Stopping Tests in the Sequential Estimation for Multiple Structural Breaks *

Christian de Peretti †
GREQAM
Université de le Méditerranée

Giovanni Urga ‡
Faculty of Finance
Cass Business School

March 22, 2004

Abstract

In this paper, we propose the use of bootstrapping methods to obtain correct critical values for dating breaks. Following the procedure proposed in Banerjee, Lazarova and Urga (1998), we consider the case of estimating a system with two or more marginal processes and a conditional process. First, the location of the breaks in marginal models is estimated. Next, the marginal models are imposed on the conditional model to form a reduced form system. The conditional model with its own breaks is then estimated. The estimation of the break dates is sequential. Break dates are estimated via two alternative procedures: including estimated break dates one by one or splitting the sample. Inclusion of additional breaks or splitting samples are repeated until a criterion for stopping is satisfied. In this paper we propose bootstrap tests as criterion for stopping sequential search. This procedure allows to improve the estimators to avoid excessive bias and prove to be stable in the case of both stationary and non-stationary series. Finally, we illustrate the methods by modelling the money demand in United Kingdom.

Keywords: Structural Breaks, Sequential Testing, Bootstrap.
JEL Classification: C10, C12, C13, C15.

*We wish to thank participants to the International Conference on: "Modelling Structural Breaks, Long Memory and Stock Market Volatility" (Cass Business School, London, 6-7, December 2002) for discussions and comments. C. de Peretti thanks Sandrine Lardic and Valérie Mignon for the invitation to present the paper at the *2e Journée d'économétrie:* "Développement récent de l'économétrie appliquée à la finance" (Paris X – Nanterre, January 16^{th} 2003) and Catherine Bruneau for her very useful comments at the conference. He also thanks the organiser of the *LXXXIth International Conference of the AEA*: "Exchange Rate" (GREQAM, Marseille, 2003) and Mohamed Boutahar for his invitation to present the paper at the GREQAM seminar, Marseille. Stepana Lazarova provided very useful comments and suggestions which helped to improve the paper. However, the usual disclaimer applies. G. Urga wishes to thank ESRC under grant number R000238145 for funding this research.

†GREQAM, Centre de la Vieille Charité, rue de la Charité, 13002 Marseille, France. Tel.: +33 (0)4 91 140 727. Fax: +33 (0)4 91 900 227. E-mail: peretti@ehess.cnrs-mrs.fr, http://durandal.cnrs-mrs.fr/greqam/

‡Cass Business School, Faculty of Finance, 106 Bunhill Row, London EC1Y 8TZ (U.K.). Tel: +/44/20/70408698. Fax: +/44/20/70408881. E-mail: g.urga@city.ac.uk. http://www.staff.city.ac.uk/ ~giourga.

1 Introduction

In recent work, methods for dating break using the supremum and other functional forms of processes based on standard statistics have been established (Andrews (1993; 2003); Andrews and Ploberger (1994); Hansen (n.d.)). The works published recently by Bai (1997; 1999), Bai and Perron (1998) and Culver and Papell (1997) among others have extended Perron's (1989) analysis not only to the case where the break date is unknown but to a scenario where the series may be broken, both in trend and in mean, more than once. One of the main difficulties posed by this literature is the calculation of the critical values of a significance test for the break for a particular data generation process (DGP), especially for a system of equations. The paper by MacKinnon (1994), among others, has emphasised the usefulness of response surfaces that may be used to recompute critical values under changes to the DGP such as sample size, signal-noise ratio, unconditional mean etc. Here, we propose the use of bootstrapping methods as another way of overcoming the difficulty of obtaining the useful critical values [1].

Our work originated and expands a companion work by Banerjee, Lazarova and Urga (1998) where the authors estimated a system of equations consisting of a conditional process (the variable that we want to explain) and two marginal processes (explanatory variables). They propose the following general estimating procedure to detect breaks. The breaks are defined as exogenous changes in the mean and in the trend of the processes. They can be modelled with dummy variables. First, the locations of the breaks in marginal models are estimated. Next, the marginal models are integrated into the conditional model to create a reduced form of the system. Following the estimation of the marginal models, the conditional model with its own breaks is estimated. The estimation of the break dates is sequential. Break dates are estimated by two alternative procedures: one method involves including estimated break dates one by one by including dummies, estimating each of them over the full sample, while the other requires splitting the sample before each subsequent estimation. Inclusion of breaks or splitting the sample are repeated until a criterion for stopping is satisfied. The final stage of the procedure is to impose the break dates from the marginal models into the conditional model and repeat the sequential research for the conditional model. What emerges at the end are congruent marginal and conditional models with the breaks in all the series and the relations of interest properly identified having easily interpretable coefficients. However, the main unresolved issue in that paper was that no criterion for stopping the sequential search for breaks was developed, so the searching was run until the exhaustion of the degrees of freedom.

Here we fill this gap. We propose bootstrap stopping (parametric and nonparametric) tests for the cases of stationary and non stationary series (when the series are integrated of order one (I(1)) and when the regressors are nonstationary in the sense of I(1) and/or having structural breaks that cause difficulties in calculating the statistic distributions that are not pivotal). Before presenting the test procedures, we describe two break point estimators that we use, given that the test statistics are based on these estimators and also to improve the performance of the estimators to avoid excessive bias. One is asymptotic,

[1] A valuable account of the usefulness of bootstrap methods in time series econometrics has appeared in a recent issue of Econometric Reviews (1996) and provides useful background for the discussion in our paper. There is an extensive literature on bootstrapping both in statistics and econometrics. In addition to the journal cited above, see, for example, Efron and Tibshirani (1993) and references therein, Freedman and Peters (1984a; 1984b) and Basawa, Mallik, McCormick, Reeves and Taylor (1991) and Horowitz (2001).

and the other bootstrap. For each date, compute the F-statistic ,F_t, for testing the null hypothesis of no break. The asymptotic break point estimator is defined as the date at which the F_t-statistics reach their maximum. The bootstrap break point estimator is defined as the date where the bootstrap P-values reach their minimum. Two types of bootstrap are presented, one parametric and another non-parametric. However, in some cases, the estimations of the first dates are not consistent. The inconsistency is due to the model misspecification because of the lack of knowledge of the other(s) break date(s) during the first estimations. To solve this problem, a possible solution is a simultaneous estimation of two or more breaks, combined with bootstrap methods. Unfortunately this procedure is too computational intensive. In order to overcome this problem and to reduce the estimation bias, we may re-estimate the first breaks after the estimation of the next ones. In this way, the re-estimation of the first break will be done conditionally on the knowledge of the next breaks. These second estimations will have considerably less bias than the first ones without taking account other breaks. One the other hand, we also propose a modified bootstrap estimator which will identify an optimal number of bootstrap replications to reduce the computation time.

We then propose tests for stopping the sequential search. We want to stop when the last estimated break is not significant, that is when the null of $i-1$ breaks against the alternative of i breaks is not rejected. The conventional test of this hypothesis is to compare the computed F-statistic against the 0.05 critical value of the relevant asymptotic (or even bootstrap) F-distribution. However, this is not appropriate in our case as the time of break is not determined exogenously. To obtain a critical value that takes into account that the breaks are chosen endogenously on the basis of the maximum F-value or minimum bootstrap P-value, there are two approaches: the Ploberger's (asymptotic) approach and the bootstrap approach. The bootstrap errors terms are generated in the same way than for the estimate, parametrically or non-parametrically. This time, however, we perform the whole estimating procedure for building a test statistic for each bootstrap sample, such that the bootstrap P-value can be computed. We propose two test statistics for discriminating the hypotheses, they are based on the two estimators cited above. The combination of the bootstrap approach for the test with the bootstrap estimation of the break point can be called a *double bootstrap*.

We provide Monte Carlo experiments that study the finite sample performances of the statistic tests and of the estimators in the context of stationary and nonstationary series, I(1) and with structural breaks in the regressors. We provide the bias and the standard deviations of the estimators, and the P-value plots and the power-size curves of the tests.

Finally, we apply our methods to the case of money demand in United Kingdom explained by interest rate and income.

The remaining of the paper is organised as follows. Section 2 introduces the procedures for dating breaks, while in Section 3 we describe the break point estimators, both the asymptotic and the parametric and nonparametric bootstrap versions (3.1), the modified bootstrap estimator (3.2) and a procedure to improve the estimators to avoid excessive bias (3.3). In Section 4 we propose bootstrap tests as criterion for stopping sequential search and in Section 5 we report the results of the evaluation of the performance of the estimators and the tests using Monte Carlo experiments tuned to account for the characteristics of the real series of money demand, income and interest rate for UK used in Section 6, where we report an illustration of the methods proposed by modelling the money demand function for UK. Section 7 concludes.

2 The main methods: the sequential estimation procedures

In this section, we report the two procedures for dating breaks as proposed in Banerjee, Lazarova and Urga (1998). For an exposition of the proposed methods, we use a very simple system. Our justification for doing so is twofold. It keeps the analysis simple and at the same time provides easily interpretable results.

Consider estimating the following system:

$$
\begin{aligned}
y_t &= \mu_0 + \delta_0 t + \sum_{i=1}^{L_0} \rho_{0,i} y_{t-i} + \sum_{i=1}^{I_0} (\alpha_{0,i} M_{0,i,t} + \beta_{0,i} D_{0,i,t}) + \\
&\quad + \sum_{n=1}^{N} (a_n x_{n,t-1} + \sum_{i=1}^{I_n} (\eta_{n,i} M_{n,i,t} + \nu_{n,i} D_{n,i,t})) + u_t, \qquad (1a)
\end{aligned}
$$

$$
x_{n,t} = \mu_n + \delta_n t + \sum_{i=1}^{L_n} \rho_{n,i} x_{n,t-i} + \sum_{i=1}^{I_n} (\alpha_{n,i} M_{n,i,t} + \beta_{n,i} D_{n,i,t}) + e_{n,t}, \qquad (1b)
$$

for $t = 1, 2, \ldots, T$ and $n = 1, 2, \ldots, N$, where y_t is a variable denoting the conditional process, here taken to be the real money demand, $x_{n,t}$ is a variable denoting either of the marginal processes, here taken to be income and interest rates. u_t and e_{it} represent mutually uncorrelated white noise processes. The dummy variables, designed to capture breaks in mean and linear trend, are defined as follows:

$$
\begin{aligned}
M_{n,i,t} &= \mathbf{I}(t \geq b_{n,i}), & n = 1, 2, \ldots, N, \quad i = 1, 2, \ldots, I_n, \\
D_{n,i,t} &= (t - b_{n,i} + 1)\mathbf{I}(t \geq b_{n,i}), & n = 1, 2, \ldots, N, \quad i = 1, 2, \ldots, I_n,
\end{aligned}
$$

where $b_{n,i}$ stands for the date of break.

The structure and notations used in the system reflect the working of the procedure. First, the locations of the breaks in marginal models are estimated. Next, the marginal models are imposed on conditional model to create a reduced form of the system. *i.e.* the marginal processes are used as explanatory variables in the conditional regression as well as the breaks found in the marginal processes in addition to the own constant and trend terms and the own lag variables of the conditional process. That means that the break dummies of the marginal models can appear with different coefficients in the conditional model. Following estimation of the marginal models, the conditional model with its own breaks is estimated.

Break dates are estimated by two alternative procedures. One method involves including break dates one by one, estimating each of them in full sample, while the other require splitting the sample before each subsequent estimation. These two different but related methods are now described.

2.1 Including the breaks one by one

Under this method, a break date in a given process is estimated first. Further, a dummy for the break is included in the specification of the model and another break date is estimated, always using the whole sample. Inclusion of additional breaks is repeated until a criterion for stopping is satisfied, as we will describe in section 4.

For example, consider the following marginal model:

$$x_t = \mu + \delta t + \rho x_{t-1} + \sum_{i=1}^{I} (\alpha_i M_{i,t} + \beta_i D_{i,t}) + e_t, \tag{2}$$

for $t = 1, 2, \ldots, T$, where

$$\begin{aligned} M_{i,t} &= \mathbf{I}(t \geq b_i), & i = 1, 2, \ldots, I, \\ D_{i,t} &= (t - b_i + 1)\mathbf{I}(t \geq b_i), & i = 1, 2, \ldots, I, \end{aligned}$$

We employ the following sequential procedure. In each step i, a break date is estimated, namely $\widehat{b}_i$. The first step is carried out without any break in the model. In the next step, $i + 1$, two dummies:

- $\widehat{M}_{i,t} = I(t \geq \widehat{b}_i)$ and
- $\widehat{D}_{i,t} = (t - b_i + 1)I(t \geq \widehat{b}_i)$,

are included into the regression and the estimation procedure is repeated to obtain another break [2]. The procedure continues in this manner until a criterion for stopping is satisfied.

2.2 Splitting the sample

Under the method of splitting sample, the first step is carried out precisely as in the method of adding breaks one by one. Our next step is to repeat the estimation procedure in each of the two subsamples created by splitting the sample at the estimated break point and to continue this procedure until a criterion for stopping is satisfied [3].

2.3 The conditional process

The final stage of the procedure is to impose the break dates from the marginal models on the conditional model and repeat the sequential research for the conditional model. What emerges at the end are congruent marginal and conditional models with the breaks (potentially) in all the series and relations of interest properly identified.

3 The break point estimators and our modifications

In this section, we describe the break point estimators. This will allow us to introduce our modifications dealing with some problems encountered in practice and it will be useful to understand the test statistics that will be introduced in Section 4.

[2]It is worth noticing that we cannot test the presence of a break date near a previous break date because the regressor matrix X has not numerically full rank, so we must exclude any dates too close to the previous break dates depending the numerical invertibility of the $X'X$ matrix. Indeed, there is no loss of effectiveness, since in practice, we cannot distinguish two break dates that are too close. Moreover, in real data, a break does not necessarily come instantaneously at a date b, but it can come over time a little more progressively or with a little pre-break. So, there would be reason to believe that any breaks found within such a short neighbourhood might in fact be reflective of the same break and repeated procedure would prove to be misleading in its finesse. Thus, this including procedure can exclude too many dates that are in fact the realisation of only one break date.

[3] The same remark than in the previous footnote concerning the fact that the regressor matrix X has not numerically full rank holds, since near the border of the dates the $X'X$ matrix is not numerically invertible but this should be solved by trimming.

3.1 The break point estimators

We estimate the regression for each admissible date of break t, namely for $t = 3, \dots, T-1$ except for the dates of the breaks already included in the regression, (in fact, for a small neighbourhood around the dates), using the first observation x_1 as an initial condition. For each t, compute the F-statistic, F_t, for testing the null hypothesis of no break, $\alpha_i = \beta_i = 0 | b_i = t$.

3.1.1 Asymptotic estimate

A break point estimator $\widehat{b_i}$ is defined as the time t at which the F_t-statistic reaches its maximum.

$$\widehat{b_i} = \text{argmax}\{F_t\} = \text{argmin}\{1 - \hat{F}_{as}(F_t)\} \qquad \text{for all admissible } t \ , \tag{3}$$

where $\hat{F}_{as}$ is the relevant asymptotic estimation of the F-distribution under the null of no more break.

3.1.2 Parametric and non parametric bootstrap estimates

We bootstrap the set of F_t-values adopting the following approach. The regression (2) under the null of $i-1$ breaks imposed is run,

$$x_t = \mu + \delta t + \rho x_{t-1} + \sum_{j=1}^{i-1} (\alpha_j M_{j,t} + \beta_j D_{j,t}) + e_t^0 \tag{4}$$

The basic idea of this procedure is that the modified residuals from the null regression are resampled with replacement. Then the estimated parameters together with bootstrap error terms are used to create a recursive bootstrap sample x_t^* following the equation

$$\begin{aligned} x_t^* &= \hat{\mu} + \hat{\delta} t + \hat{\rho} x_{t-1}^* + \sum_{j=1}^{i-1} (\hat{\alpha}_j M_{j,t} + \hat{\beta}_j D_{j,t}) + e_t^*, \qquad t \geq 2, & (5) \\ x_1^* &= x_1, & (6) \end{aligned}$$

where e_t^* denotes the bootstrap error terms and where parameters with a hat have been estimated in the regression (4). The resampling is repeated B_{est} times, so that B_{est} bootstrap samples are obtained. For each of the bootstrap samples the regression (2) with the break date fixed at t for all admissible t is estimated and the set of F_t-statistics of the hypothesis $\alpha_i = \beta_i = 0$ is computed. The bootstrap F-distributions (obtained in the manner of empirical distribution) for each time of break t are used to get the set of P values for each of the F_t-statistics (Christiano (1992)). A break point estimator $\widehat{b_i}$ is defined as the t at which the bootstrap P values attain their minimum.

$$\widehat{b_i} = \text{argmin}\{1 - \hat{F}_{bs,t}(F_t)\} \qquad \text{for all admissible } t \ , \tag{7}$$

where $\hat{F}_{bs,t}$ is the bootstrap estimation of the F-distribution for testing no more break against one more break at the date t, under the null of no more break.

For each bootstrap P value, the same set of random number is used to reduce the experimental error in the comparison of the methods.

For the *parametric bootstrap*, we draw the bootstrap error term from a $N(0, \hat{\sigma}_0^2)$, where $\hat{\sigma}_0^2$ is the estimate of the variance of the error terms under the null.

In this case of *non-parametric bootstrap* the e_t^* are generated by re-sampling from the vector with typical element $\tilde{e}_t$ constructed as follows:

1. let d_t be the t^{th} diagonal element of $P_{<X>}$, the matrix projecting onto the space spanned by the regressor matrix X,
2. divide each element of the residuals $\hat{e}$ by $\sqrt{1-d_t}$,
3. re-centre the resulting vector,
4. re-scale it so that it has variance $\hat{\sigma}_e^2$.

This type of procedure is advocated in Weber (1984).

3.2 The number of bootstrap replications

A natural question we pose at this point is what are the consequences if there are no sufficiently large number of bootstrap replications. It is well known that the main consequence is that the bootstrap P value can be non-distinguishable from 0 in some circumstances (if the P value is small, close to zero), depending on the data and particularly on the error distribution and on the autoregressive parameter.

In the context of a test, it is not very important if the number of bootstrap replications is not sufficiently large to distinguish the bootstrap P value from 0 in the case of typical tests at levels 0.05 or 0.01. So, if we obtain a numerical P value equal to zero, we could conclude that the true P value is smaller than the significance level of the test and so reject the null.

However, in the case of our estimation method, we must take the *argmin* of the set of the bootstrap P values. If the number of bootstrap if not sufficient, there is a set of dates t for which the associated P values are numerically equal to zero. Thus, the *argmin* of this set is not unique, and we cannot determine the estimate.

With real data, when a large number of bootstrap replications is required for the estimation, one can verify that the number is sufficient by regarding if there is at least a computed P value that is equal to zero. If it is the case, the researcher must do again the estimation procedure with a larger number of bootstrap until there is no zero P value. As the last step, we suggest to do again the estimation with more bootstrap to avoid random effect due to the fact that the previous bootstrap number is just sufficient to distinguish a P value from zero.

There are two other cases where taking a too large bootstrap number can be problematic - the case of Monte Carlo experiments and the case of double bootstrap test that is presented later in the paper (Subsection 4.2). In both the cases, the problem is that we do a second loop containing the bootstrap estimation loop, and it can be computationally costly. For the Monte Carlo experiments, there is a loop over the simulated samples, and for each sample, there is a loop over the bootstrap samples used to compute the bootstrap estimate. Moreover, the bootstrap number can vary depending on the sample, and we must verify that it is correct for each sample, that can take more time again. For the double bootstrap test, there is a loop over the bootstrap samples used to compute the bootstrap P value, and for each of the bootstrap sample, there is a loop over a second set of bootstrap samples used to compute the test statistic because this statistic is based on the same principle than the bootstrap estimation. As for the Monte Carlo case, we must verify that the bootstrap number is correct for each bootstrap sample used for P value.

Modified bootstrap estimator To avoid the problem due to insufficient bootstrap replications, we propose the following break date estimator: First, we take a not too small bootstrap number to make the method not too poor. We must take more bootstrap than for a test, but less than what is needed normally. Second, we compute the set

$$\widehat{B_i} = \text{argmin}\{1 - \hat{F}_{bs,t}(F_t)\}$$

for all admissible t. If the bootstrap number is sufficiently large, $\widehat{B_i} = \{\widehat{b_i}\}$, where $\widehat{b_i}$ is the previous bootstrap estimator. Otherwise, $\widehat{B_i}$ is not unique, and the estimate is not well defined. We can then define a new estimator $\widehat{\widehat{b_i}}$ from $\widehat{B_i}$. We propose here simply the average of the break dates in $\widehat{B_i}$:

$$\widehat{\widehat{b_i}} = \frac{1}{\#\widehat{B_i}} \sum_{1}^{\#\widehat{B_i}} t \in \widehat{B_i},$$

$\#(\, . \,)$ represents the number of elements in a set.

This estimator is consistent because we have

$$\widehat{B_i} \overset{T\to\infty}{\longrightarrow} \{\widehat{b_i}\} \overset{T\to\infty}{\longrightarrow} \{b_i\},$$

and thus

$$\widehat{\widehat{b_i}} \overset{T\to\infty}{\longrightarrow} \widehat{b_i} \overset{T\to\infty}{\longrightarrow} b_i.$$

Of course, $\widehat{\widehat{b_i}}$ is a less efficient estimator than $\widehat{b_i}$, and it must not be used with true data for which the exact detection of a break is very important. Nevertheless, for Monte Carlo experiments, where we are interested only by the average behaviour of the methods, the use of $\widehat{\widehat{b_i}}$ instead of $\widehat{b_i}$ can be sufficient to compare the bootstrap method to the asymptotic one or to obtain a good idea of the performances of the bootstrap. For the double bootstrap test, it is more delicate, since the less precise is the underlying estimate, the less efficient is the resulting test. But exact estimation of the break date is less crucial since we are more interested in the distribution of a test statistic to make inference. Thus, we can accept a small loss in efficiency of the test to gain in computation time.

3.3 A simple simulation exercise to explore the problem of bias

In this section we illustrate a way to improve the estimators to avoid excessive bias.

3.3.1 The bias

Let us estimate the following DGP:

$$\begin{aligned}
x_t &= \delta t + \rho x_{t-1} + \beta_1 D_{1,t} + \beta_2 D_{2,t} + e_t && (8)\\
e_t &\sim \mathrm{N}(0,1)\\
t &\in \{1, \ldots, T\}\\
D_{1,t} &= (t - T_1 + 1)\mathbf{I}(t \geq b_1)\\
D_{2,t} &= (t - T_2 + 1)\mathbf{I}(t \geq b_2)
\end{aligned}$$

where the parameter values are:

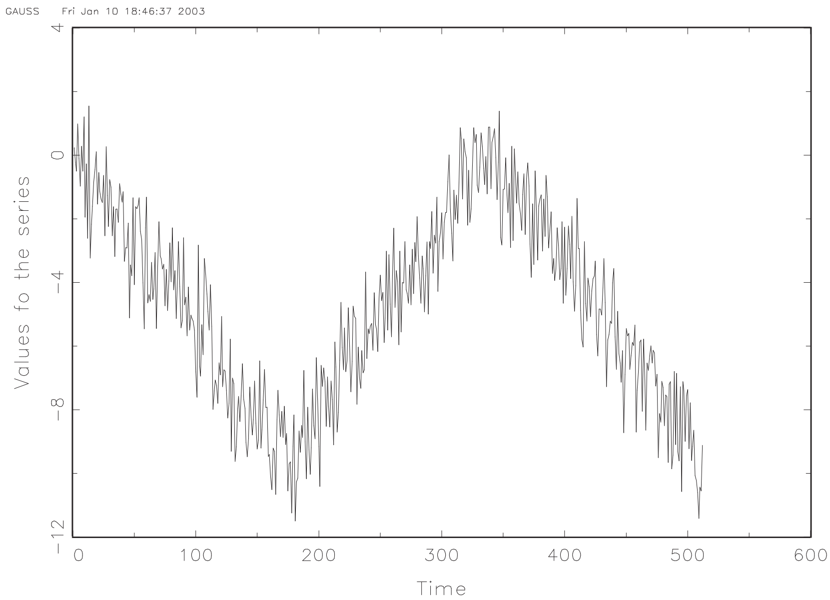
Parameters	Values
δ	-0.05
ρ	0.1
T	512

Parameters	Values
β_1	0.1
b_1	$[T/3]+1$

Parameters	Values
β_2	-0.1
b_2	$[2T/3]+1$

and where [.] is the integer part. We choose the Gaussian distribution for the errors and $\rho = 0.1$ for the autoregressive parameter to show that the problem does not come from the leptokurtic characteristic of the data or from a distortion of the statistic distributions. From a close inspection of figure 1 it is evident where the breaks are but the estimators are not able to pick them up despite a sample size of $T = 512$. Figure 2 shows the F_t-

Figure 1: The simulated time series



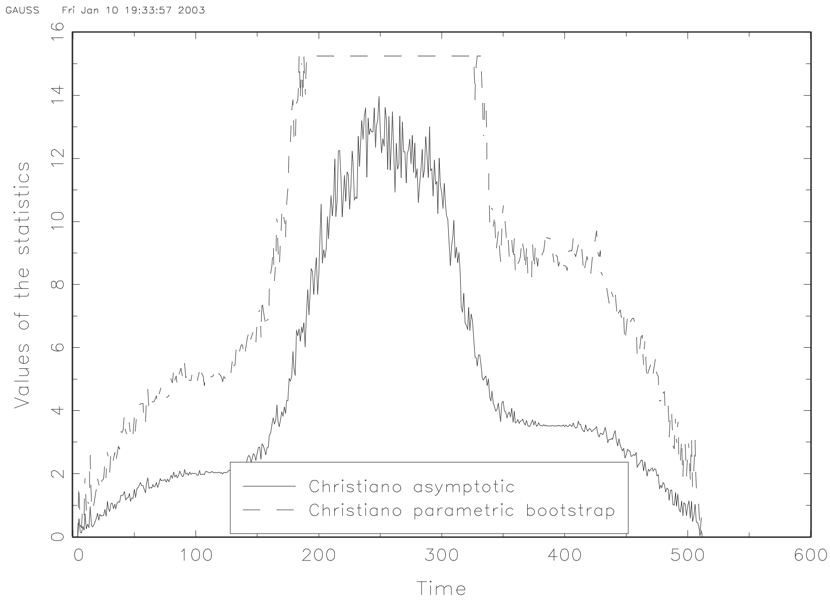
statistics of Christiano (1992) method using first the asymptotic distribution, and second the parametric bootstrap distribution with $B_{est} = 1024$. The bootstrap F_t statistics are calculated as follows

$$\hat{F}_{as}^{-1}(1 - p_{bs,t}),$$

where $p_{bs,t}$ are the bootstrap P values. The zero bootstrap P values are replaced by the *min* of the P values. The aim of this transformation is to compare both the asymptotic and the bootstrap statistics. One can see that the estimate (the *max* of the curve) is totally spurious. Figure 3 shows the four sets of F_t-statistics (for each step) of sequential search procedure using the asymptotic distribution. One can see that the first estimate (in fact, it is equal to Christiano (1992) first estimate) is totally spurious. The second estimate detects correctly one of the two breaks (and even the second). The third estimate detects very well the second break. The time series geometry leads the first estimation $\hat{b}_1$ to choose the break date such as $b_1 = [T/2] + 1$, even for large sample size ($T \leq 1024$).

Perron (1989) showed that if the estimation of only one break is processed when there are two breaks, the estimator converges to one of the two breaks. But in finite sample, this can appear differently. In our situation, none of the breaks dominate the other (they have equal amplitude and symmetric location). Thus, the first estimation must choose one of the breaks randomly (and the second estimation must choose the other break). But the presence of the second break when estimating the first one biases the estimation. Since the location of the breaks is symmetric, the first estimation is biased and the

Figure 2: The Christiano method



result is somewhat between the two breaks. We can conclude that this first estimate is inconsistent since the second and the third estimations detect correctly the first and the second breaks.

3.3.2 A solution illustrated via a simple experiment

The inconsistency is due to the model misspecification because of the lack of knowledge of the other(s) break date(s) during the first estimations. A simultaneous estimation of two or more breaks, combined with bootstrap methods, is too complex and computationally expensive. So, a basic idea is to re-estimate the previous breaks with the knowledge of the following breaks. In what follows we illustrate the procedure with a simple example with two breaks: we estimate a first break, then, a second break, as in the classical sequential search. But in the new search, we re-estimate the first break. Figure 4 shows the F_t-statistics of the new sequential search using the asymptotic distribution. We only re-estimate the first break once. If there are only two breaks, we can adopt a *two step method*, and re-estimate the second break, and then the first until a precision criterion stop the procedure. If there are three breaks, we can do the following:

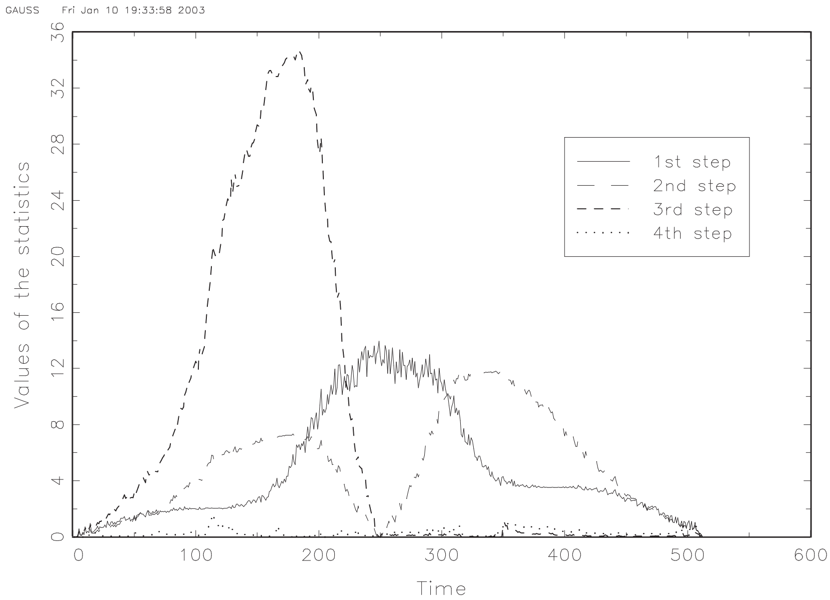
1. estimate the first break,
2. estimate the second break, and then re-estimate the first break,
3. estimate the third break, re-estimate the second break, and re-estimate the first break.

If there are more than two breaks, a *two step method* can also be constructed, but only one iteration can be sufficient.

And what about the computation time? Recall that the number of breaks is denoted by I. Let us assume that the stopping test performs correctly. For the classical sequential search, the number of break estimations, let it be denoted $\hat{I}_{classic}$, is equal to I, so

$$\hat{I}_{classic} = \mathrm{O}(I).$$

Figure 3: The sequential search



For the new sequential search, the number of estimations, say $\hat{I}_{new}$, is equal to $(I+1)I/2$, so

$$\hat{I}_{new} = \mathrm{O}(I^2).$$

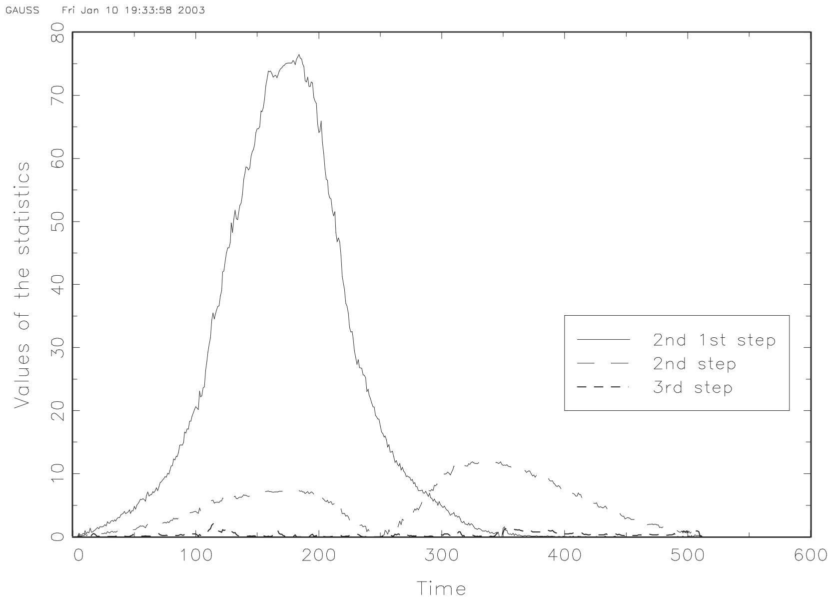
If the number of breaks I is not too large, the loss of computing time is not too critical. This is a reasonable assumption, since, if I is too large compared to the sample size T, it is more suitable to use another approach to model the time series, as, for example, a random threshold process that switches between regimes, or a long range dependence process that can approximate (under certain conditions) a structural break process, see Diebold and Inoue, (2001). For comparison with the time computing of a simultaneous estimation of all the breaks, let this be denoted by $\hat{I}_{sim}$, we present now the computing time in terms of the number of statistics calculated at a date t:

$$\begin{aligned} \hat{I}_{classic} &= T \times I, \\ \hat{I}_{new} &= T \times \frac{(I+1)I}{2}, \\ \hat{I}_{sim} &= T^I. \end{aligned}$$

Moreover, one must assume that one knows the number of breaks. Algorithms for optimising a search in a multidimensional space can reduce the computing time of the last type of estimation, but it stays exponential, and thus, dominates the others when I increases.

The problems of the Christiano (1992)'s method are that, since there is only one step in its search, in which it estimates all the breaks, all the estimations can be non consistent and/or some break dates can be difficult to detect. The advantage of the sequential search is that, since the model is more flexible for each step, the following estimations will detect the break dates with less bias at each step. Thus, in the estimated date set, there will be all the true break dates, but also spurious dates. If there are spurious dates, they correspond to the first estimations, which should be the most significant ! The new sequential search solves this problem. In an extreme case, if the time series are fractal, all the estimations can be spurious. But in this case, the distinction between

Figure 4: The sequential search with reestimation



this sort of structural breaks and long memory (fractal) property can have no sense. Thus we can restrict our analysis to the reasonable case of when the sample is finite, there is a finite number of breaks over. For comparison, we programmed a simple kernel estimation. We used a window in which we estimate an eventual break date in the same way than Christiano (1992). This method is more robust but it uses less information than Christiano's method and than the sequential search method, and thus, it can be less powerful. In fact, the sequential search uses all the information in the sample. One cannot find a method that uses more information. As for all kernel estimation, we are faced to the problem of the choice of the window. When the window grows, the method converges to the Christiano (1992)'s method with its disadvantages. If the window is too small, the detection will be too random (not powerful), and it cannot detect the breaks.

4 Tests for stopping the sequential search

As already mentioned earlier in the paper, we want to stop the sequential serach when the last estimated break is not significant, that is when the null of $i-1$ breaks against the alternative of i breaks is not rejected.

About the consistency of the estimation of the number of breaks, Bai and Perron (1998) provide some theoretical proofs for the consistency of the estimation of the number of breaks by using sequential tests. One of their remarks is that asymptotically, the significance level of the tests must decrease to 0 when $T \longrightarrow \infty$, otherwise, the number of breaks will be asymptotically over-estimated. Conversely, in finite samples, the significance level must increase to gain power for that the number of breaks is not under-estimated. However, since the significance level cannot be reasonably increased too much, in practice, the number of breaks can be over-estimated. We think that there is no solution to this problem: it is a very basic problem in econometrics, if there is no data enough, the significance of any variables cannot be detected. Each break is in fact specified as a variable, thus, it is possible that it cannot be detected.

The conventional test of this hypothesis is to compare the computed F-statistic with the 0.05 critical value of the relevant asymptotic F-distribution. However, this is not appropriate in our case as the time of break is not determined exogenously. Specifically, the standard F-distribution critical values appear to be too low, causing excessive rejection of the null hypothesis of structural stability. Even the bootstrapped F-distribution has this problem, if the time of break is considered to be determined exogenously.

If we consider that b_i is selected exogenously, the asymptotic P-value is

$$1 - \hat{F}_{as}(\max\{F_t\}) = \min\{1 - \hat{F}_{as}(F_t)\} \qquad \text{for all admissible } t,$$

where $\hat{F}_{as}$ is the relevant asymptotic estimation of the F-distribution (under the null); and the bootstrap P-value is

$$\min\{1 - \hat{F}_{bs,t}(F_t)\} \qquad \text{for all admissible } t,$$

where $\hat{F}_{bs,t}$ is the bootstrap estimation of the F-distribution at the break date t. These P-values turn out to be too conservative, bringing about a huge loss of power. Christiano (1992) has come to the same conclusion in a similar case of search for breaks.

To obtain a critical value that takes into account the structure of our model, there are two approaches: the Andrews and Ploberger's approach and the bootstrap approach. In these methods, we take into account the fact that the break was chosen endogenously on the basis of the maximum F-value or minimum bootstrap P-value.

4.1 The Andrews and Ploberger's approach (asymptotic)

If we use the asymptotic estimate of b_i, Andrews (1993; 2003) and Andrews and Ploberger (1994) tabulate critical values for the asymptotic distribution of the supremum of Gaussian or χ^2 set of variables, i.e., taking into account the endogenous character of the selecting mechanism. If we use the bootstrap estimate of b_i, there is no method in the literature that computes the asymptotic P-value of the test. Moreover, in the case of unit root series, or conditional series explained by a marginal series containing one or more breaks, the statistics are neither pivotal nor asymptotically pivotal. The asymptotic distribution would be very difficult to calculate. Therefore, we do not develop this approach.

4.2 The bootstrap approach

We bootstrap the set of F_t-values adopting the following approach.

To generate the bootstrap samples, we follow the same way as for the estimation. We run regression (4) under null of $i-1$ breaks imposed, such that the two dummies for last (i^{th}) estimated break are not included. The bootstrap errors terms are generated in the same way as for the estimate (parametrically or non-parametrically). Then the estimated parameters together with bootstrap error terms are used to create a recursive bootstrap sample x_t^* following Equation 6. The resampling is repeated B_{test} times, so that B_{test} bootstrap samples are obtained. B_{test} is not necessarily equal to B_{est}. B_{test} must be chosen as for classical bootstrap tests, depending on the distortions, especially on the skewness and the kurtosis of the statistics.

This time, however, we perform the whole testing procedure for each of these samples. We compute the bootstrap P-value as follows. For each of the bootstrap samples the test statistic τ is computed. The test statistics can be either or both the following :

$$\tau = \max\{F_t\} \text{ equivalent to the use of } \tau = \min\{1 - \hat{F}_{as}(F_t)\}, \quad (9)$$
$$\tau = \min\{1 - \hat{F}_{bs,t}(F_t)\} \qquad \text{for all admissible } t\ . \quad (10)$$

The bootstrap distributions of the statistic (obtained in the manner of empirical distribution) for each admissible time of break t are used to get the P-values. This procedure follows closely Christiano (1992).

The combination of the bootstrap approach for the test with the bootstrap estimation of the break point can be called a *double bootstrap*. This type of bootstrap is computationally costly and we do not use it for Monte Carlo experiments, but we use it for real data.

5 Monte Carlo experiments

In this section we evaluate the performances of the statistic tests and of the estimators using Monte Carlo experiments. We provide the bias and the standard deviations of the estimators, and the P value plots and the power–size curves of the tests.

5.1 Monte Carlo design

Each of the experiments contains $S = 1024$ replications. Each replication consists of generating a simulated series on which we apply the methods. The data generating processes (DGP) that we use are the set of processes defined by the equation 1b or 1a. We use two types of DGP to generate the simulated series. For the first type of DGPs, we do not impose a unit root hypothesis. For the second type of DGPs, we impose this hypothesis [4]. To be realistic, we choose the parameter values as the estimates from real data set that we eventually modify (see Appendix A).

For the bootstrap methods, we choose: $B_{test} = 99$ and $B_{est} = 99$. We saw that even with a small number of bootstrap replications, chosen to save computing time for simulations, the results are quiet robust. For the application with real data, we use a more substantial number of replications. Since the convergence of the estimators is very fast, we use only 3 reestimations for the "exogenous" asymptotic tests and 1 for the endogenous bootstrap tests. In our experiments, the methods with splitting lead to results close to the ones without splitting, and a little worse. So, we present results only for methods without splitting. However, the methods with splitting could be useful when the variance changes greatly before and after a break and if one does not want to specify this feature in the model (it will be better if we do). It can be also useful to verify that the method with splitting is stable.

At the end of a Monte Carlo experiment of S simulations, we obtain a set of S P-values for each statistic test, and a set of S estimations for each estimator. Note that the tests are run for each potential break date, so we obtain a set of P values for each of the tests but also for each of the break date. We also obtain a set of estimations for each of the estimators and for each potential break dates. We use the same set of random number for each experiment to reduce the experimental error in the comparison of the methods.

[4]We use this DGP to have a first idea about the average performance of the method over a large set of values for the parameters.

The estimates of the bias and the standard deviations of the estimators are computed as follows:

$$\widehat{\text{bias}} = \hat{\mathrm{E}}(\hat{b}_i) - b_i = \frac{1}{S}\sum_{s=1}^{S}\hat{b}_i - b_i,$$

$$\text{standard error: } \frac{1}{S-1}\sum_{s=1}^{S}(\hat{b}_i - \hat{\mathrm{E}}(\hat{b}_i))^2.$$

In case of multiple breaks, we suppose that the closest break is the one estimated by the estimator. But the bias can lead to wrong conclusions concerning the effectiveness of the estimators since: if a break is not in the middle of the time range, there are more possible realisations for the estimator in a side of the break rather than the other. This feature leads to a bias depending on the location of the break in the time range whereas the histograms of the estimators display maxima almost exactly at the break. Then, we decide to provide the maximum of the histogram of the estimators.

Remark 1: The choice of the sample size T and the standard deviation σ_e of the model

Generally, in Monte Carlo experiments, σ_e is only a scale parameter and it does not have any impact on the performances of the methods, and one presents the results for various sample sizes. But in the case of structural breaks, σ_e is not a scale parameter compared to the size of the break (intuitively, if σ_e is large, a break is less distinguishable from a random shock). It is impossible to present here the results for all the combinations of (T, σ_e), but we do not need so, since an experiment for a large T is equivalent (in the sense of the information contained in the series) to a smaller T with a smaller σ_e. So, we decide to present the results only for the sample size and the variance of our real data (as reported later in this paper), that are realistic parameters for general macroeconomic series. For information, we did simulations for $T \in \{256, 512, 1024\}$ and various σ_e (between 0.1 and 10). Of course the performances of the methods become better as T and σ_e increase, even though the results do not differ qualitatively. Thus, what we present is sufficient to illustrate the behaviour of the methods [5].

Remark 2: How to calculate the first values of the simulated series

In the unit root case, the starting value of the series belong to the parameter set, thus, to be close to real data, we put the simulated initial value equal to the first observation of the real series. In the stationary case, the first observations of the real series cannot be chosen because they are not necessarily generated from the distribution of the DGP with our imposed hypothesis. If one does so, the consequence is that the estimators do not distinguish these unlikely points with structural breaks. To solve this problem, we calculate the distributions of the first observations, when it is easy, and we generate simulated values from these distributions. But with exogenous regressors, it can be very difficult. Therefore, we calculate only approximations of these distributions in this case, then generate some additional points before time $t = 1$ for allowing the simulated series to approach the true distributions, and then we truncate the series to obtain observations between 1 and T.

[5] The full set of results is available on request from the authors.

5.2 The case of a stationary marginal process

In the case of stationary time series, theory says that bootstrap method performs asymptotically better than the asymptotic method. But in finite sample, there are some rare examples where the bootstrap is unstable. Moreover, in general, it can suffer from size distortions. Nevertheless, the bootstrap tests in our Monte Carlo experiments have optimal performances.

For calibrating our simulations, we can note, from our bootstrapped ADF tests, that the UK interest rate is I(0). We then use it to guide us in the choice of the parameters in our simulations.

5.2.1 Hypothesis of no break in the series

The estimated model of the UK interest rate under the hypothesis of no break is

$$\begin{aligned}
x_t &= -0.7600 + 0.0158\ t + 1.2898\ x_{t-1} + -0.4385\ x_{t-2} + e_t \\
e_t &\sim \mathrm{D}(0, 1.8114, -1.2163, 6.8540) \\
t &\in \{1, \ldots, 110\}
\end{aligned}$$

where D is the modified empirical distribution of the residuals with (in brackets) its four moments, i.e. mean, standard error, skewness, kurtosis respectively. In this case, the performances of the estimators are not applicable since there is no break to estimate. For the tests, we present the probabilities of rejecting the null hypothesis of "no break" against "one break". Since "no break" is the true, the probabilities are the true probabilities of rejecting the null. These are the sizes of the tests. We decide to present also the performances of the tests for testing "one break" against "two breaks". That is, if the first break is accepted (that arises in any cases), the true probabilities of rejecting the null of "no more break (more than the one that was found)", so, it is also sizes of the tests. Figure 5 reports the results.

5.2.2 Hypothesis of one break in the series

The estimated model of the UK interest rate under the hypothesis of one break is

$$\begin{aligned}
x_t &= -0.7124 - 0.0097\ t + 1.2284\ x_{t-1} - 0.4512\ x_{t-2} + 2.4520\ M_t + 0.0004\ D_t + e_t \\
e_t &\sim \mathrm{D}(0, 1.7430, -1.3214, 7.0716) \\
t &\in \{1, \ldots, 110\} \\
M_t &= \mathbf{I}(t \geq 47) \\
D_t &= (t - 47 + 1)\mathbf{I}(t \geq 47)
\end{aligned}$$

Since this break is not necessarily significant in the original time series, we should generate series that look breakless, but in this case we will not be able to measure the power of the tests. Thus, we increased the size of the amplitude of the break compared to the series by reducing by 2 the standard deviation of the error terms in the DGP for simulations:

$$e_t \sim \mathrm{D}(0, 1.7430/2, -1.3214, 7.0716).$$

To do this, we divide by 2 each component of the sample in which we draw the error terms.

Figure 5: Stationary case: No and One Break

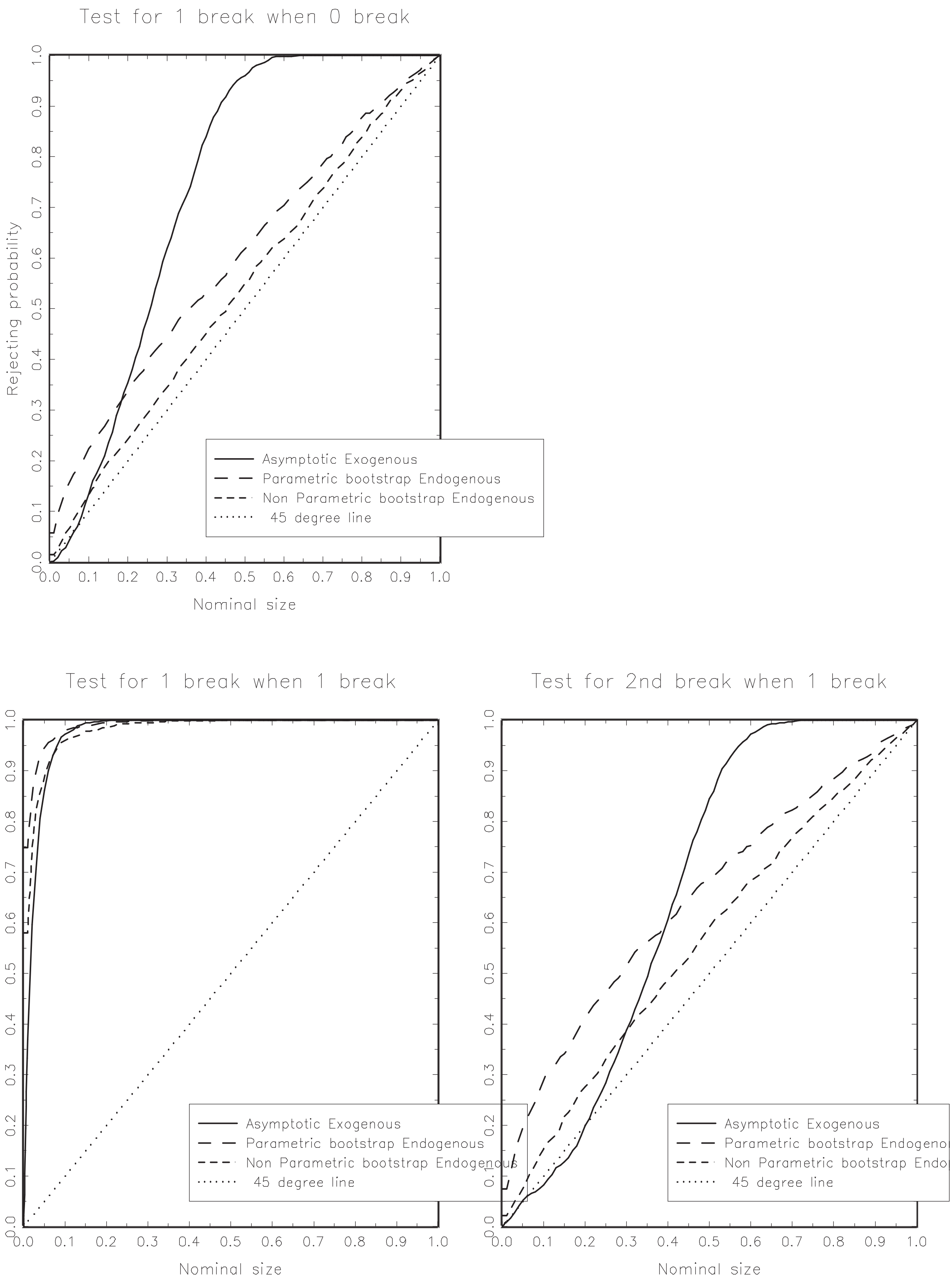


Table 1: One Break, Break Estimators

Estimator	Mean	Bias	Standard Deviation	Density maximum
Asymptotic	50.5068	3.5068	18.9138	46.4375
Parametric bootstrap	51.7818	4.7818	18.4119	46.4375
Nonparametric Bootstrap	50.9116	3.9116	17.1371	46.4375

The performances of the estimators are applicable for the first break. The results are presented in Table 1. Since there is only one break, there is no difference between with or without splitting the sample.

The performances of the tests are assessed for the following hypothesis considerations:

1. "no break" against "one break" (power of the tests),
2. "one break" against "two breaks" (size of the tests),
3. "two breaks" against "three breaks" (size of the tests).

Since for the first break, the null hypothesis of no break is false and the alternative of one break is true, the probabilities of rejecting the null are not the size of the tests but the probabilities of accepting the alternative when it is true, what are called the powers of the tests. See Figures 5 for the results.

5.2.3 Hypothesis of two breaks in the series

The estimated model of the UK interest rate under the hypothesis of two breaks is

$$\begin{aligned}
x_t &= -1.7662 + 0.1404\ t + 1.0713\ x_{t-1} - 0.3956\ x_{t-2} \\
&\quad +1.8697\ M_{1,t} - 5.0104\ D_{1,t} - 0.1583\ M_{2,t} + 0.0056\ D_{2,t} + e_t \\
e_t &\sim \mathrm{D}(0, 1.5789, -0.5210, 3.4676) \\
t &\in \{1, \ldots, 110\} \\
M_{1,t} &= \mathbf{I}(t \geq 48) \\
D_{1,t} &= (t - 48 + 1)\mathbf{I}(t \geq 48) \\
M_{2,t} &= \mathbf{I}(t \geq 20) \\
D_{2,t} &= (t - 20 + 1)\mathbf{I}(t \geq 20)
\end{aligned}$$

For the same reason as above, we reduce by 2 the standard deviation of the error terms in the DGP for simulations.

The performances of the estimators are applicable for the two first breaks. The results are presented in Table 2.

The performances of the tests are assessed for the following hypothesis considerations:

1. "no break" against "one break": power of the tests,
2. "one break" against "two breaks": size of the tests,
3. "two breaks" against "three breaks": size of the tests,

Table 2: Two Breaks, Break Estimators

Estimator without splitting for the 1st break	Mean	Bias	Standard Deviation	Density maximum
Asymptotic	51.9414	3.9414	12.2672	48.0625
Parametric bootstrap	52.5583	4.5583	11.7295	51.3125
Nonparametric Bootstrap	53.4069	5.4069	10.9326	53.6984
Estimator without splitting for the 2nd break	**Mean**	**Bias**	**Standard Deviation**	**Density maximum**
Asymptotic	20.7803	0.7803	5.8156	20.4375
Parametric bootstrap	22.6985	2.6985	8.4471	20.4375
Nonparametric Bootstrap	22.2532	2.2532	8.7553	20.6825

4. “three break” against “four breaks”: size of the tests.

See Figure 6 for the results.

5.3 The case of a marginal process with a unit root

In our knowledge, in the case of I(1) processes, there is a gap of bootstrap theory. Only Monte Carlo experiments can give an idea of the performance of bootstrap in this situation.

From our bootstrapped ADF tests, we can conclude that the UK income is I(1). We then use it to guide our simulations in the case of unit root processes.

5.3.1 Hypothesis of no break in the series

The estimated model of the UK income under the hypothesis of no break is

$$\begin{aligned} x_t &= 0.0056 + x_{t-1} + e_t \\ e_t &\sim \mathrm{D}(0, 0.0110, 0.3501, 5.5339) \\ t &\in \{1, \ldots, 110\} \end{aligned}$$

See Figure 7 for the results.

5.3.2 Hypothesis of one break in the series

The estimated model of the UK income under the hypothesis of one break is

$$\begin{aligned} x_t &= -0.0001 + 0.0012\ t + \ x_{t-1} - 0.0160\ M_t - 0.0012\ D_t + e_t \\ e_t &\sim \mathrm{D}(0, 0.0107, 0.1693, 4.4626) \\ t &\in \{1, \ldots, 110\} \\ M_t &= \mathbf{I}(t \geq 16) \\ D_t &= (t - 16 + 1)\mathbf{I}(t \geq 16) \end{aligned}$$

As in the stationary case, since the break is not necessarily significant in the original time series, we increased the size of the break by dividing it by 2 the standard deviation

Figure 6: Stationary case: Two Breaks

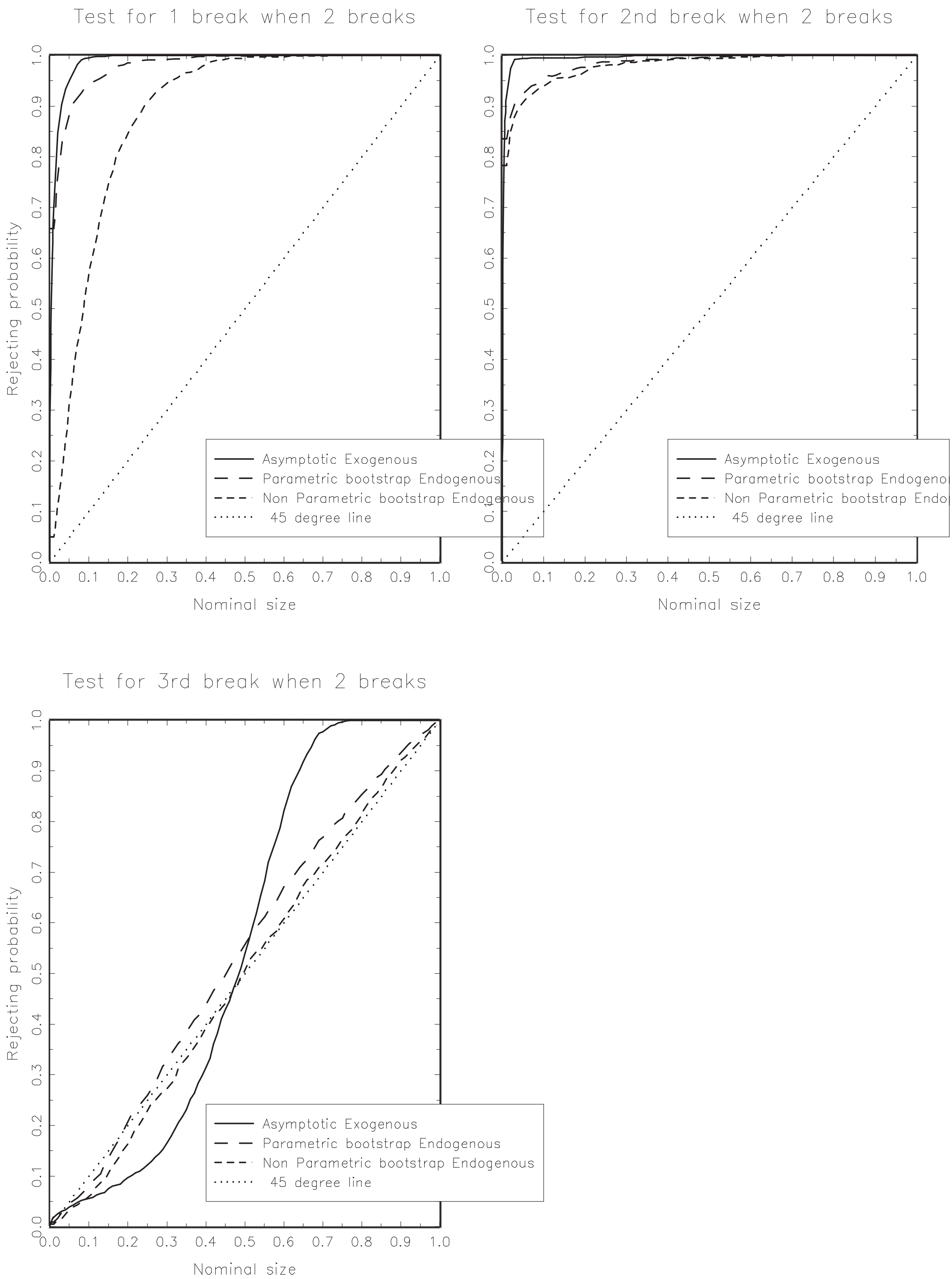


Figure 7: Integrated case: No and One Break

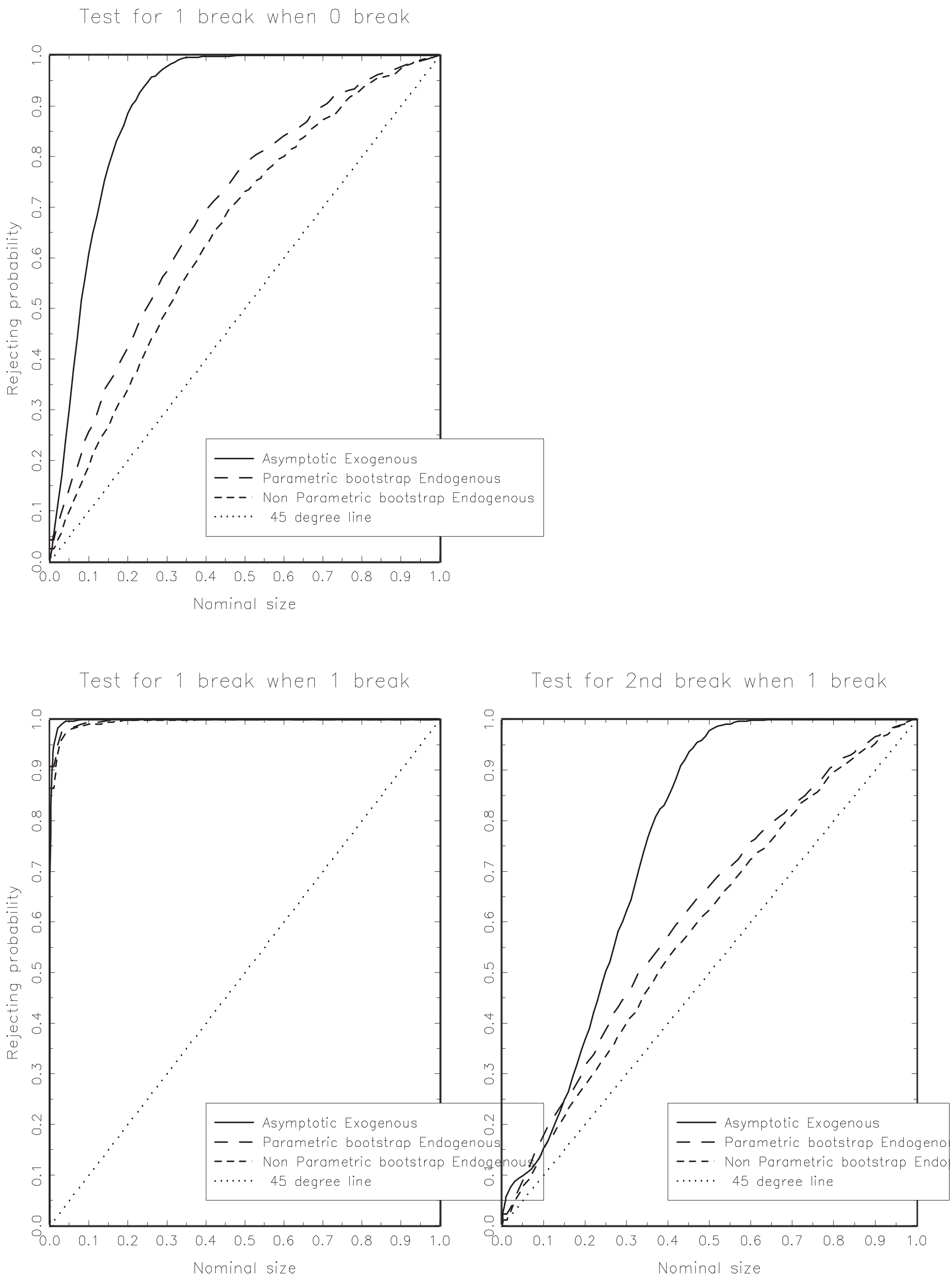


Table 3: One Break, Break Estimators

Estimator	Mean	Bias	Standard Deviation	Density maximum
Asymptotic	30.9482	14.9482	26.5415	**16.3047**
Parametric bootstrap	30.3136	14.3136	28.1061	**16.3047**
Nonparametric Bootstrap	30.1608	14.1608	26.9022	**16.3047**

Table 4: One Break, Break Estimators in the fixed parameter case

Estimator without splitting for the 1st break	Mean	Bias	Standard Deviation	Density maximum
Asymptotic	41.9814	25.9814	28.3791	16.3047
Parametric bootstrap	43.9686	27.9686	27.2796	16.3047
Nonparametric Bootstrap	44.4705	28.4705	26.7002	16.4091
Estimator without splitting for the 2nd break	**Mean**	**Bias**	**Standard Deviation**	**Density maximum**
Asymptotic	32.9346	19.9346	29.4630	14.6641
Parametric bootstrap	29.6262	16.6262	27.0769	14.6641
Nonparametric Bootstrap	31.2554	18.2554	28.3191	14.5000

of the error terms in the DGP for simulations:

$$e_t \sim \mathrm{D}(0, 0.0107/2, 0.1693, 4.4626).$$

The results are presented in Table 3. See Figures 7 for the results dealing with the performances of the tests.

5.3.3 Hypothesis of two breaks in the series

The estimated model of the UK income under the hypothesis of two breaks is

$$\begin{aligned}
x_t &= -1.7662 + 0.1404\ t + 1.0713\ x_{t-1} - 0.3956\ x_{t-2} \\
&\quad +1.8697\ M_{1,t} - 5.0104\ D_{1,t} - 0.1583\ M_{2,t} + 0.0056\ D_{2,t} + e_t \\
e_t &\sim \mathrm{D}(0, 1.5789, -0.5210, 3.4676) \\
t &\in \{1, \ldots, 110\} \\
M_{1,t} &= \mathbf{I}(t \geq 48) \\
D_{1,t} &= (t - 48 + 1)\mathbf{I}(t \geq 48) \\
M_{2,t} &= \mathbf{I}(t \geq 20) \\
D_{2,t} &= (t - 20 + 1)\mathbf{I}(t \geq 20)
\end{aligned}$$

For the same reason as above, we divide by 2 the standard deviation of the error terms in the DGP for simulations. The performances of the estimators are applicable for the two first breaks. The results are presented in table 4. The biases are very large in this case

Table 5: One Break, Break Estimators

Estimator	Mean	Bias	Standard Deviation	Density maximum
Asymptotic	78.6855	0.6855	2.2866	78.1172
Parametric bootstrap	82.5974	4.5974	2.5379	83.9575
Nonparametric Bootstrap	82.7128	4.7128	2.5011	83.5960

because the breaks are at the beginning of the series. But we can see, by looking at the density maxima, that the estimators performs well. See Figures 8 for the results dealing with the performances of the tests.

5.4 The conditional process

For these simulations, we add marginal variables in the regressor matrix to explain the conditional variable. When the regressors are not stationary, especially with structural breaks, the statistics are very far from pivotal. The asymptotic distributions are very difficult to calculate, and the bootstrap could encounter problems. Our Monte Carlo experiments show the opposite.

For the two marginal series, we choose the real interest rate and the real income. We then generate simulated conditional series. For the choice of the model parameters, we are inspired from the UK money demand.

5.4.1 Hypothesis of no break in the series

Under this hypothesis, the series seem to be not cointegrated, leading to the invalidation of the estimation. So, simulations are not provided in the context. However, it can seen in the Section 6 how to treat this situation.

5.4.2 Hypothesis of one break in the series

The estimated model of the UK Money Demand under the hypothesis of one break is

$$\begin{aligned}
y_t &= 0.1520 - 0.0049\ t + 0.6729\ y_{t-1} + 0.5478\ x_{1,t} + 0.0008\ x_{2,t} - 0.0281\ M_t + 0.0044\ D_t + e_t \\
e_t &\sim \mathrm{D}(0, 0.0109, 0.4649, 4.5921) \\
t &\in \{1, \ldots, 110\} \\
M_t &= \mathbf{I}(t \geq 78) \\
D_t &= (t - 78 + 1)\mathbf{I}(t \geq 178)
\end{aligned}$$

where x_1 is the income and x_2 is the interest rate. We reduce by 2 the standard deviation of the error terms in the DGP for simulations. The results are presented in table 5. See Figures 9 for the results dealing with the performances of the tests.

Figure 8: Integrated case: Two Breaks

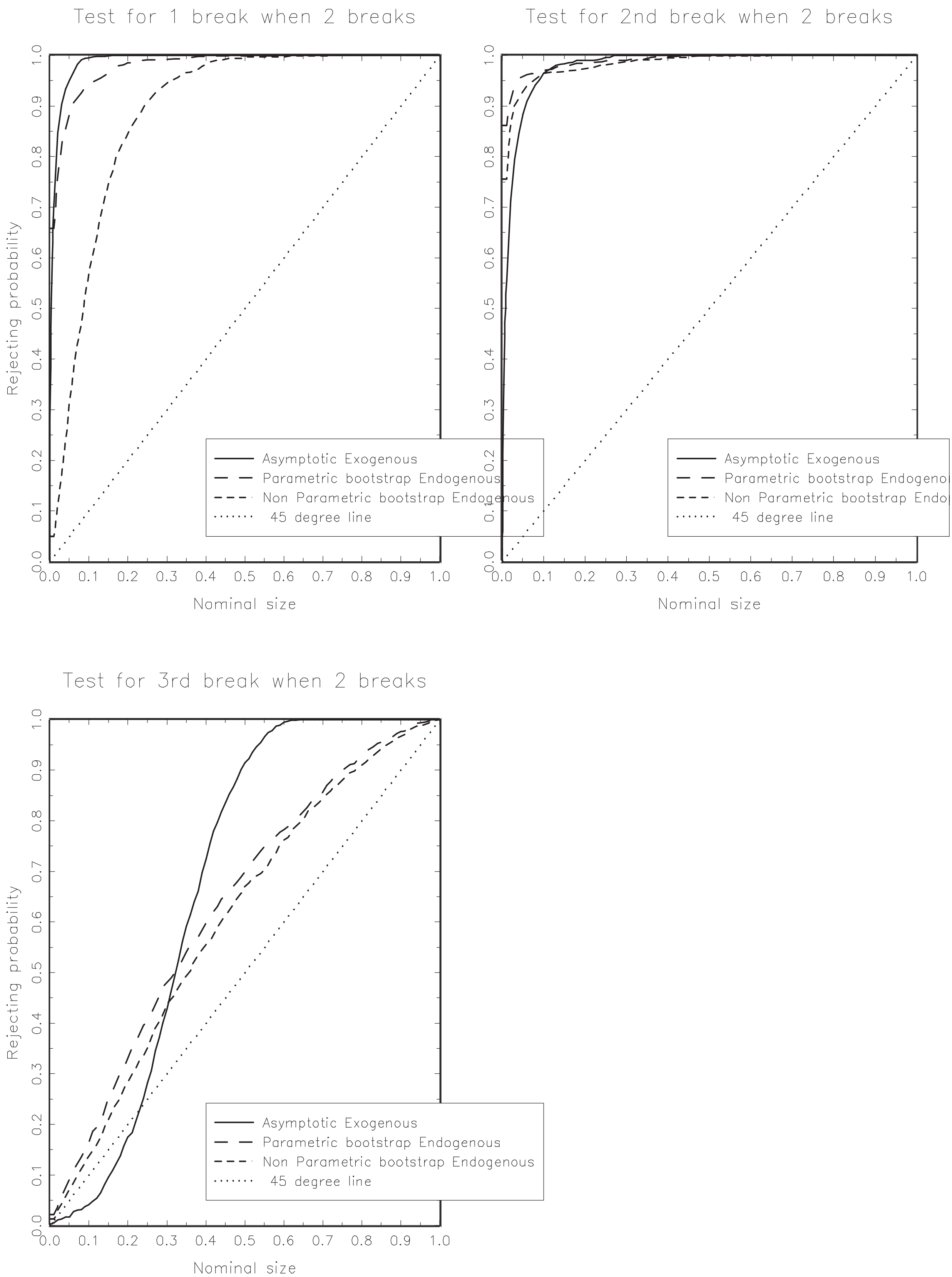


Figure 9: Conditional series: One Break

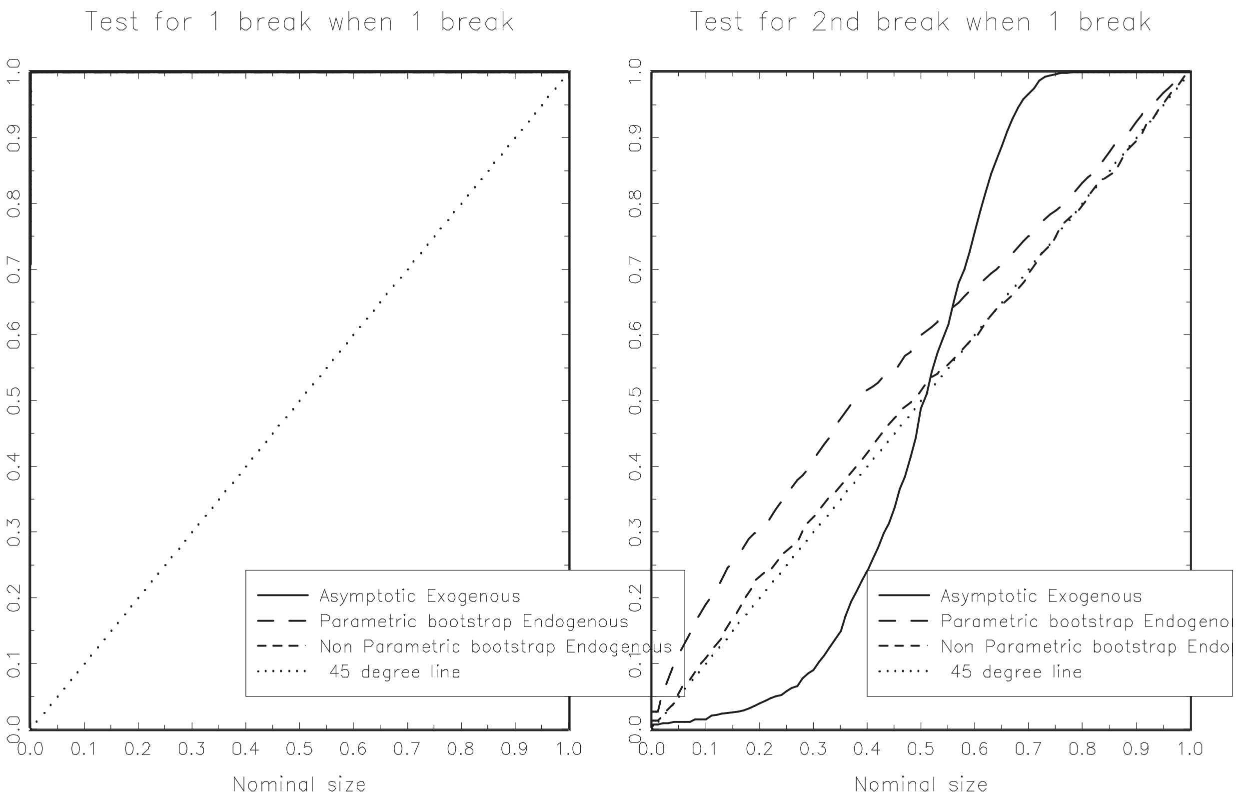


Table 6: One Break, Break Estimators in the fixed parameter case

Estimator without splitting for the 1st break	Mean	Bias	Standard Deviation	Density maximum
Asymptotic	78.7803	0.7803	3.2443	77.3125
Parametric bootstrap	82.9564	4.9564	5.4773	85.4375
Nonparametric Bootstrap				
Estimator without splitting for the 2nd break	**Mean**	**Bias**	**Standard Deviation**	**Density maximum**
Asymptotic	43.2725	3.2725	17.1408	39.9375
Parametric bootstrap	46.1376	6.1376	16.5890	39.9375
Nonparametric Bootstrap				

5.4.3 Hypothesis of two breaks in the series

The estimated model of the UK money demand under the hypothesis of two breaks is

$$
\begin{aligned}
y_t &= 0.4245 - 0.0039\ t + 0.6964\ y_{t-1} + 0.4591\ x_{1,t} + 0.0012\ x_{2,t} \\
&\quad -0.0282\ M_{1,t} - 0.0194\ D_{1,t} + 0.0038\ M_{2,t} - 0.0001\ D_{2,t} + e_t \\
e_t &\sim \mathrm{D}(0, 0.0104, 0.1959, 4.7722) \\
t &\in \{1, \ldots, 110\} \\
M_{1,t} &= \mathbf{I}(t \geq 78) \\
D_{1,t} &= (t - 78 + 1)\mathbf{I}(78) \\
M_{2,t} &= \mathbf{I}(t \geq 40) \\
D_{2,t} &= (t - 40 + 1)\mathbf{I}(t \geq 40)
\end{aligned}
$$

We divide by 2 the standard deviation of the error terms in the DGP for simulations. The performances of the estimators are applicable for the two first breaks. The results are presented in table 6. See Figures 10 for the results dealing with the performances of the tests.

6 Empirical application: modelling UK money demand

To show the importance of taking into account the presence of breaks in macroeconomic time series, we present two modellings of the relation between the UK money demand and the real income and the interest rate. First, a simple model is used where the money demand is directly explained by the real income and the interest rate. Second, a more complicated model is used where the marginal models (with their breaks) are integrated into the conditional model to create a reduced form of the system. The graphs of the series are reported in figure 11.

Figure 10: Conditional series: Two Break

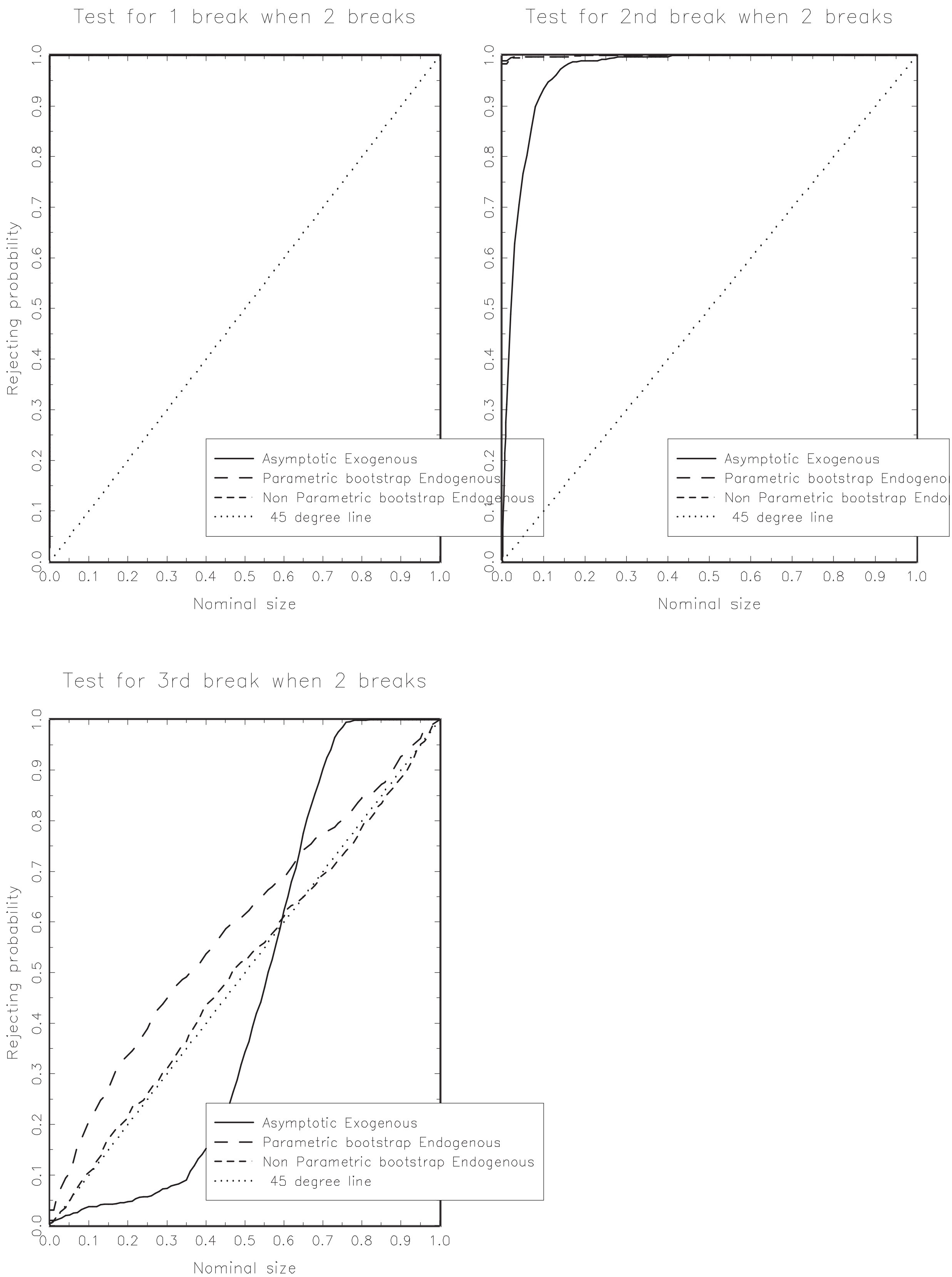
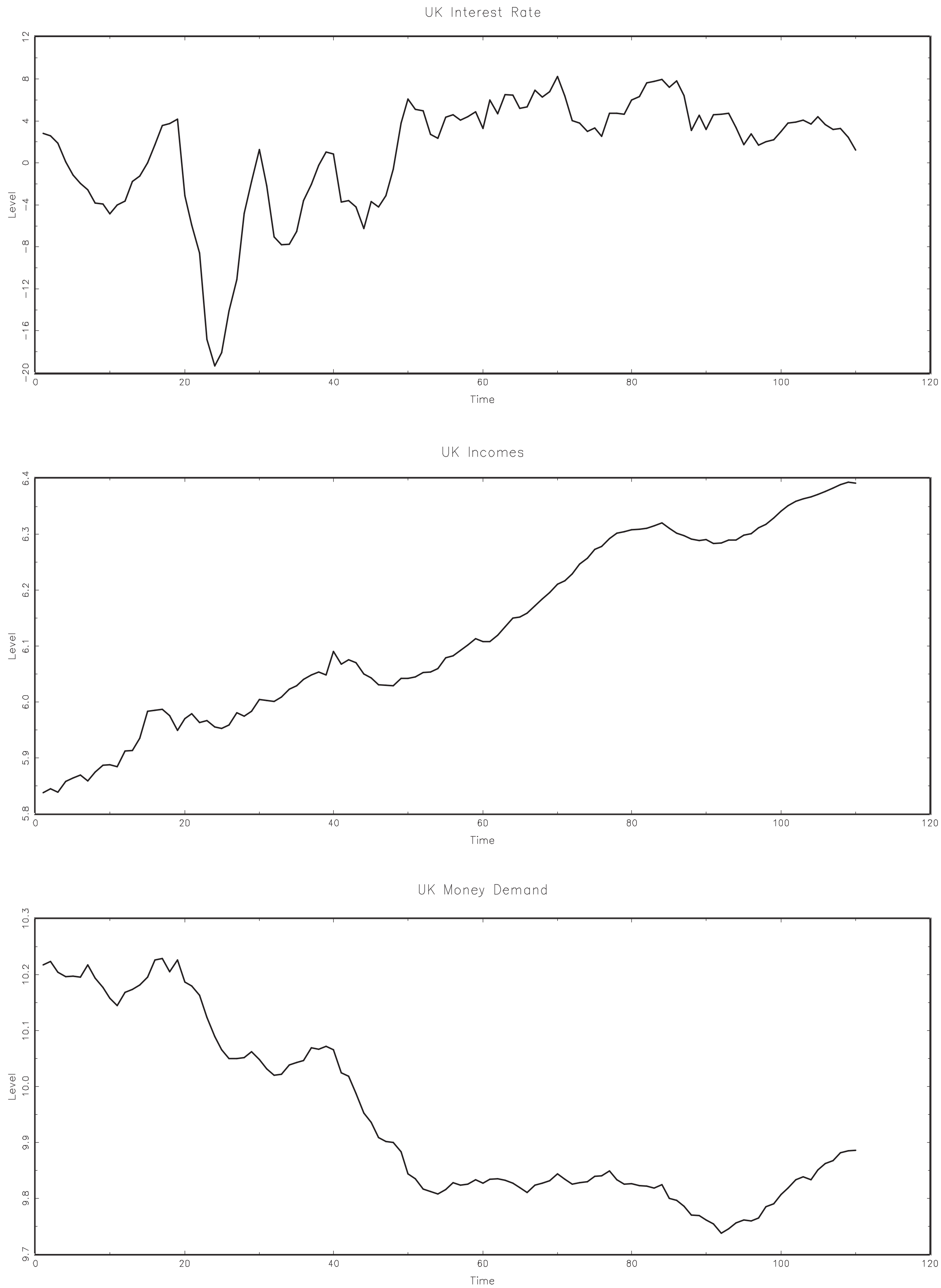


Figure 11: The three series



6.1 First modelling: cointegration

We start with a simple model. We explain the money demand by a constant term, a trend term, the real income and the interest rate. Structural breaks in this relation will be added in a second step. Since both the money demand and the incomes are I(1) series, we will test for cointegration among the series. We do not consider lags of money demand, otherwise, they will capture the unit root feature in case of noncointegration. If there is cointegration, the error terms of the regression described above should be I(0). For testing the integration order of the error terms, parametric and nonparametric bootstrapped versions of the augmented of Dickey-Fuller (ADF) tests are applied on the residuals, with unilateral and bilateral bootstrap P values (see appendix B). The number of bootstrap replications being 999. For selecting the number of augmentations, the residuals from the ADF regressions are tested for serial correlation using Ljung-Box and Box-Pierce tests (from 1 to 8 lags) until they look like white noise. The number of augmentations is 0 here. All tests have been conducted by using our own programs via Gauss software.

6.1.1 Modelling without break

Table 7 provides the bootstrap P values of the ADF tests in the relation without break. The results suggest that the variables are strongly noncointegrated.

Table 7: bootstrap P values of the ADF tests on the residuals

parametric bootstrap (Gaussian)

Regressions	Unilateral P values	Bilateral P values
AR(1)	0.5275	0.8048
AR(1) + const	0.8098	0.4024
AR(1) + const + trend	0.8158	0.3864

nonparametric bootstrap

Regressions	Unilateral P values	Bilateral P values
AR(1)	0.5265	0.7888
AR(1) + const	0.8208	0.3844
AR(1) + const + trend	0.8218	0.3744

6.1.2 Estimation of the breaks

We estimate and test for one, two, and three breaks in the UK money demand. We do not split the sample, and apply three reestimations for the breaks for the estimation as well as for the tests. Since the bootstrap estimators leads to very similar results to those for asymptotic estimators, we provide only the results for this latter method. A large number of bootstrap replications must be used for estimation. Concerning the tests, we provide the results for the exogenous asymptotic test, and both the endogenous parametric and nonparametric bootstrap tests using $B_{Test} = 999$. Table 8 provides the estimations of the location of the breaks. A great stability can be remarked for the estimates, in this case,

Table 8: Asymptotic Estimates

Number of simultaneous estimations	1rt break	2nd break	3rd break
1	85	—	—
2	85	40	—
3	85	40	30

a reestimation is not necessary. Table 9 provides the P values for the presence of the last estimated break. These results suggest us the presence of only one break at $t = 85$.

Table 9: P values

Break	Asymptotic	Parametric bootstrap	non parametric bootstrap
1st: $b_1 = 85$	0.001	0.0000	0.0000
2nd: $b_2 = 40$	0.4080	0.2332	0.3413
3rd: $b_3 = 30$	0.4699	0.2192	0.3363

For verifying the stability of these results, we split the sample and apply again the tests on the first subsample: $t = 1$ to $t = 84$. The results with splitting confirm the previous results.

6.1.3 Modelling with one break

Now, we consider the same model, but we add one break at $t = 85$ (it was determined previously). We rerun the bootstrapped ADF tests and we obtain the results in table 10. Exceptionally, we take 9999 bootstrap replications in order to gain precision because the P values are very small. The number of augmentations is 0 again. By adding only one break, we obtain a dramatically different conclusion: now the variables appear to be strongly cointegrated.

6.2 Second modelling: co-breaking

We use now a more complicated model. First, the location of the breaks in marginal models is estimated. Next, the marginal models are integrated into the conditional model to create a reduced form of the system: the conditional model is estimated by replacing income and the interest rate by their explanatory variables, *i.e.* by their lags and their breaks. Following the estimation of the marginal models, the conditional model with its own breaks is estimated.

For the estimations of the break locations, we use the asymptotic estimator that is sufficient in our situation. Estimations are done without splitting (including breaks one by one). Three re-estimations are processed. The nonparametric bootstrap test is used with 999 bootstrap replications. For deciding whether a break is significant, a significance level of 0.10 is chosen. 0.10 is larger than the usual significance level of 0.05. However,

Table 10: ADF tests with one break

parametric bootstrap (Gaussian)

Regressions	Unilateral P values	Bilateral P values
AR(1)	0.0000	0.0000
AR(1) + const	0.0000	0.0000
AR(1) + const + trend	0.0001	0.0002

nonparametric bootstrap

Regressions	Unilateral P values	Bilateral P values
AR(1)	0.0000	0.0000
AR(1) + const	0.0000	0.0000
AR(1) + const + trend	0.0000	0.0000

since the significance level must decrease to zero when the sample size goes to infinite for not over-estimating the break number, the level must also increase (reasonably) when the sample size is small for not under-estimating the break number (because the tests may be not very powerful). The stars in the tables indicate the P values that are significant.

6.2.1 Breaks in income

For the past information set, the first lag of the variable is chosen. The first part of table 11 provides the sequential estimates for two breaks (with re-estimations), three breaks, four breaks, and eight breaks. The estimation in the beginning of the series seems unstable. Perhaps this is due to a progressive structural change. Probably there are biases induced by other breaks. Nevertheless, we must be careful when estimating the first breaks. In a first time, we run tests with three re-estimations (see first line in the second part of table 11). Only the third break seems significant. Since it can disturb the detection of the two first breaks, we rerun the tests fixing the third break: 85 (see second line if the second part of table 11). Thus, we rerun the sequential tests by putting the third break (85) as known (see second line of second part of table 11). The second break (44) now appears significant. However, the first break (18) is not significant, but, when a third break estimation with re-estimation is done, the break date 18 becomes 20 and 15, that are jointly significant. The same feature arises when the date 44 is imposed. We can conclude that the date 18 is spurious and explained by the presence of two true break dates that are too close: 20 and 15. There is a second justification for this choice: when more than four estimations are done (with re-estimations, and thus, less biased) 18 disappears in favour to 20 and 15. We can conclude that the four breaks are significant: at dates 41, 85, 20, and 15.

6.2.2 Breaks in interest rate

For the past information set, the two first lags of the variable are chosen. The second line of table 12 provides the sequential estimates (with re-estimations). The estimation is relatively stable, we run tests with only one re-estimations, that is sufficient here (see

Table 11: Breaks in Income

Break No	1	2	3	4
Two locations	16	13		
Three locations	18	44	85	
Four locations	18	41	85	39
Eight locations	20	41	85	39
P value	0.5976	0.7077	0.0440*	0.4004
P value	0.2943	0.0711*	known	0.0420*
Associated breaks	44,18	44	85	44,20,15
P value	0.2773	known	0.0220*	0.0551*
Associated breaks	85,18	44	85	85,20,15

third line table 12). Only the second and the fourth breaks seem significant. We rerun the tests fixing the second break: 20 (see fourth line table 12) this allows to see that the first break (48) is in fact significant. We also rerun the sequential tests by putting the fourth break (23) as known (see fifth line of table 12). Now 31 appears jointly significant with 48 and 20. We can conclude that the fourth first breaks are significant.

Table 12: Breaks in Interest Rate

Break No	1	2	3	4
Break locations	48	20	31	23
P value	0.5445	0.0521*	0.1111	0.0130*
P value	0.0080*	known	0.1221	0.0070*
P value	0.0200*	0.0651*	0.0310*	known
Associated breaks	31,20,48	29,20	25	23

6.2.3 Breaks in the money demand as a marginal

The money demand is explained only by its past and dummies. For the past information set, the two first lags of the variable are chosen [6]. The second line of table 13 provides the sequential asymptotic estimates of the break locations in the money demand series. These estimates are very stable: the re-estimations do not change the estimates. Thus, to save computing time, tests are processed without re-estimations. After a first test sequence, the second and the third breaks seem significant, but not the first (see third line of table 13). But as for the estimation where the next breaks can bias the estimates of the first breaks, the inference can suffer from the same problem: since for testing for the first breaks, the regression is not well specified, the first breaks can appear not significant (since the test interprets the shape of the tested break as "normal" and not as a break according to the rest of the series since there are other (non-detected) breaks in the rest of the series). Thus, we rerun the sequential tests by putting the second break (20) as

[6] The choice was done as previously, using tests for residual independence and the BIC.

known (see fourth line of table 13). The first break (41) now appears significant. We can conclude that the three first breaks are significant [7].

Table 13: Breaks in the money demand as marginal

Break No	1	2	3	4
Break location	41	20	85	8
P value	0.4024	0.0180*	0.0551*	0.2943
P value	0.0320*	known	0.0350*	0.2743

6.2.4 Co-breaking analysis

Table 14 reports the estimated breaks in the three series and it is clear that the money demand and the income have the same breaks. But since the money demand was estimated as a marginal, the breaks of the explanatory variables (that are not specified in a regressive model) will appear in the money demand. Thus, we cannot know whether the detected breaks in the money demand are totally explained by the explanatory variables or if the events that provoked these structural changes have their own effect on the money demand. Then, we estimate the breaks in the money demand conditionally to its past, income and interest rate.

Table 14: Breaks in the three series

Break No	1	2	3	4
Income	41	85	20	15
Interest rate	48	20	31	23
Money demand	41	20	85	

6.2.5 Breaks in money demand conditionally to income and interest rate

Finally, we estimate the conditional model by replacing income and the interest rate by their explanatory variables: we replace the income by its first lag and its breaks and the interest rate by its two first lags and its breaks (except at date 20 which is common with the income). The ability to detect co-breaking at a date between money demand and an explanatory variable disappears since the breaks in the explanatory variables are now specified. Only the own breaks of the money demand will appear (see table 15).

For seeing whether specifying these breaks is really useful in the UK money demand modelling, the Ljung-Box and Box-Pierce tests are run on the residuals of the regression without and with these two breaks. Table 16 presents the P values of these tests. The stars * correspond to the significant P values at the 0.05 level. Table 16 shows that two other breaks, in addition to the breaks coming from the marginal processes, have to be defined for specifying much better the regression model.

[7] Some other explorations was done on the fourth break, but it does not appear significant.

Table 15: Own Breaks in Money Demand

Break No	1	2	3	4
Break locations	39	59	10	90
P value	0.0485*	0.0566*	0.3131	0.1697

Table 16: P values of the tests for independence of the residuals of the UK money demand regression

Regression without the own breaks of the money demand				
Lag(s)	1	2	3	4
Ljung-Box test	0.0090*	0.0213*	0.0440*	0.0041*
Box-Pierce test	0.0100*	0.0238*	0.0490*	0.0054*
Regression with the own breaks of the money demand				
Lag(s)	1	2	3	4
Ljung-Box test	0.1027	0.2578	0.0576	0.0022*
Box-Pierce test	0.1075	0.2676	0.0656	0.0031*

7 Conclusion

In this paper we propose bootstrap tests as criterion for stopping sequential search of breaks. Monte Carlo experiments show that the bootstrap tests have remarkably good performances, even for integrated series (where there is a gap of theory for the moment), and for nonstationary regressors in the sense of break (where tests can suffer from large distortion generally). The only problem comes from the fact that the first estimations can be disturbed (biases for the estimators, and less significant for the test statistics) due to the misspecification of the rest of the series. This problem is independent from the bootstrap methodology, and seems unavoidable for all the procedures existing in the literature for determining the number of breaks. The solution we propose, which we included in the bootstrap methodology, is a re-estimation of the first breaks knowing next breaks. Finally, we show in a simple application the importance of taking into account breaks for determining co-integration, and the usefulness of these procedures for determining co-breaking.

References

Andrews, D.W.K. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61**, 821–856. 1, 4.1

Andrews, D.W.K. 2003. Tests for parameter instability and structural change with unknown change point: a corrigendum. *Econometrica*, **71**, 395–397. 1, 4.1

Andrews, D.W.K., & Ploberger, W. 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, **62**, 1383–1414. 1, 4.1

Bai, J. 1997. Estimating Multiple Breaks One at a Time. *Econometric Theory*, **13**, 315–352. 1

Bai, J. 1999. Likelihood ratio tests for multiple structural changes. *Journal of Econometrics*, **91**, 299–323. 1

Bai, J., & Perron, P. 1998. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, **66**, 47–78. 1, 4

Banerjee, Lazarova, & Urga. 1998. Bootstrap Sequential Tests for Multiple Structural Breaks. *CASS Working Paper*. (document), 1, 2

Basawa, I. V., A. K. Mallik W. P. McCormick J. H. Reeves, & Taylor, R. L. 1991. Bootstrapping Unstable First-Order Autoregressive Processes. *The Annals of Statistics*, **19**, 1096–1101. 1

Christiano, L. J. 1992. Searching for a Break in GNP. *Journal of Business and Economic Statistics*, **10**, 237–250. 3.1.2, 3.3.1, 3.3.1, 3.3.2, 4, 4.2

Culver, S. E., & Papell, D. H. 1997. Is There a Unit Root in the Inflation Rate? Evidence from Sequential Break and Panel Data Models. *Journal of Applied Econometrics*, **12**, 435–444. 1

Davidson, R., & MacKinnon, J. 1993. *Estimation and inference in economics*. Oxford University Press. New York. 7c

Dekker, Marcel. 1996. New York. *Econometric Reviews*, **15(2)**, 435–444. 1

Diebold, F.X., & Inoue, A. 2001. Long Memory and Regime Switching. *Journal of Econometrics*, **105**, 131–159. 3.3.2

Efron, B., & Tibshirani, R.J. 1993. *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability 57, Chapman and Hall. London. 1

Freedman, D. A., & Peters, S. C. 1984a. Bootstrapping a Regression Equation: Some Empirical Results. *Journal of the American Statistical Association*, **79**, 97–106. 1

Freedman, D. A., & Peters, S. C. 1984b. Bootstrapping an Econometric Model: Some Empirical Results. *Journal of Business and Economic Statistics*, **2**, 150–158. 1

Hansen, B. E. Testing for structural change in conditional models", journal=. 1

Horowitz, J.L. 2001. *The bootstrap*. (eds) J.J. Heckman and E. Leamer. in Handbook of Econometrics, Chapter 52, Vol. 5, North Holland. 1

MacKinnon, J. G. 1994. Approximate Asymptotic Distribution Functions for Unit Root and Cointegration Tests. *Journal of Business and Economic Statistics*, **12**. 167–176. 1

Perron, P. 1989. The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis. *Econometrica*, **57**, 1361–1401. 1, 3.3.1

Weber, N.C. 1984. On resampling techniques for regression models. *Statistics and Probability Letters*, **2**, 275–278. 3.1.2

A Monte Carlo design: the choice of the parameter values

We choose the parameter values for the simulation experiments as the estimates from real data set that we eventually modify. The estimates are constructed in the following way:

1. Parametric and nonparametric bootstrapped versions of the augmented of Dickey-Fuller (ADF) tests are performed , with unilateral and bilateral bootstrap P values (see appendix B), to see whether the series can be considered I(1) or not. The ADF regressions are augmented until the residuals are white noise using Ljung-Box and Box-Pierce tests (with 1 to 8 lags). The number of bootstrap replications is 999.

2. We choose the hypothesis under which we want to assess the performances of the estimators and tests: 0 break, 1 break, 2 breaks, etc.

3. We estimate the model using the real data, under the hypothesis that the series are I(1) or I(0), and that there is/are n break(s).

4. The breaks are estimated by the asymptotic sequential research, with three reestimations. The number of lag of the dependent variable that we take into account is determined by the number of augmentations in the ADF tests.

5. The parameters are estimated using OLS procedure. In the case of cointegration, the estimators are "superconsistent". The number of lags is determined by the Ljung-Box and Box-Pierce tests (with 1 to 8 lags) and the BIC (the results are quite similar with both the methods).

We must specify the distribution of the error terms e_t, that are determined using the sample distribution of the residuals and corrected in the same way that nonparametric bootstrap. The residuals are eventually modified for emphasising the breaks (see details in section 5). For the starting points, since they are part of model parameters in the I(1) case, we put them equal to the values of the corresponding observations of the real series. In the I(0) case, we try to compute the marginal law (see section 5).

Note that since the number of breaks is imposed, the estimation of the model is not necessarily correct, but the goal of this estimation is to obtain a DGP following the hypothesis that we impose and the nearest to the real data.

B Bootstrapped ADF tests

The following variables have to be defined:

- B, the number of bootstrap replications,
- p, the number of augmentations in the ADF regressions.

B has to be chosen as large as possible, depending on the characteristics of the computer. The choice of p is more difficult. We recall that the ADF regressions are:

$$\begin{array}{llll}\Delta y_t = & & & \alpha y_{t-1} + \beta_1 \Delta y_{t-1} + \ldots + \beta_p \Delta y_{t-p} + e_t, \\ \Delta y_t = & \textit{constant} + & & \alpha y_{t-1} + \beta_1 \Delta y_{t-1} + \ldots + \beta_p \Delta y_{t-p} + e_t, \\ \Delta y_t = & \textit{constant} + & \textit{trend} + & \alpha y_{t-1} + \beta_1 \Delta y_{t-1} + \ldots + \beta_p \Delta y_{t-p} + e_t,\end{array}$$

where y_t is the time series, and e_t are the error terms, t goed from 1 to T. We propose the following procedure for chosing p: The procedure starts at $p = 0$. The residuals of each ADF regressions with p augmentations are tested for independence using both the Ljung-Box's and Box-Pierce's tests. The number of autocorrelation coefficients taken into account for the Ljung-Box's and Box-Pierce's tests go from 1 to 8. If the residuals are not independent, p is incremented by 1 until the residuals look independent.

The steps of the bootstrapped ADF test are the following:

1. The Student test statistics for α for each ADF regression are computed. Let the statistics be denoted t_α. At this step, the residuals can be kept to be tested for independence.

2. The boostrap procedure needs a DGP for generating simulated samples under the null. This DGP is determined by estimating the model under the null using the data and the OLS procedure.

3. The bootstrap loop starts now. The simulated error terms, denoted e_t^b, are generated for a sample. There are four ways for generating the simulated error terms:

   (a) Parametric bootstrap: The simulated error terms are drawn from the normal distribution
   $$e_t^b \sim N(0, s^2),$$
   where s is the standard error of the error terms estimated from the ADF regression using the data.

   (b) Basic nonparametric boostrap: The simulated error terms are drawn by ...

   (c) Nonparametric bootstrap with corrected degree of freedom: since $\mathrm{E}(\hat{e}_t^2) \neq \mathrm{E}(e_t^2)$ where $\hat{e}_t^2)$ are the residuals of the ADF regression, but $\mathrm{E}(\hat{e}_t^2) =$

   For our program, the parametric and the second nonparametric bootstrap are chosen.

4. The simulated time series under the null, denoted $(y_t^b)_t$, is generated recursively using both the following steps:

   (a) first, define Δy_t^b recursively:
   $$\begin{array}{rlll} \Delta y_t^b = & & & \hat{\beta}_1 \Delta y_{t-1}^b + \ldots + \hat{\beta}_p \Delta y_{t-p}^b + e_t^b, \\ \Delta y_t^b = & \widehat{constant} + & & \hat{\beta}_1 \Delta y_{t-1}^b + \ldots + \hat{\beta}_p \Delta y_{t-p}^b + e_t^b, \\ \Delta y_t^b = & \widehat{constant} + & \widehat{trend} + & \hat{\beta}_1 \Delta y_{t-1}^b + \ldots + \hat{\beta}_p \Delta y_{t-p}^b + e_t^b, \end{array}$$
   The p first values for y_t^b can be chosen equal to the the p first values of y_t, interpreted as initial conditions. (Another way is to choose them randomly.)

   (b) second, compute y_t^b:
   $$y_t^b = y_1 + \sum_{i=2}^{t} y_i^b.$$
   y_1 is an initial condition.

5. The Student test statistics for α for each ADF regression are computed using the simulated series $(y_t^b)_t$. Let the statistics be denoted t_α^b.

6. The steps 3–5 are done again B times. A set of statistics t_α^b, $b = 1, \ldots, B$, is then obtained for each the three ADF regressions, and for each both the parametric and nonparametric bootstraps (thus there are six statistics).

7. The bootstrap P value is finally computed depending on the test hypothesis:

   (a) If the null hypothesis $H_0 : \alpha = 0$ is tested against the alternative hypothesis $H_1 : \alpha < 0$, the classical P value is

   $$p_{\text{uni}} = \frac{1}{B} \sum_{b=1}^{B} I(t_\alpha^b \leq t_\alpha),$$

   where I is the indicator function. This P value corresponds to an unilateral test.

   (b) If the null hypothesis $H_0 : \alpha = 0$ is tested against the alternative hypothesis $H_1 : \alpha \neq 0$, the classical bootstrap P value is

   $$p_{\text{bil}_{\text{sym}}} = \frac{1}{B} \sum_{b=1}^{B} I(|t_\alpha^b| \geq |t_\alpha|).$$

   This P value corresponds to a bilateral test.

   (c) In the case where the null hypothesis $H_0 : \alpha = 0$ is tested against the alternative hypothesis $H_1 : \alpha \neq 0$, we also propose the following bootstrap P value:

   $$p_{\text{bil}_{\text{asym}}} = 2 \min\{p_{\text{uni}}, 1 - p_{\text{uni}}\}.$$

   This P value also corresponds to a bilateral test, but it takes into account the asymmetry of the statistic distribution in addition. This P value can be found in Davidson and MacKinnon (1993), chapter 5, in the context of confidence regions.

   In our program, the two last P values are used.

8. Finally, a significance level is chosen and compared to the P values. If a P value is lower to the significance level, H_1 is retained, otherwise H_0 is retained.