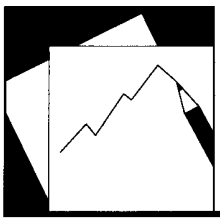
WP/08/75



IMF Working Paper

Testing for Structural Breaks in Small Samples

Sergei Antoshin, Andrew Berg, and Marcos Souto

INTERNATIONAL MONETARY FUND

 WP/08/75

IMF Working Paper

African Department

Testing for Structural Breaks in Small Samples

Prepared by Sergei Antoshin, Andrew Berg, and Marcos Souto[1]

March 2008

Abstract

In a recent paper, Bai and Perron (2006) demonstrate that their approach for testing for multiple structural breaks in time series works well in large samples, but they found substantial deviations in both the size and power of their tests in smaller samples. We propose modifying their methodology to deal with small samples by using Monte Carlo simulations to determine sample-specific critical values under the null each time the test is run. We draw on the results of our simulations to offer practical suggestions on handling serial correlation, model misspecification, and the use of alternative test statistics for sequential testing. We show that, for most types of data generating processes in samples with as low as 50 observations, our proposed modifications perform substantially better.

JEL Classification Numbers: C29, C39, C59

Keywords: Structural breaks, small samples, Monte Carlo simulation.

Author's E-Mail Address: santoshin@imf.org, aberg@imf.org, msouto@imf.org.

[1] We thank without implication for numerous useful conversations with Jonathan Ostry and Jeromin Zettelmeyer, who suggested the sequential UDmax procedure presented here, and for useful comments from Jushan Bai, Sam Ouliaris, Zhongjun Qu, and Pierre Perron. Zhongjun also guided us through Perron's code on structural breaks, which we have modified and used extensively in this study.

Contents Page

I. INTRODUCTION

In a series of influential papers, Bai and Perron (1998, 2003a and 2003b, henceforth BP) developed a methodology for finding multiple structural breaks in time series and testing for their statistical significance. The simulation analysis conducted in BP (2006) demonstrates that the size and power of their tests can be significantly distorted by several factors, such as: 1) a small sample size, 2) a small break size, 3) a small segment size and breaks clustering, and 4) the use of heteroskedasticity and autocorrelation corrections.

In this paper, we extend the BP methodology in several directions, all aimed at improving small-sample (time series with as low as 50 observations) performance. First, in tests for significance of structural breaks, we propose to use critical values that are specific to the time series in question, instead of relying on the asymptotic critical values (i.e., bootstrap). The asymptotic critical values in BP are generated for Wiener (white Gaussian noise) processes with a large number of observations, and can cause considerable distortions in the test size and power for small samples with a non-Wiener data generating process. We instead estimate a "mimicking process" from the data under the null and bootstrap critical values at each step of the sequential procedure, under the corresponding null hypothesis. The use of bootstrapped segment-specific residuals allows us to: calculate sample-size-specific critical values; relax the assumption of the normal distribution of the residuals; and account for segmental heteroskedasticity.

Second, we address the issue of misspecification of the data generating process. In the presence of serial correlation, BP consider two alternative approaches to modeling the underlying data generation process. The first approach is to model the process explicitly (e.g., as an AR(1)), so that the error terms are independently identically distributed (iid). The second approach is to model the process in a simple way (e.g., as a Wiener process), and to use a heteroskedasticity-autocorrelation-consistent (HAC) correction. In the general case, when the nature of the process is unknown, the first approach may yield over-specified tests, while the second may yield underspecified statistics in small samples.

Third, we examine the small-sample performance of the two statistics put forward by BP for testing for an unknown number of multiple breaks. After finding a first break, BP suggest testing sequentially for two breaks versus a null of one break by testing for the existence of one break in each of the two segments formed by the initial break (the sequential supF test), and so on, until the null hypothesis is not rejected. We compare this approach to a variant that uses another BP statistic to test for any number of breaks in each segment (the sequential Dmax test). We show (?) that the performance of the sequential supF test can be poor when the segment size becomes small.

We focus on a sample size of 50 observations, where the true number of breaks is as high as two. This case is partly inspired by a companion paper (Berg, Ostry and Zettelmeyer,

2006) which uses the techniques presented here to characterize and analyze breaks in annual per capita GDP growth for a broad sample of countries.

The rest of the paper is organized as follows. In Section II, we briefly review the BP methodology, focusing on the empirical procedure and simulation analysis. In Section III, we outline our strategy on how to modify and apply the BP methodology for small samples. The results from our Monte Carlo simulations are presented in Section IV. Section V includes discussion and concluding remarks.

II. THE BP METHODOLOGY

Drawing heavily on Bai and Perron (1998, 2001), we summarize the main elements of their methodology for estimating and testing linear models for multiple structural changes, focusing on the ones that are most relevant to our analysis in Section III.

The BP methodology can be disentangled in two separate and independent parts. First, one can identify any number of breaks in a time series, regardless of statistical significance. Second, once the breaks have been identified, BP propose a series of statistics to test for the statistical significance of these breaks, using asymptotic critical values. As we shall see in more details below, these statistics can yield significant deviations in both size and power, especially when dealing with small time series (with as low as 50 observations).

It is worth stressing one finite-sample complication involved in testing for the statistical significance of a set of breaks, which forms the second part of the BP methodology. The usual method is to use the F ratio that compares the SSR for the restricted versus the unrestricted model. For example, in testing for the presence of one break, the F ratio is the ratio between the SSR for 0 breaks over the SSR for one break. Because the breaks are found through a global minimization procedure, there are instances when the set of t breaks is not a subset of $t+1$ breaks. In this case, the hypothesis of $t+1$ breaks does not nest the hypothesis of t breaks, and the SSR_{t+1} / SSR_t ratio does not have the property of asymptotic convergence to the F-distribution. In particular, its asymptotic distribution depends on sample-specific parameters, such as the size of the break.

BP propose to overcome this problem by always testing for the presence of one break versus 0 breaks in the segments between breaks, thus avoiding the issue of non-nested hypotheses. But this solution comes at a price, particularly when dealing with an already small time series: the segments will be even smaller and the statistics will need to be computed/calculated with just a few observations.

One important advantage of the BP framework is its capability of allowing for autocorrelation and heteroskedasticity in the time series, as compared to other breaks selection procedures that cannot accommodate these features (e.g., the Bayesian Information Criteria by Yao (1988) and the modified Schwarz criterion proposed by Liu et al. (1997)). This feature is of particular importance in BP methodology, as their statistics utilize

asymptotic critical values that are generated for a Wiener process. To deal with autocorrelation in a non-parametric fashion, BP propose to correct the time series residuals either through a Newey-West procedure or by including the lag of the time series as one of the regressors in the projection model.

BP provide tables with asymptotic critical values for all statistics (at main confidence levels), for a Wiener process. When dealing with smaller time series, BP recommend using a larger segment size, relative to the sample size. BP also suggest using the autocorrelation and heteroskedasticity correction only when there is a strong prior that the correction is necessary.

A. The model

BP adopt the following model:

$$y_t = x_t'\beta + z_t'\delta_j + u_t, \tag{1}$$

for $j = 1,\ldots,m+1$, where m is the number of breaks, y_t is the dependent variable, x_t and z_t are vectors of covariates, β and δ_j are the corresponding vectors of coefficients, and u_t is the disturbance term.

This model has some interesting features. First, it allows for joint the estimation of the regression coefficients, through the term $x_t'\beta$, along with the identification of structural changes, captured through the term $z_t'\delta_j$, which may be useful for several applications. Second, equation (1) represents a partial structural model, since the parameter vector β is not subject to shifts and is estimated using the entire sample. Dropping the term $x_t'\beta$ from equation (1) results in a pure structural change model, where all coefficients are subject to change, and is the model used for the analysis in this paper. Finally, u_t can be non iid under the null.

For locating the breaks, BP propose two approaches using (1). In the first, global, approach, each partition m, where m is the number of breaks, is obtained as the one that minimizes the sum of square residuals (SSR). In other words, the break locations T_i, $i = 1,...,m$, are determined so as to minimize $\sum_{i=1}^{m+1}\sum_{t=T_{i-1}+1}^{T_i}[y_t - x_t'\beta - z_t'\delta_j]^2$. BP use a dynamic programming algorithm so as to optimize the computational time when finding the global SSR-minimizing breaks.

In the second approach, breaks are determined sequentially, starting with the single break that minimizes the SSR. Then, for each resulting partition, the single break that minimizes the SSR is determined. The second break is the one with the minimum SSR between the two. This process is repeated sequentially to find further breaks. The search for

the breaks that minimize SSR is implemented regardless of whether these breaks are statistically significant or not. As it turns out, the test for the existence of breaks can be done separately, which will be discussed below.

The procedure of global minimization has the advantage of assuring that only the biggest breaks (i.e., those that cause the biggest reduction in the SSR) will be selected (as opposed to the sequential breaks selection), at least asymptotically. This distinguishes the approach from others that proceed sequentially (e.g. Altissimo and Corradi (2003)).[2] The main disadvantage, as we shall see, which is related to the fact that, for a particular time series, the biggest *n* breaks may not all be included among the biggest *n+1* breaks. This issue poses significant challenges for sequentially testing for the significance of the breaks, as the tested hypotheses will in general be non-nested.

B. Testing for the existence of breaks

The statistics proposed by BP for multiple breaks are generalizations of Andrews (1993) test for the single structural change case, and are shown to be robust to serial correlation and heterogeneity of the residuals under the null.

B.1. Zero versus a fixed number of breaks

In this case, one wants to test the null hypothesis of no breaks against the alternative of a known number of breaks k. The test is calculated as the usual F-ratio between the SSE for the null ('unrestricted' SSE) and the SSE for the alternative hypothesis ('restricted' SSE). In other words, it is simply the conventional test of the null $\delta_1 = ... = \delta_{k+1}$ against the alternative $\delta_i \neq \delta_{i+1}$, for some i, where δ is the vector of coefficients attached to the covariate z in the pure structural change model. For the global minimized breaks, this test is referred to as the sup F(0, m).

One problem with this formulation relates to the estimation of the variance-covariance matrix for δ, which is part of the formula for the F-statistic and which may become quite cumbersome to compute in the presence of autocorrelation and heteroskedasticity in the error term. To overcome this problem, BP propose to estimate a much simpler variance-covariance matrix for δ that is equivalent asymptotically. However, it has been shown that this simplification can introduce a source of potential size and power distortions, particularly when this test is used in small time series.

[2] Sequential methodologies first find the single break that minimizes the SSR. If this break is found to be statistically significant, then they move to find the second break, given the existence and location of the first break, that minimize the SSR, and so forth.

B.2. Zero versus an unknown number of breaks

The number of breaks is often not known, and the standard F-statistic becomes insufficient for testing for the existence of breaks. In this case, BP propose variations of the sup F(0, m) test, which are called double maximum tests and are defined as:

$$D_{\max} = \max_{n=1,\ldots,m} \left(a_n \sup F(0,n) \right), \tag{2}$$

where the weights a_n can be equal to 1, for all $n = 1,\ldots,m$. In this case, the $D_{\max}$ statistic is called $UD_{\max}$ test. More generally, a_n can be a function of the asymptotic critical values for the $\sup F(0,n)$, so as to make the marginal p-values equal across the values of *n*, in which case the $D_{\max}$ statistic is called $WD_{\max}$ test. It is important to note that since the $D_{\max}$ statistics are based on the sup F(0, m), finite sample distortions in the estimation of the variance-covariance matrix for δ will also affect the size and power of the $UD_{\max}$ and $WD_{\max}$ tests..

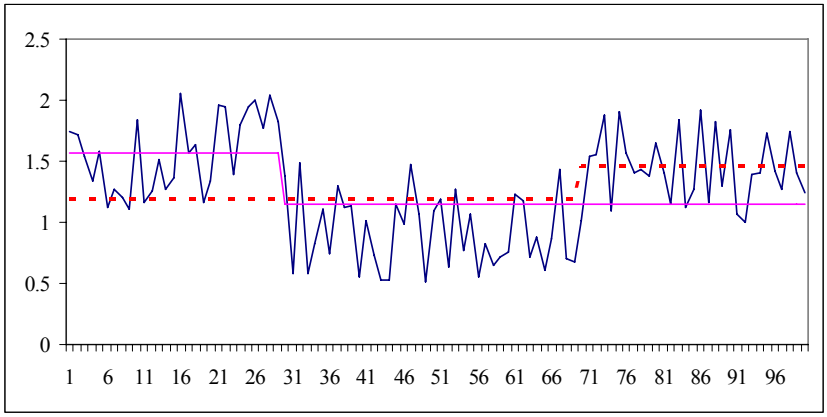
B.3. l versus $l+1$ breaks

Similarly to the F(0, m) ratio, the F($l+1|l$) ratio also relates the 'unrestricted' SSE (for l breaks), to the 'restricted' SSE(for $l+1$ breaks). Calculating the F($l+1|l$) ratio is equivalent to estimating $l+1$ tests of the null of zero breaks against the alternative of a single break. More specifically, the test decides in favor of the null whenever the sum of SSE for the optimal $l+2$ partitions (or $l+1$ breaks) is sufficiently larger than that for $l+1$ partitions (or l breaks). A complicating factor is that the critical values of the statistic under the null $l+1$ depend on sample-specific factors, such as the break size and the properties of the residual. BP propose an alternative approach that uses the sup F(0,1) (testing for the presence of one significant break) in each of the partitions. If the null of 0 breaks can be rejected against the alternative of one break in at least one of the $l+1$ partitions, then BP approach establishes that l+1 breaks are statistically significant.

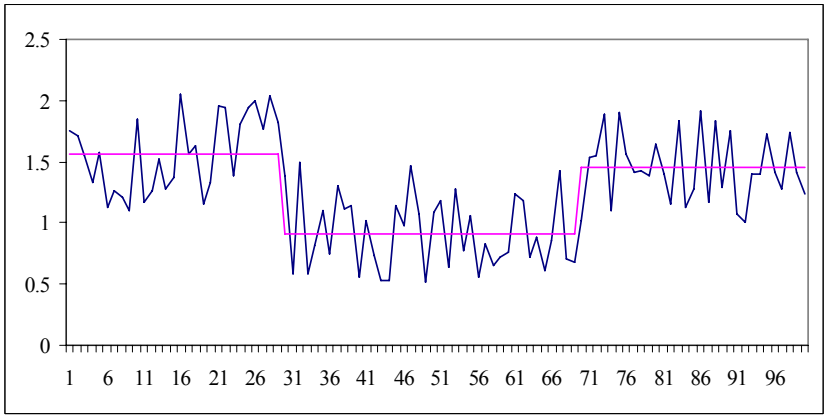
B.4. Criteria for finding the number of breaks

The number of significant breaks can be found via information criteria, such as the Bayesian Information Criterion (BIC), proposed by Yao (1988); and the modified Schwarz criterion (LWZ), proposed by Liu et al. (1997). It is also possible to determine the number of breaks by estimating a sequence of sup F statistics, as suggested by BP. The basic steps would include testing for the presence of one break via the sup F(0,1) and moving forward to test for the presence of $l+1$ breaks, via the F($l+1|l$) ratio, stopping when the null is not rejected. The variance-covariance of δ embedded in these tests, is robust to heteroskedasticity and auto-correlation. Thus, the BP approach accounts for these features, unlike the information criteria-based approaches.

The BP approach may, however, incorrectly estimate the number of significant breaks in some situations, particularly when time series have more than one break and the regimes switch up and down. To illustrate this point, consider Figures 1a and 1b.



(a) One break.



(b) Two breaks.

Figure 1: The problem of sequential testing for determining the number of breaks.

In the situation depicted on Figure 1, a sequence of the sup F($l+1|l$) tests may fail to detect the correct number of breaks. The test of one break against the null of zero breaks may lack power, because the alternative of one break is badly misspecified. According to the BP sequential algorithm, there is no test of whether the sup F(1,2) will reject the null of one break in favor of the alternative of two breaks. This problem can be reduced by using the Dmax statistics in the first step (as proposed by BP (2001)), but the identification problem will still persist when there may be a greater number of breaks.

III. A MODIFIED BP METHODOLOGY FOR SMALL SAMPLES

To address the issues mentioned above, we propose to extend the BP approach for testing the significance of the breaks in two main dimensions:[3]

1. Sample-specific critical values. Each time we run a particular test on a given time series, we perform Monte Carlo simulations to estimate sample-specific critical values. To carry out this procedure, we define a "mimicking process" with structural parameters estimated from the data, under the null hypothesis. We then simulate this process with bootstrapped residuals to infer appropriate critical values for the underlying data generation process. For example, suppose we have rejected the null of 0 breaks. We then identify the size and location of the largest break. In testing for 2 breaks against 1, the mimicking process includes the estimated single break. Thus, for the test of 2 breaks against the null of 1 break, the critical values take into account the size and location of this break, as well as the length and variance of the series in each segment. More generally, the critical values are tailored to the specific series for which the breaks are being investigated, both in terms of sample size, the nature of the process, and heteroskedasticity across segments. This allows us to achieve the following objectives: 1) correction for the small sample size; 2) reduction in misspecification bias[4]; and 3) control for segment-specific heteroskedasticity.

2. Treatment of serial correlation. We also investigate how best to adjust for serial correlation in small samples, by contrasting two different approaches:

Parametric: If one suspects that the time series possesses serial correlation and its order can be guesstimated, then it is possible to model the process explicitly (e.g. an AR(1) process). The BP approach allows one to model the process parametrically, by adding y_{t-1} as one regressor in the break estimation (then breaks regression will no longer be a pure structural model).

HAC robust errors: An alternative to explicit modeling is a robust-error correction, when the nature of serial correlation is complex or unknown. In this case, the BP methodology implies modeling the process as a Wiener process and uses an HAC estimate of the variance-covariance matrix.

[3] Appendix I outlines the algorithms in more detail. GAUSS programs are available from the authors on request.

[4] Whether using asymptotic or sample-specific critical values, there is always the risk of misspecifying the underlying process. However, we find the sample-specific critical values to produce better test results when compared to asymptotic critical values even with some misspecification, as discussed below.

The parametric method should be superior when the order of serial correlation is known.[5] However, even in this situation, the estimation of at least two parameters instead of only one, as in the second method, may reduce the power of the first method, if the sample size is small owing to a poorly estimated model. This disadvantage may become especially severe when a sequential procedure is used, and statistics are estimated in the partitions with even fewer observations available.

Another potential problem relates to the degree of serial correlation. For example, for an AR(1) process with a large β and at least one significant break, the power erodes as β increases, because the F-statistics under the null will be contaminated by the fitted AR(1) process.[6]

In what follows, we examine which method performs better in small samples in terms of the size and the power, and under what specific conditions.

3. Alternative statistics for finding multiple breaks: As mentioned above, the standard statistic used in the BP approach for selecting the number of significant breaks is the sequential sup F test. We propose the use of two alternative statistics:

Sequential UDmax: The sequential UDmax procedure is similar to the sequential sup F test. However, instead of testing the null of 0 breaks against 1 break, we propose to test the null of 0 breaks against the alternative of an unknown number of breaks, in order to avoid the potential power distortions depicted in Figure 1 above. The UDmax(n) test, as defined above, is the maximum of {supF(0,1), supF(0,2),…, supF(0,n)}.

The sequential UDmax should be able to capture more breaks in the situation when the alternative of 1 break is not accepted, even though the alternative of 2 breaks would be accepted. A potential problem with the UDmax is that its distribution is unknown, because the UDmax(n) can be in fact be any of the sup F(0,i), i=1,n. We avoid this issue by bootstrapping the critical values.

Global UDmax: For the sequential procedures described above, the UDmax statistic needs to be estimated with respect to each segment. When the time series sample size is small, having to estimate UDmax in a segment may be even more prone to distortions, when

[5] Potential problems with parametric modeling are: (i) that one may overspecify the process, and (ii) the number of introduced parameters may be too large relative to the sample size. Both problems lead to a reduction in the power. When the order of autocorrelation is known, the process will not be overspecified the process; when the order of autocorrelation is low, the number of parameters is low relative to the sample size.

[6] Suppose an AR(1) process with $\beta = \beta_1$ and one large break. When we fit an AR(1) process into the data, without the break, the estimated beta will be $\hat{\beta} = \beta_2 \geq \beta_1$. The larger the difference $\beta_1 - \beta_2$, the larger is the difference between $supF_1$ and $supF_2$ (with $supF_2 > supF_1$).

contrasted to asymptotic critical values. Moreover, as noted above, in finite samples the globally SSR-minimizing n+1 breaks will not generally nest the globally-SSR-minimizing n breaks. This may lead to low power when applying sequential break selection methods such as the sequential UDmax.

We propose here to make the full use of the whole sample and its properties by using what we call the global UDmax (GUDMax). The corresponding procedure is also sequential, but instead of partitioning the series into n segments at the n-th step, we propose to embed n-1 breaks at the n^{th} step and to use the modified series to generate critical values for the UDmax statistics (with n-1 breaks embedded). We then test the null of these specific n-1 breaks against the alternative of <u>at least</u> n breaks.

The distribution of such a test statistic depends on the size of each of the n-1 breaks, the standard error of the series, and other sample-specific parameters. Moreover, the text is non-nested, insofar as the n breaks need not include the n-1 breaks under the null. Thus, we do not know the analytic distribution of the test statistic, and critical values cannot be usefully determined, even using simulation methods. However, we propose estimating the critical values "on the fly" for each specific application of the test. This is feasible for a number of applications, if still somewhat time-consuming.

IV. RESULTS

We carry out size and power tests. The purposes of the size tests are: (i) to assess the magnitude of improvement in the test size owing to sample-specific critical values; and (ii) to infer which type of modeling—parametric or non-parametric—when serial correlation is present. In the section on power below, we examine the extent to which the procedures find the right number of breaks when there are in fact some breaks. We investigate which type of modeling and which statistics perform better for different break structures.

A. Size tests

The essence of the size test is to determine the proportion of the times when the estimation procedure finds at least one break, when in fact there are no breaks.[7] These size tests are one-sided, in that we focus on a null of zero breaks.

We begin with the assessment of improvement in the test size owing to sample-specific critical values for a Wiener data generating process (DGP), when the modeling approach also assumes no serial correlation or heteroskedasticity. The selection of a Weiner process allows us to isolate the effect of using sample-specific, and in particular sample-size-specific, critical values. Table 1 shows the actual size for tests with a 10-percent theoretical

[7] See Appendix 1A for the algorithms for the size tests.

size for the asymptotic and sample-specific critical values for different sample sizes. For example, for the sample size of 50, the actual size for the asymptotic critical values is 4-5 percent for the sup F's and is 8 percent for the UDmax(3). The sample-specific procedure outperforms the asymptotic one, yielding 8-10 percent for the sup F's and 8 percent for the UDmax, respectively.[8]

We next examine test sizes when (A) there is serial correlation in the DGP; and (B) when there is not but the modeling procedure assumes that there is.

Case A. Serial correlation is present:

Parametric estimation: Here, the test procedures model the processes (correctly) as autocorrelated of order 1. For high degrees of autocorrelation, only the sample-specific approach performs adequately (Table 2). The asymptotic procedure for AR(1) processes has a very poor size in our sample of 50 observations with serial correlation; for a theoretical size of 10 percent, the actual size is in the range of 8-15 percent for an autoregressive parameter of 0.5 and of 26-44 percent for 0.9. Although this result is no entirely surprising, it is worth mentioning that the sample-specific test does much better, with actual sizes of 9-11 percent and 14-15 percent, respectively.[9]

Non-parametric, with robust error: When variance-covariance matrices are calculated so as to be robust to serial correlation, in small samples in the presence of serial correlation, our procedure can yield tests of somewhat more accurate size, though both are poor. Table 3 shows that: (1) both the asymptotic and sample-specific procedures perform poorly (the UDmax(3) p-values are 0.87 percent and 0.85 percent, respectively) when the autocorrelation coefficient is high; (2) the sample-specific procedure somewhat outperforms the asymptotic one when serial correlation is moderate (0.5) as the p-values for UDmax are 0.41 percent and 0.34 percent respectively (BP 2003b, Table 1, DGP-4, section 4).

Parametric vs. Non-parametric: If serial correlation is present, size considerations argue strongly for parametric modeling, because the robust error correction does not improve the test size sufficiently. However, we shall see that the story is somewhat different when looking at power.

Case B. Serial correlation is absent:

When serial correlation is suspected to be present in the time series, a natural question arises: what happens if one assumes autocorrelation when there is none? Is the

[8] Our results for the asymptotic test are consistent with the published Bai and Perron (2003b) tables (Table 1 DGP-1 section 1), except that the UDmax has slightly higher true size.

[9] The (unnecessary in this case) application of a White heteroskedasticity correction substantially worsens the BP size results but has no important effect on the size of the sample-specific tests.

overspecification bias considerable? Table 4 shows that one does not run into a problem if the model is overspecified through an AR(1) modeling when the sample-specific critical values are used (the p-values are 8-9 percent). The asymptotic results are inferior (the p-values are 3-8 percent).

B. Power tests

To assess the power of the test, which is DGP specific, we assume that he DGP is either Wiener or AR(1), with an autoregressive parameter of either 0.5 or 0.9. The sample size is 50. The mimicking process which corresponds to the type of modeling we implement is either a Wiener with robust errors or an AR(1). We report the four statistics discussed above: (i) the asymptotic sequential supF (the Bai-Perron statistic); (ii) the sample-specific sequential supF; (iii) the sequential UDmax; and (iv) the global UDMax (GUDMax).

We embed two breaks in the original series. The break structure makes a difference, and the basic structures may be represented by two breaks. We consider three cases: (1) a symmetric U-form with a downbreak followed by an equally large upbreak (2) an asymmetric U-form, and (3) a downhill (two consecutive downbreaks). For cases 1 and 3, we have two sizes of the breaks: medium and large. The location of the breaks is constant and is such that the two breaks divide the sample into three approximately equal sub-samples. The standard deviation is assumed constant across the three segments.[10] The break sizes are measured in multiples of standard deviations. For simplicity, we denote the break structure in terms of the segment means. For example, $\{1,0,1\}$ means that two breaks of the size of 1 standard deviation are embedded to form a symmetric U-shape.

A measure of the test power is the probability of finding two breaks. A more complete measure of performance takes into account that missing both breaks is worse than missing one, while finding four is worse than finding three. Thus, we define a loss function $L = \frac{1}{M}\sum_{i=1}^{M}\left|\hat{N}_i - N\right|$, where N is the true number of breaks, $\hat{N}_i$ is the estimated number of breaks in the i^{th} run of the simulation and there are a total of M runs.[11] This function gives equal weight to type I and type II errors but would be easy to generalize.

The results of the simulations reported in Table 5 permit several conclusions. First, the sample-specific critical values procedures broadly and considerably outperform the

[10] It is possible to relax this assumption.

[11] This loss function is equivalent to $L = \sum_{all\ \hat{N}} P(\hat{N})\left|\hat{N} - N\right|$, where $P(\hat{N})$ is the frequency with which the procedure finds $\hat{N}$ breaks during the simulation. This formulation shows that the value of the loss function can be quickly calculated from the information in Table 5.

asymptotic one. Second, in most cases and in contrast to the results on test size, the Wiener modeling with robust errors deals with autocorrelation better than the AR(1) modeling. Thus, if the goal is to find the right number of breaks rather than to avoid finding false breaks, using robust errors is generally the better approach. Third, among the sample-specific approaches, in a majority of the cases, the sequential UDmax is the best-performing statistic. Fourth, the GUDmax statistic performs best when the DGP has high autocorrelation. It is substantially more conservative, in that most errors represent an underestimation of the correct number of breaks, while the sequential UDmax is more likely to overestimate. Finally, the power of the sequential UDmax and GUDmax are still better than that of the BP asymptotic statistic, even for a large degree of autocorrelation. If one suspects there is a high degree of autocorrelation in the series, than using the non-parametric approach may yield a better power.

V. SUMMARY AND CONCLUSION

In this paper, we focus on the application to small samples of the techniques of Bai and Perron (1998, etc. etc.) for finding multiple breaks in time series. We focus on a sample size of 50 observations, where the true number of breaks is as high as two. This case is partly inspired by a companion paper (Berg et al. 2007, not cited in the references) which uses the techniques presented here to characterize and analyze breaks in annual per capita GDP growth in a broad sample of countries.

Our first innovation is a sample-specific approach to finding critical values for structural break tests, derived from the asymptotic procedures of Bai and Perron. Rather than tabulate critical values, we develop processes that mimic the null hypothesis, and then estimate critical values using Monte-Carlo simulations. This approach can be computationally intensive, particularly because the procedure must be repeated many times to determine the number of breaks in one series. However, we show that the it produces substantially superior results. In particular, the use of sample-specific critical values substantially improves the test size and power in small samples for nearly all tested specifications of the data generating process.

Second, we investigate the small-sample properties of various methods of dealing with serial correlation. We find that using robust errors deals with autocorrelation better than parametrically modeling the time series. In particular, if the goal is to find the right number of breaks, rather than to avoid finding false breaks, we suggest that using robust errors is generally the better approach.

Lastly, we examine the small-sample performance of the procedure suggested by BP for determining the correct number of multiple breaks, based on the sequential Sup F test. We propose two alternatives, one very similar to the suggested procedure and based on a sequential UDmax test, and one more novel, based on the global UDmax statistic. Both the proposed alternatives generally perform substantially better than the BP proposal. For most

cases we examine, the sequential UDmax is the best-performing statistic. We find that the global UDmax statistic performs best when the DGP has high autocorrelation. It is substantially more conservative, in that most errors represent an underestimation of the correct number of breaks, while the sequential UDmax is more likely to overestimate.

References

Altissimo, F. and V. Corradi, 2003, "Strong rules for detecting the number of breaks in a time series," *Journal of Econometrics* 117, 207-244.

Andrews, D. W. K., I. Lee, and W. Ploberger, 1996, "Optimal changepoint tests for normal linear regression," *Journal of Econometrics* 70, 9-38.

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

———, 2003a, "Critical values for multiple structural change tests," *Econometrics Journal* 6, 72-78.

———, 2003b, "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics* 18, 1-22.

———, 2006, "Multiple Structural Change Models: A Simulation Analysis," in *Econometric Theory and Practice: Frontier of Analysis and Applied Research (Essays in Honor of Peter Phillips)*, ed. by Corbae D., S. Durlauf and B.E. Hansen, Cambridge University Press.

Bai, J., 1997, "Estimating Multiple Breaks One at a Time," *Econometric Theory* 13, 315-352.

———, 1999, "Likelihood ratio tests for multiple structural changes," *Journal of Econometrics* 91, 299-323.

Campos, J., N. R. Ericsson, and D. F. Hendry, 1996, "Cointegration tests in the presence of structural breaks," Journal of Econometrics 70, 187-220.

Diebold, F. X. and C. Chen, 1996, "Testing structural stability with endogenous breakpoint: A size comparison of analytic and bootstrap procedures," Journal of Econometrics 70, 221-241.

Garcia, R. and P. Perron, 1996, "An Analysis of the real Interest Rate Under Regime Shifts," *Review of Economics and Statistics* 78, 111-125.

Heston, A., R. Summers and B. Aten, Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002.

Krishnaiah, P. R. and B. Q. Miao, 1988, "Review about Estimation of Change Points," in *Handbook of Statistics*, Vol. 7, ed. By P. R. Krishnaiah and C. R. Rao. New York: Elsevier.

Liu, J., S. Wu, and J. V. Zidek, 1997, "On Segmented Multivariate Regression," *Statistica Sinica* 7, 497-525.

Perron, P., 1989, "The great crash, the oil price shock and the unit root hypothesis," *Econometrica* 57, 1361-1401.

Pesaram, H. and A. Timmermann, 2000, "Model instability and the choice of observations window," Mimeo, UCSD and University of Cambridge.

Yao, Y.-C. 1988. "Estimating the number of change-points via Schwarz' criterion." *Statistics and Probability Letters* 6, 181-189.

Zacks, S., 1983, "Survey of Classical and Bayesian Approaches to the Change-Point Problem: Fixed and Sequential Procedures of Testing and Estimation," in *Recent Advances in Statistics*, ed. By M. H. Rivzi, J. S. Rustagi, and D. Sigmund. New York: Academic Press, 245-269.

APPENDIX

Algorithms for the test statistics

A. Size tests

0. Select the model parameters. There are many parameters to set; the most important ones are: the nature of the generated process (Wiener or AR(1)), the nature of the mimicking process (Wiener, Wiener with the robust option or AR(1)), the maximum number of breaks, the significance level.

1. Simulate N_1 initial processes (further, *generated processes*) with no breaks.

2. Use the global minimizer procedure to determine the location of the potential breaks, the size of the potential breaks. Obtain the statistics for the generated processes.

3. Estimate each of the generated processes according to the nature of the mimicking process, under the null of no breaks for the whole series, extract the residuals.

4. Use the bootstrap to simulate N_2 mimicking processes for each of the generated processes (use the residuals and the estimated parameters from step 3).

5. Obtain the statistics for the mimicking processes, sort the statistics, and find the critical values, in accordance with the significance level.

6. Determine if the statistic of the generated process is significant.

7. If significant, count this case as at least one break is found; if not significant, count this case as no breaks are found.

8. The size of the test is the proportion of the times the procedure identifies at least one break.

B. Power tests for the sequential UDmax statistic

0. Select the model parameters. There are many parameters to set; the most important ones are: the nature of the generated process (Wiener or AR(1)), the nature of the mimicking process (Wiener with the robust option or AR(1)), the maximum number of breaks, the significance level, the structure of the breaks to be embedded.

1. Simulate N_1 initial processes (further, *generated processes*) with the break structure defined in step 0.

2. Use the global minimizer procedure to determine the location of the potential breaks, the size of the potential breaks. Obtain the statistic for the generated processes.

3. Estimate each of the generated processes according to the nature of the mimicking process, under the null of no breaks for the whole series, extract the residuals.

4. Use the bootstrap to simulate N_2 mimicking processes for each of the generated processes (use the residuals and the estimated parameters from step 3).

5. Obtain the statistics for the mimicking processes, sort the statistics, and find the critical values, in accordance with the significance level.

6. Determine if the statistic of the generated process is significant.

7. If significant, continue; if not significant, stop, and count this case as no breaks are found.

8. The result in step 7 says there is at least one break. Using the location of the 1st break from step 2, estimate the generated process—on each of the two segments—according to the nature of the mimicking process, under the null of no breaks at each of the two segments. Extract segment-specific residuals.

9. Repeat steps 4-6 on each of the two segments.

10. If one rejects the null on at least one of them, assume there are two breaks (and three segments).

11. Continue the hypothesis-testing process on the segments, until one either reaches the maximum number of breaks or fails to reject the null on every segment.

12. For the generated processes, count how many times, one finds 1 break, 2 breaks, ..., N breaks.

13. The power of the test is the proportion of the times the procedure identifies the number of breaks correctly.

C. Power tests, the global UDmax

0. Select the model parameters. There are many parameters to set; the most important ones are: the nature of the generated process (Wiener or AR(1)), the nature of the mimicking process (Wiener with the robust option or AR(1)), the maximum number of breaks, the significance level, the structure of the breaks to be embedded.

1. Simulate N_1 initial processes (further, *generated processes*) with the break structure defined in step 0.

2. Use the global minimizers procedure to determine the location of the potential breaks, the size of the potential breaks. Obtain the statistics for the generated processes.

3. Estimate each of the generated processes according to the nature of the mimicking process, under the null of no breaks for the whole series, extract the residuals.

4. Use the bootstrap to simulate N_2 mimicking processes for each of the generated processes (use the residuals and the estimated parameters from step 3).

5. Obtain the statistics for the mimicking processes, sort the statistics, and find the critical values, in accordance with the significance level.

6. Determine if the statistic of the generated process is significant.

7. If significant, continue; if not significant, stop, and count this case as no breaks are found.

8. The result in step 7 says there is at least one break. Using the location and size of the 1st break from step 2, construct N_2 new mimicking processes, based the segment-specific bootstrapped residuals and the embedded break.

9. Repeat steps 5-6 on each for the new mimicking series. (Note the statistic for the generated series is still based on the whole sample).

10. If one rejects the null of no breaks using the new critical values, assume there are two breaks.

11. Embed the second break in yet other N_2 new mimicking processes. Continue the hypothesis testing process based on the series with the embedded breaks, until one either reaches the maximum number of breaks or fails to reject the null.

12. For the generated processes, count how many times, one finds 1 break, 2 breaks, ..., N breaks.

13. The power of the test is the proportion of the times the procedure identifies the number of breaks correctly.

Table 1: Size Tests—Weiner DGP With No Breaks

	Asymptotic[1/]	Sample Specific
SupF(0,1)	0.05	0.08
SupF(0,2)	0.05	0.09
SupF(0,3)	0.04	0.10
UDmax(3)	0.08	0.08

1/ Bai and Perron (1998)

Notes:

True size is 10 percent. 500 Monte Carlo (MC) simulations were performed. Sample specific critical values were generated for the DGP process (50 observations), with 500 MC simulation runs within each of the 500 MC runs performed.

Table 2: Size Tests with Autocorrelated DGP with No Breaks and Parametric Estimation

Panel A: True model: $y_t = 0.9*y_{t-1} + u_t$, $u_t \sim N(0,1)$ and i.i.d. Sample specific critical values are estimated using an AR(1) process with standard errors that are not robust to serial correlation, and no heretoskedasticity correction.

	Asymptotic[1/]	Sample Specific
SupF(0,1)	0.26	0.15
SupF(0,2)	0.36	0.14
SupF(0,3)	0.41	0.15
UDmax(3)	0.44	0.14

Panel B: True model: $y_t = 0.5*y_{t-1} + u_t$, $u_t \sim N(0,1)$ and i.i.d. Sample specific critical values are estimated using an AR(1) process with standard errors that are not robust to serial correlation, and no heretoskedasticity correction.

	Asymptotic[1/]	Sample Specific
SupF(0,1)	0.08	0.10
SupF(0,2)	0.10	0.11
SupF(0,3)	0.09	0.09
UDmax(3)	0.15	0.10

1/ Bai and Perron (1998)

Notes:

True size is 10 percent. 500 Monte Carlo (MC) simulations were performed. Sample specific critical values were generated for the DGP process (50 observations), with 500 MC simulation runs within each of the 500 MC runs performed.

Table 3: Size Tests with Autocorrelated DGP with No Breaks and Standard Errors Robust to Serial Correlation

Panel A: True model: $y_t = u_t$, $u_t = 0.9*u_{t-1} + e_t$, $e_t \sim$ N(0,1) and i.i.d. Sample specific critical values are estimated using a Wiener process with standard errors robust to serial correlation and no heretoskedasticity correction.

	Asymptotic[1/]	Sample Specific
SupF(0,1)	0.56	0.59
SupF(0,2)	0.81	0.81
SupF(0,3)	0.89	0.89
UDmax(3)	0.87	0.85

Panel B: True model: $y_t = u_t$, $u_t = 0.5*u_{t-1} + e_t$, $e_t \sim$ N(0,1) and i.i.d. Sample specific critical values are estimated using a Wiener process with standard errors robust to serial correlation and no heretoskedasticity

	Asymptotic[1/]	Sample Specific
SupF(0,1)	0.20	0.26
SupF(0,2)	0.32	0.34
SupF(0,3)	0.34	0.34
UDmax(3)	0.41	0.34

1/ Bai and Perron (1998)

Notes:

True size is 10 percent. 500 Monte Carlo (MC) simulations were performed. Sample specific critical values were generated for the DGP process (50 observations), with 500 MC simulation runs within each of the 500 MC runs performed.

Table 4: Size Tests with DGP with No Breaks and Over- and Under-Specification of Degree of Autocorrelation

Case of overspecification

Panel A: True model: $yt = u_t$, $u_t \sim N(0,1)$ and i.i.d. Sample specific critical values are estimated using an AR(1) process with no serial correlation, and no heretoskedasticity correction.

	Asymptotic[1/]	Sample Specific
SupF(0,1)	0.06	0.09
SupF(0,2)	0.03	0.09
SupF(0,3)	0.04	0.09
UDmax(3)	0.08	0.08

Case of underspecification

Panel B: True model: $yt = 0.9*y_{t-1} + u_t$, $u_t \sim N(0,1)$ and i.i.d. Sample specific critical values are estimated using a Wiener process with no serial correlation, and no heretoskedasticity correction.

	Asymptotic[1/]	Sample Specific
SupF(0,1)	0.90	0.94
SupF(0,2)	0.96	0.97
SupF(0,3)	0.96	0.98
UDmax(3)	0.97	0.97

Panel C: True model: $yt = 0.5*y_{t-1} + u_t$, $u_t \sim N(0,1)$ and

	Asymptotic[1/]	Sample Specific
SupF(0,1)	0.38	0.52
SupF(0,2)	0.55	0.66
SupF(0,3)	0.54	0.68
UDmax(3)	0.61	0.60

1/ Bai and Perron (1998)

Notes:

True size is 10 percent. 500 Monte Carlo (MC) simulations were performed. Sample specific critical values were generated for the DGP process (50 observations), with 500 MC simulation runs within each of the 500 MC runs performed.

Table 5: Power Tests

Panel A: True model: $y_t = u_t$, $u_t \sim N(0,1)$ and i.i.d.

Averages (in s.d.)		Mimicking Process: Wiener with heteroskedasticity and serial correlation corrections				Mimicking Process: AR(1), no heteroskedasticity correction			
		BP asypt.[1/]	S supF[2/]	SUDmax[3/]	AUDmax[4/]	BP asypt.[1/]	S supF[2/]	SUDmax[3/]	AUDmax[4/]
1 0 1	Prob k=0	0.68	0.11	0.03	0.03	0.77	0.16	0.05	0.05
	Prob k=1	0.12	0.10	0.10	0.34	0.14	0.12	0.16	0.41
	Prob k=2	0.18	0.63	0.70	0.55	0.09	0.57	0.62	0.50
	Prob k≥3	0.01	0.16	0.16	0.08	0.00	0.15	0.17	0.04
	Loss value[5/]	1.51	0.49	0.33	0.48	1.68	0.58	0.43	0.55
2 0 2	Prob k=0	0.73	0.05	0.00	0.00	0.55	0.00	0.00	0.00
	Prob k=1	0.00	0.00	0.00	0.03	0.00	0.02	0.03	0.05
	Prob k=2	0.24	0.78	0.82	0.83	0.43	0.83	0.81	0.87
	Prob k≥3	0.03	0.17	0.18	0.14	0.02	0.15	0.16	0.08
	Loss value[5/]	1.48	0.28	0.18	0.17	1.12	0.17	0.19	0.13
2 1 0	Prob k=0	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	Prob k=1	0.63	0.29	0.32	0.71	0.81	0.43	0.45	0.73
	Prob k=2	0.33	0.56	0.53	0.23	0.18	0.49	0.45	0.25
	Prob k≥3	0.04	0.15	0.16	0.06	0.01	0.09	0.10	0.02
	Loss value[5/]	0.67	0.44	0.47	0.77	0.83	0.51	0.55	0.75
4 2 0	Prob k=0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Prob k=1	0.01	0.00	0.00	0.19	0.06	0.01	0.02	0.25
	Prob k=2	0.85	0.79	0.79	0.68	0.89	0.91	0.93	0.67
	Prob k≥3	0.14	0.21	0.21	0.13	0.05	0.08	0.05	0.08
	Loss value[5/]	0.15	0.21	0.21	0.32	0.11	0.09	0.07	0.33
2 0 1	Prob k=0	0.05	0.00	0.00	0.00	0.12	0.00	0.00	0.00
	Prob k=1	0.28	0.10	0.11	0.55	0.46	0.19	0.22	0.55
	Prob k=2	0.61	0.74	0.73	0.38	0.40	0.66	0.63	0.42
	Prob k≥3	0.06	0.16	0.15	0.07	0.02	0.15	0.15	0.04
	Loss value[5/]	0.43	0.26	0.27	0.62	0.71	0.34	0.37	0.58

1/ Sequential testing using asymptotic critical values (Bai and Perron (1998)).
2/ Sequential sup F testing using sample specific critical values.
3/ Sequential UDmax testing using sample specific critical values.
4/ The adjusted global UDmax methodology.
5/ Using the loss function defined for power tests (page 14).

Notes:
True size is 10 percent. 500 Monte Carlo (MC) simulations were performed. Sample specific critical values were generated for the DGP process (50 observations), with 500 MC simulation runs within each of the 500 MC runs performed. Both true underlying process and the simulated model have two breaks. First one of them is located at the observation number 17 and the second one at the observation number 34. Within each break interval, time series dynamics are governed by a Wiener process and the averages are a multiple of the unitary standard deviation as indicated in the first column of the table.

Table 5: Power Tests (cont.)

Panel B: True model: $y_t = u_t$, $u_t = 0.5*u_{t-1} + e_t$, $e_t \sim N(0,1)$ and i.i.d.

Averages (in s.d.)		Mimicking Process: Wiener with heteroskedasticity and serial correlation corrections				Mimicking Process: AR(1), no heteroskedasticity correction			
		BP asypt.[1/]	S supF[2/]	SUDmax[3/]	AUDmax[4/]	BP asypt.[1/]	S supF[2/]	SUDmax[3/]	AUDmax[4/]
1 0 1	Prob k=0	0.70	0.26	0.06	0.06	0.82	0.31	0.18	0.18
	Prob k=1	0.23	0.19	0.15	0.58	0.14	0.34	0.41	0.49
	Prob k=2	0.07	0.33	0.41	0.32	0.04	0.28	0.32	0.28
	Prob k≥3	0.01	0.22	0.38	0.04	0.00	0.07	0.09	0.04
	Loss value[5/]	1.63	0.94	0.66	0.75	1.78	1.03	0.86	0.91
2 0 2	Prob k=0	0.74	0.26	0.00	0.00	0.74	0.16	0.08	0.08
	Prob k=1	0.10	0.06	0.04	0.42	0.16	0.24	0.34	0.33
	Prob k=2	0.14	0.39	0.49	0.54	0.09	0.48	0.46	0.49
	Prob k≥3	0.03	0.29	0.47	0.04	0.01	0.12	0.12	0.10
	Loss value[5/]	1.60	0.87	0.51	0.46	1.65	0.68	0.62	0.59
2 1 0	Prob k=0	0.09	0.02	0.00	0.00	0.30	0.03	0.03	0.03
	Prob k=1	0.68	0.41	0.30	0.90	0.62	0.68	0.70	0.79
	Prob k=2	0.21	0.35	0.35	0.08	0.08	0.23	0.21	0.14
	Prob k≥3	0.02	0.22	0.34	0.01	0.00	0.06	0.06	0.04
	Loss value[5/]	0.88	0.67	0.65	0.92	1.22	0.79	0.82	0.89
4 2 0	Prob k=0	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	Prob k=1	0.36	0.14	0.10	0.78	0.71	0.57	0.62	0.61
	Prob k=2	0.53	0.52	0.50	0.21	0.26	0.38	0.32	0.29
	Prob k≥3	0.11	0.34	0.39	0.01	0.02	0.05	0.06	0.10
	Loss value[5/]	0.47	0.48	0.50	0.79	0.75	0.62	0.68	0.71
2 0 1	Prob k=0	0.49	0.16	0.01	0.01	0.65	0.16	0.06	0.06
	Prob k=1	0.32	0.23	0.16	0.69	0.29	0.41	0.56	0.55
	Prob k=2	0.15	0.35	0.43	0.26	0.06	0.37	0.32	0.33
	Prob k≥3	0.04	0.26	0.40	0.04	0.00	0.06	0.05	0.06
	Loss value[5/]	1.34	0.81	0.58	0.75	1.58	0.79	0.74	0.73

1/ Sequential testing using asymptotic critical values (Bai and Perron (1998)).
2/ Sequential sup F testing using sample specific critical values.
3/ Sequential UDmax testing using sample specific critical values.
4/ The adjusted global UDmax methodology.
5/ Using the loss function defined for power tests (page 14).

Notes:
True size is 10 percent. 500 Monte Carlo (MC) simulations were performed. Sample specific critical values were generated for the DGP process (50 observations), with 500 MC simulation runs within each of the 500 MC runs performed. Both true underlying process and the simulated model have two breaks. First one of them is located at the observation number 17 and the second one at the observation number 34. Within each break interval, time series dynamics are governed by a Wiener process and the averages are a multiple of the unitary standard deviation as indicated in the first column of the table.

Table 5: Power Tests (cont.)

Panel C: True model: $y_t = u_t$, $u_t = 0.9*u_{t-1} + e_t$, $e_t \sim N(0,1)$ and i.i.d.

Averages (in s.d.)		Mimicking Process: Wiener with heteroskedasticity and serial correlation corrections				Mimicking Process: AR(1), no heteroskedasticity correction			
		BP asypt.[1/]	S supF[2/]	SUDmax[3/]	AUDmax[4/]	BP asypt.[1/]	S supF[2/]	SUDmax[3/]	AUDmax[4/]
1 0 1	Prob k=0	0.39	0.11	0.01	0.01	0.59	0.20	0.08	0.08
	Prob k=1	0.31	0.17	0.07	0.62	0.32	0.49	0.67	0.47
	Prob k=2	0.21	0.28	0.22	0.30	0.08	0.22	0.21	0.22
	Prob k≥3	0.09	0.44	0.70	0.07	0.00	0.08	0.04	0.23
	Loss value[5/]	1.18	0.83	0.79	0.70	1.51	0.98	0.86	0.85
2 0 2	Prob k=0	0.37	0.12	0.00	0.00	0.64	0.22	0.10	0.10
	Prob k=1	0.29	0.17	0.06	0.63	0.28	0.45	0.64	0.48
	Prob k=2	0.23	0.25	0.23	0.29	0.07	0.28	0.20	0.24
	Prob k=3	0.11	0.46	0.71	0.08	0.00	0.06	0.05	0.17
	Loss value[5/]	1.15	0.87	0.78	0.71	1.57	0.94	0.90	0.86
2 1 0	Prob k=0	0.38	0.13	0.01	0.01	0.59	0.14	0.08	0.08
	Prob k=1	0.34	0.16	0.06	0.62	0.34	0.53	0.75	0.52
	Prob k=2	0.23	0.26	0.24	0.29	0.07	0.24	0.15	0.23
	Prob k≥3	0.05	0.44	0.69	0.09	0.00	0.09	0.02	0.17
	Loss value[5/]	1.14	0.87	0.76	0.72	1.52	0.90	0.93	0.85
4 2 0	Prob k=0	0.36	0.09	0.00	0.00	0.47	0.13	0.07	0.07
	Prob k=1	0.33	0.17	0.05	0.59	0.42	0.54	0.67	0.47
	Prob k=2	0.24	0.27	0.21	0.31	0.10	0.25	0.19	0.19
	Prob k≥3	0.07	0.47	0.73	0.09	0.01	0.09	0.07	0.27
	Loss value[5/]	1.12	0.82	0.79	0.69	1.37	0.88	0.88	0.88
2 0 1	Prob k=0	0.38	0.13	0.01	0.01	0.64	0.23	0.10	0.10
	Prob k=1	0.31	0.18	0.06	0.67	0.30	0.42	0.64	0.44
	Prob k=2	0.21	0.25	0.21	0.24	0.05	0.27	0.19	0.23
	Prob k≥3	0.09	0.45	0.72	0.08	0.01	0.09	0.08	0.23
	Loss value[5/]	1.17	0.88	0.80	0.76	1.59	0.96	0.91	0.86

1/ Sequential testing using asymptotic critical values (Bai and Perron (1998)).
2/ Sequential sup F testing using sample specific critical values.
3/ Sequential UDmax testing using sample specific critical values.
4/ The adjusted global UDmax methodology.
5/ Using the loss function defined for power tests (page 14).

Notes:
True size is 10 percent. 500 Monte Carlo (MC) simulations were performed. Sample specific critical values were generated for the DGP process (50 observations), with 500 MC simulation runs within each of the 500 MC runs performed. Both true underlying process and the simulated model have two breaks. First one of them is located at the observation number 17 and the second one at the observation number 34. Within each break interval, time series dynamics are governed by a Wiener process and the averages are a multiple of the unitary standard deviation as indicated in the first column of the table.