^{-1} Q_{XZ}' Q_{XX}^{-1}.$$

Now plug (A9) into (A7) to conclude that

$$\sqrt{T}(\mathbf{R}\hat{\beta} - \mathbf{r}) \stackrel{H_0}{=} \sqrt{T}\mathbf{R}(\hat{\beta} - \beta) \Rightarrow \mathbf{R}_1 Q_{XX}^{-1} \Lambda_1 \left(\frac{1}{\lambda} W_{p+q}(\lambda) + \frac{1}{1-\lambda} (W_{p+q}(\lambda) - W_{p+q}(1)) \right). \quad \square$$

The following lemma is used in the proof of Lemma 3.

Lemma 3. Let $K = Q_{xz}Q_{ZZ}^{-1}Q'_{xz}$. Then it holds that $Q_{xx}^{-1}KP = P - Q_{xx}^{-1}KQ_{xx}^{-1}$.

Proof of Lemma 3. One can easily show

$$Q_{\tilde{X}\tilde{X}} = \begin{pmatrix} \lambda Q_{xx} & 0 \\ 0 & (1-\lambda)Q_{xx} \end{pmatrix} - \begin{pmatrix} \lambda^2 Q_{xz}Q_{ZZ}^{-1}Q'_{xz} & \lambda(1-\lambda)Q_{xz}Q_{ZZ}^{-1}Q'_{xz} \\ \lambda(1-\lambda)Q_{xz}Q_{ZZ}^{-1}Q'_{xz} & (1-\lambda)^2 Q_{xz}Q_{ZZ}^{-1}Q'_{xz} \end{pmatrix}.$$

The desired result comes from the identity $Q_{\tilde{X}\tilde{X}}Q_{\tilde{X}\tilde{X}}^{-1} = I$ by substituting Equation (A9) for $Q_{\tilde{X}\tilde{X}}^{-1}$. \square

Proof of Lemma 2. First note that implicit in the proof of Proposition 2 is the result that $p \lim \hat{Q}_{\tilde{X}\tilde{X}}^{-1} = Q_{\tilde{X}\tilde{X}}^{-1}$. For $\mathbf{R} = (\mathbf{R}_1, -\mathbf{R}_1)$, it follows that

$$p \lim \mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} = \mathbf{R}_1 \left(\frac{1}{\lambda} Q_{xx}^{-1}, -\frac{1}{1-\lambda} Q_{xx}^{-1} \right) \quad (\text{A10})$$

using (A9). The scaled partial sum process is given by

$$\begin{aligned} T^{-1/2} \tilde{S}_{[rT]}^x &= T^{-1/2} \sum_{t=1}^{[rT]} \tilde{X}_t \hat{u}_t \\ &= T^{-1/2} \sum_{t=1}^{[rT]} \tilde{X}_t u_t - T^{-1} \sum_{t=1}^{[rT]} \tilde{X}_t \tilde{X}_t' \sqrt{T}(\hat{\beta} - \beta) - T^{-1} \sum_{t=1}^{[rT]} \tilde{X}_t z_t' \left(\frac{Z'Z}{T} \right)^{-1} \left(T^{-1/2} Z'u \right). \end{aligned} \quad (\text{A11})$$

For $0 \leq r < \lambda$, the first term in (A11) satisfies

$$T^{-1/2} \sum_{t=1}^{[rT]} \tilde{X}_t u_t \Rightarrow \begin{pmatrix} \Lambda_1 W_{p+q}(r) - \lambda Q_{xz}Q_{ZZ}^{-1}\Lambda_2 W_{p+q}(r) \\ -(1-\lambda)Q_{xz}Q_{ZZ}^{-1}\Lambda_2 W_{p+q}(r) \end{pmatrix}. \quad (\text{A12})$$

Hence with $\mathbf{R} = (\mathbf{R}_1, -\mathbf{R}_1)$, from (A10) and (A12), it follows that

$$\begin{aligned} &\mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} T^{-1/2} \sum_{t=1}^{[rT]} \tilde{X}_t u_t \\ &\Rightarrow \mathbf{R}_1 \left(\frac{1}{\lambda} Q_{xx}^{-1}, -\frac{1}{1-\lambda} Q_{xx}^{-1} \right) \begin{pmatrix} \Lambda_1 W_{p+q}(r) - \lambda Q_{xz}Q_{ZZ}^{-1}\Lambda_2 W_{p+q}(r) \\ -(1-\lambda)Q_{xz}Q_{ZZ}^{-1}\Lambda_2 W_{p+q}(r) \end{pmatrix} = \frac{1}{\lambda} \mathbf{R}_1 Q_{xx}^{-1} \Lambda_1 W_{p+q}(r). \end{aligned}$$

For the first part of the second term in (A11), it follows that

$$\begin{aligned} T^{-1} \sum_{t=1}^{[rT]} \tilde{X}_t \tilde{X}_t' &\Rightarrow \begin{pmatrix} rQ_{xx} & \mathbf{0}_{p \times p} \\ \mathbf{0}_{p \times p} & \mathbf{0}_{p \times p} \end{pmatrix} - \begin{pmatrix} r\lambda Q_{xz}Q_{ZZ}^{-1}Q'_{xz} & r(1-\lambda)Q_{xz}Q_{ZZ}^{-1}Q'_{xz} \\ \mathbf{0}_{p \times p} & \mathbf{0}_{p \times p} \end{pmatrix} \\ &- \begin{pmatrix} r\lambda Q_{xz}Q_{ZZ}^{-1}Q'_{xz} & \mathbf{0}_{p \times p} \\ r(1-\lambda)Q_{xz}Q_{ZZ}^{-1}Q'_{xz} & \mathbf{0}_{p \times p} \end{pmatrix} + r \begin{pmatrix} \lambda^2 Q_{xz}Q_{ZZ}^{-1}Q'_{xz} & \lambda(1-\lambda)Q_{xz}Q_{ZZ}^{-1}Q'_{xz} \\ \lambda(1-\lambda)Q_{xz}Q_{ZZ}^{-1}Q'_{xz} & (1-\lambda)^2 Q_{xz}Q_{ZZ}^{-1}Q'_{xz} \end{pmatrix} \\ &= \begin{pmatrix} rQ_{xx} + (r\lambda^2 - 2r\lambda)K & -r(1-\lambda)^2 K \\ -r(1-\lambda)^2 K & r(1-\lambda)^2 K \end{pmatrix}, \end{aligned}$$

where $K = Q_{xz}Q_{ZZ}^{-1}Q'_{xz}$. Hence with $\mathbf{R} = (\mathbf{R}_1, -\mathbf{R}_1)$,

$$\mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} T^{-1} \sum_{t=1}^{[rT]} \tilde{X}_t \tilde{X}_t' \Rightarrow \mathbf{R}_1 \left(\frac{1}{\lambda} Q_{xx}^{-1}, -\frac{1}{1-\lambda} Q_{xx}^{-1} \right) \times \begin{pmatrix} rQ_{xx} + (r\lambda^2 - 2r\lambda)K & -r(1-\lambda)^2 K \\ -r(1-\lambda)^2 K & r(1-\lambda)^2 K \end{pmatrix}$$

$$= r\mathbf{R}_1 \left(\frac{1}{\lambda} I - Q_{xx}^{-1} K, \frac{\lambda-1}{\lambda} Q_{xx}^{-1} K \right),$$

which combined with (A7) and Lemma 3 immediately yields

$$\begin{aligned} & \mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} \left(T^{-1} \sum_{t=1}^{[rT]} \tilde{X}_t \tilde{X}_t' \right) \sqrt{T}(\hat{\beta} - \beta) \\ \Rightarrow & r\mathbf{R}_1 \left(\frac{1}{\lambda} I - Q_{xx}^{-1} K, \frac{\lambda-1}{\lambda} Q_{xx}^{-1} K \right) \times \begin{pmatrix} \frac{1}{\lambda} Q_{xx}^{-1} + P & P \\ P & \frac{1}{1-\lambda} Q_{xx}^{-1} + P \end{pmatrix} \\ & \times \begin{pmatrix} \Lambda_1 W_{p+q}(\lambda) - \lambda Q_{xZ} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1) \\ \Lambda_1 (W_{p+q}(1) - W_{p+q}(\lambda)) - (1-\lambda) Q_{xZ} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1) \end{pmatrix} \\ = & r\mathbf{R}_1 \left(\frac{1}{\lambda^2} Q_{xx}^{-1}, \mathbf{0}_{p \times p} \right) \times \begin{pmatrix} \Lambda_1 W_{p+q}(\lambda) - \lambda Q_{xZ} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1) \\ \Lambda_1 (W_{p+q}(1) - W_{p+q}(\lambda)) - (1-\lambda) Q_{xZ} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1) \end{pmatrix} \\ = & \frac{r}{\lambda^2} \mathbf{R}_1 Q_{xx}^{-1} \Lambda_1 W_{p+q}(\lambda) - \frac{r}{\lambda} \mathbf{R}_1 Q_{xx}^{-1} Q_{xZ} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1). \end{aligned}$$

Finally, premultiplying the third term in (A11) by $\mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1}$ gives

$$\begin{aligned} & \mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} T^{-1} \sum_{t=1}^{[rT]} \tilde{X}_t z_t' \left(\frac{Z'Z}{T} \right)^{-1} \left(T^{-1/2} Z'u \right) \\ = & \mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} T^{-1} \sum_{t=1}^{[rT]} \left(X_t - X'Z(Z'Z)^{-1} z_t \right) z_t' \left(\frac{Z'Z}{T} \right)^{-1} \left(T^{-1/2} Z'u \right) \\ \Rightarrow & \mathbf{R}_1 \left(\frac{1}{\lambda} Q_{xx}^{-1}, -\frac{1}{1-\lambda} Q_{xx}^{-1} \right) \times \begin{pmatrix} r(1-\lambda) Q_{xZ} \\ -r(1-\lambda) Q_{xZ} \end{pmatrix} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1) \\ = & \frac{r}{\lambda} \mathbf{R}_1 Q_{xx}^{-1} Q_{xZ} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1). \end{aligned}$$

Combining the results for the three terms gives

$$\begin{aligned} & \mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} T^{-1/2} \hat{S}_{[rT]}^{\tilde{X}} \Rightarrow \mathbf{R}_1 Q_{xx}^{-1} \Lambda_1 \frac{1}{\lambda} W_{p+q}(r) - \mathbf{R}_1 Q_{xx}^{-1} \Lambda_1 \frac{r}{\lambda^2} W_{p+q}(\lambda) \\ & + \mathbf{R}_1 Q_{xx}^{-1} Q_{xZ} Q_{ZZ}^{-1} \Lambda_2 \frac{r}{\lambda} W_{p+q}(1) - \mathbf{R}_1 Q_{xx}^{-1} Q_{xZ} Q_{ZZ}^{-1} \Lambda_2 \frac{r}{\lambda} W_{p+q}(1) \\ = & \mathbf{R}_1 Q_{xx}^{-1} \Lambda_1 \left(\frac{1}{\lambda} W_{p+q}(r) - \frac{r}{\lambda^2} W_{p+q}(\lambda) \right) = \mathbf{R}_1 Q_{xx}^{-1} \Lambda_1 \frac{1}{\lambda} F_{p+q}^{(1)}(r, \lambda). \end{aligned} \quad (\text{A13})$$

Similar results can be obtained for $\lambda \leq r \leq 1$:

$$\begin{aligned} & \mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} T^{-1/2} \sum_{t=1}^{[rT]} \tilde{X}_t u_t \Rightarrow \mathbf{R}_1 Q_{xx}^{-1} \Lambda_1 \left(\frac{1}{\lambda} W_{p+q}(\lambda) - \frac{1}{1-\lambda} (W_{p+q}(r) - W_{p+q}(\lambda)) \right), \\ & \mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} \left(T^{-1} \sum_{t=1}^{[rT]} \tilde{X}_t \tilde{X}_t' \right) \sqrt{T}(\hat{\beta} - \beta) \\ \Rightarrow & \mathbf{R}_1 Q_{xx}^{-1} \Lambda_1 \left(\frac{1}{\lambda} W_{p+q}(\lambda) + \frac{r-\lambda}{(1-\lambda)^2} W_{p+q}(\lambda) - \frac{r-\lambda}{(1-\lambda)^2} W_{p+q}(1) \right) - \mathbf{R}_1 Q_{xx}^{-1} Q_{xZ} Q_{ZZ}^{-1} \Lambda_2 \frac{1-r}{1-\lambda} W_{p+q}(1), \\ \text{and} \\ & \mathbf{R} \hat{Q}_{\tilde{X}\tilde{X}}^{-1} T^{-1} \sum_{t=1}^{[rT]} \tilde{X}_t z_t' \left(\frac{Z'Z}{T} \right)^{-1} \left(T^{-1/2} Z'u \right) \Rightarrow \frac{1-r}{1-\lambda} \mathbf{R}_1 Q_{xx}^{-1} Q_{xZ} Q_{ZZ}^{-1} \Lambda_2 W_{p+q}(1). \end{aligned}$$

Now combining the results for the three terms gives $\mathbf{R}\hat{\mathbf{Q}}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}^{-1}T^{-1/2}\hat{\mathbf{S}}_{[rT]}^{\tilde{\mathbf{z}}} \Rightarrow -\mathbf{R}_1\mathbf{Q}_{xx}^{-1}\Lambda_1 \cdot \frac{1}{1-\lambda}F_{p+q}^{(2)}(r, \lambda)$. Thus, we obtain for $r \in [0, 1]$,

$$\mathbf{R}\hat{\mathbf{Q}}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}^{-1}T^{-1/2}\hat{\mathbf{S}}_{[rT]}^{\tilde{\mathbf{z}}} \Rightarrow \mathbf{R}_1\mathbf{Q}_{xx}^{-1}\Lambda_1 \left(\frac{1}{\lambda}F_{p+q}^{(1)}(r, \lambda) - \frac{1}{1-\lambda}F_{p+q}^{(2)}(r, \lambda) \right).$$

□

Proof of Theorem 2. To save space, the proof for this Theorem is provided only for the case of the Bartlett kernel with $M = T$ (i.e., $b = 1$). However, the proof given here is applicable to other kernels and different values of b . Note that with $b = 1$, the HAC estimator can be rewritten as (see Kiefer and Vogelsang (2002a) [2]) $\hat{\Omega}_{b=1}^{(F)} = \frac{2}{T} \sum_{t=1}^{T-1} T^{-1/2}\hat{\mathbf{S}}_t^{\tilde{\mathbf{z}}}T^{-1/2}\hat{\mathbf{S}}_t^{\tilde{\mathbf{z}'}}$. With this HAC estimator, the term within the inverse in (24) is given by

$$\begin{aligned} & \frac{2}{T} \sum_{t=1}^{T-1} \left\{ \mathbf{R} \left(T^{-1} \sum_{s=1}^T \tilde{\mathbf{X}}_s \tilde{\mathbf{X}}_s' \right)^{-1} T^{-1/2} \hat{\mathbf{S}}_t^{\tilde{\mathbf{z}}} T^{-1/2} \hat{\mathbf{S}}_t^{\tilde{\mathbf{z}'}} \left(T^{-1} \sum_{s=1}^T \tilde{\mathbf{X}}_s \tilde{\mathbf{X}}_s' \right)^{-1} \mathbf{R}' \right\} \\ & \Rightarrow \mathbf{P} \left(1, \mathbf{R}_1\mathbf{Q}_{xx}^{-1}\Lambda_1 \left(\frac{1}{\lambda}F_{p+q}^{(1)}(r, \lambda) - \frac{1}{1-\lambda}F_{p+q}^{(2)}(r, \lambda) \right) \right) \end{aligned}$$

where the limit is obtained directly from Lemma 3 and the continuous mapping theorem. The result for (24) can be obtained by using similar arguments as those used in Theorem 1 where we use the transformation: $\mathbf{R}_1\mathbf{Q}_{xx}^{-1}\Lambda_1 W_{p+q}(r) \stackrel{d}{=} \Xi \cdot W_l(r)$, $0 \leq r \leq 1$ for a p.d. matrix $\Xi_{l \times l}$ satisfying $\Xi\Xi' = \mathbf{R}_1\mathbf{Q}_{xx}^{-1}\Lambda_1\Lambda_1'\mathbf{Q}_{xx}^{-1}\mathbf{R}_1'$.

References

1. Kiefer, N.M.; Vogelsang, T.J. A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests. *Econom. Theory* **2005**, *21*, 1130–1164.
2. Kiefer, N.M.; Vogelsang, T.J. Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica* **2002**, *70*, 2093–2095.
3. Kiefer, N.M.; Vogelsang, T.J. Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size. *Econom. Theory* **2002**, *18*, 1350–1366.
4. Hashimzade, N.; Vogelsang, T.J. Fixed- b Asymptotic Approximation of the Sampling Behavior of Nonparametric Spectral Density Estimators. *J. Time Ser. Anal.* **2008**, *29*, 142–162.
5. Jansson, M. The Error Rejection Probability of Simple Autocorrelation Robust Tests. *Econometrica* **2004**, *72*, 937–946.
6. Sun, Y.; Phillips, P.C.B.; Jin, S. Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing. *Econometrica* **2008**, *76*, 175–194.
7. Gonçalves, S.; Vogelsang, T.J. Block Bootstrap HAC Robust Tests: The Sophistication of the Naive Bootstrap. *Econom. Theory* **2011**, *27*, 745–791.
8. Sun, Y. *Fixed-Smoothing Asymptotics in a Two-Step GMM Framework*; Working Paper; Department of Economics, University of California, San Diego: La Jolla, CA, USA, 2013.
9. Andrews, D.W.K. Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* **1993**, *61*, 821–856.
10. Andrews, D.W.K.; Ploberger, W. Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative. *Econometrica* **1994**, *62*, 1383–1414.
11. Bai, J.S.; Perron, P. Estimating and Testing Linear Models with Multiple Structural Breaks. *Econometrica* **1998**, *66*, 47–78.
12. Perron, P. Dealing with structural breaks. *Palgrave Handb. Econom.* **2006**, *1*, 278–352.
13. Banerjee, A.; Urga, G. Modelling structural breaks, long memory and stock market volatility: An overview. *J. Econom.* **2005**, *129*, 1–34.
14. Aue, A.; Horváth, L. Structural breaks in time series. *J. Time Ser. Anal.* **2013**, *34*, 1–16.

15. Bai, J.; Perron, P. Computation and analysis of multiple structural change models. *J. Appl. Econom.* **2003**, *18*, 1–22.
16. Andrews, D.W.K. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* **1991**, *59*, 817–854.
17. Davidson, J. *Stochastic Limit Theory*; Oxford University Press: New York, NY, USA, 1994.
18. Phillips, P.C.B.; Durlauf, S.N. Multiple Regression with Integrated Processes. *Rev. Econom. Stud.* **1986**, *53*, 473–496.
19. Phillips, P.C.B.; Solo, V. Asymptotics for Linear Processes. *Ann. Stat.* **1992**, *20*, 971–1001.
20. Wooldridge, J.M.; White, H. Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econom. Theory* **1988**, *4*, 210–230.
21. DeJong, R.M.; Davidson, J. Consistency of Kernel Estimators of Heteroskedastic and Autocorrelated Covariance Matrices. *Econometrica* **2000**, *68*, 407–424.
22. Hansen, B.E. Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes. *Econometrica* **1992**, *60*, 967–972.
23. Jansson, M. Consistent Covariance Estimation for Linear Processes. *Econom. Theory* **2002**, *18*, 1449–1459.
24. Newey, W.K.; West, K.D. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* **1987**, *55*, 703–708.
25. Cho, C.K. Essays on Time Series Econometrics. Ph.D. Thesis, Department of Economics, Michigan State University, East Lansing, MI, USA, 2014.
26. Andrews, D.W. Tests for parameter instability and structural change with unknown change point: A corrigendum. *Econometrica* **2003**, *71*, 395–397.
27. Cho, C.K.; Vogelsang, T.J. *Fixed-b Inference for Testing Structural Change in a Time Series Regression*; Working Paper; Department of Economics, Michigan State University: East Lansing, MI, USA, 2014.
28. Crainiceanu, C.; Vogelsang, T.J. Non-monotonic Power for Tests of Mean Shift in a Time Series. *J. Stat. Comput. Simul.* **2007**, *77*, 457–476.
29. Schott, J.R. *Matrix Analysis for Statistics*; Wiley InterScience Publication: New York, NY, USA, 1997.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).