# Testing for Breaks Using Alternating Observations

Helle Bunzel
Iowa State University

Emma M. Iglesias
Michigan State University

March 30, 2009*

Abstract. This paper proposes several new tests for structural change in the multivariate linear regression model. One of the most popular alternatives are Sup-Wald type tests along the lines of Bai, Lumsdaine and Stock (1998), which Bernard, Idoudi, Khalaf and Yélou (2007) show to have very large size distortions, especially for high dimensional systems. They propose the use of Monte Carlo type tests to control for size in finite samples along the lines of Dufour and Khalaf (2002) and Dufour (2006). In this paper we propose several procedures that find a balance between the two previous approaches. We first estimate the break point using alternating observations, and then use the estimated breakpoint to create a test statistic either with the whole sample or with the observations not used for the breakpoint estimation. We show that these tests are optimal in the sense that it is possible to obtain the same local asymptotic power as we would obtain if the breakpoint was known. In addition, when observations used to estimate the breakpoint are not re-used for the testing, it is possible to use Monte Carlo methods to control size perfectly. In contrast to the Sup-Wald type tests, which have non-standard asymptotic distributions, we show that our tests are asymptotically distributed Chi-square using methods similar to those in Andrews (2004). Additionally, our tests stay asymptotically valid even when the distributional assumption made for the Monte Carlo adjustments is incorrect. We illustrate the new test statistics in the univariate context of discount rates and changes in the interest rates, and also in the multivariate setting of the Capital Asset Pricing Model.

*The second author acknowledges support from the MSU Intramural Research Grants Program. We are very grateful for useful comments from Donald Andrews, James Davidson, Miguel Delgado, Yanqin Fan, Jesus Gonzalo, Michael Jansson, Lynda Khalaf, Soumen Lahiri, Yoonseok Lee, Serena Ng, Carlos Velasco, Tim Vogelsang and participants at the 2006 North American Summer Meeting of the Econometric Society, the 2007 European Meeting of the Econometrics Society, the CREATES opening conference and seminars at Michigan State University, University of British Columbia, the University Carlos III, University of Michigan, University of Århus, University of Essex and Vanderbilt University. All remaining errors are our own.

## 1. Introduction

In this paper we consider tests for a single structural change in the multivariate linear regression model. Qu and Perron (2007) consider tests for multiple structural changes, however in this paper, for simplicity reasons, we will focus on a single break, although our methodology could also be extended to the multiple setting. Currently one of the most commonly used tests of structural change in these models is the procedure introduced in Bai, Lumsdaine and Stock (1998).[1] The test statistics used are sup-Wald tests and exponential Wald type tests which have non-standard but pivotal asymptotic distributions. Recently, Bernard et al (2007) have demonstrated that the sup-Wald tests can have severe size distortions, especially for high dimensional systems. To alleviate the size distortions, they propose using Monte Carlo type tests along the lines of Dufour and Khalaf (2002). While this approach does indeed provide excellent control for size, it requires knowledge of the finite sample distribution of the errors, and the test is not robust to incorrect assumptions about the distribution.

We propose new likelihood-ratio-based-procedures that find a balance between the two previous approaches. As Dufour and Khalaf (2002) note, there are some cases in which Monte Carlo tests will be asymptotically valid even under failure of the distributional assumptions. For that purpose, we need that: (1) "the assumptions used to derive an asymptotic distribution include as a special case the parametric distributional assumptions imposed in order to perform the Monte Carlo tests", and also that (2) "the asymptotic distribution of the test statistic does not involve unknown nuisance parameters". We design a test, *LRpart*, to exactly conform to these two desiderata.

Our proposed procedure is to first find an estimate of the break point with alternating observations, and then use it to create a likelihood ratio test either with the whole sample (*LRall*) or only with the observations not used in the first step (*LRpart*). Both the resulting test statistics are asymptotically Chi-squared distributed. The advantage of *LRall*

[1] This procedure builds on a long literature on breaks in the parameters of regression models, going back to Quandt (1960) and followed by, among many others, Davies (1977), Banerjee, Lumsdaine and Stock (1992), Zivot and Andrews (1992), Hansen (1992), Andrews (1993), Bai (1997) and Bai and Perron (1998).

is increased power resulting from the additional data used in the second step. In addition, our simulations show that *LRall* has much less size distortions than the tests in Bai et al (1998). The advantage of *LRpart* is that it allows the use of Monte Carlo tests to control the size of the test; this is needed mainly in the multivariate setting. Both these new test statistics provide significantly less size distortion than the sup-Wald test. While *LRpart* has significantly reduced power in moderately sized samples, the *LRall* test has power quite similar to the standard sup-Wald test.

One of the key arguments that we use to develop our new test procedures has been applied before in the literature (Andrews (2004)). As Andrews (2004, page 675) points out, in the context of the block bootstrap where $N$ is the sample size, $l$ is the size of the block and $\pi \in (0,1)$, we can design a setting where "...The last nonzero summand in one block is separated from the first summand in the next block by $[\pi l]$ time periods, where $[\pi l] \to \infty$ as $N \to \infty$. In consequence, for an asymptotically weakly dependent time series...the blocks are asymptotically independent". We use the same type of argument to create and prove the asymptotic independence that we need in our setting. In our case, in the first stage, we employ a shrinking function $\pi$ of the observations to obtain an initial estimate of the breakpoint, and use this estimate for Chow (1960)-type tests of the stability hypothesis in the second stage. The fact that the initial estimate is based on a shrinking fraction of the overall sample ensures asymptotic independence of the first stage estimator from the randomness of the statistic in the second stage. One thus obtains standard asymptotic chi-squared distributions of the test statistic under the null hypothesis. Because the tests can be interpreted as Chow tests with asymptotically known breakpoints, it is possible to show that they are optimal, and in fact simulations demonstrate that they can obtain the same local asymptotic power as would obtain if the breakpoint was known.

One important remark is the simplicity of our test procedures versus alternative ones, and the generality of our setting. For example, Elliott and Müller (2006) provide conditions under which the precise form of unstable processes is asymptotically irrelevant. However, they focus their attention in linear models with Gaussian errors. Andrews and Ploberger (1994) and Sowel (1996) have also derived optimal break tests, but they are also sup-type

tests, and as such, they too are subject also to the large size distortions in finite samples reported in Bernard et al (2007). Moreover, an important final remark is that our likelihood ratio based tests allow for standard trending regressors of polynomial form, and they still have the asymptotic chi-squared distribution.

We perform extensive simulations to evaluate the performance of the new test statistics under various circumstances. Relying on these simulations, we recommend that in the single equation environment, the researcher should use *LRall*. In multidimensional systems, given the large size distortions, the researcher should use *LRpart* combined with a Monte Carlo type procedure. The *LRpart* statistic combined with the Monte Carlo procedure does require that we assume a distribution on the errors, but we have demonstrated that if the distributional assumption is wrong, the test statistic will stay asymptotically valid. In fact, simulations indicate that *LRpart* has excellent size control even when the distribution differs from the assumed one.

In addition to the likelihood-ratio based statistics, we also verify the asymptotic distribution of the equivalent Wald tests with alternating observations. These have the advantage of allowing the use of nonparametric heteroscedasticity and autocorrelation consistent (HAC) estimators (see e.g. Newey and West (1987)) if serial correlation is present.

We apply our test statistic to two empirical examples, one univariate and one multivariate. The univariate illustration replicates part of the work of Bai (1997) in which he examines the impact of changes in the discount rate on the market interest rate. The multivariate illustration examines, for the first time, simultaneous breaks in a 5-variate CAPM model. Our main finding here is that our statistics definitely find fewer breaks than the statistic of Bai et al (1998), as is expected considering the inflated size of the sup-Wald statistic. In the univariate framework our test statistic and that of Bai et. al (1998) are not too different in their behavior, but when we move to the multivariate framework, the sup-Wald type test finds 15 breakpoints in 954 observations, whereas we find only 6 with the LR-based tests. The breaks found with the sup-Wald statistic are spread over the entire series, whereas the breaks we locate are all in the first quarter of the sample, where the data is clearly more unstable. Furthermore, *LRpart* and *LRall* are in agreement in

spite of their differing size and power properties.

The outline of the paper is as follows. Section 2 sets up the model, provides the assumptions and the asymptotic distributions of the test statistics, Section 3 provides simulation results, Section 4 provides several empirical applications and Section 5 concludes. Proofs are relegated to the appendices.

## 2. Theory

**2.1. The Framework.** We consider the multiple linear regression model

$$Y = XB + U \tag{1}$$

where $Y = [Y_1, Y_2, ..., Y_T]'$ is a $T \times n$ matrix of $T$ observations on $n$ dependent variables, $X$ is a $T \times k$ full column rank matrix of regressors, where we can allow for lagged dependent variables and $U = [U_1, ..., U_T]'$ is a $T \times n$ matrix of error terms.

Bai, Lumsdaine and Stock (1998), as a generalization of Bai (1997), use the following augmented version of (1) to test for change points

$$Y = XB + D^s\Delta^s + U = Z^s\Theta^s + U, \qquad Z^s = \begin{bmatrix} X & D^s \end{bmatrix}, \quad \Theta^s = \begin{bmatrix} B \\ \Delta^s \end{bmatrix}, \tag{2}$$

where $D^s$ is a matrix with typical row equal to $D_t^s\overline{X}_t'$, where $D_t^s$ is the dummy variable given by

$$\begin{aligned} D_t^s &= 1, \quad t > s \\ &= 0, \quad t \leq s, \end{aligned}$$

and $\overline{X}_t'$ is the $t'$th row of $\overline{X} = XQ_X$ with $Q_X$ being the $k \times q_X$ selection matrix (of zeros and ones) which specifies which regression coefficients are tested for constancy. Finally $s \in [[\tau T] + 1 : T - [\tau T]]$ is the break date and $\tau < \frac{1}{2}$ is a trimming parameter typically between 0.05 and 0.3. We will denote the true breakdate $s^0$ and assume that there exists a $\lambda_0$ such that $0 < \lambda_0 < 1$ and $s^0 = [\lambda_0 T]$ ([.] is the integer part function). Finally, we denote $Z^{s^0}$ by $Z^0$.

To test for the presence of a change point in this model, we have to test the null hypotheses

$$H_0^{s*} : \Delta^s = 0 \Longleftrightarrow R^* \Theta^s = 0,$$

where $R^* = \begin{bmatrix} 0_{q_X \times k} & I_{q_X} \end{bmatrix}$ and $0_{l \times m}$ and $I_{q_X \times k}$ denotes an $l \times m$ matrix of zeros and the identity matrix respectively. Combining these hypotheses into a single null, they can be written as

$$H_0^* : \Delta^s = 0 \ \forall s \Longleftrightarrow \cap_{s \in [T_* + 1 : T - T_*]} \left( H_0^{s*} \right).$$

There are a number of test statistics currently available to test this hypothesis. Their expressions are given below.

First the likelihood ratio based test can be written as

$$\Lambda^* = \sup_{s \in [[\tau T] + 1 : T - [\tau T]]} \left\{ -T \ln \left( \Lambda^s \right) \right\}, \qquad \Lambda^s = \left| S^s \right| / \left| S^0 \right|,$$

where

$$S^s = \widehat{U}^s \widehat{U}^s, \qquad S^0 = \widehat{U}^0 \widehat{U}^0,$$

and $\widehat{U}^0$ and $\widehat{U}^s$ are the ordinary least squares (OLS) residuals from (1) and (2) respectively. Similarly, the test based on the Wald statistic can be written as

$$\mathcal{F}^* = \sup_{s \in [[\tau T] + 1 : T - [\tau T]]} \left\{ \mathcal{F}^s \right\},$$

where

$$\mathcal{F}^s = T \cdot trace \left( \left( S^s \right)^{-1} \left( S^0 - S^s \right) \right),$$

or, for a more standard Wald representation that allows for serial correlation and heteroscedasticity in the data

$$W^s = T \left( R^* \hat{\Theta}^s \right)' \left( R^* \left( T^{-1} \sum_{t=1}^{T} Z_t^s \left( \hat{\Sigma}^s \right)^{-1} \left( Z_t^s \right)' \right)^{-1} \left( R^* \right)' \right)^{-1} \left( R^* \hat{\Theta}^s \right),$$

and $\hat{\Sigma}^s$ is the estimator of $\Sigma$, the variance covariance matrix of $U$, based on OLS residuals under the alternative hypothesis, given $s$, $Z_t^s$ refers to the $t'$th row of the matrix $Z\left(\hat{\Sigma}^s\right)^{-1}$ containing the regressors and

$$W^* = \sup_{s\in[[\tau T]+1:T-[\tau T]]} \{W^s\}.$$

Bernard et al (2007) suggest using a test statistic which can be written as

$$LP(\Lambda) = -2 \sum_{s\in[[\tau T]+1:T-[\tau T]]} \ln\left(pv\left[\Lambda^s\right]\right),$$

where

$$pv\left[\Lambda^s\right] = G_F\left(\left[\frac{1-\left(\Lambda^s\right)^{1/m_2}}{\left(\Lambda^s\right)^{1/m_2}}\frac{m_1m_2-2m_3}{nq_X}\right]|nq_X, m_1m_2-2m_3\right),$$

$$m_1 = T-(k+q_X)-\frac{n-q_X+1}{2}, \quad m_2 = \sqrt{\frac{n^2q_X^2-4}{n^2+q_X^2-5}}, \quad m_3 = \frac{nq_X-2}{4},$$

and $G_F(x|v_1,v_2)$ is the survival function, evaluated at point $x$, of the $F$ distribution with $(v_1,v_2)$ degrees of freedom.

Bernard et al (2007) examine the properties of the Bai et. al. (1998) tests and show that these tests have very large size distortions. Therefore, they propose that exact versions of the test statistics should be used. Specifically, Bernard et al (2007) propose Monte Carlo (MC) exact tests of $\Lambda^*$ and $LP(\Lambda)$ (which we will denote as $\Lambda^*_{MC}$ and $LP(\Lambda_{MC})$ respectively), where we have to draw $N$ realizations under the null hypothesis from the distributional assumption we impose on the residuals, and then we derive the (empirical) p-values:

$$\widehat{p}_N(.) = \frac{N\widehat{G}_N(.)+1}{N+1}$$

where $N\widehat{G}_N(.)$ is the number of simulated values greater than or equal to the observed value of the test statistics. See Dufour (2006) for more details.

The proposal of Bernard et al (2007) provides a large improvement when the distributional assumptions about the errors are correct. As they show in their simulations, the

gains of controlling for size in this setting can be huge in relation to Bai et al (1998) test. However, the Bernard et al (2007) proposal have one main drawback: if their distributional assumptions are wrong, then in general their tests are not valid. We overcome this problem by introducing tests that are valid asymptotically, even if the assumption about the distribution is incorrect. In the next section we present our test statistics and their asymptotic distributions.

**2.2. The test statistics.** In this paper we propose a set of test statistics, that can be considered a "middle position" between Bai et al (1998) and Bernard et al (2007) statistics. We design two types of split-sample LR-based tests. Both tests follow a chi-square distribution asymptotically; additionally one allows control for size in small samples. The two tests (denoted *LRpart* and *LRall*, depending on whether or not we use the whole sample in the second stage) are defined below, but first we need the following notation: Select $J$ observations by picking out alternating observations that correspond to a fraction of $\pi_T^1$ of the total sample size (for example, if $\pi_T^1 = 1/3$ and $T = 99$, pick observations $3, 6, 9, ..., 99$ for a total of $J = 33$ observations). Denote this set of observations $\mathbb{N}_J$. Now let $\mathbb{N}_R$ denote the remaining $R = T - J$ observations. Then we construct the following statistics:

**LRpart**. Use the observations in $\mathbb{N}_J$ to get a consistent estimate of $\lambda = [sJ]$. With this estimate of the break, $\left(\hat{\lambda}_J\right)$, apply a traditional LR test using the remaining $R$ observations, such that

$$LRpart\left(S_{\mathbb{N}_R}\left(\hat{\lambda}_J\right), S^0_{\mathbb{N}_R}\right) = -R\ln\left(\frac{\left|S^{\hat{\lambda}_J}_{\mathbb{N}_R}\right|}{\left|S^0_{\mathbb{N}_R}\right|}\right),$$

where the $\mathbb{N}_R$ in the subscript signifies that the observations in $\mathbb{N}_R$ are used for calculating the sum of squared residuals.

**LRall**. Use the observations in $\mathbb{N}_J$ to get a consistent estimate of $\lambda = [sJ]$. With this estimate of the break, $\left(\hat{\lambda}_J\right)$, apply a traditional LR test using all the observations, such

that

$$LRall\left(S\left(\hat{\lambda}_J\right), S^0\right) = -T\ln\left(\frac{\left|S^{\hat{\lambda}_J}\right|}{|S^0|}\right),$$

where we do not use any subscript on $S^0$ and $S^{\hat{\lambda}_J}$ since all the observations are used.

In addition to these two test statistics, we introduce a third test statistic, *LRblock*, which is constructed purely for use in the proofs of Theorem 2 and 3.

**LRblock**. Use the observations in $\mathbb{N}_J$ to get a consistent estimate of $\lambda = [sJ]$. Now calculate the LR statistic on a subset $\mathbb{N}_C$ of $\mathbb{N}_R$. Let $0 < \pi_T^2 \leq \frac{1}{2}$. Then $\mathbb{N}_C$ is the observations in $\mathbb{N}_R$, except for the $\left[R\pi_T^2\right]$ observations which are closest (before and after) to the observations in $\mathbb{N}_J$ (see Figure 1 for an example of how to choose $\mathbb{N}_J$, $\mathbb{N}_R$ and $\mathbb{N}_C$). Thus *LRblock* is defined as

$$LRblock\left(S_{\mathbb{N}_C}\left(\hat{\lambda}_J\right), S^0_{\mathbb{N}_C}\right) = -\left(R - \left[R\pi_T^2\right]\right)\ln\left(\frac{\left|S^{\hat{\lambda}_J}_{\mathbb{N}_C}\right|}{\left|S^0_{\mathbb{N}_C}\right|}\right). \tag{3}$$

$W_{block}$, $W_{all}$ and $W_{part}$ are defined in a parallel way and their exact expressions can be found in Appendix C.

To obtain asymptotic distributions and use the finite sample Monte Carlo methods, we need various combinations of the following assumptions:

A1 a) $\{U_t\}_{t=1}^T$ is iid with finite-valued covariance terms.

A1 b) With $\{\mathcal{F}_i : i = 1, 2, ...\}$ a sequence of increasing $\sigma$-fields assume that $\{U_i, \mathcal{F}_i\}$ forms a $L^r$-mixingale sequence with $r = 4 + \delta$ for some $\delta > 0$ (McLeish (1975) and Andrews (1993)) with finite-valued covariance terms.

A2 Assume that there exists an $l_0 > 0$ such that for all $l > l_0$, the minimum eigenvalues of $A_{1l} = \frac{1}{l}\sum_{t=1}^{l} X_t X_t'$, $A_{1lJ} = \frac{1}{l}\sum_{t=1,\ t\in\mathbb{N}_J}^{l} X_t X_t'$, $A_{1lR} = \frac{1}{l}\sum_{t=1,\ t\in\mathbb{N}_R}^{l} X_t X_t'$, $A_{2l} = \frac{1}{l}\sum_{t=[\lambda_0 T]}^{[\lambda_0 T]+l} X_t X_t'$, $A_{2lJ} = \frac{1}{l}\sum_{t=[\lambda_0 T],\ t\in\mathbb{N}_J}^{[\lambda_0 T]+l} X_t X_t'$, $A_{2lR} = \frac{1}{l}\sum_{t=[\lambda_0 T],\ t\in\mathbb{N}_R}^{[\lambda_0 T]+l} X_t X_t'$ are bounded away from zero in probability. Finally we assume that the matrix $B_{lk} =$

$\sum_l^k \overline{X}_t \overline{X}_t{}'$ is invertible for $l-k \geq q_X$. Furthermore, if a breakpoint exists, let $\bar{Z}^0 = diag(Z_1^0, Z_2^0)$, where $Z_1^0$ and $Z_2^0$ partitions $Z$ at the true breakpoint. We assume that $\left(Z_1^0\right)' \left(Z_1^0\right) / [\lambda T]$ and $\left(Z_2^0\right)' \left(Z_2^0\right) / (T - [\lambda T])$ as well as the corresponding matrices formed with alternating observations converge in probability to non-random positive definite matrices. Finally, we assume that the errors $U_t$ are independent of the regressors $X_s$ for all $t, s$.[2]

A3 $J = \left[\pi_T^1 T\right]$, $R = \left[\left(1-\pi_T^1\right) T\right]$, $0 < \pi_T^1 < 1$, $(J, R) \to \infty$.

A4 $\left[T \ln(1+\pi_T^1)\right] \to 0$, $\left[\pi_T^1 T\right] \to \infty$.

A5 $\left[\frac{R\pi_T^2}{J}\right] \to \infty$, $0 < \pi_T^2 \leq 1/2$ and $\pi_T^2 \to 0$.

A6 The (Quasi-) log-likelihood function is regular.[3]

A7 $T^{-1} \sum_{t=1}^{[vT]} \overline{X}_t \Sigma^{-1} \overline{X}_t' \xrightarrow{p} Q(v)$ and $T^{-1} \sum_{t=1}^{[vT]} Z_t \Sigma^{-1} Z_t' \xrightarrow{p} Q'(v)$ uniformly in $v \in [0,1]$, where $Q(v)$ and $Q'(v)$ are positive definite for $v > 0$ and strictly increasing in $v$.

A8 $\Delta^{s^0}$ depends on $T$ and can be written as $\Delta^{s^0} = \Delta_0 \cdot v_T$ where $v_T$ is a positive number such that $v_T \longrightarrow 0$ and $\left[\pi_T^1 T\right]^{(1/2-\alpha)} v_T \to \infty$ for some $\alpha \in (0, 1/2)$ and $\Delta_0 \neq 0$.

Assumption 1 specifies the assumptions on the errors. A1 a) is typically not valid for time series data, but the proof using the simple assumption is instructive and it is an assumption often made when using finite sample Monte Carlo methods. The alternative assumption, A1 b) follows Bai and Perron (1998). This assumption is fairly general in that it allows for broad ranges of serial correlation and heteroscedasticity, and we can also allow for lagged dependent variables in the regression. Note that the crucial assumption is that we need to have an error term that is weakly dependent. For a detailed definition of the $L^r$-mixingale see Bai and Perron (1998). Assumption A2 contains the standard assumptions

[2] Note that this assumption along with A1 b) can be easily replaced by an assumption that the errors form a martingale difference sequence, along with an imposition of a minimum partion length. See Bai and Perron (1998, Assumption A4 and related discussion).

[3] See for example Greene (2003) for specific conditions.

on the regressors of a multivariate regression model as well as assumptions ensuring that there is enough data surrounding the breakpoint for the breakpoint to be identified. A3 states that $J$ and $R$ go to infinity and that both $J$ and $R$ are fractions of the sample size. In the proofs we may allow for going to infinity sequentially, although this assumption can be relaxed due to the asymptotic independence. A4 ensures that the number of observations used to estimate the breakpoint goes to infinity, but sufficiently slowly to ensure asymptotic irrelevance. A5 ensures that the number of observations left out when calculating *LRblock* goes to infinity, but sufficiently slowly to allow the test statistic to remain consistent. A6 is required to ensure that we can apply the standard Taylor expansion to the various (quasi) log-likelihood functions. A7 is required to ensure that the Wald statistic converges to the correct limit when the observations are not iid. This is required because, in that case, the Wald statistic cannot be based purely on residuals, and as a result we need restrictions on our regressors. Note that this assumption allows for trending regressors written as any function of the time trend $g(t/T)$ as in Bai (1997). Finally A8 is needed to obtain the rate of convergence of the breakpoint estimator and, as a result, root $T$ convergence of the regression parameter estimates when a break is present.

We now consider the following theorems, where the first is for independent identically distributed data and the second and third are more general. Note that "$\Rightarrow$" denotes convergence in distribution.

**Theorem 1.** ***(a)*** *Under (1), and under A1a), A2, A3 and A6 the LRpart test*

$$LRpart_{J,R}\left(\hat{\lambda}_J\right) \Rightarrow \chi^2(q_X)$$

*where $\chi^2(q_X)$ denotes a chi-square distribution with $q_X$ degrees of freedom.*

***(b)*** *Under (1) and under A1a), A2, A3, A4 and A6 the LRall test*

$$LRall_{J,T}\left(\hat{\lambda}_J\right) \Rightarrow \chi^2(q_X)$$

*where $\chi^2(q_X)$ denotes a chi-square distribution with $q_X$ degrees of freedom.*

***(c)*** *Under (1) and if $U_t = HW_t$ where $t = 1, ..., T$ and $H$ is unknown, nonsingular and the distribution of the error $w = vec(W_1, ...W_T)$ is known, a Monte Carlo version of the*

*LRpart test based on the ratio of residual sums of squares will be invariant to the choice of the parameters in $B$ and it will be exact in small samples.*

**Proof.** Given in Appendix A. ■

Part (a) of this Theorem is very clear. Since the observations used to obtain a consistent estimate of the breakpoint and the observations used to test for the presence of a breakpoint are independent, it is no surprise that we obtain a Chi-squared distribution asymptotically. Part (b) is proven simply by verifying that *LRpart* and *LRall* are asymptotically identical in a probabilistic sense. Part (c) states the invariance property of LRpart and that this test has all the characteristics required to be exact when MC is applied. We now present the parallel theorem for the dependent case. We make use of a Feasible Generalized Least Squares (FGLS) procedure that can take into account possible serial correlation and heteroskedasticity.

**Theorem 2.** ***(a)*** *Under (1) estimated with Feasible Generalized Least Squares (FGLS)*[4] *and A1b), A2, A3, A4, A5 and A6, the LRblock test*

$$LRblock_{J,R-\left[R\pi_T^2\right]}\left(\hat{\lambda}_J\right) \Rightarrow \chi^2\left(q_X\right)$$

*where $\chi^2\left(q_X\right)$ denotes a chi-square distribution with $q_X$ degrees of freedom.*

***(b)*** *Under (1) estimated with FGLS, and under A1b), A2, A3, A4 and A6, the LRpart test*

$$LRpart_{J,R}\left(\hat{\lambda}_J\right) \Rightarrow \chi^2\left(q_X\right)$$

*where $\chi^2\left(q_X\right)$ denotes a chi-square distribution with $q_X$ degrees of freedom.*

***(c)*** *Under (1) estimated with FGLS and under A1b), A2, A3, A4 and A6, the LRall test*

$$LRall_{J,T}\left(\hat{\lambda}_J\right) \Rightarrow \chi^2\left(q_X\right)$$

*where $\chi^2\left(q_X\right)$ denotes a chi-square distribution with $q_X$ degrees of freedom.*

[4] For a thorough description of FGLS see Greene (2003, Chapter 15)

**Proof.** Given in Appendix B. ■

In part (a) of this Theorem it is again easy to obtain the asymptotic Chi-squared distribution. By construction, the number of observations separating the observations used for estimating the breakpoint and the observations used to calculate *LRblock* goes to infinity. As a result, the observations in $\mathbb{N}_C$ and $\mathbb{N}_J$ are asymptotically independent, and the Chi-square distribution follows easily. Parts (b) and (c) are then proven by verifying that the test statistics are asymptotically identical in a probabilistic sense. An important observation is that we do not require A7 in Theorems 1 and 2, and thus we allow for standard trending regressors of polynomial form.

It is worth noting in the theorem above that we require the use of FGLS. This is because we are limiting ourselves to the LR form of the statistic. In the next theorem, we provide the asymptotic distributions of the Wald form of the test statistic. The Wald form is important because it allows use of non-parametric covariance matrix estimates such that the serial correlation can be of completely unknown form (allowing for HAC estimation). On the other hand, the Wald form of the statistic only allows for trending regressors of the form $g(t/T)$, hence the LR form has an advantage in that dimension. Note also that this statistic is what the literature calls the robust likelihood-ratio-based test (see for example Stock and Watson (1996)).

**Theorem 3.** ***(a)*** *Under (1) and A1b), A2, A3, A4, A5 and A7, the $W_{block}$ test defined in Appendix C*

$$W_{block\ J,R-[R\pi_T^2]}\left(\hat{\lambda}_J\right) \Rightarrow \chi^2\left(q_X\right)$$

*where $\chi^2\left(q_X\right)$ denotes a chi-square distribution with $q_X$ degrees of freedom.*

***(b)*** *Under (1), and under A1b), A2, A3, A4 and A7,, the $W_{part}$ test defined in Appendix C*

$$W_{part\ J,R}\left(\hat{\lambda}_J\right) \Rightarrow \chi^2\left(q_X\right)$$

*where $\chi^2\left(q_X\right)$ denotes a chi-square distribution with $q_X$ degrees of freedom.*

***(c)*** *Under (1), and under A1b), A2, A3, A4 and A7,, the* $W_{all}$ *test defined in Appendix C*

$$W_{all\ J,T}\left(\hat{\lambda}_J\right) \Rightarrow \chi^2\left(q_X\right)$$

*where* $\chi^2\left(q_X\right)$ *denotes a chi-square distribution with* $q_X$ *degrees of freedom.*

**Proof.** Given in Appendix C. ■

The theorems above confirm that *LRpart*, *LRall*, $W_{part}$, and $W_{all}$ provide test statistics with standard asymptotic distributions under fairly general assumptions. The finite sample Monte Carlo procedure will produce exact results for *LRpart* in the independent case, providing that we know the correct distributional assumptions of the disturbances. In some cases, we can also get exact finite sample results when we apply the MC procedure to *LRpart* with dependent data. What is required in that case is that we must choose $\mathbb{N}_J$ in such a way that *conditional* on $\mathbb{N}_J$, the observations in $\mathbb{N}_R$ are independent. For example, if we have an AR(1) process, but we use the odd observations for the first stage and the even observations in the second stage, then we will still have an exact test after applying the MC method (we can adopt the results of Dufour and Jasiak (2001) and Dufour and Kiviet (1996) directly). It is important to remember, however, that even if we have a type of dependence in the data that does not produce an exact test, the test will still be asymptotically valid when we apply the MC method.

The main advantages of the new test statistics are the following: (1) they follow an asymptotic $\chi^2$ distribution. (2) *LRpart* allows a Monte Carlo version that will control for size in case the distributional assumption that we impose on the errors is correct; and more importantly, in case the distributional assumption that we impose is incorrect, we will fall back on the asymptotic distribution if we do not have any additional nuisance parameters. (3) *LRall* has better size control than the Bai et al test (1998) both in univariate and multivariate settings with similar power (see next section for this result).

Finally, in relation to asymptotic power, note that under Assumption A8 we trivially have power against alternatives of order $\left(\pi_T^1 T\right)^{-\frac{1}{2}}$. It takes a little more to demonstrate that we have power against the $(T)^{-\frac{1}{2}}$ alternative, which is standard in the literature. The

reason is that we only obtain consistent estimates of the breakpoint if $\alpha$ as defined in A8 is restricted to $(0, 1/4)$. In this case it is straightforward to verify that the tests presented in this paper have power against alternatives of order $(T)^{-\frac{1}{2}}$ *and* in fact that the tests are asymptotically locally optimal. These results are gathered in the following Theorem, which is proven in the appendix. In the next section, we provide a simulation study of the power and size properties of these procedures.

**Theorem 4.** ***(a)*** *Under (1) estimated with Feasible Generalized Least Squares (FGLS) and A1b), A2, A3, A4, A5, A6 and A8 where $\alpha \in (0, 1/2)$, the LRblock test has power greater than size against alternatives of order $(T)^{-\frac{1}{2}}$ and is asymptotically locally optimal when $\alpha \in (0, 1/4)$ such that the estimate of the break point in the first stage is consistent.*

***(b)*** *Under (1) estimated with FGLS, and under A1b), A2, A3, A4, A6 and A8 where $\alpha \in (0, 1/2)$, the LRpart test has power greater than size against alternatives of order $(T)^{-\frac{1}{2}}$ and is asymptotically locally optimal when $\alpha \in (0, 1/4)$ such that the estimate of the break point in the first stage is consistent.*

***(c)*** *Under (1) estimated with FGLS and under A1b), A2, A3, A4, A6 and A8 where $\alpha \in (0, 1/2)$, the LRall test has power greater than size against alternatives of order $(T)^{-\frac{1}{2}}$ and is asymptotically locally optimal when $\alpha \in (0, 1/4)$ such that the estimate of the break point in the first stage is consistent.*

**Proof.** Given in Appendix D. ■

## 3. Simulation Results

Note that in the asymptotics we require $\pi_T^1$ be chosen such that $\left[\pi_T^1 T\right] \to \infty$ and $\pi_T^1 \to 0$. In practice we select $\pi_T^1 \leq 1/2$. For the purpose of these simulations, we choose $\pi_T^1 = 1/3$ (see how Dufour and Jasiak (2001) and Dufour and Iglesias (2008) also select a smaller sample size for the first part of their split sample procedure). The simulation results are produced using GAUSS with 5000 repetitions for the general simulations and 99 artificial datasets for the Monte Carlo simulations. For all simulations and procedures, the trimming parameter, $\tau$, is set to 0.15. The tables containing the simulation results can be found in Appendix E. Tables 1 and 2 correspond to Tables 1 and 2 in Bernard et al (2007). The

model considered is a special case of (2) where only an intercept and a time trend (Table 1) and an intercept and a standard normal variate (Table 2) are present. The break may occur in the regression intercept, and the parameter $\xi_0$ controls the magnitude of the break. The values that we consider are 1.5, 5 and 10. The regression errors are drawn as standard multivariate normal in all the experiments except in Table 11 where we use a t-distribution. In all our simulations, we consider a one time break at dates $s_0 = [.5T] + 1$, $[.85T]$ and $[.95T]$, where [.] is the integer part function. Tables 1 and 2 clearly show the dangers of applying the Bai et al (1998) test, especially for $n > 1$, and how $\Lambda^*_{MC}$ of Bernard et al (2007) allow for full control of the size in finite samples. We have also augmented those tables to show the performance of the *LRpart* statistic with MC finite sample adjustments, when the distributional assumption imposed in the errors is correct, and as expected this statistic allows for full control of size. Both in the case of *LRpart* and *LRall*, we estimate the breakpoint with ordinary least squares with the first part of the sample.

Tables 3 and 4 show the power results comparing the procedure of Bernard et al (2007) and our *LRpart* with the Monte Carlo procedure applied. Basically, *LRpart* needs around 180 observations, to start to have similar power to that of Bernard et al (2007) with 80 observations. However, note that with 180 observations the asymptotic Bai et al (1998) test still has very large size distortions (Tables 1 and 2) and the results of Bernard et al (2007) are not robust to failures in the distributional assumption. Specifically, the power gains of Bernard et al (2007) demonstrated in Tables 3 and 4 can be viewed as the power that is gained from using the knowledge of the finite sample distribution. Clearly, any procedure which does not presume such knowledge cannot hope to obtain similar power. As demonstrated by Tables 1-4, our procedure finds a balance between the finite sample approach, which assumes knowledge of the finite sample distribution and the purely asymptotic approach, which leads to severe finite sample size distortions.

Tables 5 and 6 show that *LRpart* (w/o the MC correction) and *LRall* produce much less size distortions than the procedure of Bai, Lumsdaine and Stock (1998) (compare with Table 2). But even so, size distortions are non-trivial, especially for $n > 1$, which justifies the need for the MC procedure in finite samples. Therefore, our recommendation is to use

*LRall* for systems where $n = 1$ and *LRpart* with the MC method for $n > 1$.[5]

Tables 7 and 8 show the asymptotic power results of *LRpart* and *LRall*. We have already noted that the size control of *LRpart* and *LRall* is much better than the Bai et al (1998) test. Table 7 shows that in general Bai et al (1998) test has a virtually identical power performance to *LRall*. Note also that Bai et al test (1998) requires the use of different critical values depending on whether we have a time trend or a normal regressor.[6] The distribution of our LR tests remains the same in both cases.

To explore the effect of choosing different $\pi_T^1$ values, in Table 9 we show power results of *LRall* and *LRpart* with different values of $\pi_T^1$. Here we have chosen a sample size of 1000 for several reasons. First, when we have very few observations, $\pi_T^1$ must be $\frac{1}{2}$ or $\frac{1}{3}$. If it is any smaller, there will not be enough observations left to estimate the breakpoint, therefore the choice of $\pi_T^1$ only becomes interesting when we have more observations. 1000 was chosen because our second empirical application in Section 4.2 has 950 observations, so simulations with 1000 observations could potentially provide some guidance about which value of $\pi_T^1$ to choose. The table demonstrates that the power does not vary much with $\pi_T^1$, and hence the applied econometrician does not need to worry too much about the specific value of $\pi_T^1$. This is a very nice result, since the theory cannot guide our choice of $\pi_T^1$ for any individual sample size.

Another issue we examine is what happens when we assume an incorrect distribution of the errors. In our simulations reported in Table 10, we assume that the disturbance follows a N(0,1) distribution when in fact it follows a t-distribution. This assumption affects the statistic when we use the MC procedure. The first thing to note is that, at least for this example, the $\Lambda_{MC}^*$ of Bernard et al (2007) becomes highly conservative. We run three cases for samplesizes 40, 80, 120, 180, they are: $n = 1$ and $t(5)$ errors, $n = 5$ and $t(5)$ errors, and $n = 5$ and $t(35)$ errors. For each of those cases, even when we use the $t(35)$ errors which are reasonable close to a normal distribution, the $\Lambda_{MC}^*$ has actual size 0. The standard sup-Wald statistic, $\Lambda^*$, actually has somewhat better size in small samples than when the

[5] If sample size is high, it might be sensible to use LRall for slightly higher $n$ as well.

[6] Critical values for models without trending regressors are tabulated in Andrews (1993, 2003). To our knowledge tables of critical values for models with trending regressors are not available.

errors are normal, indicating that with the $t-$distribution we have chanced upon a data generating process where the asymptotic distribution is a better approximation than with the normal distribution data generating process. The actual size of $\Lambda^*$ seems not to change much as the degrees of freedom in the $t-$distribution changes, but it still does substantially better when the system is unidimensional. The *LRall* statistic, which also relies solely on the asymptotic distribution, performs similarly or slightly worse than the $\Lambda^*$ statistic. *LRpart* with the MC procedure applied performs perfectly with the $t(35)$ distribution, and for small sample sizes and $n = 5$, $t(5)$ it performs better than the asymptotic statistics. It is noteworthy however that the convergence to the nominal size is slower as sample size increases. When $n = 1$, the purely asymptotic tests have a slight advantage. In conclusion, Table 10 demonstrates that *LRpart* strikes the balance we were hoping for: When the distribution is sufficiently close to the one we assume for the simulations we get very, very good size performance, and even in the cases where the distribution is wrong, the asymptotics kick in and the performance is still decent.

Table 11 considers what happens when lagged dependent variables are present in the system. It shows the asymptotic size of *LRpart* when we use the critical values from a $\chi^2$ distribution and we use a model with an intercept, a normal regressor and a lagged dependent variable (an autoregressive process of order 1: AR(1)). The simulations demonstrate that for a nominal size of 5%, even with 80 observations *LRpart* does not present large size distortions. Note that to deal with moving average processes of finite order, the block-statistics defined in Section 2 would be useful in practice.

Finally, in Figure 2 we consider the local asymptotic power of our *LRpart* test. Theorem 4 in the previous section states the optimal properties of the local asymptotic power of our tests. Figure 2 shows the local asymptotic power characterized using an alternative of the form $CONSTANT/\sqrt{T}$ of the SupWald test (SUPWALD), the test with known breakpoint (KNOWNBREAK) and the value of $\pi_T^1$ (denoted PI) that gives same power for the *LRpart* as the SupWald test. As expected, the power envelope for the test with known break point is larger than that of the Sup-Wald test. The true break is simulated in the middle of the sample. The number of replications is 10000 and $T = 1000$. The nominal size in the

simulations is 5%. As we can see from Figure 2, the value of $\pi_T^1$ is mostly around 1/3. This indicates that using 1/3 in practice may be a good rule of thumb if one is satisfied with the level of power of the Sup Wald test. 1/3 will provide the same asymptotic power, but improved size properties. The value of CONSTANT in Figure 2 is from 0 to 15.[7]

Our general recommendations for application of the statistics based on the simulation results presented above is as follows: Since *LRpart* involves loses in power in finite samples in relation to *LRall*, and according to our simulations, for $n = 1$, *LRall* provides good size control in finite samples, we advice that *LRall* be used in practice for $n = 1$, while *LRpart* with the Monte Carlo procedure be used for higher dimensional systems.

## 4. Empirical Illustrations

We consider two examples. The first example is a univariate framework, where we re-estimate the breaks in the data of Bai (1997). He estimates the relations between changes in discount rates and changes in the market interest rate. In the other example, we estimate the Capital Asset Pricing Model (CAPM) on 5 different return series and test for simultaneous breaks in these 5 series.

**4.1. Interest rate changes.** We consider the empirical example in Bai (1997) where the following linear regression describes the relationship between the change in the discount rate for the *ith* observation ($\Delta DR_i$) and the change in the market interest rate ($\Delta TB_i$)

$$\Delta TB_i = \alpha + \beta \Delta DR_i + \varepsilon_i$$

Bai (1997) considers the same data as given in Dueker (1992), where the sample period covers from 1973 to 1989 and there are in total 56 observations. Bai (1997) applies his Wald type test detecting breaks at positions (in terms of observation numbers) 28, 38 and 42 at the 10% nominal size.

Since we have the reference of the breakpoints detected by Bai (1997), our objective in this section is to find out if we can verify them with our LR-based tests. We therefore apply our *LRall* and *LRpart* tests to the same sample (the critical value from the $\chi^2$ distribution

[7] We are very grateful to Tim Vogelsang his guidance with these simulations.

is 4.60 at 10% nominal size). Given the very small sample size, we use $\pi_T^1 = 1/2$, and given that in this case $n = 1$ we simply apply the asymptotic versions of our tests.

When we use the whole sample size, we get a value for *LRpart* equal to 0.0439 and *LRall* 0.1987, with the most likely breakdate being 14 and therefore, we are not able to detect any break. Bai (1997) obtains 28 as the most likely but insignificant breakdate, and thus also is not able to verify a break on the whole sample. He explains this rejection with the fact that his Wald-test (and the same in this case with the LR-based tests) has low power when multiple breaks exist. Bai (1997) continues his analysis assuming a break at position 28, and detects a significant break at 38. Bai then looks at the data from 1 to 37 and detects a significant break at position 28. Finally he detects a significant break at 42 if he examines the data from 38 to 56. If we continue from the insignificant break we found at position 14, we do not eventually detect any breaks in the data. If, however, we follow Bai (1997) and analyze the data from position 28 onwards, we do detect a break at position 33 since *LRpart* and *LRall* take on the values 6.6134 and 5.4665 in this period. Moreover, when we run our tests from observations 37 until 56, our tests detect a clear break at position 42 with *LRpart* equal to 8.2894 and *LRall* 9.5722.

Therefore, if we assume that indeed there is a break at position 28, both the Bai (1997) test and our LR-ratio tests give very similar answers since the detection of break 42 happens in both cases, and while Bai (1997) test detects two breaks at 28 and 38, we detect a break in the same area around 33. In the next section we will analyze a multivariate model where it turns out that Bai (1997) and our LR-tests give very different results.

**4.2. The CAPM model.** In this section we will test for breaks in the CAPM model. Parameter constancy has been an issue in the finance literature for a while, and both models of continuously changing parameters (see Ang and Chen (2007)) and models incorporating discrete breaks (see Huang and Cheng (2005)) have been examined. To our knowledge, however, we are the first to consider simultaneous breaks in all the series of the multivariate model. Clearly if parameter changes are due to international, political or market structure changes, it is natural to expect that all the series would break simultaneously, and hence the multivariate CAPM model is perfectly suited to illustrate the methodology in this

paper.

In Section 3, we demonstrated that in the multivariate framework, the Bai, Lumsdaine and Stock (1998) test is badly oversized. Since this test tends to overreject, we would expect the Bai, Lumsdaine and Stock (1998) test to find more breaks than our LR-based tests.

We consider the framework given in Gibbons (1982) where, if $r_{it}$, $r_{mt}$ and $r_{ft}$ are the returns of asset $i$, the market portfolio $m$ and the risk free rate at time $t$ respectively, then

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \tag{4}$$

where $R_{it} = r_{it} - r_{ft}$ and $R_{mt} = r_{mt} - r_{ft}$. $R_{mt}$ is the excess return on the market portfolio at time $t$.

We consider monthly data from July 1926 until December 2005 for five portfolios sorted according to size.[8] They are constructed at the end of each June using the June market equity and NYSE breakpoints. $r_{mt}$ is the return on the market portfolio which is the value-weighted return on all NYSE, AMEX and NASDAQ stocks and $r_{ft}$ is the one month Treasury Bill rate and is a proxy for the risk free interest rate. The portfolios for July of year $t$ to June of $t+1$ include all NYSE, AMEX, and NASDAQ stocks. Figure 3 in Appendix E shows a graph of the six time series with a sample size of 954 observations.

We proceed to apply Bai et al (1998) test to the 5-dimensional system given in (4), and it finds significant breaks at positions 72, 158, 166, 200, 235, 300, 432, 495, 576, 625, 650, 706, 768, 878 and 900 (15 breaks in all). If we apply our LR-based tests, both *LRpart* with the MC procedure and *LRall* find breaks only at positions 58, 90, 117, 145, 172, and 256 (6 breaks in all). From Figure 3, it is observable that all the breaks detected by our test statistics are at the beginning of the sample where the data seems more unstable. As our theory predicts, the Bai et al (1998) test finds too many breaks in the 5-dimensional system. In the reported results we used $\pi_T^1 = 1/5$, but we also tested for breaks using $\pi_T^1 = 1/3$, 1/4 and 1/6, and the results are qualitatively similar.

[8] The data is publicy available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

## 5. Conclusion

In this paper we construct new tests for structural tests based on alternating observations. While the commonly used sup-Wald tests have non-standard asymptotic distributions, we prove that our new tests have the same asymptotic distribution as regular tests when the breakpoint is known (namely, a chi-square). Moreover both in a univariate and a multivariate framework, we show that in finite samples, *LRpart* combined with Monte Carlo can be constructed to be exact, while *LRall* is shown in simulations to have much less size distortions than Bai et al (1998) and similar power. Asymptotically, both tests follow a chi-square distribution.

We show in a practical application how, since Bai et al (1998) test does not have very important size distortions in the univariate case (even though *LRall* has even less size distortions than Bai et al (1998)), our LR-based tests produce a very similar outcome. However, we also show in an application with the CAPM, how in this case Bai et al (1998) test finds many more breaks than our LR-based tests, which is a clear consequence of the very high over-rejections of Bai et al (1998) test.

# Appendix

## A. Proof of Theorem 1

**(a)** First note that to obtain an asymptotic $\chi^2$ distribution it is sufficient that the breakpoint is *asymptotically* independent of the data used to calculate the test statistic. This implies that the proof of (a) is trivial since the data is iid and $\hat{\lambda}_J$ is calculated using $\mathbb{N}_J$,while the $LR$ statistic is calculated using $\mathbb{N}_R$.

**(b)** We will prove (b) by verifying that

$$plim_{T\to\infty}\left(\lim_{J\to\infty} LRpart_{J,R}\left(\hat{\lambda}_J\right) - \lim_{J\to\infty} LRall_{J,T}\left(\hat{\lambda}_J\right)\right) = 0.$$

For notational purposes, we write $plim_{J\to\infty}\left(LRpart_{J,T}\left(\cdot\right)\right) = LRpart_{J,T}\left(\cdot\right)$ to indicate that in the first stage we let $J\to\infty$, and in the second stage we allow $T$ to go to infinity**.** Note that the estimate of the breakpoint used in $LRpart$ and $LRall$ is identical, and that by Andrews (1988, 1993) it converges to a random variable $\tilde{\lambda}$ *with* support $[\tau,(1-\tau)]$, under the null hypothesis. This implies that we can write the previous limit as

$$\begin{aligned}
& plim_{T\to\infty}\left(LRpart_{\infty,R}\left(\tilde{\lambda}\right) - LRall_{\infty,T}\left(\tilde{\lambda}\right)\right) \\
= \; & -2plim_{T\to\infty}\left(R\ln\frac{\left|S^{\tilde{\lambda}}_{\mathbb{N}_R}\right|}{\left|S^{0}_{\mathbb{N}_R}\right|} - T\ln\frac{\left|S^{\tilde{\lambda}}\right|}{\left|S^{0}\right|}\right) \\
= \; & -2plim_{T\to\infty}\left(\frac{R}{T}\right)plim_{T\to\infty}\left(T\ln\frac{\left|S^{\tilde{\lambda}}_{\mathbb{N}_R}\right|}{\left|S^{0}_{\mathbb{N}_R}\right|}\right) + 2plim_{T\to\infty}\left(T\ln\frac{\left|S^{\tilde{\lambda}}\right|}{\left|S^{0}\right|}\right).
\end{aligned}$$

And since

$$plim_{T\to\infty}\left(\frac{R}{T}\right) = plim_{T\to\infty}\left(\frac{T-\pi^1_T T}{T}\right) = 1$$

we have

$$\begin{aligned}
& -2plim_{T\to\infty}\left(\frac{R}{T}\right)plim_{T\to\infty}\left(T\ln\frac{\left|S^{\tilde{\lambda}}_{\mathbb{N}_R}\right|}{\left|S^{0}_{\mathbb{N}_R}\right|}\right) + 2plim_{T\to\infty}\left(T\ln\frac{\left|S^{\tilde{\lambda}}\right|}{\left|S^{0}\right|}\right) \\
= \; & -2plim_{T\to\infty}\left(T\ln\frac{\left|S^{\tilde{\lambda}}_{\mathbb{N}_R}\right|}{\left|S^{0}_{\mathbb{N}_R}\right|}\right) + 2plim_{T\to\infty}\left(T\ln\frac{\left|S^{\tilde{\lambda}}\right|}{\left|S^{0}\right|}\right) \\
= \; & -2plim_{T\to\infty}\left(T\ln\frac{\left|S^{\tilde{\lambda}}_{\mathbb{N}_R}\right|\left|S^{0}\right|}{\left|S^{\tilde{\lambda}}\right|\left|S^{0}_{\mathbb{N}_R}\right|}\right).
\end{aligned}$$

Now, note that $S^{\tilde{\lambda}} = S^{\tilde{\lambda}}_{\mathbb{N}_R} + \sum_{j\in\mathbb{N}_J} \hat{U}_j^{\tilde{\lambda}\prime}\hat{U}_j^{\tilde{\lambda}}$ and $S^0 = S^0_{\mathbb{N}_R} + \sum_{j\in\mathbb{N}_J} \hat{U}_j^{0\prime}\hat{U}_j^0$ . We can now write

$$\begin{aligned}
& plim_{T\to\infty}\left(LRpart_{\infty,R}\left(\tilde{\lambda}\right) - LRall_{\infty,T}\left(\tilde{\lambda}\right)\right) \\
= \; & -2plim_{T\to\infty} T\ln\left(\left(\frac{\left|S^{\tilde{\lambda}}_{\mathbb{N}_R}\right|}{\left|S^{\tilde{\lambda}}_{\mathbb{N}_R} + \sum_{j\in\mathbb{N}_J}\hat{U}_j^{\tilde{\lambda}\prime}\hat{U}_j^{\tilde{\lambda}}\right|}\right)\left(\frac{\left|S^0_{\mathbb{N}_R} + \sum_{j\in\mathbb{N}_J}\hat{U}_j^{0\prime}\hat{U}_j^0\right|}{\left|S^0_{\mathbb{N}_R}\right|}\right)\right) \\
= \; & -2plim_{T\to\infty} T\ln\left(\left(\frac{\left|\frac{1}{T}S^{\tilde{\lambda}}_{\mathbb{N}_R}\right|}{\left|\frac{1}{T}S^{\tilde{\lambda}}_{\mathbb{N}_R} + \frac{1}{T}\sum_{j\in\mathbb{N}_J}\hat{U}_j^{\tilde{\lambda}\prime}\hat{U}_j^{\tilde{\lambda}}\right|}\right)\left(\frac{\left|\frac{1}{T}S^0_{\mathbb{N}_R} + \frac{1}{T}\sum_{j\in\mathbb{N}_J}\hat{U}_j^{0\prime}\hat{U}_j^0\right|}{\left|\frac{1}{T}S^0_{\mathbb{N}_R}\right|}\right)\right)
\end{aligned}$$

Since $\mathbb{N}_J$ contains only $J$ observations, by A4,

$$\begin{aligned}
\frac{1}{T}\sum_{j\in\mathbb{N}_J}\hat{U}_j^{\tilde{\lambda}\prime}\hat{U}_j^{\tilde{\lambda}} &= \frac{J}{T}\frac{1}{J}\sum_{j\in\mathbb{N}_J}\hat{U}_j^{\tilde{\lambda}\prime}\hat{U}_j^{\tilde{\lambda}} = O_P\left(\pi_1^T\right), \text{ and} \\
\frac{J}{T}\frac{1}{J}\sum_{j\in\mathbb{N}_J}\hat{U}_j^{0\prime}\hat{U}_j^0 &= \frac{1}{T}\sum_{j\in\mathbb{N}_J}\hat{U}_j^{0\prime}\hat{U}_j^0 = O_P\left(\pi_1^T\right),
\end{aligned}$$

and therefore

$$\ln\left(\left(\frac{\left|\frac{1}{T}S^{\tilde{\lambda}}_{\mathbb{N}_R}\right|}{\left|\frac{1}{T}S^{\tilde{\lambda}}_{\mathbb{N}_R} + \frac{1}{T}\sum_{j\in\mathbb{N}_J}\hat{U}_j^{\tilde{\lambda}\prime}\hat{U}_j^{\tilde{\lambda}}\right|}\right)\left(\frac{\left|\frac{1}{T}S^0_{\mathbb{N}_R} + \frac{1}{T}\sum_{j\in\mathbb{N}_J}\hat{U}_j^{0\prime}\hat{U}_j^0\right|}{\left|\frac{1}{T}S^0_{\mathbb{N}_R}\right|}\right)\right) = O_P\left(\ln(\pi_1^T)\right)$$

and by A4

$$plim_{T\to\infty}\left(LRpart_{\infty,R}\left(\tilde{\lambda}\right) - LRall_{\infty,T}\left(\tilde{\lambda}\right)\right) = 0. \tag{5}$$

By (a) and (5), the proof of (b) is complete.

**(c)**Following Bernard et al (2007), the *LRpart* test is invariant to the parameter values in the multiple linear regression model. Therefore, *LRpart* can be obtained in the second stage without having to estimate the coefficients of the multiple linear regression model in the first stage. The invariance results of the *LRpart* test when the test is a function of residual sum of squares is proved in Bernard et al (2007), so this result applies directly to our case with the split sample. We can adopt this invariance result in Bernard et al (2007) only under A1a) and when the *LRpart* test is a ratio of residual sum of squares.■

## B. **Proof of Theorem 2**

**(a)** Note that the number of observations separating data points used to estimate $\lambda$ and observations used to construct the test statistic is $\left[\frac{R\pi_T^2}{2J}\right]$, and that by A5 $\left[\frac{R\pi_T^2}{J}\right] \to \infty$.

The mixing property implies independence between any two fixed blocks of data that are separated by an increasing number of observations. As a result, the two sets of observations are asymptotically independent. This establishes the asymptotic independence of the estimate of the breakpoint from the data used to calculate the test statistic. As a result,

$$LRblock\left(\hat{\lambda}_J\right) \Rightarrow \chi^2\left(q_X\right).$$

**(b)** To complete the proof of (b) it is sufficient to establish that

$$plim_{T\to\infty}\left(\lim_{J\to\infty} LRpart_{J,R}\left(\hat{\lambda}_J\right) - \lim_{J\to\infty} LRblock_{J,T}\left(\hat{\lambda}_J\right)\right) = 0. \tag{6}$$

Note that the estimate of the breakpoint used in *LRpart* and *LRblock* is identical, and that by Andrews (1988, 1993) it converges to a random variable $\tilde{\lambda}$ *with* support $[\tau, (1-\tau)]$, under the null hypothesis. This implies that we can write the previous limit as

$$\text{plim}_{T\to\infty}\left(LRpart_R\left(\tilde{\lambda}\right) - LRblock_R\left(\tilde{\lambda}\right)\right)$$

By definition, if $L^N$ is the likelihood under the null, and $L^A$ is under the alternative, and the subscript provides the set of observations on which the test statistic is calculated,

$$\begin{aligned} & \text{plim}_{T\to\infty}\left(LRpart_R\left(\tilde{\lambda}\right) - LRblock_R\left(\tilde{\lambda}\right)\right) \\ = \quad & -2\text{plim}_{T\to\infty}\left(\ln\frac{L^N_{\mathbb{N}_C}}{L^A_{\mathbb{N}_C}} - \ln\frac{L^N_{\mathbb{N}_R}}{L^A_{\mathbb{N}_R}}\right) \end{aligned}$$

so we need to prove that $\ln\frac{L^N_{\mathbb{N}_C}}{L^N_{\mathbb{N}_R}} \xrightarrow{p} 0$ and $\ln\frac{L^A_{\mathbb{N}_R}}{L^A_{\mathbb{N}_C}} \xrightarrow{p} 0$. First consider

$$\ln\frac{L^N_{\mathbb{N}_C}\left(\hat{B}_{\mathbb{N}_C}, \tilde{\lambda}\right)}{L^N_{\mathbb{N}_R}\left(\hat{B}_{\mathbb{N}_R}, \tilde{\lambda}\right)} = \ln L^N_{\mathbb{N}_C}\left(\hat{B}_{\mathbb{N}_C}, \tilde{\lambda}\right) - \ln L^N_{\mathbb{N}_R}\left(\hat{B}_{\mathbb{N}_R}, \tilde{\lambda}\right),$$

where $\hat{B}_{\mathbb{N}_C}$ and $\hat{B}_{\mathbb{N}_R}$ are the parameter estimates of $B$ from model (1) using the observations in $\mathbb{N}_C$ and $\mathbb{N}_R$ respectively. Now by A6 and the mean value theorem

$$\begin{aligned} \ln\frac{L^N_{\mathbb{N}_C}\left(\hat{B}_{\mathbb{N}_C}, \tilde{\lambda}\right)}{L^N_{\mathbb{N}_R}\left(\hat{B}_{\mathbb{N}_R}, \tilde{\lambda}\right)} \quad = \quad & \ln L^N_{\mathbb{N}_C}\left(B_0, \tilde{\lambda}\right) - \ln L^N_{\mathbb{N}_R}\left(B_0, \tilde{\lambda}\right) \\ & + \left(\hat{B}_{\mathbb{N}_C} - B_0\right)\mathbf{H}^N_{\mathbb{N}_C}\left(\bar{B}, \tilde{\lambda}\right) - \left(\hat{B}_{\mathbb{N}_R} - B_0\right)\mathbf{H}^N_{\mathbb{N}_R}\left(\check{B}, \tilde{\lambda}\right), \end{aligned} \tag{7}$$

where $\mathbf{H}^N_{\mathbb{N}_C}$ and $\mathbf{H}^N_{\mathbb{N}_R}$ are the respective Hessians and $\bar{B}$ lies between $\hat{B}_{\mathbb{N}_C}$ and $B_0$ while $\check{B}$ lies between $\hat{B}_{\mathbb{N}_R}$ and $B_0$. Note that since the number of observations in $\mathbb{N}_C$ and $\mathbb{N}_R$ both go to infinity as $T \to \infty$

$$\text{plim} \ln L^N_{\mathbb{N}_C}\left(B_0, \tilde{\lambda}\right) - \text{plim} \ln L^N_{\mathbb{N}_R}\left(B_0, \tilde{\lambda}\right) = 0$$

and

$$\text{plim}\left(\hat{B}_{\mathbb{N}_R} - B_0\right) = \text{plim}\left(\hat{B}_{\mathbb{N}_C} - B_0\right) = 0 \tag{8}$$

Finally, by (8), plim$\bar{B}$ =plim$\check{B} = B_0$, and therefore

$$\text{plim}\mathbf{H}^N_{\mathbb{N}_C}\left(\bar{B}, \tilde{\lambda}\right) = \text{plim}\mathbf{H}^N_{\mathbb{N}_R}\left(\check{B}, \tilde{\lambda}\right).$$

Using these results in (7), we get

$$\text{plim} \ln \frac{L^N_{\mathbb{N}_C}\left(\hat{B}, \tilde{\lambda}\right)}{L^N_{\mathbb{N}_R}\left(\hat{B}, \tilde{\lambda}\right)} = 0.$$

By similar arguments it is possible to show that

$$\text{plim} \ln \frac{L^A_{\mathbb{N}_R}}{L^A_{\mathbb{N}_C}} = 0.$$

Therefore (6) holds and by Theorem 2 (a) the proof is complete.

**(c)** To prove this part, we need to prove that

$$\text{plim}_{T\to\infty}\left(LRpart_R\left(\tilde{\lambda}\right) - LRall_R\left(\tilde{\lambda}\right)\right) = 0.$$

For that

$$\begin{aligned} & \text{plim}_{T\to\infty}\left(LRall_R\left(\tilde{\lambda}\right) - LRpart_R\left(\tilde{\lambda}\right)\right) \\ = \quad & -2\text{plim}_{T\to\infty}\left(\ln\frac{L^N}{L^A} - \ln\frac{L^N_{\mathbb{N}_R}}{L^A_{\mathbb{N}_R}}\right) = -2\text{plim}_{T\to\infty}\ln\frac{L^N L^A_{\mathbb{N}_R}}{L^A L^A_{\mathbb{N}_R}}. \end{aligned}$$

By arguments similar to those employed in part (b) this limit is 0 and the result holds.■

## C. Proof of Theorem 3

$W_{part}$, $W_{block}$ and $W_{all}$ correspond to the Wald statistics with the HAC covariance matrix estimator (see Newey and West (1987)) using $\mathbb{N}_R$ and $\mathbb{N}_C$ and all the observations respectively. For simplicity reasons, we refer in the proofs to the traditional variance covariance matrix, but the extension of the proofs to the HAC context is straightforward.

**(a)** Note that the number of observations separating data points used to estimate $\lambda$ and observations used to construct the test statistic is $\left[\frac{R\pi_T^2}{2J}\right]$, and that by A5 $\left[\frac{R\pi_T^2}{2J}\right] \to \infty$. Combining this with A1 b), the two sets of observations ($\mathbb{N}_J$ and $\mathbb{N}_C$) are asymptotically independent. This establishes the asymptotic independence of the estimate of the breakpoint from the data used to calculate the test statistic. As a result,

$$W_{block\ J,R-\left[R\pi_T^2\right]}\left(\hat{\lambda}_J\right) \Rightarrow \chi^2\left(q_X\right).$$

**(b)** To complete the proof of (b) it is sufficient to establish that

$$plim_{T\to\infty}\left(\lim_{J\to\infty} W_{part\ J,R}\left(\hat{\lambda}_J\right) - \lim_{J\to\infty} W_{block\ J,R-\left[R\pi_T^2\right]}\left(\hat{\lambda}_J\right)\right) = 0. \tag{9}$$

Note that the estimate of the breakpoint used in *LRpart* and *LRblock* is identical, and that by Andrews (1988, 1993) it converges to a random variable $\tilde{\lambda}$ with support $[\tau, (1-\tau)]$ under the null hypothesis. This implies that we can write the previous limit as

$$plim_{T\to\infty}\left(W_{part\ \infty,R}\left(\tilde{\lambda}\right) - W_{block\ \infty,R-\left[R\pi_T^2\right]}\left(\tilde{\lambda}\right)\right)$$

Recall that

$$\begin{aligned} & W_{block}\left(\tilde{\lambda}\right) \\ = \ & \left(R-\left[R\pi_T^2\right]\right)\left(R^*\hat{\Theta}_{\mathbb{N}_C}\right)'\left(R^*\left(\frac{1}{R-\left[R\pi_T^2\right]}\sum_{t\in\mathbb{N}_C} Z_t\hat{\Sigma}_{\mathbb{N}_C}^{-1}Z_t'\right)^{-1}(R^*)'\right)^{-1}\left(R^*\hat{\Theta}_{\mathbb{N}_C}\right) \end{aligned}$$

where $\hat{\Theta}_{\mathbb{N}_C}$ and $\hat{\Sigma}_{\mathbb{N}_C}^{-1}$ correspond to the estimates of $\Theta_s$ and $\hat{\Sigma}_s^{-1}$ in relation to (2) and in Section 2 using the set of observations $\mathbb{N}_C$, and

$$W_{part}\left(\tilde{\lambda}\right) = R\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right)'\left(R^*\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\hat{\Sigma}_{\mathbb{N}_R}^{-1}Z_t'\right)^{-1}(R^*)'\right)^{-1}\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right) \tag{10}$$

Note that $R^*$ is the restriction matrix, while $R$ is the number of observations in $\mathbb{N}_R$. Now, by A2, A5 and A7

$$\text{plim}\left(\frac{1}{R-\left[R\pi_T^2\right]}\sum_{t\in\mathbb{N}_C} Z_t\hat{\Sigma}_{\mathbb{N}_C}^{-1}Z_t'\right) = \text{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\hat{\Sigma}_{\mathbb{N}_R}^{-1}Z_t'\right) = Q$$

for some p.d. matrix $Q$,

$$\text{plim}\hat{\Sigma}_{\mathbb{N}_C}^{-1} = \text{plim}\hat{\Sigma}_{\mathbb{N}_R}^{-1} = \Sigma^{-1}$$

and

$$\text{plim}\left(\frac{1}{R-\left[R\pi_T^2\right]}\sum_{t\in\mathbb{N}_C} Z_t\hat{\Sigma}_{\mathbb{N}_C}^{-1}Z_t'\right) = \text{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\hat{\Sigma}_{\mathbb{N}_R}^{-1}Z_t'\right).$$

Thus, we have

$$\begin{aligned}
& \text{plim}\left(W_{block}\left(\tilde{\lambda}\right) - W_{part}\left(\tilde{\lambda}\right)\right) \\
= \; & \text{plim}\left[\left(R-\left[R\pi_T^2\right]\right)\left(R^*\hat{\Theta}_{\mathbb{N}_C}\right)'\left(R^*Q^{-1}\left(R^*\right)'\right)^{-1}\left(R^*\hat{\Theta}_{\mathbb{N}_C}\right)\right] \\
& -\text{plim}\left[R\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right)'\left(R^*Q^{-1}\left(R^*\right)'\right)^{-1}\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right)\right] \\
= \; & \text{plim}\left[\frac{\left(R-\left[R\pi_T^2\right]\right)}{R}R\left(R^*\hat{\Theta}_{\mathbb{N}_C}\right)'\left(R^*Q^{-1}\left(R^*\right)'\right)^{-1}\left(R^*\hat{\Theta}_{\mathbb{N}_C}\right)\right] \\
& -\text{plim}\left[R\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right)'\left(R^*Q^{-1}\left(R^*\right)'\right)^{-1}\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right)\right] \\
= \; & \text{plim}\left[R\left(R^*\hat{\Theta}_{\mathbb{N}_C}\right)'\left(R^*Q^{-1}\left(R^*\right)'\right)^{-1}\left(R^*\hat{\Theta}_{\mathbb{N}_C}\right)\right] \\
& -\text{plim}\left[R\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right)'\left(R^*Q^{-1}\left(R^*\right)'\right)^{-1}\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right)\right]
\end{aligned}$$

Therefore, what remains to be shown is that

$$\text{plim}\left(\sqrt{R}\;\hat{\Theta}_{\mathbb{N}_C} - \sqrt{R}\;\hat{\Theta}_{\mathbb{N}_R}\right) = 0$$

$$
\begin{aligned}
& \operatorname{plim}\left(\sqrt{R}\hat{\Theta}_{\mathbb{N}_R}-\sqrt{R}\hat{\Theta}_{\mathbb{N}_C}\right) \\
= & \operatorname{plim}\left(\sqrt{R}\left(\hat{\Theta}_{\mathbb{N}_R}-\Theta_0\right)-\sqrt{R}\left(\hat{\Theta}_{\mathbb{N}_C}-\Theta_0\right)\right) \\
= & \operatorname{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\hat{\Sigma}_{\mathbb{N}_R}^{-1}Z_t'\right)^{-1}\left(\frac{1}{\sqrt{R}}\sum_{t\in\mathbb{N}_R} Z_t\hat{\Sigma}_{\mathbb{N}_R}^{-1}U_t\right) \\
& -\sqrt{R}\operatorname{plim}\left(\frac{1}{R-\left[R\pi_T^2\right]}\sum_{t\in\mathbb{N}_C} Z_t\hat{\Sigma}_{\mathbb{N}_C}^{-1}Z_t'\right)^{-1}\left(\frac{1}{R-\left[R\pi_T^2\right]}\sum_{t\in\mathbb{N}_C} Z_t\hat{\Sigma}_{\mathbb{N}_C}^{-1}U_t\right) \\
= & \operatorname{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\Sigma^{-1}Z_t'\right)^{-1}\left(\frac{1}{\sqrt{R}}\sum_{t\in\mathbb{N}_R} Z_t\Sigma^{-1}U_t\right) \\
& -\operatorname{plim}\left(\frac{R}{R-\left[R\pi_T^2\right]}\right)\operatorname{plim}\left(\frac{1}{R-\left[R\pi_T^2\right]}\sum_{t\in\mathbb{N}_C} Z_t\Sigma^{-1}Z_t'\right)^{-1}\left(\frac{1}{\sqrt{R}}\sum_{t\in\mathbb{N}_C} Z_t\Sigma^{-1}U_t\right) \\
= & \operatorname{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\Sigma^{-1}Z_t'\right)^{-1}\operatorname{plim}\left[\frac{1}{\sqrt{R}}\sum_{t\in\mathbb{N}_R} Z_t\Sigma^{-1}U_t-\frac{1}{\sqrt{R}}\sum_{t\in\mathbb{N}_C} Z_t\Sigma^{-1}U_t\right] \\
= & \operatorname{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\Sigma^{-1}Z_t'\right)^{-1}\operatorname{plim}\left[\frac{1}{\sqrt{R}}\sum_{t\in\mathbb{N}_R/\mathbb{N}_C} Z_t\Sigma^{-1}U_t\right] \\
= & \operatorname{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\Sigma^{-1}Z_t'\right)^{-1}\operatorname{plim}\left(\sqrt{\frac{\left[R\pi_T^2\right]}{R}}\right)\operatorname{plim}\left[\frac{1}{\sqrt{\left[R\pi_T^2\right]}}\sum_{t\in\mathbb{N}_R/\mathbb{N}_C} Z_t\Sigma^{-1}U_t\right]
\end{aligned}
$$

Now, looking at the three terms separately,

$$
\operatorname{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\Sigma^{-1}Z_t'\right)^{-1}=O_P\left(1\right),
$$

$$
\frac{1}{\sqrt{\left[R\pi_T^2\right]}}\sum_{t\in\mathbb{N}_R/\mathbb{N}_C} Z_t\Sigma^{-1}U_t=O_P\left(1\right),
$$

and

$$
\sqrt{\frac{\left[R\pi_T^2\right]}{R}}=O\left(\sqrt{\pi_T^2}\right).
$$

Since by A5 $\pi_T^2\to 0$, we get that

$$
\operatorname{plim}\left(\sqrt{R}\hat{\Theta}_{\mathbb{N}_C}-\sqrt{R}\hat{\Theta}_{\mathbb{N}_C}\right)=0,
$$

which verifies (9) so that by Theorem 3 (a) the result holds.

**(c)** To complete the proof of (c) it is then sufficient to establish that

$$\text{plim}_{T\to\infty}\left(\lim_{J\to\infty} W_{part\ J,R}\left(\hat{\lambda}_J\right) - \lim_{J\to\infty} W_{all\ T}\left(\hat{\lambda}_J\right)\right) = 0.$$

Note that the estimate of the breakpoint used in *LRpart* and *LRblock* is identical, and that by Andrews (1988, 1993) it converges to a random variable $\tilde{\lambda}$ *with* support $[\tau, (1-\tau)]$ under the null hypothesis. This implies that we can write the previous limit as

$$\text{plim}_{T\to\infty}\left(W_{part}\left(\tilde{\lambda}\right) - W_{all}\left(\tilde{\lambda}\right)\right)$$

Recall that

$$W_{all}\left(\tilde{\lambda}\right) = T\left(R^*\hat{\Theta}\right)'\left(R^*\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\hat{\Sigma}^{-1}Z_t'\right)^{-1}(R^*)'\right)^{-1}\left(R^*\hat{\Theta}\right)$$

and $W_{part}\left(\tilde{\lambda}\right)$ is given by (10). Now, by A2, A5 and A7

$$\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\hat{\Sigma}^{-1}Z_t'\right) = \text{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\hat{\Sigma}_{\mathbb{N}_R}^{-1}Z_t'\right) = Q$$

for some p.d. matrix $Q$,

$$\text{plim}\hat{\Sigma}^{-1} = \text{plim}\hat{\Sigma}_{\mathbb{N}_R}^{-1} = \Sigma^{-1},$$

and

$$\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\hat{\Sigma}_{all}^{-1}Z_t'\right) = \text{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\hat{\Sigma}_{part}^{-1}Z_t'\right).$$

Thus, we have

$$\begin{aligned}
& \text{plim}\left(W_{part}\left(\tilde{\lambda}\right) - W_{all}\left(\tilde{\lambda}\right)\right) \\
= \ & \text{plim}\left[R\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right)'\left(R^*Q^{-1}(R^*)'\right)^{-1}\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right)\right] \\
& -\text{plim}\left(T\left(R^*\hat{\Theta}\right)'\left(R^*Q^{-1}(R^*)'\right)^{-1}\left(R^*\hat{\Theta}\right)\right) \\
= \ & \text{plim}\left(\frac{R}{T}\right)\cdot\text{plim}\left[T\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right)'\left(R^*Q^{-1}(R^*)'\right)^{-1}\left(R^*\hat{\Theta}_{\mathbb{N}_R}\right)\right] \\
& -\text{plim}\left(T\left(R^*\hat{\Theta}\right)'\left(R^*Q^{-1}(R^*)'\right)^{-1}\left(R^*\hat{\Theta}\right)\right)
\end{aligned}$$

Now, by A3 and A4, $\text{plim}\left(\frac{R}{T}\right) = 1$, so all that remains to be shown is that

$$\text{plim}\left(\sqrt{T}\ \hat{\Theta}_{\mathbb{N}_R} - \sqrt{T}\ \hat{\Theta}\right) = 0$$

$$\begin{aligned}
& \text{plim}\left(\sqrt{T}\ \hat{\Theta}_{\mathbb{N}_R} - \sqrt{T}\ \hat{\Theta}\right) \\
= \ & \text{plim}\left(\sqrt{T}\left(\hat{\Theta}_{\mathbb{N}_R} - \Theta_0\right) - \sqrt{T}\left(\hat{\Theta} - \Theta_0\right)\right) \\
= \ & \text{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\hat{\Sigma}^{-1}_{\mathbb{N}_R}Z_t'\right)^{-1}\left(\frac{\sqrt{T}}{R}\sum_{t\in\mathbb{N}_R} Z_t\hat{\Sigma}^{-1}_{\mathbb{N}_R}U_t\right) \\
& -\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\hat{\Sigma}^{-1}Z_t'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\hat{\Sigma}^{-1}U_t\right) \\
= \ & \text{plim}\left(\frac{T}{R}\right)\text{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\Sigma^{-1}Z_t'\right)^{-1}\text{plim}\left(\frac{1}{\sqrt{T}}\sum_{t\in\mathbb{N}_R} Z_t\Sigma^{-1}U_t\right) \\
& -\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\Sigma^{-1}Z_t'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\Sigma^{-1}U_t\right)
\end{aligned}$$

Now, since $\text{plim}\left(\frac{T}{R}\right) = 1$ and

$$\text{plim}\left(\frac{1}{R}\sum_{t\in\mathbb{N}_R} Z_t\Sigma^{-1}Z_t'\right)^{-1} = \text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\Sigma^{-1}Z_t'\right)^{-1}$$

we get

$$\begin{aligned}
& \text{plim}\left(\sqrt{T}\ \hat{\Theta}_{\mathbb{N}_R} - \sqrt{T}\ \hat{\Theta}\right) \qquad (11) \\
= \ & \text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\Sigma^{-1}Z_t'\right)^{-1}\text{plim}\left[\frac{1}{\sqrt{T}}\sum_{t\in\mathbb{N}_R} Z_t\Sigma^{-1}U_t - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\Sigma^{-1}U_t\right] \\
= \ & -\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\Sigma^{-1}Z_t'\right)^{-1}\text{plim}\left[\frac{1}{\sqrt{T}}\sum_{t\in\mathbb{N}_J} Z_t\Sigma^{-1}U_t\right] \\
= \ & -\text{plim}\left(\frac{\sqrt{J}}{\sqrt{T}}\right)\text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\Sigma^{-1}Z_t'\right)^{-1}\text{plim}\left[\frac{1}{\sqrt{J}}\sum_{t\in\mathbb{N}_J} Z_t\Sigma^{-1}U_t\right]
\end{aligned}$$

Now by A3 and A4

$$\text{plim}\left(\sqrt{\frac{J}{T}}\right) = 0.$$

Also, we know that

$$\operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t \Sigma^{-1} Z_t'\right)^{-1} = O_P(1)$$

and using a CLT, $\frac{1}{\sqrt{J}}\sum_{t\in\mathbb{N}_J} Z_t \Sigma^{-1} U_t = O_P(1)$.

Plugging these three into (11) we immediately get that $\operatorname{plim}\left(\sqrt{T}\,\hat{\Theta}_{\mathbb{N}_R} - \sqrt{T}\,\hat{\Theta}\right) = 0$ and the proof is complete.■

## D. Proof of Theorem 4

We prove parts (a), (b) and (c) at the same time. First note that when an alternative cannot be detected by the test statistic, we return to the situation under the null, where the estimated breakpoint is a draw from the distribution described in Andrews (1993). In this case our test statistic is simply a Chow-test of a break point which is drawn from the distribution described in Andrews (1993). Since it is drawn independently of the observations used for testing, however, we still retain power against this alternative.

Now we turn our attention to the question of which alternatives can be detected, in the sense that they provide consistent estimates of the breakpoint. Recall that we assumed in Theorem 3 that

A8 $\Delta^{s^0}$ depends on $T$ and can be written as $\Delta^{s^0} = \Delta_0 \cdot v_T$ where $v_T$ is a positive number such that $v_T \longrightarrow 0$ and $(\pi_1 T)^{(1/2-\alpha)} v_T \to \infty$ for some $\alpha \in (0, 1/2)$ and $\Delta_0 \neq 0$

Then we have

$$\hat{k} = k_0 + O_P\left(\|v_T\|^{-2}\right)$$

So, if we define a $\frac{1}{\sqrt{T}}$ alternative as in Andrews (1993) such that

$$\left(\Delta^t - \Delta_0\right) = \eta\left(\frac{t}{T}\right)\frac{1}{\sqrt{T}},$$

where $\eta(\cdot)$ is some bounded function on $[0, 1]$ such that $\eta$ is not equal to a constant almost everywhere on $[\tau, 1-\tau]$, we have

$$\begin{aligned}(\pi_1 T)^{(1/2-\alpha)}\left(\Delta^t - \Delta_0\right) &= \eta\left(\frac{t}{T}\right)(\pi_1 T)^{(1/2-\alpha)}\frac{1}{\sqrt{T}} = \eta\left(\frac{t}{T}\right)(\pi_1)^{(1/2-\alpha)} T^{-\alpha} \\ &= \eta\left(\frac{t}{T}\right)\frac{(\pi_1)^{(1/2-\alpha)}}{T^{\alpha}},\end{aligned}$$

For A8 to be satisfied, we need

$$\frac{(\pi_1 T)^{(1/2-\alpha)}}{\sqrt{T}} \to \infty. \tag{12}$$

Recall however, that $\pi_1 \to 0$. It is clear that $\pi_1 \to 0$ and (12) can both be satisfied by choosing $\pi_1 = O\left(T^{-\frac{\alpha}{1/2-\alpha-\varepsilon}}\right)$ with $\alpha$ less than $\frac{1}{4}$ and $\varepsilon > 0$. So if we limit $\alpha$ to be less than $\frac{1}{4}$, since we are using a consistent estimate of the breakpoint in the first stage, we will inherit the optimality properties of Chow-type tests based on the correct break point (see e.g. Rossi (2005)).■

## E. Tables

Table 1: Empirical Size. Model with intercept and time trend. Nominal size = 5%

| | $n=1$ | | | $n=2$ | | | $n=5$ | | | $n=10$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | $\mathcal{F}^*$ | $\Lambda^*_{MC}$ | $\mathrm{LRp}_{MC}$ | $\mathcal{F}^*$ | $\Lambda^*_{MC}$ | $\mathrm{LRp}_{MC}$ | $\mathcal{F}^*$ | $\Lambda^*_{MC}$ | $\mathrm{LRp}_{MC}$ | $\mathcal{F}^*$ | $\Lambda^*_{MC}$ | $\mathrm{LRp}_{MC}$ |
| 35 | 11 | 4 | 5 | 17 | 4 | 5 | 47 | 5 | 5 | 92 | 6 | 5 |
| 40 | 11 | 5 | 5 | 16 | 6 | 5 | 36 | 4 | 5 | 83 | 5 | 5 |
| 50 | 10 | 5 | 5 | 13 | 5 | 6 | 29 | 5 | 5 | 68 | 5 | 5 |
| 60 | 10 | 5 | 5 | 10 | 4 | 5 | 23 | 5 | 5 | 52 | 5 | 5 |
| 80 | 9 | 5 | 5 | 11 | 5 | 5 | 20 | 4 | 5 | 36 | 4 | 5 |
| 100 | 8 | 5 | 5 | 10 | 5 | 5 | 16 | 5 | 5 | 30 | 5 | 5 |
| 120 | 8 | 5 | 5 | 11 | 6 | 5 | 14 | 4 | 6 | 24 | 5 | 5 |
| 140 | 8 | 5 | 5 | 9 | 5 | 5 | 13 | 6 | 5 | 20 | 4 | 5 |
| 180 | 7 | 4 | 5 | 8 | 4 | 5 | 11 | 5 | 5 | 18 | 6 | 5 |

Note: $\mathcal{F}^*$ is the asymptotic Bai et al (1998) test, $\Lambda^*_{MC}$ is the exact procedure of Bernard et al (2007) and $\mathrm{LRp}_{MC}$ is *LRpart* with the finite sample correction employed.

Table 2: Empirical Size. Model with intercept and normal regressor. Nominal size = 5%

| | $n=1$ | | | $n=2$ | | | $n=5$ | | | $n=10$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | $\mathcal{F}^*$ | $\Lambda^*_{MC}$ | $\mathrm{LRp}_{MC}$ | $\mathcal{F}^*$ | $\Lambda^*_{MC}$ | $\mathrm{LRp}_{MC}$ | $\mathcal{F}^*$ | $\Lambda^*_{MC}$ | $\mathrm{LRp}_{MC}$ | $\mathcal{F}^*$ | $\Lambda^*_{MC}$ | $\mathrm{LRp}_{MC}$ |
| 35 | 10 | 6 | 5 | 15 | 5 | 5 | 41 | 5 | 5 | 88 | 6 | 5 |
| 40 | 10 | 6 | 5 | 12 | 6 | 5 | 31 | 5 | 5 | 76 | 5 | 5 |
| 50 | 6 | 4 | 5 | 9 | 5 | 5 | 23 | 4 | 5 | 59 | 4 | 5 |
| 60 | 6 | 5 | 5 | 9 | 5 | 5 | 17 | 5 | 5 | 45 | 5 | 5 |
| 80 | 6 | 6 | 5 | 9 | 6 | 5 | 14 | 5 | 5 | 31 | 6 | 5 |
| 100 | 5 | 5 | 5 | 6 | 5 | 5 | 10 | 4 | 5 | 26 | 6 | 4 |
| 120 | 6 | 5 | 4 | 8 | 7 | 5 | 10 | 6 | 5 | 19 | 5 | 5 |
| 140 | 5 | 5 | 5 | 7 | 5 | 5 | 8 | 5 | 5 | 16 | 5 | 6 |
| 180 | 4 | 4 | 5 | 6 | 6 | 5 | 7 | 5 | 5 | 13 | 6 | 5 |

Note: $\mathcal{F}^*$ is the asymptotic Bai et al (1998) test, $\Lambda^*_{MC}$ is the exact procedure of Bernard et al (2007) and $\mathrm{LRp}_{MC}$ is *LRpart* with the finite sample correction employed.

Table 3: Empirical Power. Model with intercept, trend and normal variate. Nominal Size = 5%

| | | $\xi_0$ | 1.5 | | | 5 | | | 10 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | $n$ | $s_0$ | $\Lambda^*_{MC}$ | LP($\Lambda$) | LRp$_{MC}$ | $\Lambda^*_{MC}$ | LP($\Lambda$) | LRp$_{MC}$ | $\Lambda^*_{MC}$ | LP($\Lambda$) | LRp$_{MC}$ |
| 40 | 1 | $[.5T]+1$ | 21 | 12 | 8 | 100 | 84 | 20 | 100 | 100 | 25 |
| | | $[.85T]$ | 37 | 43 | 9 | 100 | 100 | 42 | 100 | 100 | 66 |
| | | $[.95T]$ | 12 | 12 | 5 | 91 | 70 | 4 | 100 | 99 | 3 |
| | 3 | $[.5T]+1$ | 39 | 16 | 7 | 100 | 92 | 12 | 100 | 100 | 12 |
| | | $[.85T]$ | 66 | 68 | 10 | 100 | 100 | 26 | 100 | 100 | 29 |
| | | $[.95T]$ | 20 | 17 | 4 | 99 | 80 | 4 | 100 | 99 | 4 |
| | 10 | $[.5T]+1$ | 59 | 19 | 6 | 100 | 70 | 8 | 100 | 93 | 8 |
| | | $[.85T]$ | 86 | 80 | 8 | 100 | 100 | 10 | 100 | 100 | 11 |
| | | $[.95T]$ | 25 | 19 | 5 | 96 | 55 | 4 | 100 | 65 | 4 |
| 80 | 1 | $[.5T]+1$ | 48 | 27 | 9 | 100 | 100 | 36 | 100 | 100 | 59 |
| | | $[.85T]$ | 69 | 73 | 18 | 100 | 100 | 87 | 100 | 100 | 100 |
| | | $[.95T]$ | 23 | 24 | 6 | 100 | 100 | 15 | 100 | 100 | 33 |
| | 3 | $[.5T]+1$ | 88 | 48 | 12 | 100 | 100 | 24 | 100 | 100 | 26 |
| | | $[.85T]$ | 97 | 97 | 25 | 100 | 100 | 83 | 100 | 100 | 97 |
| | | $[.95T]$ | 49 | 40 | 7 | 100 | 100 | 11 | 100 | 100 | 13 |
| | 10 | $[.5T]+1$ | 100 | 69 | 11 | 100 | 100 | 14 | 100 | 100 | 14 |
| | | $[.85T]$ | 100 | 100 | 25 | 100 | 100 | 44 | 100 | 100 | 47 |
| | | $[.95T]$ | 75 | 56 | 7 | 100 | 100 | 8 | 100 | 100 | 8 |

Note: $\Lambda^*_{MC}$ is the exact procedure of Bernard et al (2007), LP($\Lambda^*$) is the new test introduced by Bernard et al (2007) and LRp$_{MC}$ is *LRpart* with the finite sample correction employed.

Table 4: Empirical Power. Model with intercept, trend and normal variate. Nominal Size = 5%

| | | $\xi_0$ | 1.5 | 5 | 10 | | | 1.5 | 5 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | $n$ | $s_0$ | $LRp_{MC}$ | $LRp_{MC}$ | $LRp_{MC}$ | T | n | $LRp_{MC}$ | $LRp_{MC}$ | $LRp_{MC}$ |
| 100 | 1 | $[.5T]+1$ | 11 | 53 | 85 | 140 | 1 | 14 | 67.16 | 95.52 |
| | | $[.85T]$ | 22 | 95 | 100 | | | 31 | 99.32 | 100.0 |
| | | $[.95T]$ | 6 | 19 | 35 | | | 8 | 38.18 | 78.80 |
| | 3 | $[.5T]+1$ | 15 | 44 | 55 | | 3 | 19 | 61.16 | 79.78 |
| | | $[.85T]$ | 35 | 96 | 100 | | | 51 | 99.94 | 100.0 |
| | | $[.95T]$ | 7 | 13 | 16 | | | 9 | 33.98 | 50.34 |
| | 10 | $[.5T]+1$ | 15 | 21 | 22 | | 10 | 19.84 | 32.00 | 33.70 |
| | | $[.85T]$ | 38 | 69 | 74 | | | 63.14 | 96.92 | 99.04 |
| | | $[.95T]$ | 6 | 8 | 9 | | | 9.46 | 16.98 | 18.44 |
| 120 | 1 | $[.5T]+1$ | 11 | 51 | 83 | 180 | 1 | 16.58 | 81.20 | 99.34 |
| | | $[.85T]$ | 26 | 98 | 100 | | | 40.30 | 100.0 | 100.0 |
| | | $[.95T]$ | 7 | 30 | 60 | | | 8.72 | 52.16 | 90.10 |
| | 3 | $[.5T]+1$ | 15 | 42 | 51 | | 3 | 23.44 | 82.34 | 96.54 |
| | | $[.85T]$ | 43 | 99 | 100 | | | 69.38 | 100.0 | 100.0 |
| | | $[.95T]$ | 8 | 24 | 31 | | | 11.38 | 52.32 | 74.36 |
| | 10 | $[.5T]+1$ | 14 | 21 | 22 | | 10 | 27.02 | 49.62 | 53.10 |
| | | $[.85T]$ | 50 | 99 | 99 | | | 83.32 | 99.92 | 100.0 |
| | | $[.95T]$ | 8 | 13 | 14 | | | 13.50 | 29.54 | 32.78 |

Note: $LRp_{MC}$ is *LRpart* with the finite sample correction employed.

Table 5: Empirical Size: Model with intercept and normal regressor. Nominal size = 5%

| | $n = 1$ | | $n = 5$ | | $n = 10$ | | $n = 15$ | | $n = 20$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| T | *LRpart* | *LRall* | *LRpart* | *LRall* | *LRpart* | *LRall* | *LRpart* | *LRall* | *LRpart* | *LRall* |
| 40 | 6 | 7 | 9 | 12 | 14 | 25 | 26 | 55 | 45 | 89 |
| 50 | 6 | 6 | 7 | 10 | 12 | 18 | 20 | 36 | 31 | 62 |
| 60 | 4 | 5 | 7 | 9 | 10 | 16 | 15 | 26 | 25 | 45 |
| 80 | 5 | 5 | 7 | 8 | 8 | 12 | 11 | 18 | 17 | 28 |
| 100 | 5 | 6 | 6 | 7 | 7 | 10 | 10 | 14 | 12 | 21 |
| 120 | 5 | 6 | 6 | 7 | 7 | 9 | 8 | 12 | 10 | 16 |
| 140 | 5 | 5 | 6 | 7 | 7 | 8 | 7 | 10 | 10 | 14 |
| 180 | 4 | 5 | 6 | 6 | 6 | 7 | 7 | 9 | 8 | 11 |
| 1000 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 4 | 6 |

Table 6: Empirical Size: Model with intercept and trend regressor. Nominal size = 5%

| | $n = 1$ | | $n = 5$ | | $n = 10$ | | $n = 15$ | | $n = 20$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| T | *LRpart* | *LRall* | *LRpart* | *LRall* | *LRpart* | *LRall* | *LRpart* | *LRall* | *LRpart* | *LRall* |
| 40 | 7 | 5 | 13 | 8 | 26 | 15 | 54 | 26 | 88 | 44 |
| 50 | 7 | 6 | 11 | 8 | 12 | 19 | 36 | 19 | 61 | 31 |
| 60 | 6 | 6 | 9 | 7 | 14 | 10 | 27 | 15 | 44 | 24 |
| 80 | 5 | 5 | 8 | 7 | 12 | 8 | 17 | 11 | 28 | 16 |
| 100 | 6 | 5 | 7 | 6 | 10 | 8 | 13 | 9 | 20 | 12 |
| 120 | 6 | 5 | 6 | 6 | 9 | 7 | 12 | 8 | 16 | 10 |
| 140 | 5 | 5 | 6 | 5 | 8 | 6 | 10 | 8 | 13 | 9 |
| 180 | 5 | 5 | 6 | 5 | 8 | 6 | 9 | 7 | 11 | 8 |
| 1000 | 5 | 4 | 5 | 5 | 5 | 4 | 6 | 5 | 6 | 5 |

Table 7: Power. Model with intercept and normal regressor. Nominal size = 5%.
$s_0 = [.85T]$

| $\xi_0$ | | 1.5 | 5 | 10 | 1.5 | 5 | 10 | 1.5 | 5 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | $n$ | *LRpart* | | | *LRall* | | | $\mathcal{F}^*$ | | |
| 40 | 1 | 29 | 43 | 96 | 66 | 98 | 100 | 84 | 100 | 100 |
| | 5 | 33 | 53 | 100 | 35 | 79 | 100 | 35 | 84 | 100 |
| | 10 | 48 | 56 | 100 | 51 | 65 | 100 | 48 | 64 | 100 |
| | 15 | 73 | 64 | 100 | 76 | 66 | 100 | 75 | 67 | 100 |
| | 20 | 94 | 73 | 100 | 94 | 74 | 100 | 94 | 73 | 100 |
| 80 | 1 | 55 | 75 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 5 | 71 | 94 | 100 | 99 | 100 | 100 | 100 | 100 | 100 |
| | 10 | 69 | 93 | 100 | 84 | 100 | 100 | 85 | 100 | 100 |
| | 15 | 69 | 90 | 100 | 78 | 97 | 100 | 78 | 98 | 100 |
| | 20 | 74 | 86 | 100 | 78 | 94 | 100 | 80 | 94 | 100 |
| 120 | 1 | 71 | 90 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 5 | 90 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 10 | 88 | 100 | 100 | 99 | 100 | 100 | 100 | 100 | 100 |
| | 15 | 85 | 99 | 100 | 93 | 100 | 100 | 94 | 100 | 100 |
| | 20 | 82 | 99 | 100 | 89 | 100 | 100 | 89 | 100 | 100 |
| 160 | 1 | 86 | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 5 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 10 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 15 | 98 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 20 | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Note: $\mathcal{F}^*$ is the asymptotic Bai et al (1998) test.

Table 8: Power: Model with intercept and trend. 5% nominal size.

| | | $\xi_0$ | 1.5 | | 5 | | 10 | |
|---|---|---|---|---|---|---|---|---|
| $T$ | $n$ | $s_0$ | *LRpart* | *LRall* | *LRpart* | *LRall* | *LRpart* | *LRall* |
| 40 | 1 | $[.5T]+1$ | 8 | 9 | 13 | 23 | 9 | 30 |
| | | $[.85T]$ | 13 | 14 | 57 | 79 | 89 | 99 |
| | | $[.95T]$ | 7 | 6 | 6 | 11 | 4 | 17 |
| | 3 | $[.5T]+1$ | 10 | 11 | 13 | 17 | 11 | 15 |
| | | $[.85T]$ | 18 | 21 | 47 | 78 | 56 | 95 |
| | | $[.95T]$ | 9 | 7 | 7 | 9 | 6 | 8 |
| | 10 | $[.5T]+1$ | 27 | 18 | 30 | 23 | 30 | 23 |
| | | $[.85T]$ | 35 | 30 | 48 | 52 | 53 | 59 |
| | | $[.95T]$ | 25 | 16 | 23 | 16 | 24 | 16 |
| 80 | 1 | $[.5T]+1$ | 9 | 12 | 27 | 57 | 40 | 91 |
| | | $[.85T]$ | 21 | 28 | 91 | 99 | 100 | 100 |
| | | $[.95T]$ | 7 | 7 | 18 | 24 | 41 | 61 |
| | 3 | $[.5T]+1$ | 11 | 16 | 19 | 49 | 18 | 63 |
| | | $[.85T]$ | 30 | 45 | 95 | 100 | 100 | 100 |
| | | $[.95T]$ | 8 | 8 | 15 | 20 | 15 | 28 |
| | 10 | $[.5T]+1$ | 17 | 22 | 20 | 33 | 22 | 35 |
| | | $[.85T]$ | 45 | 65 | 73 | 98 | 77 | 100 |
| | | $[.95T]$ | 15 | 12 | 16 | 15 | 16 | 16 |

Table 8: (cont.)

| | | $\xi_0$ | 1.5 | | 5 | | 10 | |
|---|---|---|---|---|---|---|---|---|
| $T$ | $n$ | $s_0$ | *LRpart* | *LRall* | *LRpart* | *LRall* | *LRpart* | *LRall* |
| 120 | 1 | $[.5T]+1$ | 13 | 18 | 59 | 83 | 92 | 100 |
| | | $[.85T]$ | 28 | 41 | 99 | 100 | 100 | 100 |
| | | $[.95T]$ | 8 | 9 | 35 | 54 | 67 | 94 |
| | 3 | $[.5T]+1$ | 18 | 25 | 51 | 85 | 66 | 99 |
| | | $[.85T]$ | 48 | 69 | 100 | 100 | 100 | 100 |
| | | $[.95T]$ | 10 | 12 | 31 | 57 | 39 | 84 |
| | 10 | $[.5T]+1$ | 23 | 33 | 33 | 57 | 34 | 61 |
| | | $[.85T]$ | 65 | 90 | 98 | 100 | 100 | 100 |
| | | $[.95T]$ | 14 | 16 | 22 | 34 | 24 | 38 |
| 160 | 1 | $[.5T]+1$ | 18 | 24 | 75 | 94 | 99 | 100 |
| | | $[.85T]$ | 39 | 53 | 100 | 100 | 100 | 100 |
| | | $[.95T]$ | 9 | 11 | 44 | 70 | 84 | 99 |
| | 3 | $[.5T]+1$ | 24 | 35 | 75 | 98 | 96 | 100 |
| | | $[.85T]$ | 64 | 85 | 100 | 100 | 100 | 100 |
| | | $[.95T]$ | 10 | 15 | 43 | 79 | 63 | 99 |
| | 10 | $[.5T]+1$ | 28 | 46 | 50 | 85 | 51 | 90 |
| | | $[.85T]$ | 84 | 98 | 100 | 100 | 100 | 100 |
| | | $[.95T]$ | 15 | 21 | 29 | 57 | 33 | 65 |

Table 9: Power: Model with intercept and normal regressor. $T = 1000$. $s_0 = [0.85T]$.

| $n = 5$ | | | | | | |
|---|---|---|---|---|---|---|
| $\xi_0$ | 0.05 | 0.1 | 0.15 | 0.05 | 0.1 | 0.15 |
| $\pi_T^1$ | *LRpart* | | | *LRall* | | |
| 1/3 | 7 | 22 | 52 | 48 | 78 | 89 |
| 1/4 | 8 | 25 | 55 | 49 | 76 | 90 |
| 1/5 | 10 | 26 | 60 | 48 | 78 | 89 |
| 1/6 | 8 | 24 | 59 | 48 | 77 | 89 |

Table 10: Size: Model with intercept and normal regressor. $U$ follows a t-distribution with $df$ degrees of freedom;
To calculate $LRpart_{MC}$ and $\Lambda^*_{MC}$ a N(0,1) distribution is assumed.

| | $n = 1$, $df = 5$ | | | | $n = 5$, $df = 5$ | | | | $n = 5$, $df = 35$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $T$ | $\mathrm{LRp}_{MC}$ | *LRall* | $\mathcal{F}^*$ | $\Lambda^*_{MC}$ | $\mathrm{LRp}_{MC}$ | *LRall* | $\mathcal{F}^*$ | $\Lambda^*_{MC}$ | $\mathrm{LRp}_{MC}$ | *LRall* | $\mathcal{F}^*$ | $\Lambda^*_{MC}$ |
| 40 | 7 | 6 | 5 | 0 | 7 | 9 | 9 | 0 | 5 | 9 | 8 | 0 |
| 80 | 7 | 5 | 4 | 0 | 6 | 7 | 6 | 0 | 5 | 7 | 6 | 0 |
| 120 | 6 | 5 | 4 | 0 | 6 | 7 | 5 | 0 | 5 | 6 | 5 | 0 |
| 160 | 5 | 4 | 5 | 0 | 6 | 6 | 5 | 0 | 5 | 6 | 5 | 0 |

Table 11: Size: Model with intercept, normal regressor and an AR(1). $T = 1000$. $s_0 = [0.85T]$. $n = 1.\pi_T^1 = 1/3$.

| $T$ | *LRpart* |
|---|---|
| 80 | 3.30 |
| 160 | 3.82 |
| 500 | 4.34 |
| 1000 | 4.48 |

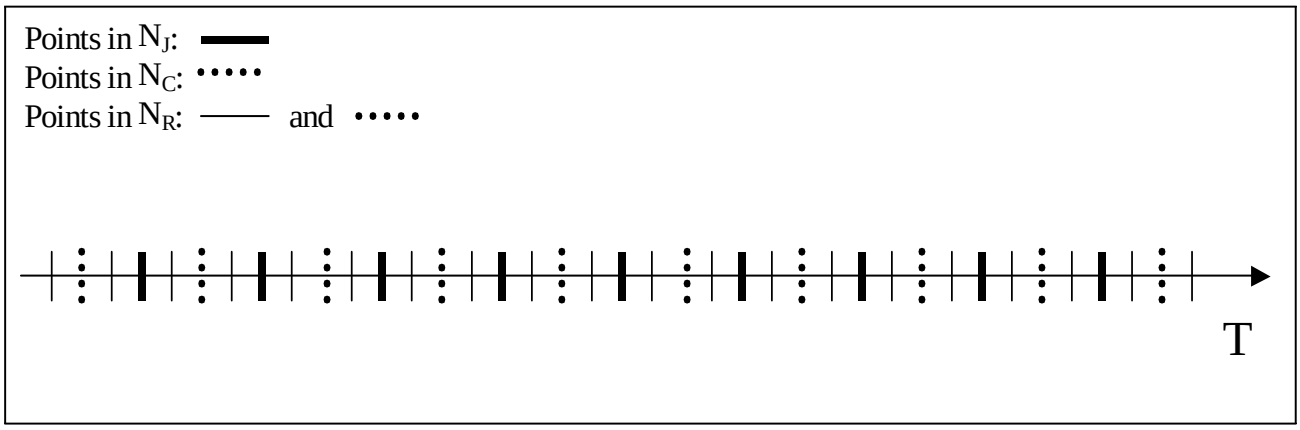
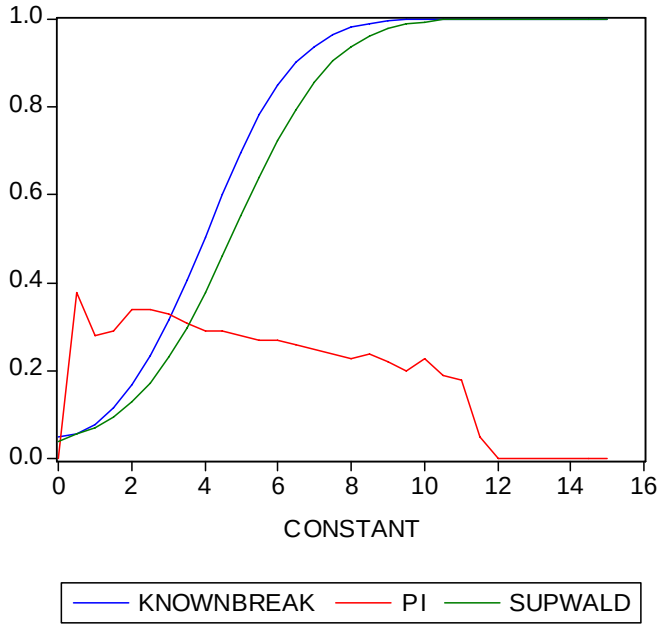



Figure 1: Definitions of sets

Figure 2: Local asymptotic power characterized as $\left(CONSTANT/\sqrt{T}\right)$ of the SupWald test (SUPWALD), the test with the break to be known (KNOWNBREAK) and value of $\pi_T^1$ (denoted PI) that gives same power for the LRpart as the SupWald test. The true break is simulated in the middle of the sample. Number of replications=10000 and $T$=1000. 5% nominal size.



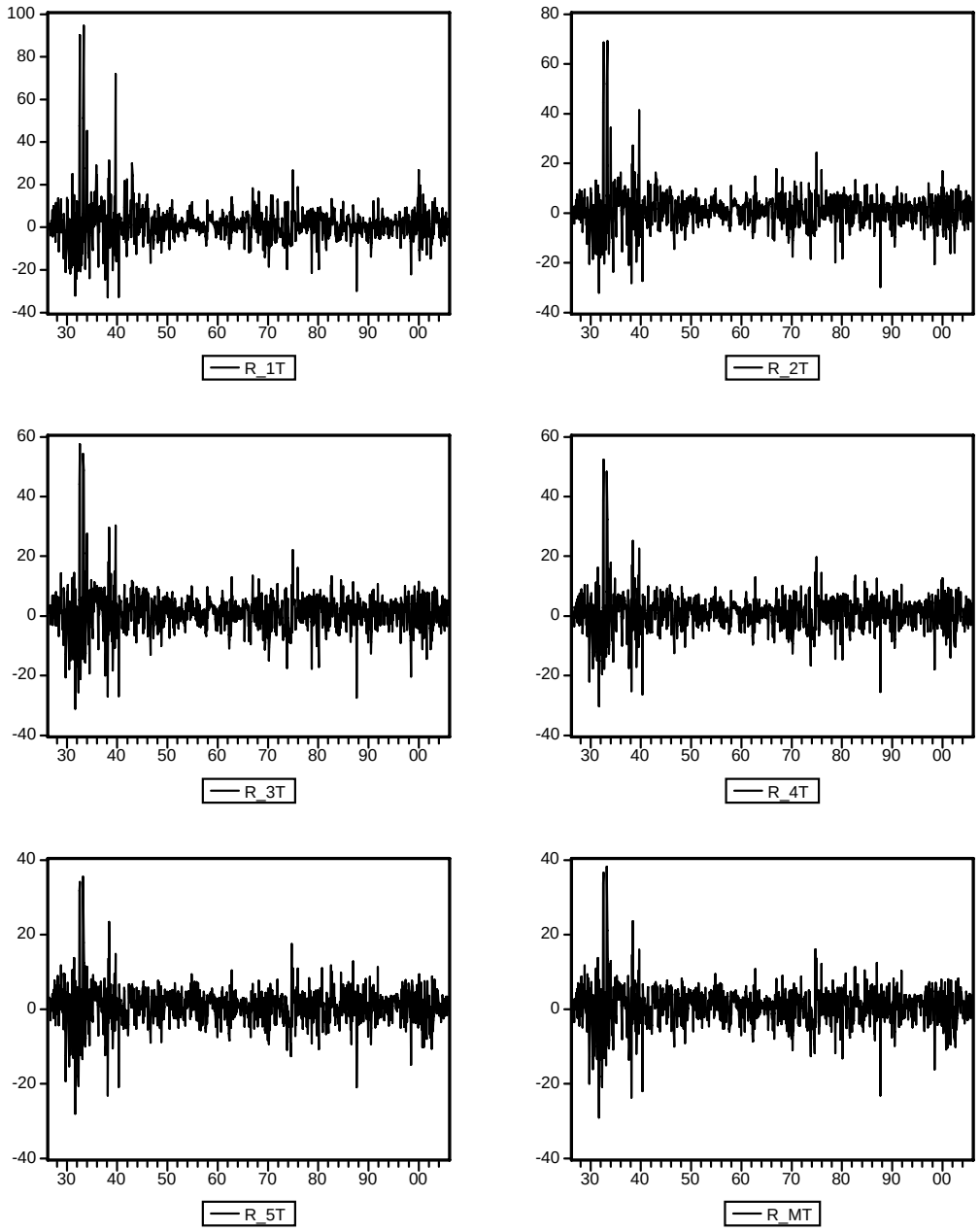



Figure 3: Five portfolios sorted according to size (R_1T,..., R_5T) and excess return on the market portfolio (R_MT)

# References

Andrews, D. W. K. (1988), Laws of large numbers for dependent non-identically distributed random variables, *Econometric Theory,* 4, 458-467.

Andrews, D. W. K. (1993), Test for Parameter Instability and Structural Change with Unknown Change Point, *Econometrica* 61, 821-856.

Andrews, D. W. K. (2003), Test for Parameter Instability and Structural Change with Unknown Change Point: A Corrigendum, *Econometrica* 71, 395-397.

Andrews, D. W. K. (2004), The Block-Block Bootstrap: Improved Asymptotic Refinements, *Econometrica* 72, 3, 673-700.

Ang, A. and J. Chen (2007), CAPM over the long run: 1926-2001, *Journal of Empirical Finance* 14, 1, 1-40.

Andrews and Ploberger (1994), Optimal Tests when a Nuisance Parameter is present only Under the Alternative, *Econometrica* 62, 6, 1383-1414.

Bai, J. (1997), Estimation of a Change Point in Multiple Regression Models, *Review of Economic and Statistics* 79, 4, 551-563.

Bai, J., R. L. Lumsdaine and J. H. Stock (1998), Testing and Dating Common Breaks in Multivariate Time Series, *The Review of Economic Studies* 65, 3, 395-432.

Bai, J. and P. Perron (1998), Estimating and Testing Linear Models with Multiple Structural Changes, *Econometrica* 66, 1, 47-78.

Banerjee, A ., R. L. Lumsdaine and J. H. Stock (1992), Recursive and Sequential Tests for a Unit Root: Theory and International Evidence, *Journal of Business and Economic Statistics*, 10, 271-287.

Bernard, J.-T., N. Idoudi, L. Khalaf and C. Yélou (2007), Finite Sample Multivariate Structural Change Tests with Application to Energy Demand Models, *Journal of Econometrics* 141, 1219-1244.

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions, *Econometrica* 28, 3, 591-605.

Davies, R. B. (1977), Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative, *Biometrika*, 64, 247-254.

Dueker, M. J. (1992), The Response of Market Interest Rates to Discount Rate Changes, *Federal Reserve Bank of St. Louis Review* 74:4, 78-91.

Dufour, J.-M. and E. M. Iglesias (2008), Finite sample and optimal adaptive inference in possibly nonstationary general volatility models with gaussian or heavy-tailed errors, Working paper, University of Montreal.

Dufour, J.-M. and J. Jasiak (2001), Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors, *International Economic Review* 42, 815-843.

Dufour, J.-M. and L. Khalaf (2002), Simulation Based Finite and Large Sample Tests in Multivariate Regressions, *Journal of Econometrics* 111, 303-322.

Dufour, J. M. and J. Kiviet (1996), Exact Tests for Structural Change in First-Order Dynamic Models, *Journal of Econometrics* 70, 39-68.

Dufour, J.-M. (2006), Monte Carlo Tests with Nuisance Parameters : A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics, *Journal of Econometrics* 133, 2, 443-477.

Elliott, G. and U. Müller (2006), Efficient Tests for General Persistent Time Variation in Regression Coefficients, *Review of Economic Studies* 73, 907-940.

Gibbons, M. (1982), Multivariate Tests of Financial Models: A New Approach, *Journal of Financial Economics*, 10, 3-28.

Greene, W. H. (2003), *Econometric Analysis*, 5th Edition, Prentice Hall.

Hansen B. E. (1992), Tests for Parameter Instability in Regressions with I(1) Processes, *Journal of Business and Economic Statistics*, 10, 321-335.

Huang, H.-C. and W.-H. Chen (2005), Tests of the CAPM under structural changes, *International Economic Journal* 19, 4, 523-541.

Mcleish, D. (1975), A maximal inequality and dependent strong laws, *The Annals of Probability* 3, 829–839.

Qu, Z. and P. Perron (2007), Estimating and Testing Structural Changes in Multivariate Regressions, *Econometrica* 75, 459–502.

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

Rossi, B. (2005), Optimal Tests for Nested Model Selection with underlying Parameter Instability, *Econometric Theory* 21(5), 962-990.

Sowell, F. (1996), Optimal Tests for parameter Instability in the Generalized Method of. Moments Framework, *Econometrica* 64,1085–1107.

Stock, J. and M. Watson (1996), Evidence on Structural Instability in Macroeconomic Time Series Relations, *Journal of Business and Economic Statistics* 14, 1, 11-30.

Newey, W.K. and K. D. West (1987), A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55, 703–708.

Zivot, E. and D. W. K. Andrews (1992), Further Evidence on the Great Crash, the Oil PriceShock and the Unit Root Hypothesis, *Journal of Business and Economic Statistics*, 10, 251-270.