Detecting Multiple Structural Breaks: Dummy Saturation vs Sequential Bootstrapping. With an Application to the Fisher Effect for US*

Michele Bergamelli†
(Cass Business School, UK)

Giovanni Urga ‡
(Cass Business School, UK and Bergamo University, Italy)

6th November 2013

Abstract

In this paper, we propose two novel approaches to detect multiple structural breaks affecting the deterministic component of a linear system. The first is an extension of the so called *Dummy Saturation* developed by Hendry et al. (2008), Johansen and Nielsen (2009), Ericsson (2011) and Castle et al. (2012) and it involves impulse, step and steptrend dummies. The second approach considers a *Sequential Bootstrapping* procedure based on the sup-F statistic of Andrews (1993). Through an extensive Monte Carlo exercise, we evaluate the performance of both procedures under several DGPs including stationary, nonstationary and cointegrated models. We explore the ability of the two approaches to detect the correct number of breaks and break dates. Finally, we apply the two procedures to investigate the stability of the Fisher relationship in the US over the period 1985:1-2012:3.

Keywords: Dummy Saturation, Sequential Bootstrap, Structural Breaks, Monte Carlo, Autometrics.
JEL Classifiction: C52, C22, C15.

*We wish to thank partecipants in the 12th OxMetrics User Conference (Cass, 3-4 September 2012), in particular Bent Nielsen and Jurgen Doornik, for useful comments and helpful suggestions. Special thanks to David Hendry and Felix Pretis for useful comments on a preliminary version of the paper, and Lynda Khalaf for very insighful discussions and suggestions. The usual disclaimer applies. This paper draws from Michele Bergamelli PhD research project sponsored by the "PhD Scholarship in Memory of Ana Timberlake". We acknowledge financial support from Timberlake Consultants and the Centre for Econometric Analysis.

†Centre for Econometric Analysis, Cass Business School, City University London, 106 Bunhill Row EC1Y 8TZ London (UK). email: Michele.Bergamelli.2@cass.city.ac.uk

‡Centre for Econometric Analysis, Cass Business School, City University London, 106 Bunhill Row, EC1Y 8TZ, London (UK) and University of Bergamo (Italy). email: g.urga@city.ac.uk.

1 Introduction

Since the seminal contributions by Perron (1989) and Rappoport and Reichlin (1989), the literature has produced a comprehensive set of results on the break-point problem in a time series framework. Useful surveys are Stock (1994), Banerjee and Urga (2005), Perron (2006) and Aue and Horváth (2013), while Andreou and Ghysels (2009) consider structural breaks involving not only the mean process but also higher order moments as well as changes affecting the whole distribution of financial time series.

Focusing on the problem of estimating and testing for the presence of structural breaks, Andrews (1993) derives the asymptotic distribution of a class of "sup" type statistics based on the Quandt (1960) statistic to test for the null of stability against the alternative of an unknown structural break. For the multiple unknown structural breaks case, Bai and Perron (1998, 2003) propose to estimate break dates by a minimum least squares approach and to test the significance of the resulting break date estimates by means of three different tests all based on the sup-F statistic derived in Andrews (1993). In a cointegrated framework, Hansen (1992) is the first to consider tests for intercept and slope stability. In the context of cointegrated VAR processes, Seo (1998) provides testing framework for an unknown single break case while Hansen (2003) provides a modelling framework to analyse multiple structural changes though occurring at known dates. For the unknown multiple breaks case, the large sample theory of the Bai and Perron (1998, 2003) framework has been recently generalized to the case of nonstationary models by Kejriwal and Perron (2008, 2010). However, despite the sounding theoretical framework, the Kejriwal and Perron (2008, 2010) procedure presents a series of critical practical limitations. First, the asymptotic distributions of the tests to asses the statistical significance of the break dates are not pivotal, thus requiring a derivation of the critical values case by case via numerical simulation. Second, the three complementary tests (sup-F, UDMax and Sequential sup-F) proposed often provide contradictory results about the number of breaks to include in the model, making the inference difficult.

In this paper, we consider two novel approaches to detect multiple structural breaks in a wide range of linear models including the cointegration set up. The first procedure is an extension of the so called *Dummy Saturation* (DS) developed by Hendry et al. (2008), Johansen and Nielsen (2009), Ericsson (2011) and Castle et al. (2012) that involves not only impulse dummies but also step and steptrend dummies as originally mentioned in Banerjee et al. (1998), De Peretti and Urga (2005) and Ericsson (2011). Generally, the DS consists in detecting structural breaks affecting a linear system by saturating a regression with dummies and then removing the non-significant ones through a *general-to-specific* approach. The second methodology consists in a *Sequential Bootstrapping* (SB) procedure based on the sup-F statistic of Andrews (1993). We evaluate the performance of both the extended DS and SB procedures under several DGPs including stationary, nonstationary and cointegrated

models. In particular, we assess the two procedures according to the following two criteria: first, the ability to pick up the correct number of structural breaks, and second, the ability to correctly locate them. To this extent, we set up an extensive Monte Carlo exercise in order to compute the empirical retention frequencies of the DS dummies and the empirical rejection frequencies of the SB sup-F tests as well as to measure the goodness of the resulting break date estimates. More precisely, evaluation criteria include empirical size and power as well as the *gauge* and *potency* criteria introduced by Castle et al. (2012); the latter assess the considered method from the model selection perspective rather than as formal statistical break tests. For the DS method, evaluation criteria yield mixed results confirming the model selection nature of the procedure. In contrast, the empirical size and power of the SB suggest that the method hold clear and definitive promise. Finally, we carry out an empirical application involving an analysis of the stability of the Fisher equation in the United States in order to study the performance of the two novel procedures with a real dataset.

The remainder of the paper is organised as follows. Section 2 presents the extended DS and the SB procedures. Section 3 reports the criteria used to assess the Monte Carlo exercise, while the results of the simulations are reported and discussed in Section 4, where we also offer some useful guidelines about the implementation of the two procedures. Section 5 reports the empirical application involving the analysis of the Fisher relationship in the United States. Section 6 concludes

2 Two Novel Approaches to Detect Multiple Structural Breaks

In this section, we describe the DS and the SB procedures.

2.1 Dummy Saturation (DS)

The DS approach, as originally proposed in Hendry et al. (2008), Johansen and Nielsen (2009), Ericsson (2011) and Castle et al. (2012), is a technique to test model constancy by means of a set of dummy variables. The underlying idea is to saturate a linear model involving T observations with T dummy variables (one for each observations) to capture outliers and structural breaks. Following the "*general-to-specific*" approach, the technique starts with an initial model where an outlier/break may happen at all times and then removes the statistically insignificant dummies. The framework is very general allowing to test for the presence of multiple structural breaks in a wide range of systems.

In particular, the original approach as outlined in Hendry and Santos (2005) and Hendry et al. (2008) involves a saturation with 0-1 impulse dummies (*impulse indicator saturation*, IIS). Recently, Doornik et al. (2013) develop the theoretical properties of the step indicator saturation, where impulse dummies are replaced by partial sums of impulse dummies (step dummies). In

this paper, we propose to extend the original IIS approach to include also step dummies and double partial sums of impulse dummies (sums of step dummies or simply steptrend dummies) to capture structural breaks in the level or in the trend. Following the terminology in Ericsson (2011), we define *super dummy saturation* (SDS) the version with also step dummies, and *super-duper dummy saturation* (SDDS) the regression saturation involving also steptrend dummies.

Consider a standard regression with a constant and a linear trend

$$y_t = \alpha + \delta t + \boldsymbol{\beta}^\top \mathbf{z}_t + e_t \tag{1}$$

where $\mathbf{z}_t$ is k-dimensional vector containing exogenous regressors as well as lagged values of y_t and assume we are interested in testing the stability of the deterministic component, i.e. $\alpha + \delta t$. Then, applying the principle of the DS, we saturate (1) with dummies and we start the selection of the parsimonious representation from one of the following models

$$\text{DS} \begin{cases} \text{IIS:} & y_t = \alpha + \delta t + \boldsymbol{\beta}^\top \mathbf{z}_t + \sum_{i=1}^{T} \gamma_i B^I_{i,t} + e_t \\ \text{SDS:} & y_t = \alpha + \delta t + \boldsymbol{\beta}^\top \mathbf{z}_t + \sum_{i=1}^{T} (\gamma_i B^I_{i,t} + \psi_i B^L_{i,t}) + e_t \\ \text{SDDS:} & y_t = \alpha + \delta t + \boldsymbol{\beta}^\top \mathbf{z}_t + \sum_{i=1}^{T} (\gamma_i B^I_{i,t} + \psi_i B^L_{i,t} + \omega_i B^T_{i,t}) + e_t \end{cases}$$

where $B^I_{i,t} = \mathbf{I}_{\{t=i\}}$, $B^L_{i,t} = \mathbf{I}_{\{t \geq i\}}$ and $B^T_{i,t} = (t - i + 1)\mathbf{I}_{\{t \geq i\}}$ for $i = 1, \ldots, T$. After dropping the statistically insignificant dummies from the saturated regression, one is left with a set of dummies which can be interpreted as outliers (impulse dummies) and/or structural breaks affecting the deterministic component of the process (step and steptrend dummies)[1].

There are two problems with the specifications just introduced. For all cases, though especially in SDS and SDDS, there is multicollinearity between some of the dummies themselves and between the dummies and the deterministic component of the process. Second, the estimation of the saturated regressions is infeasible because of lack of degrees of freedom, given that N, the total number of regressors, is larger than the number of observations T. In principle, the multicollinearity problem within dummies can be solved quite easily by excluding some dummies (last step or last steptrend for instance). Typically, one sets $i = \ell_y + 1, \ldots, T - 1$, where ℓ_y is the highest order of lagged dependent variables entering the process, and excludes the first step dummy, which is exactly collinear with the set of impulse dummies, and the

[1] It is possible to use the IIS to capture breaks in the level component. Retained consecutive dummies with same sign and similar magnitude may be grouped according and can be interpreted as capturing a break in the level.

last steptrend dummy. As far as the dimensionality problem is concerned, a possible solution is to split the set of dummies in J blocks such that the number of dummies in each block (N_j) plus the number of elements in the deterministic components, the number of lagged dependent variables and exogenous regressors (k) is less than the sample size ($N_j + 2 + k < T$, for $j = 1, \ldots, J$), following the strategy introduced in Hendry and Santos (2005). More explicitly, in the general case of SDDS, assume to form J blocks of about the same size[2] such that $\mathcal{I}_1 = \{B_i^I, B_i^L, B_i^T : i = 1, \ldots, \lceil T/J \rceil\}$, $\mathcal{I}_2 = \{B_i^I, B_i^L, B_i^T : i = \lceil T/J \rceil + 1, \ldots, \lceil 2T/J \rceil\}, \ldots,$ $\mathcal{I}_J = \{B_i^I, B_i^L, B_i^T : i = \lceil T(J-1)/J \rceil + 1, \ldots, T\}$. The procedure then runs as follows:

- for $j = 1, \ldots, J$ include the $\mathcal{I}_j$ subset of dummies in the equation of interest (e.g. (1)) and estimate the partially saturated regression recording the significant dummies from each regression.
- Combine all relevant dummies from the previous iterations and re-estimate the model, assuming that the total number of the retained dummies from each subset $\mathcal{I}_j$ is less than the sample size.
- Retain the significant dummies.

This is the standard way to carry out a structural break analysis using the DS approach. As explained in Castle et al. (2012), under the null of no outliers or breaks, αT impulse indicators are retained on average (α being the level of significance). For this reason, if we fix $\alpha \leq r/T$ we control the false null retention at r dummies. This is rather satisfactory if we think that we are testing the potential relevance of number of dummies function of T and even if no breaks occur we are not losing in efficiency. This point is investigated via the Monte Carlo exercise and we show that the procedure has the correct size only for some specific choice of α.

An alternative and more convenient route to implement the DS is through the algorithm for automated model selection *Autometrics* (see Doornik, 2009a). This is accessible through the software OxMetrics© which can handle $N > T$ as well as non-orthogonal candidate regressors. Using *Autometrics* , it is possible to specify a *general unrestricted model* (GUM), in this case a regression saturated with dummies, and a tree search algorithm is able to eliminate statistically insignificant regressors. The entire procedure of block creation is carried out by the algorithm as well as the management of non-orthogonal regressors. An important aspect is that *Autometrics* provides different ways to create blocks in case of $N > T$ (i.e. sequential blocks, random blocks, cross blocks, etc.). Despite in theory, different ways of selecting the candidate regressors which enter in each block should not affect the outcome radically, Doornik

[2] For the SDDS case, we set $J > \lceil \frac{3T}{T-2-k} \rceil$ in order to have enough degrees of freedom.

(2009b) reports that the ordering of the variables as they enter the GUM may influence the final results. We investigate also this issue through simulations in the Monte Carlo exercise.

Finally, a theoretical investigation of the properties of the IIS framework only is developed by Hendry et al. (2008) and generalized under less restrictive conditions in Johansen and Nielsen (2009). On the empirical side instead, Castle et al. (2012) reports a comprehensive simulation study of the IIS approach to detect outliers and level shifts in several specifications including deterministic trends, unit roots, autoregressive processes as well as autoregressions with exogenous regressors. Ericsson (2011) also report a small empirical application of the SDS only. In this paper, we extend the application of the DS approach to detect breaks in the deterministic trend (SDDS) as well as in a cointegrated set up. Similarly, the SDS and the SDDS size and power properties are still unexplored. In the section dedicated to the Monte Carlo analysis we try to fill this gap.

2.2 Sequential Bootstrapping (SB) Procedure

The second methodology we consider is the SB approach. This method is specifically developed to estimate multiple structural breaks in systems with a conditional process and several marginal (the regresssors) processes of the form:

$$y_t = \alpha_0 + \delta_0 t + \sum_{\ell_y=1}^{q} \rho_{y,\ell_y} y_{t-\ell_y} + \sum_{i=1}^{K} (\psi_i B_{i,t}^L + \omega_i B_{i,t}^T) + \\ + \sum_{n=1}^{N} (\tilde{\rho}_n x_{n,t-1} + \sum_{j=1}^{K_n} (\tilde{\upsilon}_{n,j} B_{n,j,t}^L + \tilde{\phi}_{n,j} B_{n,j,t}^T)) + u_t \quad (2)$$

$$x_{n,t} = \gamma_{n,0} + \zeta_{n,0} t + \sum_{\ell=1}^{p_n} \rho_{x,\ell_x} x_{n,t-\ell_x} + \sum_{j=1}^{K_n} (\upsilon_{n,j} B_{n,j,t}^L + \phi_{n,j} B_{n,j,t}^T) + e_{n,t} \quad (3)$$

where (2) is the conditional process, (3) is a set of $n = 1, \ldots, N$ marginal process experiencing their own independent breaks, $B_{i,t}^L = \mathbf{I}_{\{t \geq T_i\}}$ and $B_{i,t}^T = (t - T_i + 1)\mathbf{I}_{\{t \geq T_i\}}$ indicate the dummies capturing breaks in the level and the trend respectively with T_i, $i = 1, \ldots, K$ with the i^{th} break date of the conditional process; $B_{n,j,t}^L$ and $B_{n,j,t}^T$ indicate dummies capturing breaks in the level and the trend respectively in the marginal process with T_j, $j = 1, \ldots, K_n$ not necessarily equal to T_i.

In a first instance, the algorithm estimates breaks in the marginal processes then these enter into the conditional process before the search for breaks in the conditional process starts. The estimation of the breaks is sequential in that break dates are added one by one until a stopping criterion is reached[3]. Thus, contrary to the DS approach, here the philosophy is

[3]This approach was originally proposed by Banerjee et al. (1998) and De Peretti and Urga (2005).

from *"specific-to-genera*. Both stationary and nonstationary processes can be handled.

Conceptually, the sequential procedure has two building blocks: the first deals with the estimation of the break dates, whereas the second with the statistical testing framework required for stopping the sequential search. The next two sections summarise the main features of the SB procedure.

2.2.1 Structural Breaks Estimation

Consider the case of a marginal process as in (3) and assume we are interested in estimating possible breaks affecting this process at unknown times. The procedure starts by searching for break which may occur at each date $t = 3, \ldots, T-1$ by means of a Wald type test. Thus, for each t, compute the F-statistic, F_t, for testing the null hypothesis of no more breaks at $T_{j=t}$, $H_0 : \upsilon_j = \phi_j = 0 | T_j = t$. When looking for the second break and so on for the others apart from the first one, the statistics is not computed for the already estimated break dates (more precisely for a small neighbourhood around them). In particular, F_t takes the usual form:

$$F_t = \frac{T-k-2(j+1)}{2} \frac{RSS_{T_{j-1}} - RSS_{T_j=t}}{RSS_{T_j=t}} \tag{4}$$

where $RSS_{T_{j-1}}$ is the residual sum of squares for the restricted model (imposing $\upsilon_j = \phi_j = 0$) while $RSS_{T_{j=t}}$ is the residual sum of squares of the unrestricted model (with the additional j-th break occurring at $T_{j=t}$, $j = 1, .., m$ the number of breaks). The degrees of freedom are given by $k + 2(j+1)$, the number of parameters including those corresponding to the coefficients associated to the step and steptrend dummies ($2(j+1)$) in the unrestricted model, and $2 = 2(j+1) - 2j$, the number of additional parameters resulting from adding one break date.

The estimator $\hat{T}_j$ of the break date T_j is given by

$$\hat{T}_j = \text{argmax}\{F_t\} \quad \text{for all admissable } t \tag{5}$$

This amounts to define the break date estimator as the argument which maximises the usual sup-F statistic introduced initially by Quandt (1960) [4] with non-standard asymptotic distribution derived in Andrews (1993) for the stationary case. Bai (1994) proved that the estimation of the break date obtained through the minimisation of the least squares in a linear model leads to consistent estimates, i.e. $\hat{T}_j = T_j + o_p(1)$. It is evident from (4) that

$$\hat{T}_j = \text{argmax}\{F_t\} \equiv \text{argmin}\{RSS_{T_j=t}\}$$

hence the proposed estimator is consistent.

[4] Quandt (1960) considers the supremum of a set of LR statistics which boils down to the sup-F test in case of i.i.d. and normally distributed errors.

Finally, as noted in De Peretti and Urga (2005), the sequential estimation of structural breaks is biased when there is more than one break. In other words, the dating of the first break is biased when a second break is neglected and so on for the following ones. As a solution, the authors suggest to implement a backward revision of the break dates where after estimating the j^{th} break, the previous $j-1$ breaks are re-estimated. Alternatively, after a break date is found to be significant, one could also re-estimate all the breaks simultaneously but this second option is computationally expensive. Additionally, the re-estimation procedure allows also to check the stability of the break dates identified.

2.2.2 Stopping Rule for the Sequential Search

Once a break date T_j is estimated, it is important to test whether it is statistically significant. This translates in testing for the null hypothesis of $j-1$ breaks against the alternative of j breaks using an F-test. Nevertheless, here the problem is that the conventional critical values of the F-distribution cannot be employed since the break is endogenously determined (see Andrews, 1993) and this implies non standard asymptotic distributions depending on the break fraction. Andrews (1993) derived the correct distribution when the break is endogenous for the "sup" versions of the Wald, LR and LM tests but only when the series is stationary and moreover, as it has been showed in Hansen (2000), this is also incorrect when the regressors experience their own breaks. The proposed solution is to bootstrap the critical values of the test statistics:

$$\hat{\tau} = \max\{F_t\}. \tag{6}$$

The bootstrap procedure to test for the significance of the j^{th} break runs as follows:

1. Estimate the regression under the null of $j-1$ breaks, i.e. (for a marginal model)

$$x_t = \gamma_0 + \zeta_0 t + \sum_{\ell=1}^{p} \rho_{x,\ell_x} x_{t-\ell_x} + \sum_{k=1}^{j-1} (\upsilon_k B_{k,t}^L + \phi_k B_{k,t}^T) + e_t \tag{7}$$

   and store the residuals series $\{\hat{e}_t\}_{t=1}^T$ together with the coefficients estimates.

2. For $b = 1, \ldots, B$ repeat

   (a) Draw with replacement T values from the centred residuals

$$\left\{ \hat{e}_t - \frac{1}{T} \sum_{t=1}^{T} \hat{e}_t \right\}_{t=1}^{T}$$

   to get $\{e_{t,b}^*\}_{t=1}^T$ (semi-parametric approach). Alternatively, compute the sample variance $\hat{\sigma}_c^2$ of the centred residuals and draw $\{e_{t,b}^*\}_{t=1}^T$ from $\mathcal{N}(0, \hat{\sigma}_c^2)$ (parametric

approach).

(b) Build recursively the bootstrapped counterpart of x_t denoted by $x^*_{b,t}$ by

$$x^*_{b,t} = \hat{\gamma}_0 + \hat{\zeta}_0 t + \sum_{\ell=1}^{p} \hat{\rho}_{x,\ell_x} x^*_{b,t-\ell_x} + \sum_{k=1}^{j-1} (\hat{\upsilon}_k B^L_{k,t} + \hat{\phi}_k B^T_{k,t}) + e^*_{b,t}. \tag{8}$$

(c) Compute the bootstrapped counterpart of $\hat{\tau}$, namely, $\hat{\tau}^*_b = \max\{F^*_{b,t}\}$ where $F^*_{b,t}$ is computed as in (4) but using the bootstrapped sample $x^*_{b,t}$.

3. Decide on the significance of $\hat{T}_j$ by computing the bootstrapped p-value

$$\frac{1}{B} \sum_{b=1}^{B} \mathbf{I}_{\{(\hat{\tau})^2 < (\tau^*_b)^2\}}. \tag{9}$$

The sequential search stops when two subsequent structural breaks are not significant. De Peretti and Urga (2005) show that this stopping rule is optimal and robust to biases of additional significant breaks being neglected.

3 Monte Carlo Design

We investigate the performances of the extended DS and SB procedures undertaking an extensive Monte Carlo simulation exercise. The purpose of the analysis is to assess the ability of the two procedures to detect the correct number of breaks as well as the precision of the break dates. In what follows, we introduce the criteria used to evaluate the DS and the SB procedures and the several DGPs used in the Monte Carlo experiments.

3.1 Evaluating the Detection of the Correct Number of Breaks

A structural breaks analysis is valid if the adopted procedure is able to select the correct number of structural breaks.

In general given a DGP affected by K^* breaks, the power and the size are approximated by computing the empirical rejection frequency with respect to a null hypothesis stating that the number of breaks K is $H_0 : K = K_0$ versus an alternative $H_1 : K = K_1 > K_0$. If the null hypothesis is such that $K_0 = K^*$, the empirical rejection frequency of the null hypothesis gives an approximation of the size of the test on which each methodology is built, otherwise if the number of breaks considered under the null is different from the true number of breaks ($K_0 \neq K^*$) we are approximating its power. The number of times that a procedure detects the correct number of breaks is then given by one minus the empirical size, i.e. the number of times that we do not reject a true hypothesis about the number of breaks.

As far as the DS is concerned, however, the computation of the empirical rejection frequency it is not always straightforward. This is true in particular to the case of the IIS. More specifically, it is difficult in the case of the IIS to test for instance the hypothesis of the form $H_0 : K = K_0$ versus $H_1 : K = K_1$ level breaks given the nature of the impulse dummies. When instead we saturate the model with step and/or steptrend dummies, it is easier to cast the procedure in terms of testing $H_0 : K = K_0$, given that a step or a steptrend dummy is directly related to a break in the level or the trend component. Alternatively, we may assess the performance of the procedure considering how frequently a dummy entering the saturated regression is retained across the Monte Carlo replications, thus treating the DS as a model selection procedure rather than a test for breaks.

Following Castle et al. (2012), we now introduce the concept of *retention rate*. Assume to specify a GUM saturated with N dummies of which only n, $n < N$, enter in the DGP and denote with M the number of Monte Carlo simulations. The *retention rate*, r, for each dummy can be defined as

$$\hat{r}_j = \frac{1}{M} \sum_{m=1}^{M} \mathbf{I}_{\{\hat{\beta}_{j,m} \neq 0\}}, \quad j = 1, \ldots, N \tag{10}$$

where $\mathbf{I}_{\{.\}}$ is the indicator function and $\hat{\beta}_{j,m}$ is the coefficient of the j^{th} dummy computed at the m^{th} iteration. If the dummy is significant ($\hat{\beta}_{j,m} \neq 0$), then the indicator function equals unity. Given (10), it is possible to define the *gauge* and the *potency* the procedure as

$$\text{gauge} = \frac{1}{N-n} \sum_{j=n+1}^{N} \hat{r}_j$$

$$\text{potency} = \frac{1}{n} \sum_{j=1}^{n} \hat{r}_j$$

where the gauge is the average retention rate of the non-significant dummies (according to the DGP) and the potency is the average retention rate of the significant dummies[5]. Note that though gauge and potency are related to the concepts of size and power, however they are distinct: while size and power refer to a well specified null hypothesis in terms of K_0 versus K_1 breaks, the gauge and the potency are useful criteria for the assessment of the procedure but are not clearly linked to any statistical hypothesis about the number of the breaks. Finally, note that according to the definitions above, a dummy contributes to the potency if it enters the DGP. This is straightforward when saturating with impulse dummies. On the other side, if saturating the GUM with step and/or steptrend dummies (SDS and

[5]For simulation purposes, especially when using the IIS variant, it is more convenient to work with two aggregate retention rates: one for the dummies that contribute to the gauge and one for those which contribute to the potency. The gauge is then obtained as $\frac{1}{(N-n)M} \sum_{m=1}^{M} \sum_{j=n+1}^{N} \mathbf{I}_{\{\hat{\beta}_{j,m} \neq 0\}}$ and similarly the potency.

SDDS models respectively), we have to be less restrictive in that we have to define an interval around the break date such that a step and/or steptrend dummies falling in this interval contributes to the potency. Assume for instance that a break occurs at $T_1 = 50$, it would be unreasonable to count $B^L_{52,t} = \mathbf{I}_{\{t \geq 52\}}$ as contributing to the gauge of the SDS. For this reason, when computing the gauge and the potency of the SDS or SDDS, we consider a dummy as contributing to the potency if it captures a break occurring at $T_i \pm 3$ observations. If more then one dummy falling in that interval is retained, we choose the one closest to the break date, while in the unlikely case that two dummies are equally-distant from T_i we count randomly one as contributing to the potency and one contributing to the gauge.

Moreover, in the case we saturate with step dummies and also with steptrend dummies, in addition to the gauge and potency, it is possible to compute the more common empirical rejection frequency, assessing the procedure as a statistical test. As mentioned above, when working with other dummies than impulse dummies, it is easier to see the retention of a step or steptrend dummy as an acceptance/rejection of a null hypothesis of the form $H_0 : K = K_0$ against an alternative $H_1 : K = K_1 \neq K_0$ where K_1 is a positive integer different from K_0. To decide about the acceptance/rejection of the null, it is possible to count how many step dummies are retained. Consequently, the number of times that the procedure detects the correct number of structural breaks is given by how frequently the dimension of the vector containing the step dummies equals the number of step dummies entering the DGP, i.e. the number of times we do not reject a true hypothesis about K. Note that when evaluating the procedure using the gauge and the potency these criteria already assess the detection of the correct number of breaks.

As far as the SB is concerned, we obtain estimates of the break dates and their associated p-values (parametric and non-parametric) for each iteration $m = 1, \ldots, M$. In this case, it is straightforward to compute the empirical rejection frequency. The computation of the size and the power of the procedure is based on the p-values of the bootstrapped sup-F test statistic. Under the null hypothesis of no significance of the break, a p-value approaching zero means that the break date under consideration is significant. Given the sequential nature of the procedure, it makes sense to compute the size and power counting how frequently we reject a null hypothesis of the form $H_0 : K = K_0$ versus $H_1 : K = K_1 = K_0 + 1$. In terms of the p-values of the sup-F statistic and given a significance level α, we then compute

$$\frac{1}{M} \sum_{m=1}^{M} \mathbf{I}_{\{p\text{-value}_{m,K1} < \alpha\}} \tag{11}$$

where $p\text{-value}_{m,K_1}$ represents the p-value of the K_1^{th} break date computed according to (9) at the m^{th} iteration. If the null is true ($K_0 = K^*$), (11) gives the empirical size of the procedure since it computes the number of significant breaks that do not enter the DGP. Then, for the

procedure to have the correct size, we expect that the number of times that the K_1^{th} break date is significant (its p-value is small) and of course should not exceed α. On the contrary, if the null is false, (11) represents the empirical power. To summarise, assume that the DGP is formulated with $K^* = 1$, then testing $H_0 : K = 0$ versus $H_1 : K = 1$ gives the power of the procedure while a test of $H_0 : K = 1$ versus $H_1 : K = 2$ gives the size. So, we have

$$\text{size} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{I}_{\{p\text{-value}_{m,T_2}<\alpha\}}$$
$$\text{power} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{I}_{\{p\text{-value}_{m,T_1}<\alpha\}}.$$

In practise, we fix the maximum number of breaks to be estimated in each experiment according to the number of breaks under the alternative hypothesis of the last sequential test for which the null is true.

3.2 Evaluating the Dating of Breaks

In addition to the desiderable property of selecting the correct number of structural breaks, a structural breaks procedure has to able to pick up the correct timing of when the breaks occur. This translates in assessing the goodness of the resulting break date estimates.

Let $\hat{T}_i$ be the estimated counterpart of the i^{th} break date T_i. After running a Monte Carlo experiment, we obtain a set of estimators $\hat{T}_{i,m}$ for $m = 1, \ldots, M$. The Monte Carlo estimator is then defined as the sample mean of $\{\hat{T}_{i,m}\}_{m=1}^{M}$, taken with respect to M. In order to assess the goodness of this estimator, we compute its sample bias and sample root mean squared error (RMSE) defined as:

$$\widehat{\text{bias}}[\hat{T}_i] = \mathbb{E}_M[\hat{T}_i - T_i] = \frac{1}{M}\sum_{m=1}^{M}(\hat{T}_{i,m} - T_i)$$
$$\widehat{\text{RMSE}}[\hat{T}_i] = \sqrt{\mathbb{E}_M[(\hat{T}_i - T_i)^2]} = \frac{1}{M-1}\sqrt{\sum_{m=1}^{M}(\hat{T}_{i,m} - \mathbb{E}_M[\hat{T}_i])^2}.$$

In the case of multiple structural breaks, we compute the sample bias with respect to the closest estimated break date. In what follows, we discuss some important issues related either to the specific procedure or to the shape of the distribution of the break date estimator.

For sake of simplicity, assume we have a process with one break at T_1. The first issue concerns with the DS procedure. When working with the IIS, we need to define a rule such that at each iteration of a Monte Carlo experiment we can consider a spike dummy ($B_{i,t}^I = \mathbf{I}_{\{i=t\}}$) as the reference (i.e. $\hat{T}_{m,1}$) to compute the bias. In this case, we use the

dummy belonging to the relevant set of dummies which is closest to the break date[6]. If there are multiple breaks, first we have to group the spike dummies according to their magnitude and then select the first one from each set. To reduce the complexity of our experiments, we consider the case of two breaks with the same magnitude but opposite sign. On the other hand, when working with either the SDS or the SDDS, we do not have to take into account this point since there is no more a set of dummies to be associated with a single level break but only one step or steptrend dummy. However, it may happen that in the final model selected by *Autometrics* there are more step or steptrend dummies falling in the relevant interval around T_i (gauge and potency) or more dummies than those in the DGP (size and power). In both cases, in order to compute the bias, we consider as $\hat{T}_{m,1}$ the index of the closest step dummy to the true break point. The bias and the RMSE of the estimates related to the computation of the gauge and the potency are expected to be smaller than those related to the size and the power. The main reason is that in computing the gauge and the potency we are constraining the location of the breaks to the dummies falling into $T_i \pm 3$. As far as the SB is concerned, there is no need to define such a rule since at each iteration we have only one estimate for a given break date.

The second issue is valid to both procedures and relates to the skewness of the break dates estimators. If the distribution is asymmetric, the estimator computed as the sample mean is of course biased, and thus a more appropriate criterion to judge its goodness is to consider the median of estimator's distribution. Thus, we report both the sample mean and the median.

4 Simulation Results

In this section, we present the results from our simulations. In each experiment, the number of simulations is set to $M = 1,000$ and the sample size to $T = 100$. We fix the seed of the random number generator and we build recursively $T + n$ (where n depends on the characteristics of the process) observations M times, starting from $y_{-n} = 0$ and then discarding the first n in order to create independence from the initial conditions.

4.1 Breaks in Level

We start by considering several univariate (marginal models) and bivariate (conditional models) DGPs affected only by breaks in the mean component. The general model can formulated

[6] If there is a level break occurring at $T_1 = 90$, we consider $\hat{T}_{m,1}$ being the index i of the first retained dummy belonging to $\mathcal{I}_m = \{B^I_{i,t} : \ i = 91, \dots T\}$.

as

$$y_t = \alpha + \delta t + \beta x_t + \sum_{i=1}^{K} \psi_i B_{i,t}^L + u_t \tag{12}$$

$$x_t = \gamma + \zeta t + \rho x_{t-1} + \sum_{j=1}^{K_x} \upsilon_j B_{j,t}^L + e_t \qquad t = 1, \ldots, T \tag{13}$$

where $B_{i,t}^L \equiv \mathbf{I}_{\{t>T_i\}}$ denotes a step dummy (the same applies to $B_{j,t}^L$). Moving from this simple benchmark to the alternative more general model including the breaks in trend allows us to understand the impact of the different ways to create blocks by *Autometrics* . It is reasonable to expect that, if different mechanisms to create blocks have some impact on the selection of the final model, this becomes even more relevant as we move from the simplest IIS to the SDDS.

Two parameters govern the detectability of the breaks: the magnitude of the dummy coefficients ψ_i and υ_j and the variance of the error terms. In our simulations, we allow the coefficients of the dummies to take different magnitude while we fix the variance of the error term drawing from a standard normal $N(0, 1)$.

For the DS, all the GUMs correspond to the DGPs saturated with dummies and we fix all the variables that enter the DGP except the dummies. This entails that the rejection frequencies are computed only with respect to the dummies (i.e. we do not test the ability of *Autometrics* to reject non-significant regressors apart from dummies). On the other hand, the SB size and power properties are assessed by computing both parametric and non-parametric p-values through 99 bootstrap replications. The nominal significance levels used in the simulations are 1% and 5%. All DGPs are investigated assuming that the number of breaks is $K = \{0, 1, 2\}$. When conditional models with broken marginals are considered, the number of breaks in the marginal model is fixed to one and the inference focuses then on the number of breaks in the conditional model. The break dates and the coefficients of the related dummies (ψ_i and υ_j) are reported in the tables summarizing the results of the simulations. For the IIS, we restrict the experiments to the empirical retention frequency of the dummies entering the DGP, i.e. the gauge and the potency. For the SDS, we assess also the size and power by computing the empirical rejection frequencies of the following hypothesis: $H_0 : K = 0$ for the DGP with no breaks (size only); $H_0 : K = 0$ (power) and $H_0 : K = 1$ (size) for the DGP with one breaks; and $H_0 : K = 1$ (power) and $H_0 : K = 2$ (size) for the DGP with two breaks. As far as the SB is concerned, we set the maximum number of breaks, to be sequentially estimated in each experiment, equal to the true number of breaks plus one. Specifically, we have that for the case with no breaks a test of $H_0 : K = 0$ vs $H_1 : K = 1$ (size only); for the case with one break a test of $H_0 : K = 0$ vs $H_1 : K = 1$ (power), and $H_0 : K = 1$ vs $H_1 : K = 2$ (size); and for the case with 2 breaks a test of $H_0 : K = 1$ vs $H_1 : K = 2$ (power),

and $H_0 : K = 2$ vs $H_1 : K = 3$ (size).

4.1.1 Marginal Processes

The first set of experiments involves modelling the exogenous regressors one at the time. These include location-scale models, stationary or non-stationary autoregressive processes and autoregressive processes with or without trends which are all nested in (13). Table 1 reports the model specifications used in the Monte Carlo experiments for the DGPs.

[Table 1 about here.]

We denote with "LS" the location-scale model, "ARs" the stationary autoregressive process, "ARst" the stationary autoregressive process with trend, and "ARns" the nonstationary autoregression.

Number of Breaks Detection

Table 2 reports gauge and potency as well as the empirical size and power for the processes described in Table 1. In order, we consider the IIS, the SDS and the SB based on both parametric and nonparametric resampling schemes. For the IIS and the SDS, we set the block method creation in the *Autometrics* options both *standard* (i.e. sequential) and *random*. Since both gauge/potency and empirical size/power are practically identical either using standard block method or the random block method creation, we report the results for the standard option only.

Table 2 reports the gauge/potency and the size/power of the two procedures.

[Table 2 about here.]

Dummy Saturation. Focusing on the DS, it emerges that both the IIS and the SDS have a good gauge at 1% target size, meaning that the average retention rate of the insignificant dummies is satisfactorily under control. On the opposite, when working with 5% target size the gauge is generally between 20% and 30%. Thus, in line with Castle et al. (2012), this suggest for the DS to set a 1% target size in the *Autometrics* options. The average retention rate of the non null dummies is instead mixed across the different DGPs changing from the IIS to the SDS. While the potency level for the SDS are rarely below 80%, the IIS is affected in the cases of the autoregressive processes and the location of the breaks. In particular, when working with IIS and the AR models, there is lower potency because *Autometrics* tends to select just the dummies near the break dates, without retaining dummies within the regimes. This is due probably to the fact that after the shocks have been correctly captured by the impulse dummies, the subsequent (if 1 break) or intermediate instability (if 2 breaks) is

absorbed by the autoregressive component. This may explain the loss in potency with respect to the LS model. From Table 2, we can see that the loss in potency decreases if we restrict the interval between the two break dates. Furthermore, the simulation exercise allows also to highlight that the location of the break(s) matter for *Autometrics* when applying the IIS. If the level shift is close to the end of the sample, in general is no problem. However, as we move the break in the middle of the sample, the gauge increases in particular in the case of multiple level shifts as the distance between two break dates increases. The first effect might be explained in that *Autometrics* selects the minimum number of impulse dummies that provide a good approximation of the underlying DGP. If the level shift is at the beginning of the sample, *Autometrics* selects the impulses from the first observation up to that point.

As far as the SDS is concerned, Table 2 also reports the empirical size and power about the number of structural breaks as described in Section 4.1. There is evidence of very high oversize both at 1% and 5% nominal levels. When considering the 1% nominal level, in almost 50% of the cases at least one insignificant step dummy is retained. As a consequence, the number of times that the procedure detects the correct number of breaks is around 50% as well. However, as alreday anticipated, this result is expected in that when applying the DS we are testing the significance of a huge number of dummies and this comes at the cost that some insignificant dummies are retained. For this reason, the information given by the retention rates of the "null" and "non-null" dummies is in general a better criterion to evaluate the performances of the DS. For the DS, in the next sets of simulations, we focus only on the computation of the gauge and potency, assessing the approach as a model selection procedure.
Sequential Bootstrapping. Table 2 show that the procedure has an empirical size almost identical to the nominal level at both 1% and 5% nominal levels when applied to stationary and trend-stationary models, while is slightly oversized for the non-stationary autoregressive model. The most dramatic oversize is for the multiple breaks case when we observe a 13% with a nominal of 5% and a 3% for a nominal of 1%. This implies that the SB detects the correct number of structural breaks in almost all cases. As far as the power is concerned, the empirical rejection frequencies of false null hypotheses are close to 100% for all DGPs considered in the simulations.

Break Dates Estimates

To evaluate the goodness of the break estimates, in Table 3 we report the sample mean, the mean bias, the root mean squared error and the median for each estimate.

[Table 3 about here.]

The break date estimates obtained either by applying the IIS, or the SDS or the SB are very similar. For the SDS, we report the estimates obtained when computing the gauge/potency

and those we obtaine when computing size/power. In all cases, the mean bias is sufficiently small and the estimators are almost always median unbiased. Also the standard deviations are satisfactorily small for all procedures the only exception being the case of the estimates obtained when considering the size/power of the SDS in the case of stationary autoregressions. The best performances are shown by the SDS (when considering the gauge/potency related estimates) and the SB: the SDS provides slightly smaller standard deviations but the SB is the only procedure which is median unbiased for all the cases considered in the simulation exercise.

The main conclusions from the Monte Carlo exercises can be summarised as follows. First, as expected the SDS outperforms the IIS when considering multiple breaks affecting the level component. The IIS seems to work well for capturing outliers and single level shifts but when dealing with multiple breaks in level, the SDS and the SB prove to be more appropriate. Second, using empirical rejection frequencies to assess the DS can be misleading while it is more appropriate to work with the empirical retention rates of the dummies. Third, the SB provides empirical size close to the nominal one and very good power across all simulations with the exception of the nonstationary autoregression model. Finally, both the IIS and SDS as well as the SB provide good estimates of the break dates when considering structural breaks affecting the level component. A final important point is about that to discriminate between alternative methodologies it is better to use the gauge/potency and size/power properties rather than the goodness of the break dates estimates.

4.1.2 Conditional Models

We now turn to the case of conditional models with stationary and nonstationary regressors. This setting allows us to explore how breaks in the marginal processes (regressors) affect the conditional equation[7]. Table 4 reports the DGPs considered in this second set of Monte Carlo simulations.

[Table 4 about here.]

We denote with "Cs" the conditional model with stationary variables and no breaks in the marginal equation (process for x_t), "Cci" the conditional model with non-stationary cointegrated variables and no breaks in the marginal equation, "CMs" and "CMci" the corresponding versions where also the marginal equation experiences structural breaks in its level component. In the case of broken marginal equations, we let the marginal process x_t to experience one structural break distinct from those in the conditional process for y_t.

[7]The first paper to consider this case with stationary regressors is Hansen (2000). Hall et al. (2012) propose an extension to the multiple breaks case in a stationary framework, considering breaks in the regression coefficients.

Number of Breaks Detection

Table 5 reports the gauge and the potency as well as the empirical size and the power for the conditional processes introduced above. For the IIS and the SDS, as in the previous set of simulations involving location-scale and autoregressive processes, we do not find significant differences among the size and power recorded in changing the block method and thus we report the results for the standard block method option only. Thus, the assessment of the SDS is undertaken on the basis of the gauge and potency.

[Table 5 about here.]

Dummy Saturation: Similarly to the case of location-scale and autoregressive processes commented above, the IIS and the SDS models have gauge close to the target size only at 1% level. The potency is instead above 90% in almost all cases with a slightly lower performance of the IIS for the one break "Cci" case.

Sequential Bootstrapping: As far as the SB is concerned, the procedure delivers empirical size close to nominal both at 1% and 5% nominal levels and the power is close to 100% for all DGPs. Moving from stationary to nonstationary processes does not seem to affect the performance of the IIS, the SDS and the SB, the only exception being the IIS with one break case ("Cci" and "CMci"). This is of particular interest for the SB procedure in support of the validity of the bootstrapping in a nonstationary cointegrating framework. Another important result is that both the DS and the SB are robust to the presence of a break in the marginal process. The gauge and the empirical size of the procedures are not affected by letting the marginal process experience an independent break from the ones in the conditional. As far as the DS is concerned, the Monte Carlo exercise extends the results of the simulations contained in Castle et al. (2012) for the IIS and shows the good performances of the SDS. Turning to the SB, the robustness of the procedure to broken marginal processes is even more important in light of that Hansen (2000) shows that the distribution of the sup-F statistic to test for breaks in the conditional process is affected when a break is observed in the marginal process. To avoid rejecting stability in conditional models in absence of breaks, Hansen (2000) proposes a bootstrapping approach denoted *fixed regressor bootstrap*. However, the author does not take into account the cointegrating framework while presenting an autoregressive case with a moderate level of persistence (0.5). Though the SB is based on a different bootstrapping approach[8], the results confirm once again that the adoption of the bootstrap in cointegrating equations seem to perform well (see e.g. Li and Maddala, 1997).

[8]The procedure is very similar to that suggested by Diebold and Chen (1996), where reasonable size is found even when the persistence in the process considered is high.

Break Dates Estimates

Table 6 reports the results of break date estimates. All the three procedures perform quite well in terms of goodness of the resulting estimates. The estimates of the break dates are median unbiased and the mean bias is very controlled and less than one except for the IIS in some cases. Overall, the best results in terms of unbiasness and smaller RMSE are those obtained applying the SB.

[Table 6 about here.]

4.2 Breaks in Trend

In this section, we consider processes with breaks also affecting the deterministic linear time trend thus extending (12)-(13) as follows:

$$y_t = \alpha + \delta t + \beta x_t + \sum_{i=1}^{K} (\psi_i B_{i,t}^L + \omega_i B_{i,t}^T) + u_t \tag{14}$$

$$x_t = \gamma + \zeta t + \rho x_{t-1} + \sum_{j=1}^{K_x} (\upsilon_j B_{j,t}^L + \phi_j B_{j,t}^T) + e_t \qquad t = 1, \ldots, T \tag{15}$$

where $B_{i,t}^T \equiv (t - T_i + 1)\mathbf{I}_{\{t>T_i\}}$ denotes a steptrend dummy. We consider both autoregressive and conditional processes and we let also the marginal process experience independent breaks. We restrict our analysis to the SDDS and the SB. In Table 7, we report the list of the DGPs considered.

[Table 7 about here.]

ARsTr" and "ARnsTr" denote the stationary and nonstationary autoregressive processes respectively, "CsTr" and "CciTr" the conditional model with stationary regressors and nonstationary cointegrated regressors respectively, "CMsTr" and "CMciTr" the stationary and cointegrated conditional models where the marginal processes experience independent breaks, respectively. We assume that the structural break(s) affect contemporaneously both the level and the trend components, though we let the shock to be not necessarily of the same signs between the level and the trend. This implies that a negative shock in the level might be present together with a positive shock in the trend. Similarly as for the DGPs with just breaks in level, we carry out a set of experiments for $K = \{0, 1, 2\}$. The null hypotheses are the same as those reported in Section 4.1.

Number of Breaks Detection

Table 8 reports the gauge/potency and the empirical size/power for Monte Carlo simulations involving the processes described in Table 7. As far as the SDDS is concerned, we

explore again the standard and the random methods to create blocks as also suggested in Johansen and Nielsen (2009) for the trending and unit root cases. In particular, the GUM is specified alternating step and steptrend dummies such that the standard block search algorithm of *Autometrics* takes into account the dummies referring to the same potential break date in the same block[9]. As expected, when saturating with step and steptrend dummies, the role of the different methods for creating blocks impacts on the final selected model. Indeed, when moving from standard to random blocks, although the variations in terms of potency are not dramatic, the gauge is affected. Thus, we also report the results for the random block method. Again, the assessment of the SDDS is based on the gauge and the potency only since the saturation with further dummies would make the empirical rejection frequency even less meaningful than before.

[Table 8 about here.]

Dummy Saturation: As far as the SDDS is concerned, the gauge appears to be under control setting the target size to be either 1% or 5% when working with the standard block method. When using the random blocks instead, the gauge of the SDDS tends to increase above the nominal level while the gain in potency is negligible. In particular, for the autoregressive processes, we see high average retention rates of null dummies also when setting 1% as target size. Thus, we may conclude that the use of the random blocks does not improve the performance of the SDDS. On the other side, when working with standard blocks, the controlled gauge at both 1% and 5% is especially surprising given that the number of dummies that enter the initial GUM is $T \times 3$. However, the controlled gauge comes with a considerable drop in potency. The best potency level we find does not go above 70% with an average around 50%. This results is expected since there are two dummies which should be retained for each break date.

Sequential Bootstrapping. The performance of the SB is in line with the results of the rpevious simulations reported above. In particular, the empirical size is close to nominal at both 1% and 5% nominal levels. Moreover, differently from the SDDS case, the procedure does not suffer of any loss in power, and even when considering structural breaks affecting the trend, we register a power around 100%.

Break dates estimates

Table 9 reports the structural break estimates for the SDDS and the SB procedures. For the SDDS, we report separately the estimates of the break dates in the level from those affecting the trend.

[9] We also tried to insert the step and steptrend dummies sequentially (i.e. all the step dummies first and then all the step trend dummies) but strategy did not change the outcome of the experiments.

[Table 9 about here.]

Overall, both the SDDS and the SB estimates have a controlled mean bias and small standard deviations. In particular, as far as the SDDS is concerned, we have that the level break(s) estimates are all median unbiased while for the trend break(s) estimates there is a small upward median bias and the RMSEs tend to be slightly higher than the corresponding level break(s) estimates. For the SB, the estimates are all median unbiased and the RMSEs smaller than the SDDS counterparts.

In conclusion, when considering structural breaks affecting both the level and the trend component of a time series, *Autometrics* seems to suffer the saturation with impulse, step and trend dummies. Although the break dates estimates are still satisfying, the SDDS loses in terms of potency. This implies that there is a quite high number of times when relevant dummies are not retained. Furthermore, changing the way in which *Autometrics* forms the blocks from standard to random does not improve the overall performance of the procedure, which on the contrary gets worse. On the other side, the performance of the SB does not seem to be affected by the inclusion of break in trend in the light that empirical sizes is close to the nominal one with the power levels approaching 100%.

4.3 Some Useful Guidelines to Practitioners

On the basis of the results from the extensive simulation exercise, we now offer some useful guidelines to applied economists on how to correctly implement the DS and the SB procedures.

With respect to the DS, an important role is played by the *target size*. As shown in the the Monte Carlo simulations, the DS works properly only if the 1% target size is used. A higher target size would lead to the retention of too many dummies. As far as the *block method* is concerned, the standard block method performs best. As also highlighted in Doornik (2009b), we discourage the use of the random block scheme because it leads to the retention of more insignificant dummies. As far as the *choice between the IIS and the SDS* in systems affected only by breaks in mean, what emerges from the simulations is that IIS and SDS provide satisfactory performances both in terms of gauge and potency as well as in terms of break date estimates. In particular, the SDS provides better results in terms of potency than the IIS, as also highlighted by Doornik et al. (2013). Moreover, saturating with both step and impulse dummies allows the step dummies to capture breaks in level while the impulse dummies capture possible outliers. This helps also to simplify the interpretation of the final output. When using the SDS, there is no need to group the impulse dummies according to their sign and magnitude to identify a break in level, as it happens when using the IIS. The procedure of grouping the impulse dummies may be difficult to implement as the sample size grows. Note that when considering systems affected also by *breaks in trend*, the potency of the SDDS tends to be lower than the case of only breaks in level. Despite the loss in potency, the

performance of the procedure is similar to that of the SB. In addition, since the *ordering of the variables* matters as the number of dummies grows, we suggest to try alternative ordering schemes for the step and steptrend dummies. What we learn from the experiments reported in this paper is that the best approach is to alternate one step and one steptrend dummy in a chronological order. In this way, *Autometrics* takes into account contemporaneously dummies capturing the same break date when forming the blocks.

As far as the SB is concerned, the procedure works very well in detecting breaks in both level and trend. There are only few cases when the SB tends to be oversized occurring in the nonstationary autoregressions while the procedure works quite well when dealing with cointegrating equations. An additional point is about the *stopping rule in the sequential search*. As already reported in De Peretti and Urga (2005), the optimal rule is to stop the sequential search when two consecutive break dates are statistically insignificant. This procedure is robust to the the fact that the estimator of the initial break dates is biased when further breaks are neglected.

5 Testing the Fisher Effect in the U.S. Economy

In this section, we implement the DS and the SB procedure to evaluate the Fisher effect in the United States which assumes that the nominal interest rate and the expected inflation rate are linked via the following equation

$$r_t = r_t^{real} + \mathbb{E}_t \pi_{t+1} \tag{16}$$

where r_t is the nominal interest rate, r_t^{real} is the real interest rate and $\mathbb{E}_t \pi_{t+1}$ is the expected rate of inflation for the next period at time t. In a rational expectations framework we expect that inflation is equal to the observed one plus a mean-zero Gaussian forecast error term: $\mathbb{E}_t(\pi_{t+1}) = \pi_t + e_t$ with $e_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$. The Fisher equation (16) holds if estimating the linear equation

$$r_t = \alpha + \delta t + \beta \pi_t + \varepsilon_t. \tag{17}$$

β is statistically equal to one. The fact that many empirical studies (see Beyer et al, 2009) conclude that β differs significantly from unity may be due to the presence of unmodelled level/trend shifts in the Fisher relationship. To this purpose, in this paper we implement both the SDDS and the SB procedures to equation (17).

We employ quarterly data running from 1985-Q1 to 2012-Q3 (111 observations) obtained from the Federal Reserve Bank of Saint Louis database (FRED). We use the 3 months Treasury Bill rate as short term nominal interest rate (r_t) and we compute the inflation rate (π_t) as the annual relative change in the CPI $\left(\frac{CPI_t - CPI_{t-4}}{CPI_{t-4}}\right)$. Figure 1 reports the plots of the two series.

[Figure 1 about here.]

Both series are tested for the presence of unit roots and consistently with other studies involving United States post-war data (see for instance Crowder and Hoffman, 1996) we cannot reject the hypothesis that both series are I(1). Following Beyer et al. (2009), in order to address potential endogeneity between interest rates and inflation rates, we estimate (17) by DOLS including five leads and lags of the first differenced rate of inflation.

As long as the SDDS is concerned, we set the target size at 1%, we fix all the variables entering the Fisher equation except the dummies and we explore different combinations in the ordering of the step and steptrend dummies as well as different block methods. In particular, we specify GUMs where the step and trend dummies enter sequentially (i.e. all the step dummies and then all the steptrend dummies) versus GUMs which alternate one step and one steptrend dummy. We also employ both standard and random block methods. Consistently with what emerges in the Monte Carlo study, the final model that support the economic theory and passes all misspecification tests is that obtained alternating the dummies and using the standard block method. Table 10 reports the final model selected by *Autometrics* when estimating (17) by DOLS (to control for endogeneity) and applying the SDDS:

[Table 10 about here.]

Nine breaks and one outlier are identified. All misspecification tests are satisfied and the residuals have all the desirable properties as Figure 2 shows, where we report the residuals, the ACF up to 20 lags and the QQ-plot against a $\mathcal{N}(0,1)$. Moreover, the residuals are stationary implying the presence of a cointegrated relationship as found by Crowder and Hoffman (1996).

[Figure 2 about here.]

Table 10 shows that in the Fisher effect holds being $\beta \in [0.857; 1.05]$ with probability 95%.

As far as the implmenetation of SB is concerned, we estimate the same model by DOLS stopping the sequential search when two consecutive break dates are statistically insignificant. In particular, the significance of the break dates is assessed through parametric and non-parametric bootstrapped p-values of the corresponding sup-F statistics. Table 11 reports the statistically significant break dates identified by the SB procedure.

[Table 11 about here.]

On the basis of the break dates selected by the SB, we estimate the Fisher equation by DOLS and adding the corresponding step and steptrend dummies. Table 12 reports the final selected model after removing the insignificant variables[10].

[10] We additionally run *Autometrics* over the final model with the "Large Residuals" option to control for large residuals. This is denominated "I:2000(3-4)" in Table 12.

[Table 12 about here.]

Via the implementation of the SB, we identify six significant structural breaks captured by step and steptrend dummies. Similarly to what we find for the SDDS, the selected model provides a good fit of r_t with no sign of misspecification. Figure 3 reports a plot of the residuals together with their ACF and a QQ-plot. Once more, after accounting for structural breaks, the relationship between the short term interest rate and the inflation rate cointegrates and the Fisher effect holds given that $\beta \in [0.829; 1.187]$ with probability of 95%.

[Figure 3 about here.]

Finally, to give a visual representation of the similarity of the structural breaks identified by the SDDS and the SB procedures, in Figure 4 we report the actual series of the interest rate with the structural breaks identified by vertical solid lines to denote breaks affecting the level component and vertical dashed lines to denote breaks affecting the trend component.

[Figure 4 about here.]

The dynamics which emerges from Figure 4 is that there is almost a one-to-one correspondence between the structural breaks identified by the SDDS and the SB. In particular, the differences in the breaks dates are all within one quarter when comparing a break found by the SDDS with the correspondent closest break found by the SB. We have that the two procedures validate each other and this is an important result given that they are based on two completely different logics.

6 Conclusions

In this paper, we proposed two novel procedures, the Dummy Saturation (DS) and the Sequential Bootstrapping (SB) of the sup-F statistic, to estimate and date multiple structural breaks in the deterministic components of a linear system. Through an extensive Monte Carlo simulation exercise, we evaluated the performance of the two procedures considering several data generating processes ranging from the simple location-scale model to the case of cointegrating regressions, considering both conditional and marginal processes. We evaluated the performance of the DS involving not only impulse indicators (IIS) but also step dummies (SDS) and both step dummies and steptrend dummies (SDDS). We were able to select the significant regressors from a very large set ($T \times 2$ for the SDS and $T \times 3$ for the SDDS) of candidate regressors. For the DS, we found that the retention rate of the insignificant dummies is close to the chosen target size at the 1% significant level, while the SB provided empirical sizes close to the nominal ones both at 1% and 5% significance levels. When considering structural breaks affecting the level component, the DS showed good potency and similarly to

the SB good power. In particular, we found that the SDS had desirable properties in terms of potency and gauge, and it often outperformed the IIS when applied to processes with broken level. When considering also breaks in the linear time trend component (SDDS model), the potency decreased although the gauge remained well controlled. Furthermore, we evaluated the sensibility of SDDS model to the block method options provided by *Autometrics*: the use of a random block method did not improve the overall performance of the DS consistently with what reported in Doornik (2009b). On the other hand, the SB proved to have empirical size close to nominal one both at 1% and 5% levels for all the DGPs considered, the only exception being the case of nonstationary autoregressive processes showing slightly oversizes. Opposite to the SDDS case, the SB was not affected by the inclusion of steptrend dummies in terms of both size and power. Finally, both the DS and the SB procedures showed good performance considering broken marginal processes when regressors are nonstationary.

We implemented the DS and the SB procedures to study to which extent the Fisher effect holds in the U.S. economy. The application of the SDDS and the SB procedures leads us to two important findings. First, the Fisherian hypothesis is valid only when structural breaks are properly detected and modelled. Second, the two procedures detected almost the same break dates affecting the deterministic components of the Fisher equation.

Finally, the findings in this paper suggest some further developments. First, it may be interesting to extend the framework considered in this paper to identify breaks affecting the slope coefficients. While this should not complicate the inference for the SB procedure, the computational cost may increase substantially for the DS procedure due to additional T dummies for each regressor. Second, it will be useful to compare the performance of the DS with breaks in mean and trend estimated using *Autometrics* with the robust dummy saturation estimator proposed in Johansen and Nielsen (2009), based on an M-estimator with a bias-corrected variance term. Third, in the light of our empirical results on the Fisher effect and considering the profound instabilities experienced over the last decades, detecting and modelling structural breaks may also be important to validate other well established economic relationships. The robust procedures considered in this paper certainly help to undertake correct inference. This is part of an ongoing research agenda.

References

Andreou, E. and Ghysels, E. (2009). Structural Breaks in Financial Time Series. In T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 839–870. Berlin: Springer-Verlag.

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

Aue, A. and Horváth, L. (2013). Structural Break in Time Series. *Journal of Time Series Analysis 34*, 1–16.

Bai, J. (1994). Least Squares Estimation of a Shift in a Linear Process. *Journal of Time Series Analysis 15*, 453–472.

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

——— (2003). Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics 18*, 1–22.

Banerjee, A. and Lazarova, S. and Urga, G. (1998). Bootstrapping Sequential Tests for Multiple Structural Breaks. *London Business School Discussion Papers 17*.

Banerjee, A. and Urga, G. (2005). Modelling Structural Breaks, Long Memory and Stock Market Volatiliy: an Overview. *Journal of Econometrics 129*, 1–34.

Beyer, A. and Haug, A. and Dewald, W.G. (2009). Structural Breaks, Cointegration and the Fisher Effect. *ECB Working Paper Series 1013*.

Castle, J. and Doornik, J.A. and Hendry, D.F. (2012). Model Selection when there are Multiple Breaks. *Journal of Econometrics 169*, 239–246.

Crowder, W.J. and Hoffman, D.L. (1996). The Long-run Relationship Between Nominal Interest Rates and Inflation: the Fisher Equation Revisited. *Journal of Money, Credit and Banking 28*, 102–118.

De Peretti, C. and Urga, G. (2005). Stopping Tests in the Sequential Estimation of Multiple Structural Breaks. *Unpublished Manuscript,* Centre for Econometric Analysis, Cass Business School.

Diebold, F. X. and Chen, C. (1996). Testing Structural Stability With Endogenous Break Point: A Size Comparison of Analytic and Bootstrap Procedures. *Journal of Econometrics 70*, 221–241.

Doornik, J. A. (2009a). Autometrics. In J. Castle and N. Shepard (Eds.), *The Methodology and Practice of Econometrics*, Chapter 4, pp. 88–121. Oxford: Oxford University Press.

_____ (2009b). Econometric Model Selection With More Variables Than Observations. *Working Paper*.

Doornik, J. A. and Hendry, D. F. and Pretis, F. (2013). Step Indicator Saturation. *Discussion Paper 658,* Department of Economics, University of Oxford.

Ericsson, N. R. (2011). Justifying Empirical Macro-Econometric Evidence In Practice. *Working Paper*.

Hall, A. and Han, S. and Boldea, O. (2012). Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors. *Journal of Econometrics 170*, 281–302.

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

_____ (2000). Testing for Structural Change in Conditional Models. *Journal of Econometrics 97*, 93–115.

Hansen, P. R. (2003). Structural Changes in the Cointegrated Vector Autoregressive Model. *Journal of Econometrics 114*, 261–295.

Hendry, D.F. and Johansen, S. and Santos, C. (2008). Automatic Selection of Indicators in a Fully Saturated Regression. *Computational Statistics 23*, 317–335 and Erratum 337–339.

Hendry, D. F. and Santos, C. (2005). Regression Models with Data-based Indicator Variables. *Oxford Bulletin of Economics and Statistics 67*, 571–595.

Johansen, S. and Nielsen, B. (2009). An Analisys of the Indicator Saturation Estimator as a Robust Regression Estimator. In J. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics*, Chapter 1, pp. 1–36. Oxford: Oxford University Press.

Kejriwal, M. and Perron, P. (2008). The Limit Distribution of the Estimates in Cointegrated Regression Models with Multiple Structural Changes . *Journal of Econometrics 146*, 59–73.

_____ (2010). Testing for Multiple Structural Changes in Cointegrated Regression Models. *Journal of Business & Economic Statistics 28*, 503–522.

Li, H. and Maddala, G. S. (1997). Bootstrapping Cointegrating Regressions. *Journal of Econometrics 80*, 297–318.

Perron, P. (1989). The Great Crash, the Oil Price Shock and the Unit Root Hypothesys. *Econometrica 57*, 1361–1401.

——— (2006). Dealing with Structural Breaks. In *Palgrave Handbook of Econometrics, Vol. I: Econometric Theory*. Palgrave.

Quandt, R. E. (1960). Tests of the Hypothesys that a Linear Regressions System Obeys Two Separate Regimes. *Journal of the American Statistical Association 55*, 324–330.

Rappoport, P. and Reichlin, L. (1989). Segmented Trends and Non-stationary Time Series. *The Economic Journal 99*, 168–177.

Seo, B. (1998). Tests for Structural Change in Cointegrated Systems. *Econometric Theory 14*, 222–259.

Stock, J.H. (1994). Unit Roots, Structural Breaks and Trends. In R. F. Engle and D. L. MacFadden (Eds.), *Handbook of Econometrics Vol. IV*, Chapter 46, pp. 2740–2831. Elsevier Science.

Table 1: Marginal Models DGPs.

DGP	
LS:	$x_t = 0.2 + \sum_{j=1}^{K} \upsilon_j B_{j,t}^L + u_t$
ARs:	$x_t = 0.2 + 0.6x_{t-1} + \sum_{j=1}^{K} \upsilon_j B_{j,t}^L + u_t$
ARst:	$x_t = 0.2 + 0.05t + 0.6x_{t-1} + \sum_{j=1}^{K} \upsilon_j B_{j,t}^L + u_t$
ARns:	$x_t = 0.2 + x_{t-1} + \sum_{j=1}^{K} \upsilon_j B_{j,t}^L + u_t$

Note: In order to achieve independence from the initial conditions, for the "ARs" and "ARst" models we start the recursion to generate the x_t series at $x_{-50} = 0$ for $t = -50, \ldots, 100$ and then we discard the first 50 observations. Similarly, for the "ARns" model, we start from $x_{-100} = 0$, $t = -100, \ldots, 100$, and then we discard the first 100 initial observations.

Table 2: Correct Number of Breaks Detection for Marginal Models.

		IIS		SDS		SDS			SB (param.)		SB(nonparam)		
DGP	T_i	Gauge	Potency	Gauge	Potency	Size	Power	NCB	Size	Power	Size	Power	Dummies' coeff.
LS	-	1.6	-	1.6	-	48.5	-	-	1.4	-	1.5	-	
		24.1	-	18.5	-	100.0	-	-	4.5	-	5.2	-	
	70	1.6	98.7	1.2	100.0	41.4	100.0	58.6	1.2	100.0	1.0	100.0	$v_1 = 5$
		19.1	99.9	17.6	100.0	99.5	100.0	0.5	5.3	100.0	5.4	100.0	
	20, 40	0.5	49.9	0.9	98.8	35.8	100.0	64.2	1.5	99.9	1.1	100.0	$v_1 = 3, v_2 = -3$
		16.4	91.2	20.6	99.2	98.9	100.0	1.1	5.4	100.0	5.0	100.0	
ARs	-	1.6	-	3.4	-	65.3	-	-	1.2	-	0.7	-	-
		25.7	-	24.6	-	100.0	-	-	3.5	-	3.6	-	
	70	1.7	19.2	2.5	100.0	59.1	100.0	40.9	1.3	99.4	0.9	98.4	$v_1 = 5$
		22.7	35.4	23.0	100.0	99.9	100.0	0.1	5.4	100.0	5.8	100.0	
	20, 40	2.1	14.7	1.1	83.1	39.1	99.4	60.9	1.6	99.7	0.9	99.1	$v_1 = 5, v_2 = -5$
		24.3	44.0	19.7	96.7	99.2	100.0	0.8	6.0	99.9	5.3	99.9	
	20, 30	1.2	81.9	1.3	99.8	44.3	100.0	55.7	1.2	99.9	1.2	99.9	$v_1 = 5, v_2 = -5$
		18.2	90.8	21.0	99.6	99.8	99.9	0.1	4.9	99.9	5.8	99.9	
ARst	-	1.6	-	3.4	-	68.9	-	-	0.9	-	1.3	-	-
		27.7	-	21.3	-	100.0	-	-	5.7	-	5.3	-	
	70	1.7	17.7	2.4	100.0	60.7	100.0	39.3	0.8	99.1	0.9	98.1	$v_1 = 5$
		24.2	37.9	20.1	100.0	99.8	100.0	0.2	5.2	100.0	5.4	100.0	
	20, 40	2.0	14.6	1.2	75.1	40.5	98.2	59.5	1.3	96.8	0.8	95.8	$v_1 = 5, v_2 = -5$
		26.2	45.3	17.1	95.4	99.3	99.9	0.7	5.5	99.3	5.2	99.2	
	20, 30	1.3	83.3	1.3	99.6	46.1	99.9	53.9	1.0	95.1	1.0	95.1	$v_1 = 5, v_2 = -5$
		18.5	91.2	18.2	99.4	99.8	100.0	0.2	5.0	99.4	5.4	99.3	
ARns	-	1.6	-	3.4	-	74.2	-	-	2.0	-	1.6	-	-
		26.1	-	23.8	-	100.0	-	-	7.4	-	7.6	-	
	70	0.7	38.8	1.3	100.0	44.0	100.0	56.0	2.8	100.0	2.5	100.0	$v_1 = 5$
		17.7	66.3	19.6	99.8	99.7	100.0	0.3	10.7	100.0	11.6	100.0	
	20, 40	13.4	73.1	0.9	98.6	37.0	100.0	63.0	3.0	99.2	3.4	99.3	$v_1 = 3, v_2 = -5$
		28.5	82.7	17.7	99.1	98.6	100.0	1.4	13.0	99.9	12.6	100.0	
	20, 30	4.8	83.7	1.1	99.7	44.0	100.0	56.0	3.3	94.5	3.3	93.2	$v_1 = 3, v_2 = -5$
		21.9	91.8	19.1	99.4	99.6	100.0	0.4	12.0	99.4	13.5	99.4	

Notes: The specification of the processes is given in Table 1. The numbers under the columns "gauge" and "potency" represent the empirical retention frequencies of the null and non-null dummies respectively according to the DGP, while the numbers under the columns labelled "size" and "power" are the empirical rejection frequencies of null hypothesis about the number of breaks. For a detail of the null hypotheses, see the end of Section 4.1. Gauge and potency as well as size and power for the SDS are computed only with reference to the step dummies. "NCB" denotes the percentage of times a procedure detects the correct number of structural breaks according to the DGP. Finally, "-" indicates the no break case and thus only gauge and size can be calculated.

Table 3: Break Dates Estimates for Marginal Models.

		IIS				SDS (gauge/potency)				SDS (size/power)				SB			
DGP	T_i	Mean	Bias	RMSE	Median	Mean	Bias	RMSE	Median	Mean	Bias	RMSE	Median	Mean	Bias	RMSE	Median
LS	70	70.022	0.022	0.188	70.000	70.001	0.001	0.202	70.000	70.001	0.001	0.202	70.000	69.999	-0.001	0.051	70.000
	20	21.059	1.059	1.352	21.000	19.998	-0.002	0.656	20.000	20.018	0.018	0.821	20.000	19.843	-0.157	1.833	20.000
	40	38.606	-1.394	1.502	39.000	38.645	-1.355	1.107	38.000	38.614	-1.386	1.166	38.000	40.277	0.277	2.471	40.000
Ars	70	69.783	-0.217	5.161	70.000	69.978	-0.022	0.256	70.000	69.987	-0.013	0.247	70.000	69.932	-0.068	0.658	70.000
	20	20.043	0.043	0.255	20.000	19.995	-0.005	0.306	20.000	21.759	1.759	8.016	20.000	19.983	-0.017	0.549	20.000
	40	39.401	-0.599	3.108	40.000	38.688	-1.312	1.461	40.000	38.620	-1.380	5.420	38.000	40.013	0.013	0.553	40.000
	20	20.008	0.008	0.089	20.000	19.991	-0.009	0.212	20.000	19.985	-0.015	0.270	20.000	19.951	-0.049	0.713	20.000
	30	28.842	-1.158	1.200	29.000	29.900	-0.100	0.343	30.000	29.906	-0.094	0.345	30.000	30.087	0.087	1.167	30.000
ARst	70	70.110	0.110	0.847	70.000	69.980	-0.020	0.268	70.000	69.982	-0.018	0.264	70.000	69.949	-0.051	0.618	70.000
	20	20.060	0.060	0.462	20.000	20.005	0.005	0.296	20.000	22.951	2.951	10.782	20.000	20.001	0.001	0.542	20.000
	40	39.307	-0.693	3.343	40.000	38.977	-1.023	1.365	40.000	39.336	-0.664	7.149	39.000	40.005	0.005	0.491	40.000
	20	20.010	0.010	0.109	20.000	20.006	0.006	0.261	20.000	20.011	0.011	0.368	20.000	20.009	0.009	0.870	20.000
	30	28.838	-1.162	1.191	29.000	29.827	-0.173	0.451	30.000	29.805	-0.195	0.659	30.000	30.036	0.036	0.688	30.000
ARns	70	70.092	0.092	0.340	70.000	70.019	0.019	0.234	70.000	70.022	0.022	0.264	70.000	69.826	-0.174	1.728	70.000
	20	20.160	0.160	0.537	20.000	20.089	0.089	0.797	20.000	20.183	0.183	1.492	20.000	19.006	-0.994	2.245	20.000
	40	38.906	-1.094	1.150	39.000	39.980	-0.020	0.227	40.000	39.969	-0.031	0.276	40.000	39.882	-0.118	0.995	40.000
	20	20.050	0.050	0.231	20.000	19.990	-0.010	0.642	20.000	19.998	-0.002	0.807	20.000	18.835	-1.165	2.484	20.000
	30	28.878	-1.122	0.475	29.000	29.954	-0.046	0.282	30.000	29.958	-0.042	0.265	30.000	29.958	-0.042	0.786	30.000

Note: The break date estimates for the IIS and the SDS refer to the results obtained when using 1% as target size in the *Autometrics* options (gauge and potency estimates) and when considering 1% as nominal size (size and power estimates).

Table 4: Conditional Models DGPs.

DGP	
Cs:	$y_t = 0.2 + 0.8x_t + \sum_{i=1}^{K} \psi_i B_{i,t}^L + u_t$
	$x_t = 0.6x_{t-1} + e_t$
Cci:	$y_t = 0.2 + 0.8x_t + \sum_{i=1}^{K} \psi_i B_{i,t}^L + u_t$
	$x_t = x_{t-1} + e_t$
CMs:	$y_t = 0.2 + 0.8x_t + \sum_{i=1}^{K} \psi_i B_{i,t}^L + u_t$
	$x_t = 0.5 + 0.6x_{t-1} + \upsilon B_t^{L,x} + e_t$
CMci:	$y_t = 0.2 + 0.8x_t + \sum_{i=1}^{K} \psi_i B_{i,t}^L + u_t$
	$x_t = 0.5 + x_{t-1} + \upsilon B_t^{L,x} + e_t$

Note: For the "Cs" and "CMs" models, we start the recursion to generate the x_t series at $x_{-50} = 0$ for $t = -50, \ldots, 100$ and then we discard the first 50 observations. Similarly, for the "Cci" and "CMci" models, we start from $x_{-100} = 0$, $t = -100, \ldots, 100$, and then we discard the first 100 initial observations.

Table 5: Correct Number of Breaks Detection for Conditional Models.

		IIS		SDS		SB (param.)		SB(nonparam)		
DGP	T_i	Gauge	Potency	Gauge	Potency	Size	Power	Size	Power	Dummies' coeff.
Cs	-	1.5	-	1.9	-	1.6	-	1.5	-	-
		25.9	-	22.0	-	6.0	-	5.6	-	
	70	1.6	98.2	1.3	100.0	1.4	100.0	1.1	100.0	$\psi_1 = 5$
		19.4	99.9	20.7	100.0	5.9	100.0	5.6	100.0	
	20, 40	1.0	97.4	0.8	99.8	1.3	100.0	1.2	100.0	$\psi_1 = 5, \psi_2 = -5$
		18.9	99.9	18.4	99.9	6.5	100.0	6.0	100.0	
Cci	-	1.6	-	1.8	-	1.3	-	1.2	-	-
		25.9	-	21.6	-	5.5	-	5.2	-	
	70	7.5	74.6	1.4	100.0	1.3	100.0	1.5	100.0	$\psi_1 = 5$
		24.6	79.5	20.5	100.0	5.0	100.0	5.7	100.0	
	20, 40	1.7	95.5	0.7	99.8	1.2	100.0	0.4	100.0	$\psi_1 = 5, \psi_2 = -5$
		19.2	99.3	17.8	99.9	4.7	100.0	4.3	100.0	
CMs	- (90)	1.6	-	1.8	-	1.4	-	1.0	-	$\upsilon = -2$
		26.0	-	22.0	-	6.4	-	6.0	-	
	70 (90)	1.5	92.1	1.3	100.0	1.5	100.0	1.2	100.0	$\psi_1 = 5, \upsilon = -2$
		19.2	99.0	20.0	100.0	6.8	100.0	5.9	100.0	
	20, 40 (90)	1.1	97.7	0.8	99.6	1.4	100.0	1.3	100.0	$\psi_1 = 5, \psi_2 = -5, \upsilon = -2$
		19.0	100.0	18.3	99.8	5.9	100.0	6.0	100.0	
CMci	- (90)	1.6	-	1.8	-	1.1	-	1.0	-	$\upsilon = -2$
		26.0	-	21.8	-	6.0	-	6.0	-	
	70 (90)	2.7	83.6	1.3	100.0	1.7	100.0	1.8	100.0	$\psi_1 = 5, \upsilon = -2$
		20.0	93.3	19.7	100.0	4.5	100.0	5.6	100.0	
	20, 40 (90)	1.2	97.7	0.8	99.8	1.1	100.0	0.9	100.0	$\psi_1 = 5, \psi_2 = -5, \upsilon = -2$
		19.1	99.9	18.0	99.8	4.9	100.0	5.3	100.0	

Notes: The specification of the processes is given in Table 4. The numbers in brackets in the column "T_i" represent the break dates affecting the marginal process. The numbers under the columns "gauge" and "potency" represent the empirical retention frequencies of the null and non-null dummies respectively according to the DGP, while "size" and "power" represent the empirical rejection frequencies of null hypothesis about the number of breaks. For a detail of the null hypotheses, see the end of Section 4.1. Gauge and potency for the SDS are computed only with reference to the step dummies. Finally, "-" indicates the no break case and thus only gauge and size can be calculated.

Table 6: Break Date Estimates for Conditional Models.

		IIS				SDS				SB			
DGP	T_i	Mean	Bias	RMSE	Median	Mean	Bias	RMSE	Median	Mean	Bias	RMSE	Median
Cs	70	70.027	0.027	0.190	70.000	69.996	-0.004	0.173	70.000	70.013	0.013	0.362	70.000
	20	20.028	0.028	0.203	20.000	20.062	0.062	0.311	20.000	19.988	-0.012	0.418	20.000
	40	39.947	-0.053	0.352	40.000	38.445	-1.555	0.859	38.000	40.015	0.015	0.450	40.000
Cci	70	71.312	1.312	4.061	70.000	69.994	-0.006	0.219	70.000	69.996	-0.004	0.528	70.000
	20	20.135	0.135	1.002	20.000	20.087	0.087	0.351	20.000	20.003	0.003	0.437	20.000
	40	39.883	-0.117	1.024	40.000	38.983	-1.017	1.071	40.000	40.008	0.008	0.440	40.000
CMs	70	70.070	0.070	0.383	70.000	70.029	0.029	0.237	70.000	70.014	0.014	0.332	70.000
	20	20.022	0.022	0.178	20.000	20.029	0.029	0.311	20.000	19.981	-0.019	0.428	20.000
	40	39.962	-0.038	0.266	40.000	38.319	-1.681	0.766	38.000	40.011	0.011	0.389	40.000
CMci	70	70.432	0.432	2.383	70.000	70.024	0.024	0.247	70.000	69.968	-0.032	0.543	70.000
	20	20.032	0.032	0.192	20.000	20.072	0.072	0.342	20.000	19.996	-0.004	0.436	20.000
	40	39.947	-0.053	0.684	40.000	38.570	-1.430	0.953	38.000	40.007	0.007	0.460	40.000

Note: The break date estimates for the IIS and the SDS refer to the results obtained when using 1% as target size.

Table 7: Marginal and Conditional Models DGPs with Broken Linear Trend.

DGP name	
ARsTr:	$x_t = 0.2 + 0.05t + 0.6x_{t-1} + \sum_{j=1}^{K}(\upsilon_j B_{j,t}^L + \phi_i B_{i,t}^T) + u_t$
ARnsTr:	$x_t = 0.2 + 0.01t + x_{t-1} + \sum_{j=1}^{K}(\upsilon_j B_{i,t}^L + \phi_j B_{i,t}^T) + u_t$
CsTr:	$y_t = 0.2 + 0.05t + 0.5x_t + \sum_{i=1}^{K}(\psi_i B_{i,t}^L + \omega_i B_{i,t}^T) + u_t$
	$x_t = 0.6x_{t-1} + e_t$
CciTr:	$y_t = 0.2 + 0.02t + 0.5x_t + \sum_{i=1}^{K}(\psi_i B_{i,t}^L + \omega_i B_{i,t}^T) + u_t$
	$x_t = x_{t-1} + e_t$
CMsTr:	$y_t = 0.2 + 0.05t + 0.5x_t + \sum_{i=1}^{K}(\psi_i B_{i,t}^L + \omega_i B_{i,t}^T) + u_t$
	$x_t = 0.5 + 0.02t + 0.6x_{t-1} + \upsilon B_t^{L,x} + \phi B_t^{T,x} + e_t$
CMciTr:	$y_t = 0.2 + 0.01t + 0.5x_t + \sum_{i=1}^{K}(\psi_i B_{i,t}^L + \omega_i B_{i,t}^T) + u_t$
	$x_t = 0.5 + 0.001t + x_{t-1} + \upsilon B_t^{L,x} + \phi B_t^{T,x} + e_t$

Note: For the "ARsTr", "CsTr" and "CMs" models, we start the recursion to generate the x_t (or y_t) series at $x_{-50} = 0$ for $t = -50, \ldots, 100$ and then we discard the first 50 observations. Similarly, for the "ARnsTr"CciTr" and "CMciTr" models, we start from $x_{-100} = 0$, $t = -100, \ldots, 100$, and then we discard the first 100 initial observations.

Table 8: Correct Number of Breaks Detection for Marginal and Conditional Models with Broken Trend.

		SDDS				SB (param.)		SB(nonparam)		
		standard blocks		random blocks						
DGP	T_i	Gauge	Potency	Gauge	Potency	Size	Power	Size	Power	Dummies' coeff.
ARsTr	-	0.1	-	3.6	-	0.9	-	1.3	-	-
		2.8	-	13.5	-	5.7	-	5.3	-	
	30	0.6	56.0	4.0	58.2	1.1	99.2	1.2	98.5	$\upsilon_1 = -5$
		2.2	57.8	14.4	75.0	5.1	100.0	4.7	99.9	$\phi_1 = 0.1$
	50, 80	0.6	48.5	3.5	64.4	0.5	100.0	1.0	99.9	$\upsilon_1 = -8, \upsilon_2 = -5$
		3.0	53.8	14.1	75.4	4.8	100.0	4.8	100.0	$\phi_1 = 0.2, \phi_2 = 0.4$
ARnsTr	-	0.2	-	5.5	-	3.0	-	2.7	-	-
		3.5	-	13.9	-	8.7	-	8.6	-	
	30	0.7	49.3	5.8	68.5	0.7	100.0	1.1	100.0	$\upsilon_1 = -3$
		3.8	59.2	15.8	80.8	6.8	100.0	6.6	100.0	$\phi_1 = 0.05$
	50,80	1.1	49.6	5.2	64.8	0.8	95.1	1.1	95.7	$\upsilon_1 = -3, \upsilon_2 = -5$
		3.8	57.1	15.6	79.9	5.0	99.3	6.1	99.5	$\phi_1 = 0.1, \phi_2 = 0.4$
CsTr	-	0.4	-	0.8	-	1.1	-	1.6	-	-
		2.8	-	8.1	-	5.6	-	5.4	-	
	30	0.6	57.1	1.2	56.0	1.1	100.0	0.7	100.0	$\psi_1 = -5$
		2.1	54.6	9.6	67.4	4.8	100.0	4.7	100.0	$\omega_1 = 0.1$
	50,80	0.8	67.7	1.5	65.8	0.8	100.0	0.7	100.0	$\psi_1 = -8, \psi_2 = -5$
		2.7	61.5	9.9	70.5	3.8	100.0	4.5	100.0	$\omega_1 = 0.2, \omega_2 = 0.4$
CciTr	-	0.5	-	0.7	-	1.4	-	1.0	-	-
		2.6	-	8.4	-	5.4	-	5.4	-	
	30	0.6	56.3	1.2	56.1	0.8	100.0	0.7	100.0	$\psi_1 = -5$
		2.0	56.7	9.9	67.6	3.5	100.0	4.7	100.0	$\omega_1 = 0.1$
	50,80	0.8	64.1	1.5	64.5	1.0	100.0	0.5	100.0	$\psi_1 = -8, \psi_2 = -5$
		2.7	58.6	10.4	72.0	4.6	100.0	5.0	100.0	$\omega_1 = 0.1, \omega_2 = 0.4$
CMsTr	- (90)	0.3	-	0.7	-	0.8	-	1.3	-	$\upsilon = -2, \phi = -0.1$
		2.7	-	8.7	-	6.1	-	6.1	-	
	30 (90)	0.6	57.2	1.1	57.2	0.6	100.0	0.7	100.0	$\psi_1 = -5, \omega_1 = 0.1$
		2.1	56.0	10.0	69.6	4.9	100.0	5.0	100.0	$\upsilon = -2, \phi = -0.1$
	50, 80 (90)	0.8	63.4	1.4	68.9	0.7	100.0	0.7	100.0	$\psi_1 = -8, \psi_2 = -5, \omega_1 = 0.1, \omega_2 = 0.4$
		2.6	60.6	10.3	71.3	3.9	100.0	4.3	100.0	$\upsilon = -2, \phi = -0.1$
CMciTr	- (90)	0.5	-	0.7	-	0.7	-	1.3	-	$\upsilon = -2, \phi = -0.1$
		2.4	-	9.2	-	3.9	-	4.7	-	
	30 (90)	0.5	63.3	1.1	63.4	1.1	100.0	0.8	100.0	$\psi_1 = -8, \omega_1 = 0.1$
		1.6	59.3	9.3	70.1	4.4	100.0	3.7	100.0	$\upsilon = -2, \phi = -0.1$
	50, 80 (90)	0.8	55.9	1.5	64.1	0.9	99.7	0.7	99.5	$\psi_1 = -8, \psi_2 = -5, \omega_1 = 0.1, \omega_2 = 0.4$
		2.5	58.6	10.0	71.8	4.4	100.0	4.3	99.9	$\upsilon = -2, \phi = -0.1$

Notes: The specification of the processes is given in Table 7. The numbers in brackets in the column "T_i" represent the break dates affecting the marginal process. The numbers under the columns "gauge" and "potency" represent the empirical retention frequencies of the null and non-null dummies respectively according to the DGP, while the numbers under the columns denoted "size" and "power" are the empirical rejection frequencies of null hypothesis about the number of breaks. For a detail of the null hypotheses, see the end of Section 4.1. Gauge and potency for the SDDS are computed only with reference to the step and steptrend dummies. Finally, "-" indicates the no break case and thus only gauge and size can be calculated.

Table 9: Break Dates Estimates for Marginal and Conditional Models with Broken Trend.

		SDDS								SB			
		level				trend							
DGP	T_i	Mean	Mean Bias	RMSE	Median	Mean	Mean Bias	RMSE	Median	Mean	Mean Bias	RMSE	Median
ARsTr	30	29.971	-0.029	0.632	30.000	30.724	0.724	1.470	31.000	29.977	-0.023	0.511	30.000
	50	49.738	-0.262	0.793	50.000	50.838	0.838	0.961	51.000	49.990	-0.010	0.205	50.000
	80	80.027	0.027	0.682	80.000	80.547	0.547	1.519	81.000	79.996	-0.004	0.578	80.000
ARnsTr	30	29.919	-0.081	1.092	30.000	30.572	0.572	1.379	31.000	30.245	0.245	2.458	30.000
	50	50.002	0.002	0.829	50.000	50.298	0.298	1.405	51.000	50.020	0.020	1.815	50.000
	80	80.056	0.056	0.765	80.000	80.718	0.718	1.247	81.000	80.007	0.007	1.111	80.000
CsTr	30	30.128	0.128	0.643	30.000	30.697	0.697	0.935	31.000	30.002	0.002	0.408	30.000
	50	49.834	-0.166	0.412	50.000	50.748	0.748	0.871	51.000	49.998	-0.002	0.127	50.000
	80	79.996	-0.004	0.764	80.000	80.722	0.722	1.179	81.000	80.028	0.028	0.471	80.000
CciTr	30	30.095	0.095	0.729	30.000	30.700	0.700	0.992	31.000	29.995	-0.005	0.640	30.000
	50	49.842	-0.158	0.388	50.000	50.793	0.793	0.821	51.000	50.001	0.001	0.114	50.000
	80	80.013	0.013	0.667	80.000	80.841	0.841	1.186	81.000	80.025	0.025	0.423	80.000
CMsTr	30	30.052	0.052	0.844	30.000	30.702	0.702	0.979	31.000	29.978	-0.022	0.464	30.000
	50	49.932	-0.068	0.397	50.000	50.860	0.860	0.720	51.000	49.998	-0.002	0.127	50.000
	80	79.951	-0.049	0.638	80.000	80.707	0.707	1.157	81.000	80.029	0.029	0.500	80.000
CMciTr	30	29.889	-0.111	0.617	30.000	30.660	0.660	0.905	31.000	29.996	-0.004	0.141	30.000
	50	50.006	0.006	0.512	50.000	50.742	0.742	0.878	51.000	49.999	-0.001	0.095	50.000
	80	80.062	0.062	0.562	80.000	80.650	0.650	1.196	81.000	80.039	0.039	0.605	80.000

Note: The break date estimates for the SDDS refer to the results obtained when using 1% as target size and the standard blocks.

Table 10: SDDS Procedure: Final Selected Model for the Fisher Relationship.

	Coefficient	Std.Error	t-value	t-prob
Constant	-0.5732	0.1053	-5.440	0.000
Trend	0.0253	0.0049	5.120	0.000
π_t	0.9520	0.0488	19.500	0.000
$\Delta\pi_{t-5}$	0.2351	0.0775	3.030	0.003
$\Delta\pi_{t+1}$	0.7045	0.0993	7.100	0.000
$\Delta\pi_{t+3}$	0.2638	0.0847	3.120	0.003
$\Delta\pi_{t+4}$	0.6827	0.0843	8.100	0.000
$\Delta\pi_{t+5}$	0.5917	0.0967	6.120	0.000
S:1990(4)	-0.4132	0.1477	-2.800	0.006
S:1991(1)	-0.4151	0.1526	-2.720	0.008
T:1994(2)	0.3244	0.0287	11.300	0.000
T:1995(1)	-0.3178	0.0302	-10.500	0.000
S:1998(4)	-0.3559	0.0966	-3.690	0.000
S:2000(3)	0.7176	0.1338	5.360	0.000
T:2000(3)	-0.3866	0.0304	-12.700	0.000
T:2001(4)	0.3469	0.0334	10.400	0.000
S:2005(3)	0.9773	0.1000	9.770	0.000
S:2008(1)	-0.9225	0.0901	-10.200	0.000
I:2005(3)	-0.6259	0.1451	-4.310	0.000
Adj. R^2	0.9817		AIC	-3.976
			HQ	-3.776
			SC	-3.484
AR 1-5 test:	F(5,77)	=	2.316	[0.0515]
ARCH 1-4 test:	F(4,93)	=	1.646	[0.1693]
Normality test:	$\chi^2(2)$	=	0.623	[0.7325]
Hetero test:	F(27,71)	=	1.026	[0.4489]
RESET23 test:	F(2,80)	=	0.701	[0.4991]

Note: I:YYYY(Q) indicates an impulse dummy ($B^I_{i,t} = \mathbf{I}_{\{t=\text{YYYY(Q)}\}}$), S:YYYY(Q) a step dummy ($B^L_{i,t} = \mathbf{I}_{\{t\geq\text{YYYY(Q)}\}}$) and T:YYYY(Q) a steptrend dummy ($B^T_{i,t} = (t - \text{YYYY(Q)} + 1)\mathbf{I}_{\{t\geq\text{YYYY(Q)}\}}$).

Table 11: Break Dates Selected by the SB Procedure.

Date	**1994Q2**	**2005Q4**	**2002Q1**	**2008Q1**	**1991Q1**	**1998Q4**	1995Q2	2000Q4
P p-value	0.000**	0.000**	0.000**	0.000**	0.000**	0.010**	0.404	0.141
NP p-value	0.000**	0.000**	0.000**	0.010**	0.000**	0.020*	0.343	0.242

Notes: The bold dates represent the statistically significant structural breaks. ** indicates significance at 1% while * significance at 5%. Both parametric (P) and nonparametric (NP) p-values are computed according to 99 bootstrap replications. For each significant break date a step and steptrend dummies are created in order to estimate the final model reported in Table 12.

Table 12: SB Procedure: Final Selected Model for the Fisher Relationship.

	Coefficient	Std.Error	t-value	t-prob
Constant	-0.6934	0.1749	-3.960	0.000
Trend	0.0238	0.0055	4.320	0.000
π_t	1.0080	0.0895	11.300	0.000
$\Delta\pi_{t-2}$	-0.3078	0.1380	-2.230	0.028
$\Delta\pi_{t-4}$	0.3136	0.1191	2.630	0.010
$\Delta\pi_{t+1}$	0.8419	0.1579	5.330	0.000
$\Delta\pi_{t+3}$	0.6665	0.1296	5.140	0.000
$\Delta\pi_{t+4}$	0.7949	0.1340	5.930	0.000
$\Delta\pi_{t+5}$	0.5240	0.1517	3.450	0.001
S:1991(1)	-0.6555	0.1411	-4.640	0.000
S:1994(2)	1.0446	0.1303	8.010	0.000
T:1998(4)	-0.1076	0.0118	-9.100	0.000
T:2002(1)	0.0821	0.0170	4.830	0.000
S:2005(4)	0.9405	0.1496	6.290	0.000
T:2005(4)	-0.0368	0.0169	-2.170	0.033
S:2008(1)	-0.5456	0.1839	-2.970	0.004
I:2000(3-4)	0.7510	0.1581	4.750	0.000
Adj. R^2	0.9511		AIC	-3.008
			HQ	-2.830
			SC	-2.568
AR 1-5 test:	F(5,79)	=	2.006	[0.0868]
ARCH 1-4 test:	F(4,93)	=	0.606	[0.6594]
Normality test:	$\chi^2(2)$	=	3.103	[0.2119]
Hetero test:	F(27,73)	=	1.224	[0.2455]
RESET23 test:	F(2,82)	=	2.326	[0.1041]

Note: I:YYYY(Q) indicates an impulse dummy ($B^I_{i,t} = \mathbf{I}_{\{t=\text{YYYY(Q)}\}}$), S:YYYY(Q) a step dummy ($B^L_{i,t} = \mathbf{I}_{\{t\geq\text{YYYY(Q)}\}}$) and T:YYYY(Q) a steptrend dummy ($B^T_{i,t} = (t - \text{YYYY(Q)} + 1)\mathbf{I}_{\{t\geq\text{YYYY(Q)}\}}$).

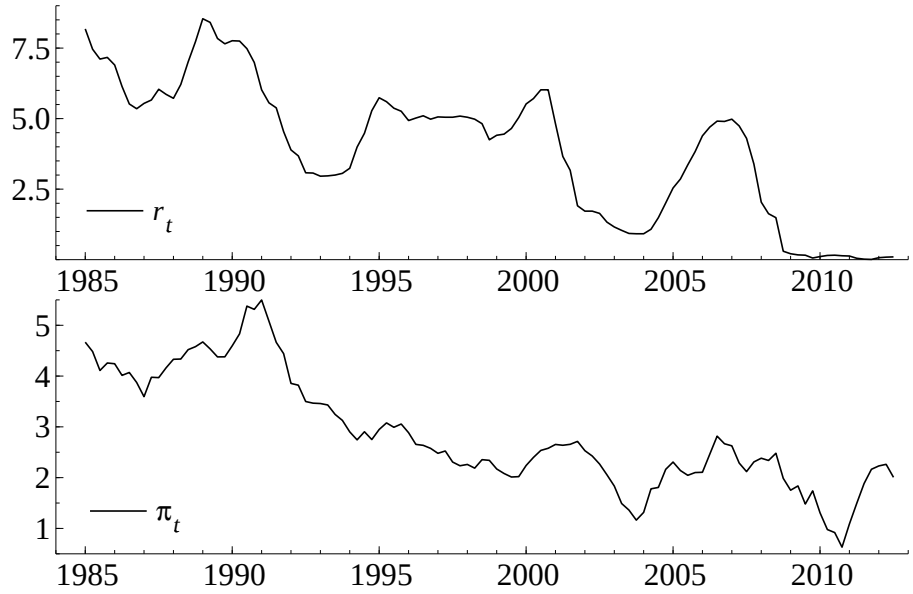

Figure 1: Three months Treasury Bill rate (r_t) and inflation rate (π_t).

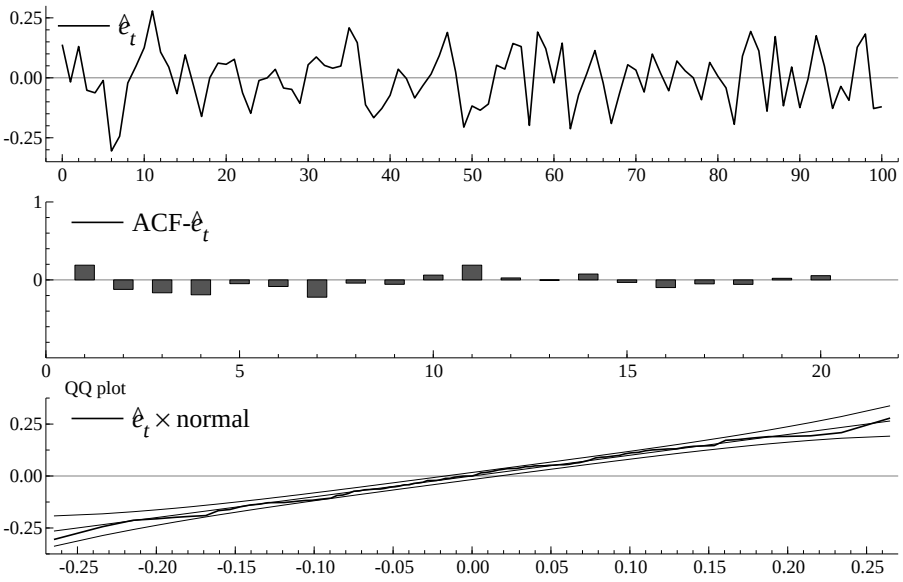

Figure 2: SDDS procedure final selected model plots. The first plot reports the residuals $\hat{e}_t$ time series. The second plot reports the autocorrelation function of $\hat{e}_t$ up to the 20^{th} lag. The third plot reports the QQ-plot of $\hat{e}_t$ against a standard normal.

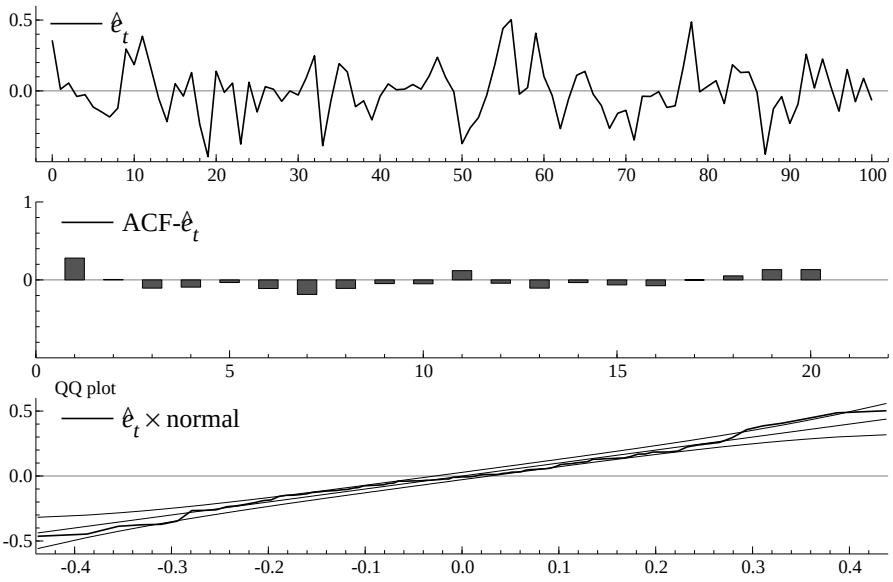

Figure 3: SB procedure final selected model plots. The first plot reports the residuals $\hat{e}_t$ time series. The second plot reports the autocorrelation function of $\hat{e}_t$ up to the 20^{th} lag. The third plot reports the QQ-plot of $\hat{e}_t$ against a standard normal.

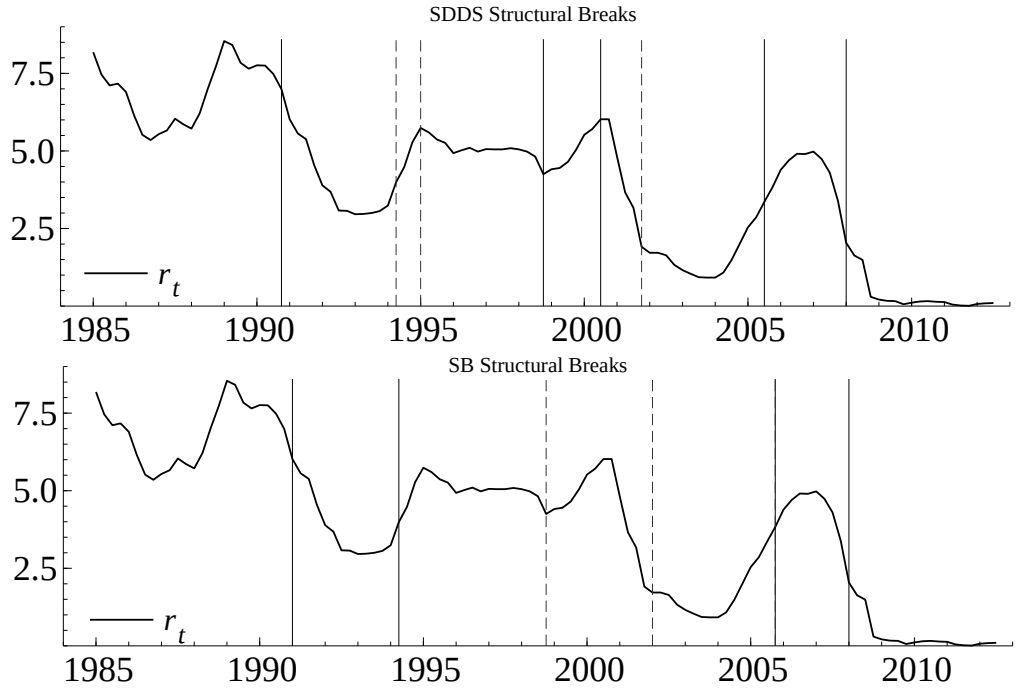

Figure 4: Structural Breaks found by the SDDS and the SB in the r_t time series. Solid lines denote breaks in level while dashed lines denote breaks in trend.