Determining the number of breaks in a piecewise linear regression model[*]

Birgit Strikholm [†]

Department of Economic Statistics and Decision Support
Stockholm School of Economics

Abstract

In this paper we propose a sequential method for determining the number of breaks in piecewise linear structural break models. An advantage of the method is that it is based on standard statistical inference. Tests available for testing linearity against switching regression type nonlinearity are applied sequentially to determine the number of regimes in the structural break model. A simulation study is performed in order to investigate the finite-sample behaviour of the procedure and to compare it with other alternatives. We find that our method works well in comparison for both single and multiple break cases.

Key words: Model specification, multiple structural breaks.

JEL Classification Code: C22, C51

[†]Department of Economic Statistics and Decision Support, Stockholm School of Economics, Box 6501, SE-113 83 Stockholm, Sweden, `email: Birgit.Strikholm@hhs.se`

1 Introduction

Models with structural breaks (SB) have been of interest to many researchers for at least the last four decades. Most of the work in this area of research has been related to the case of detecting and estimating a single break. See Chow (1960), Andrews (1993), and Bai, Lumsdaine, and Stock (1998), among others. The questions related to multiple structural changes have received somewhat less attention. Early works include Yao (1988) and Liu, Wu, and Zidek (1997) who advocated the use of the (modified) Bayesian Information Criterion and showed that the number of breaks can be estimated consistently (at least for a normal sequence of random variables with shifts in mean). In Bai (1997) it was shown that one can consistently estimate break-points, one-by-one, in a multiple break model even when the number of breaks estimated is smaller than the actual number of breaks. He also proposed a simple sequential procedure for consistently estimating the number of breaks. In a seminal paper Bai and Perron (1998) proved consistency of the estimators of the break dates, provided tests for multiple structural changes and constructed confidence intervals for the break dates. Last, but not least, they also proposed several methods (one of which is purely sequential) for determining the number of breaks and efficient algorithms for computing the estimates. In two companion papers, see Bai and Perron (2006, 2003a), the authors considered practical matters related to the methods proposed in Bai and Perron (1998): such as the behaviour of estimators and tests in finite samples, and comparisons between different methods for determining the number of breaks. Since all the tests considered have nonstandard distributions, Bai and Perron (2003b) also provided asymptotic critical values for a set of possible specifications (nominal level $\alpha = \{0.10, 0.05, 0.025, 0.01\}$, the minimum relative regime size $\epsilon_R = \{0.05, 0.10, 0.15, 0.20, 0.25\}$ and the number of regressors whose parameters are allowed to vary across regimes $q = 1, \ldots, 10$). In coming sections, we refer to papers by Bai and Perron providing the theory and simulations results, and to the methodology in general, as BP.

More recently, Prodan (2006) proposed a new procedure, including a restricted version designed to detect trend reversions, for choosing the number of breaks. This method is based on a sequence of likelihood ratio-type tests of Bai (1999) for which the critical values have to be bootstrapped. Ben Aïssa, Boutahar, and Jouini (2004) proposed a method based on the stability of the evolutionary spectral density analysis. They apply their method and Bai and Perron's method on US inflation data, but unfortunately no size and power comparisons are performed.

The main limitation of the current implementation of Bai and Perron sequential procedure is that critical values exist for a restricted number of combinations only (four significance levels, five regime sizes, up to ten regressors). If one wants to test for breaks, say, in 11 monthly dummies, then one would have to simulate the critical values. Furthermore, as documented in Prodan (2006), the asymptotic critical values obtained under the null of independent and identically distributed errors might be inadequate for relatively short (in her simulations $T = 125$) but persistent series, which in turn can cause severe size distortions. The main limitation of the procedure proposed by Prodan is that it is rather time-consuming since the critical values have to be bootstrapped for every test in the sequence.

In this paper we propose an alternative sequential procedure for determining the number of breaks in a structural break model. The technique itself is based on a sequence of parameter constancy tests in Smooth Transition Regression (STR) framework where a model with m breaks (transitions) is tested against one with $m+1$ breaks. Its advantages include the fact that standard statistical inference applies and the modeller has control over the significance level of each test. Our technique is easy to implement, it imposes no restrictions on the number of regressors, significance levels of individual tests or minimum regime size (as long as the moment matrix is well-defined). It can be applied to situations where all parameters are assumed to change over time (pure structural change model) and to the ones in which just a subset of parameters is subject to change (partial structural change model).

The plan of the paper is as follows. Section 2 contains a brief overview of smooth transition regression models and of tests against switching-type parameter nonconstancy. Section 3 describes our method step-by-step. Section 4 contains the results of a simulation study, where the size and power properties of our method are discussed and compared to results in Bai and Perron (2006). An empirical application based on the quarterly US ex-post real interest rate series can be found in Section 5, and Section 6 concludes.

2 Smooth transition regression framework

2.1 The Model

The general idea underlying our procedure is quite old. Goldfeld and Quandt (1972, pp. 263–264) considered the estimation of parameters in the switching regression model and pointed out that discontinuity of the log-likelihood complicates the esti-

mation. Their suggestion was to replace the sudden switch by a smooth transition. This removes the discontinuity, and the parameters of the resulting smooth transition regression model can be estimated by conditional maximum likelihood, using an appropriate iterative algorithm.

We will apply the idea of approximating sudden changes with smooth transitions to the regime selection problem. That, in turn, allows us to use standard inference in determining the number of regimes in a (multiple) structural break model.

A classical logistic STR (LSTR) model for a univariate time series y_t is given by

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_0 + \mathbf{x}'_t \boldsymbol{\beta}_1 G_{1t} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1}$$

where $\mathbf{x}_t = (1, x_{1t}, x_{2t}, \ldots, x_{kt})' = (1, y_{t-1}, \ldots, y_{t-p}, w_{1t}, \ldots, w_{nt})' = (1, \widetilde{\mathbf{x}}_t)'$ with $k = p+n$ is a $((k+1) \times 1)$ vector of explanatory variables, $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are $((k+1) \times 1)$ parameter vectors and $\{\varepsilon_t\}$ is a sequence of independent, normally distributed errors with zero mean and variance σ^2. The transition function G_{1t} in (1) is defined as follows:

$$G_{1t} = G_1(s_t; \gamma_1, c_1) = \left(1 + \exp\{-\gamma_1(s_t - c_1)\}\right)^{-1}, \ \gamma_1 > 0. \tag{2}$$

As $\gamma_1 \to \infty$ in (2), the logistic function G_{1t} approaches the indicator function $I[s_t > c_1]$ and the LSTR model becomes a switching regression (SR) model. The parameter c_1 is then the switch or breakpoint parameter. Thus the STR model (1) with (2) is a reasonable approximation to the SR model when γ_1 is sufficiently large. Letting the variable $s_t = t$ (or, rescaling time to be between 0 and 1, $s_t = t^* = t/T$) and $\gamma_1 \to \infty$, we obtain a single structural break model.

Analogously, we can approximate a multiple structural change model with a Multiple LSTR (MLSTR) model. For example, an MLSTR model with two transitions has the form

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_0^* + \mathbf{x}'_t \boldsymbol{\beta}_1^* G_{1t} + \mathbf{x}'_t \boldsymbol{\beta}_2^* G_{2t} + \varepsilon_t, \tag{3}$$

where the transition function $G_{2t} = G_2(t^*; \gamma_2, c_2)$ is defined as in (2). For the purposes of this paper we set $\gamma_1 = \gamma_2 = \gamma$.

2.2 Testing parameter constancy

Testing linearity (or parameter constancy) should be one of the first steps before actually fitting a more complicated nonlinear model. Lin and Teräsvirta (1994) developed a test for testing parameter constancy against continuous structural change. As structural break models are a special case of the more general time-varying autoregressive (TV-AR) models, the test developed in Lin and Teräsvirta (1994) has

power against TV-AR but also against SB ($\gamma \to \infty$) alternatives. What follows is a short description of the test.

Suppose we want to test parameter constancy of (1), where $\mathbf{x}_t = (1, \widetilde{\mathbf{x}}_t)' = (1, y_{t-1}, \ldots, y_{t-k})'$. One could achieve this by setting $\gamma_1 = 0$ in $G_{1t}(t^*; \gamma_1, c_1)$ or $\boldsymbol{\beta}_1 = \mathbf{0}$. In order to circumvent the identification problem we approximate the transition function by its Taylor expansion around $\gamma_1 = 0$. The third-order approximation can be written as $T_3 = \delta_0 + \delta_1 t^* + \delta_2 t^{*2} + \delta_3 t^{*3} + R_3(\gamma_1, c_1; t^*)$ where R_3 is the remainder and $\delta_0, \delta_1, \delta_2$ and δ_3 are constants. Substituting T_3 for G_1 in (1) and reparameterizing we obtain

$$y_t = \mathbf{x}_t'\boldsymbol{\theta}_0 + (\mathbf{x}_t t^*)'\boldsymbol{\theta}_1 + (\mathbf{x}_t t^{*2})'\boldsymbol{\theta}_2 + (\mathbf{x}_t t^{*3})'\boldsymbol{\theta}_3 + \varepsilon_t^*, \tag{4}$$

where $\varepsilon_t^* = \varepsilon_t + (\mathbf{x}_t'\boldsymbol{\beta}_1)R_3(\gamma_1, c_1; t^*)$. The parameter vectors $\boldsymbol{\theta}_j = \gamma\widetilde{\boldsymbol{\theta}}_j$, where $\widetilde{\boldsymbol{\theta}}_j \neq \mathbf{0}$, and thus our null hypothesis of linearity (parameter constancy) in (1) implies $H_0' : \boldsymbol{\theta}_j = \mathbf{0}$, $j = 1, 2, 3$, in (4). Since the auxiliary regression (4) is linear in parameters and $\varepsilon_t^* = \varepsilon_t$ under H_0', one can test this null hypothesis by a straightforward Lagrange Multiplier (LM)-type test

$$\chi_{LM}^2 = \widehat{\sigma}^{-2} \left(\sum_{t=1}^{T} \widehat{u}_t \mathbf{w}_t \right)' (\widehat{\mathbf{M}}_{11} - \widehat{\mathbf{M}}_{10}\widehat{\mathbf{M}}_{00}^{-1}\widehat{\mathbf{M}}_{01})^{-1} \left(\sum_{t=1}^{T} \mathbf{w}_t \widehat{u}_t \right), \tag{5}$$

where $\widehat{\mathbf{M}}_{00} = \sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t'$, $\widehat{\mathbf{M}}_{01} = \widehat{\mathbf{M}}_{10}' = \sum_{t=1}^{T} \mathbf{z}_t \mathbf{w}_t'$, $\widehat{\mathbf{M}}_{11} = \sum_{t=1}^{T} \mathbf{w}_t \mathbf{w}_t'$, $\widehat{\sigma}^2 = 1/T \sum_{t=1}^{T} \widehat{u}_t^2$. Here $\widehat{u}_t$ is the residual estimated under the null hypothesis, $\mathbf{z}_t = \mathbf{x}_t$ and $\mathbf{w}_t = (\mathbf{x}_t t^*, \mathbf{x}_t t^{*2}, \mathbf{x}_t t^{*3})$. Under the null hypothesis, the test statistic has an asymptotic χ^2-distribution with $3(k+1)$ degrees of freedom. This result requires the existence of all the moments implied by (5).

Following the suggestions in earlier papers, see Teräsvirta (1994), for example, an F-approximation to the χ_{LM}^2 statistic is recommended. The test can be carried out in three stages using just linear regressions:

1. Regress y_t on $\mathbf{x}_t$ and compute the residual sum of squares
$$SSR_0 = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_t^2.$$

2. Regress $\widehat{u}_t$ (or y_t) on $\mathbf{x}_t$, $\mathbf{x}_t t^*$, $\mathbf{x}_t t^{*2}$ and $\mathbf{x}_t t^{*3}$, and compute the residual sum of squares $SSR_1 = \frac{1}{T} \sum_{t=1}^{T} \widehat{v}_t^2$.

3. Compute
$$F = \frac{(SSR_0 - SSR_1)/(3(k+1))}{SSR_1/(T - 4k - 4)}.$$

4

This statistic is approximately $F_{3(k+1),T-4k-4}$ distributed under the null of linearity.

The test can also be based on the first-order Taylor approximation of G_{1t}. In that case $T_1 = \delta_0 + \delta_1 t^* + R_1(\gamma_1, c_1; t^*)$, where R_1 is the remainder and δ_0 and δ_1 are constants, and one substitutes T_1 for for G_1 in (1). The null of linearity is now $\boldsymbol{\theta}_1 = \mathbf{0}$ in (4) whereas $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_3 = \mathbf{0}$ by definition. This variant of the test is less powerful than the test of H_0' in (4) in cases where the process is returning back to its original level after the second break. We return to these issues in Section 4 where we compare the performance of different tests by simulation. For further discussion of the LM-type test, see, for example, Luukkonen, Saikkonen, and Teräsvirta (1988) or Teräsvirta (1998).

3   Determining the number of structural breaks

3.1   The sequential procedure

In this section we describe our procedure for determining the number of breakpoints in the piecewise linear structural break model. The proposed sequential testing procedure (ST-procedure) mixes the parameter constancy testing of smooth transition regression modelling framework and SB model estimation. The strategy is based on the fact that the estimators of break fractions/parameters in the SB model are superconsistent, see Bai (1997) and Bai and Perron (1998).

For illustration, consider the following three-regime SB model

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_0^\dagger I(t^* \leq c_1) + \mathbf{x}_t' \boldsymbol{\beta}_1^\dagger I(c_1 < t^* \leq c_2) + \mathbf{x}_t' \boldsymbol{\beta}_2^\dagger I(t^* > c_2) + \varepsilon_t \quad t = 1, \ldots, T, \ (6)$$

which can alternatively be written as

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_0^* + \mathbf{x}_t' \boldsymbol{\beta}_1^* I(t^* > c_1) + \mathbf{x}_t' \boldsymbol{\beta}_2^* I(t^* > c_2) + \varepsilon_t \quad t = 1, \ldots, T, \tag{7}$$

where $c_1 \neq c_2$ (for identification reasons we may assume $c_1 < c_2$). Let m denote the number of breaks. We assume that Assumptions A1-A5 in Bai and Perron (1998) are satisfied for model (6). They are required for consistent estimation of the parameters. Note, that our timevariable, t^*, is bounded between zero and one, and thus, our break points c_i correspond to the break fractions λ_i in Bai and Perron (1998).

The starting-point of the procedure is that the true model is either a linear model or a SB model (but possibly with just one break), so the first choice is between $m = 0$ (linearity) and $m = 1$ (two regimes). The ST-procedure as a whole works as follows:

1. Set $\boldsymbol{\beta}_2^* = 0$ in (7) and replace $I(t^* > c_1)$ by $G_{1t}(t^*; \gamma_1, c_1)$. Test linearity of (7) as outlined in Section 2.2; the null hypothesis $H_0 : \gamma_1 = 0$.

2. If H_0 is rejected at a predetermined significance level $\alpha(T)$, chosen such that $\alpha(T) \to 0$ as $T \to \infty$, estimate the parameters of the model (7) assuming $\boldsymbol{\beta}_2^* = 0$ (single break). According to the results in Bai and Perron (1998), $\widehat{c}_1$ is super consistent for c_1 and the following continuation suggests itself.

3. Test linearity of

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_0^\dagger I(t^* \leq \widehat{c}_1) + \mathbf{x}_t' \boldsymbol{\beta}_1^\dagger I(t^* > \widehat{c}_1) + \varepsilon_t, \qquad t = 1, \ldots, T, \qquad (8)$$

against

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_0^\dagger I(t^* \leq \widehat{c}_1) + \mathbf{x}_t' \boldsymbol{\beta}_1^\dagger I(t^* > \widehat{c}_1) + \mathbf{x}_t' \boldsymbol{\beta}_2^\dagger G_{2t} + \varepsilon_t, \qquad t = 1, \ldots, T, \quad (9)$$

where G_{2t} is a transition function satisfying the regularity conditions in Eitrheim and Teräsvirta (1996), so the Taylor expansion based test is applicable.

4. Estimate (6) if the null hypothesis is rejected at significance level $\tau\alpha(T)$, $0 < \tau < 1$. Reducing the significance level compared to the preceding test favours parsimonious models. Choosing τ is left to the modeller: in the simulations we set $\tau = 0.5$.

5. Proceed by testing linearity of (6) where c_2 is substituted for its super consistent estimator $\widehat{c}_2$.

6. The sequential estimation and testing is continued until the first non-rejection of a null hypothesis. This yields the specification for the final model.

The key point here is that the convergence rate of $\widehat{c}_1$ is faster (T) than the corresponding rate of the other estimates $(\sqrt{T})$. For the purposes of the test, $\widehat{c}_1$ can therefore be treated as a known parameter and (8) assumed to be a linear model. This assumption does not affect the asymptotic theory of the linearity test described in Section 2.2. Consequently, the test in Step 3 is just another parameter constancy test, and its asymptotic significance level is known.

It is easy to incorporate testing for partial structural breaks into this framework. The parameter constancy tests can be carried out for any subset of parameters. This is done by setting some elements (the ones we assume to be constant) $\beta_{1i} = 0$ in (1) *a priori*. This in turn means that the same elements in auxiliary regressions are assumed to be equal to zero as well.

For the testing part in our procedure to work in the univariate case we have to assume that ε_t are iid and that $x_{1t}, \ldots, x_{kt}$ are jointly stationary. In addition we require that all the cross-moments $Ew_{it}w_{jt}$ and $Ey_{t-i}w_{jt}$ exist (given that the coefficients of all explanatory variables among $\mathbf{x}_t$ are changing). Finally, the errors are assumed uncorrelated with $\mathbf{x}_t$.

3.2   TV-AR-approximation approach

It is also possible to apply TV-AR-approximation to the whole procedure[1]. The first step is identical to Step 1 in Section 3.1. If parameter constancy is rejected, a TV-AR model with a large fixed γ is estimated and the adequacy of the model tested using the appropriate misspecification tests; see, for example, Eitrheim and Teräsvirta (1996) or Teräsvirta (1998). Estimation and testing is continued until the first non-rejection of the "no parameter non-constancy"-hypothesis. This yields the specification for the final model, i.e the number of transitions is equal to the number of structural breaks, after which the parameters of the final SB model can be estimated.

An advantage of this simple method is that one obtains accurate estimates for the break parameters even when some of them lie near the smallest or largest observation in the sample. A theorethical drawback is that the asymptotic properties of those estimators are difficult to obtain. For a practitioner who is working with finite data sets it appears that there are no noticeable differences between the TV-AR approach and the consistent one outlined in Section 3.1. The finite-sample performances of the two procedures are very similar.

4   Simulation study

In this section we investigate the small-sample behaviour of our model selection procedure by simulation. This also allows us to compare the proposed technique with the one Bai and Perron developed.

Several versions of the Bai and Perron testing sequence can be constructed (and are supported in their GAUSS code) depending on the assumptions on the distribution of the covariates and the errors across segments: the errors can be serially correlated or uncorrelated, regressors are either identically distributed or are allowed

---

[1]Longer discussion using stationary random variables as transition variables can be found in Strikholm and Teräsvirta (2005).

to have heterogenous distributions across segments, and finally also heteroskedasticity of residuals can be permitted. When serial correlation and/or heteroskedasticity is present, BP use a heteroskedasticity and serial correlation consistent (HAC) estimator of the parameter covariance matrix and allow for prewhitening. The authors do not, however, give clear guidelines for when prewhitening should be applied. For the simulation study, they use prewhitening only when the data generating process involves serially correlated errors. They do not discuss the size/power properties of their procedure when prewhitening is applied to uncorrelated series or not applied to serially correlated series. Our results suggest that this issue should be addressed, because the power of the BP procedure can vary by 15 percentage points when prewhitening is erroneously applied to uncorrelated series, see the simulation study below. Some information about the size distortion implied by robustification, when the corresponding features are not present in the data, can be found in Bai and Perron (2006). Bai and Perron also note that the correction for possible serial correlation can be made, allowing the distribution of the regressors and errors to differ across regimes. In the construction of the tests, they do not consider imposing the restriction that the distribution of the regressors $\mathbf{z}_t$ be the same across segments even if they are. This means that they explicitly allow the regressors to have heterogenous distributions.

We, on the other hand, do not make use of these above-mentioned nonparametric techniques. To make our procedure comparable with the one Bai and Perron suggest, we use "parametric correction" for possible serial correlation. This amounts to first determining the AR order of the linear model using BIC (maximum lag length considered is $p = 5$) and setting up the testing sequence as we do when testing for a partial structural change and letting only the parameters of interest to change (the non-AR parameters in this study, if not noted otherwise). This is how most practitioners would cope with serial correlation. To be fair to both procedures, we report the results for both uncorrected and corrected versions (even if the data generated contain no features that have to be corrected for). The columns in Tables 1 – 8 labelled "LM$_1$" and "LM$_3$" correspond to the ST-procedure, making use of the first-order and third-order Taylor expansion, respectively. "Corrected" test sequences involve the nonparametric correction technique of BP and "parametric correction" in our method. The subscript "$_{PW}$" denotes that prewhitening has been used in BP's sequence.

4.1 Estimating the empirical size

Following Bai and Perron (2006) we simulate a number of univariate models with no structural changes and study how often the methods actually select the alternative of no breaks. The models are as follows:

	Generate data from	Estimate $y_t = \mathbf{z}_t'\boldsymbol{\delta}_j + u_t,\ j = 1, \ldots, m+1$
(a)	$y_t = e_t$	$\mathbf{z}_t = \{1\}$
(b)	$y_t = e_t + \Psi_t$	$\mathbf{z}_t = \{1, \Psi_t\}$
(c)	$y_t = 0.5y_{t-1} + e_t$	$\mathbf{z}_t = \{1, y_{t-1}\}$
(d)	$y_t = 0.5y_{t-1} + e_t$	$\mathbf{z}_t = \{1\}$
(e)	$y_t = 0.5e_{t-1} + e_t$	$\mathbf{z}_t = \{1\}$
(f)	$y_t = -0.3e_{t-1} + e_t$	$\mathbf{z}_t = \{1\}$

where $\{e_t\} \sim \text{nid}(0, 1)$, $\{\Psi_t\} \sim \text{nid}(1, 1)$ and uncorrelated with $\{e_t\}$, and $\mathbf{z}_t$ denotes the vector of covariates whose coefficients are allowed to change. For each DGP, we generate 2000 Monte Carlo replications with $T = 120$ observations[2]. Because the size of the BP sequential procedure is somewhat affected by the size of the trimming ϵ_R (minimum relative regime size), we report, following their recommendations, the results using $\epsilon_R = 0.05$ for cases with no serial correlation correction, and $\epsilon_R = 0.20$ for cases with serially correlated errors (if not otherwise noted). The nominal test size is $\alpha = 0.05$. For the ST-procedure, we choose $\tau = 0.5$, that is we halve the level of the test at each consecutive step. This has no effect on size simulations, because the first linearity test still has the correct nominal size. If we choose not to reduce the level of the test at every step, that is $\tau = 1$, the only differences would appear in probabilities $P(m = a), a \neq 0$, where m denotes the number of breaks.

Table 1 contains the results for the six DGPs listed above. The size distortion when correcting for non-existing serial correlation for our procedure is minor, see panels (a) and (b). The size of the BP sequential procedure depends heavily on whether prewhitening is used or not. Applying the prewhitening for the DGP in panel (b) can cause a size distortion as large as 10 percentage points. In the presence of serial correlation, one should try to correct for it, because ignoring its presence may lead to serious size distortions, see the results for the uncorrected versions of the tests in panels (d) and (e). Both procedures are well sized when the serial correlation in the errors is accounted for, although the BP procedure appears more oversized than our technique when the nonzero autocorrelations are positive. Prewhitening

---

[2]We first generate 200 initial observations that will be removed, to minimize the possible effect of the starting-values.

does affect the size of the BP sequential procedure even here, and not applying it when it would be necessary to do so can cause noticeable size distortion, see panel (d), for example.

A thorough discussion of size distortion in the Bai and Perron sequential procedure can be found in Prodan (2006). Simple simulations with an AR(1) model show that the size distortion becomes severe the closer we get to the unit root. Our simulations support her results. Our "corrected" procedure displays somewhat less size distortion than the BP sequential procedure, but is still oversized when the autoregressive coefficient approaches one. On nominal 5% level with $\rho = 0.9$ we reject LM_1 test in about 11% and the LM_3 test in about 21% of the cases (compared to about 22% for BP).

4.2   Simulating structural break models

To study the power of the procedures, we again replicate the experiments in Bai and Perron (2006). Even here, $\{\Psi_t\} \sim \text{nid}(1,1)$ and $\{e_t\} \sim \text{nid}(0,1)$, and these sequences are mutually uncorrelated. The minimum relative regime size for cases with no error autocorrelation correction is $\epsilon_R = 0.05$, and for cases with serial correlation $\epsilon_R = 0.2$, that is 20% of the length of the series.

4.2.1   A single break

First we look at a battery of data generating processes with a single break in the middle of the series. The model has a general form

$$
\begin{aligned}
y_t = \mu_1 + \nu_1 \Psi_t + e_t, &\qquad \text{if} \qquad t \leq [0.5T] \\
y_t = \mu_2 + \nu_2 \Psi_t + e_t, &\qquad \text{if} \qquad t > [0.5T],
\end{aligned}
\tag{10}
$$

and we are testing for the break in both parameters, i.e. $\mathbf{z}_t = \{1, \Psi_t\}$. The results appear in Table 2. In the case of a single break, the power of our procedure (either LM_1 or LM_3) is generally somewhat higher than that of the Bai-Perron sequential procedure. The test based on the first-order Taylor approximation performs a bit better than the one based on the third-order approximation because in the case of a single break, the higher-order auxiliary terms do not carry helpful extra information. Correcting for non-existing serial correlation does not seem to have a large effect on the power of our procedure. The BP procedure, on the other hand, can lose as much as 15 percentage points of its power when prewhitening is applied. When prewhitening is not applied, the results are similar to the ones of our LM_1 test, see panels (c), (f), (g) and columns labelled "BP" and "BP$_{PW}$" in Table 2.

Second, we consider two different DGPs (and two sample sizes) with one break and serially correlated errors. The model has the following general form:

$$y_t = \mu_1 + v_t, \qquad \text{if} \qquad t \le [0.5T]$$
$$y_t = \mu_2 + v_t, \qquad \text{if} \qquad t > [0.5T], \tag{11}$$

where $v_t = 0.5v_{t-1} + e_t$ and we are testing for the break in the intercept, i.e. $\mathbf{z}_t = \{1\}$. The results appear in Table 3. Again, our LM_1 test does about as well as the BP sequential procedure. It is easier to detect small breaks with procedures without serial correlation correction than with procedures with correction. A problem arises because the correction tends to partially absorb the break. This seems to be true for both methods. If the jump (the break is in the intercept) and sample size are large enough, then accounting for serial correlation pays off (panel (k)), otherwise the effect is rather the opposite. In three cases out of four, the sequential method of BP has an advantage of a few percentage points in power, but the differences are not large. Differences become somewhat larger if one did (erroneously) not apply the prewhitening technique when using BP's procedure, especially for smaller breaks and sample sizes.

4.2.2 Two breaks

To see how the procedures compare to each other in the presence of multiple structural breaks, we simulate data from the following model:

$$\begin{aligned}
y_t &= \mu_1 + \nu_1 \Psi_t + e_t, &\text{if} \quad & 1 < t \le [T/3] \\
y_t &= \mu_2 + \nu_2 \Psi_t + e_t, &\text{if} \quad & [T/3] < t \le [2T/3] \\
y_t &= \mu_3 + \nu_3 \Psi_t + e_t, &\text{if} \quad & [2T/3] < t < T.
\end{aligned} \tag{12}$$

In model (12) there are two equally-spaced breaks and all parameters of the model are potentially subject to change ($\mathbf{z}_t = \{1, \Psi_t\}$), and the errors are serially uncorrelated and homoskedastic. The results are presented in Table 4. If there are only breaks in the intercept, however large, the power loss for our procedure is substantial when correcting for non-existing serial correlation. As an example, the power may drop from 70% to 10%, see panel (g) in Table 4. This suggests that a break in the intercept can be "explained" by adding dynamic structure to the model. This, however is common practice in cases where there is no prior information about the dynamic behaviour of y_t. The loss in power is only a few percentage points when the other parameters change as well. BP for some reason often gain from correcting

11

for non-existing serial correlation (and often even more so when prewhitening the data) and their sequential procedure with (erroneous) correction is working better than ours.

In the presence of multiple breaks of opposite direction, the test based on the third-order Taylor expansion (LM_3) generally works better than the one based on first-order approximation (LM_1). Differences in the performance can be large, see panels (g) and (i), for example. This can be explained by the added flexibility in the third-order approximation that allows for nonmonotonic and asymmetric parameter nonconstancy. The sequence based on LM_1 still works better than the one based on LM_3 when the change in the parameters is gradual, see panels (c), (f), (j) – (l).

To further study the properties of the tests, we simulate data with intercept shifts and serially correlated errors. That is,

$$
\begin{aligned}
y_t &= \mu_1 + v_t, & \text{if} & & 1 < t \le [T/3] \\
y_t &= \mu_2 + v_t, & \text{if} & & [T/3] < t \le [2T/3] \\
y_t &= \mu_3 + v_t, & \text{if} & & [2T/3] < t < T,
\end{aligned}
\tag{13}
$$

where $v_t = 0.5 v_{t-1} + e_t$ with $\{e_t\} \sim \text{nid}(0, 1)$. We focus on cases where the mean returns to its old value at the second break, i.e $\mu_1 = \mu_3 = 0$. The results can be found in Table 5. When no correction is carried out for the serial correlation present in the data, our procedure (LM_3) clearly dominates and the procedure of BP selects models with $m \ge 3$ more often than parsimonious models, see panels (c) – (f). Our procedure also dominates when the autocorrelation is accounted for, except when shifts are small and samples short. As one would expect, the test sequence based on the first-order Taylor approximation has no power at all, but the one based on the third-order approximation performs well. In BP case, prewhitening improves the power when breaks are large and samples long.

We replicate two more experiments from Bai and Perron (2006). The data is generated from equation (12) but allowing the distribution of errors and variables to change across segments. That is, we use:

$$
\begin{aligned}
\Psi_t^* &\sim \text{nid}(\varsigma_1, 1), & \text{if} & & 1 < t \le [T/3] \\
\Psi_t^* &\sim \text{nid}(\varsigma_2, 1), & \text{if} & & [T/3] < t \le [2T/3] \\
\Psi_t^* &\sim \text{nid}(\varsigma_3, 1), & \text{if} & & [2T/3] < t < T
\end{aligned}
\tag{14}
$$

and

$$
\begin{aligned}
e_t &\sim \text{nid}(0, \sigma_1^2), &\text{if} \quad& 1 < t \le [T/3] \\
e_t &\sim \text{nid}(0, \sigma_2^2), &\text{if} \quad& [T/3] < t \le [2T/3] \qquad (15) \\
e_t &\sim \text{nid}(0, \sigma_3^2), &\text{if} \quad& [2T/3] < t < T
\end{aligned}
$$

in (12).

In Table 6 we report the results of one of the experiments where the fixed parameters (regression coefficients) are set as follows: $\nu_1 = 1$, $\nu_2 = 1.5$, $\nu_3 = 0.5$ and $\mu_1 = 0$, $\mu_2 = 0.5$, $\mu_3 = -0.5$. The minimum relative regime size is set to $\epsilon_R = 0.15$ as in Bai and Perron (2006). The changes in error variance and in the mean of Ψ_t^* are of the same type: starting from value one, jumping to a higher level, and then returning to value one again. That is, for every DGP in this experiment $\sigma_1^2 = \sigma_3^2 = 1$ and $\varsigma_1 = \varsigma_3 = 1$.

The results in first few columns in Table 6 concern the case where no correction for serial correlation and/or heteroskedasticity is made. Columns 6 – 9 refer to the case most likely to be encountered in practice, which is correcting for serial correlation but not accounting for changes in error variance. The last column covers the results for the correct variant of BP's test sequence. The test statistics used there account for heteroskedasticity and do not allow for serial correlation correction. That test performs exceptionally well and requires no further comment. Currently there does not exist an ST counterpart to it.

When no correction is undertaken for heteroskedasticity or serial correlation, the BP procedure excels. That is to be expected, as their test accommodates the possibility of heterogenous regressors and, allowing for that possibility, is highly recommended by BP. Although the best of the LM tests is somewhat less powerful than the sequential procedure of BP, it is able to choose $m = 2$ frequently enough, except for cases when the sample size is small and either $\varsigma_2 = 4$ or $\sigma_2^2 = 4$ or both. In situations like that, parsimonious models may get selected more often than a model with two breaks. If one corrects for the serial correlation, then results are mixed: no test clearly dominates the other and the power of both methods decreases. BP can lose even as much as 74 percentage points (see panel (d)) when prewhitening is used, and even more when data are not prewhitened. At least in half of the cases, see panels (c), (d), (g), (h), the correction (partly) absorbs the changes, and models that are too parsimonious are selected most frequently.

Table 7 reports the results of the other experiment. The DGPs are such that the intercept and slope parameters increase at breakpoints gradually, i.e. $\nu_1 = 1$,

$\nu_2 = 1.5$, $\nu_3 = 2$ and $\mu_1 = 0$, $\mu_2 = 0.5$, $\mu_3 = 1$. Even here $\sigma_1^2 = \sigma_3^2 = 1$ and $\varsigma_1 = \varsigma_3 = 1$ in all cases, whereas the second segment mean and variance differ from the values above. Again, the test with exactly correct setup has the best performance, see the last column in Table 7. The uncorrected[3] version of BP performs somewhat better than ours in some cases, because it has the advantage of explicitly allowing for heterogenous regressors. For larger samples, our procedure is able to select $m = 2$ reasonably frequently. In this case the serial correlation correction has a rather devastating effect on the procedure of BP. Our procedure loses some power as well but not nearly as much as BP and has superior power in four cases out of eight and about equal power in one more.

The results of our technique depend somewhat on the choice of the discount coefficient τ. It is clear that increasing τ makes the strategy less parsimonious. When $\tau = 1$, we retain the same nominal level for each test in the sequence and are more likely to choose less parsimonious models than if we choose $\tau < 1$. When setting $\tau = 1$, our ST-procedure can lose up to 7 percentage points in power for the current DGPs with one break compared to the case $\tau = 0.5$. The hypothesis of only one break is rejected somewhat more frequently and some probability mass is shifting from $P(\widehat{m} = 1)$ to $P(\widehat{m} = 2)$. On the other hand, for models with two breaks we can gain up to 15 percentage points in power. This is true for configurations where it was previously difficult to detect the second break. We may also lose a little in the cases where the number of breaks was estimated precisely, since now models with more than two breaks have a chance to be selected. Setting $\tau < 0.5$ has the opposite effect - parsimony is strongly preferred and finding the second break becomes more difficult.

We can also conclude that it is not necessary to set the minimum regime size equal to $15 - 20\%$ for the "corrected" cases when the ST-procedure is applied. One could easily set the minimum regime length to be $5 - 10\%$ of the total sample size, without much affecting the power of our test. Depending on the DGP, the average change would be less than one percentage point and maximum gains and losses about three percentage points.

---

[3]In their original paper BP set the minimum regime size to $\epsilon_R = 0.20$ to ensure tests with adequate sizes. In practice, the heterogeneity of error terms might not be known in advance and using a large minimum regime size with Uncorrected version of the test is not justified. By setting $\epsilon_R = 0.20$ already here BP gain up to 13 percentage points in power (for panels (a) and (h) in Table 7). The power of our procedure does not depend on the regime size in this exercise; the power does not vary by more than 0.5 percentage points.

4.2.3 Dynamic models and breaks

It is not completely clear, however, how one would in practice handle the problem of detecting breaks in the presence of autocorrelation. The design of the simulations just discussed implicitly suggests that the model builder is primarily prepared for finding a break in the intercept. The conditional mean of the simulated models has a very simple structure, and the error process is assumed to be autocorrelated. An interesting question is what would happen if the breaks were of more general character. In order to illuminate this situation, we simulated data from an AR(2) model with two structural breaks in the dynamic behaviour of the process:

$$y_t = \begin{cases} 2.7 + 0.8y_{t-1} - 0.2y_{t-2} + \varepsilon_t & \text{if} \quad 1 < t \leq [T/3] \\ 0.3 - 0.2y_{t-1} + 0.5y_{t-2} + \varepsilon_t & \text{if} \quad [T/3] < t \leq [2T/3] \\ 1 + 0.7y_{t-1} - 0.3y_{t-2} + \varepsilon_t & \text{if} \quad [2T/3] < t < T. \end{cases} \tag{16}$$

In (16) each segment is covariance stationary and an example of a generated series can be found in Figure 1.
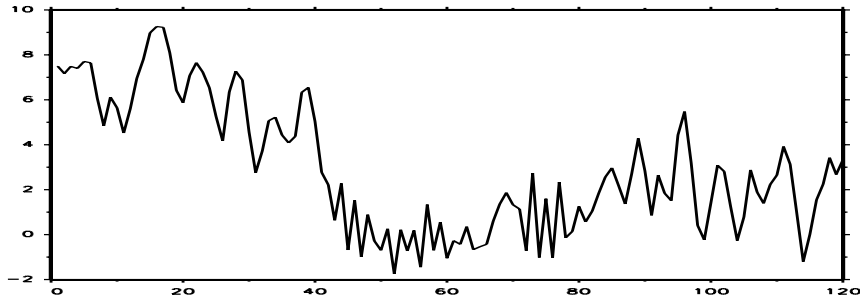
Figure 1: Data simulated according to (16)

We consider the following three strategies for proceeding that are also supported in the GAUSS code of BP:

- Strategy (1): One makes the assumption that there are breaks in the overall unconditional mean of the process but that the dynamics are not changing over time. In that case, when using BP's sequential method, one would only test for breaks in the intercept and correct for possible serial correlation in the errors using nonparametric methods.

- Strategy (2): One assumes that both the mean and the dynamics of the process are changing over time, but that the distribution of regressors should be the same in every segment. This would mean that the AR order of the model is

15

first selected by an appropriate information criterion and the model is then tested for structural breaks using the homogenous version of BPs sequential procedure.

- Strategy (3): One proceeds as in Strategy (2) but follows the recommendations of BP and relaxes the homogeneity assumption.

In Table 8 we report the model selection frequencies of the sequential procedure of BP for those three strategies and the model selection frequencies of the "practitioner's strategy" (similar to the second strategy above) when using our ST-method. Strategy (1) is using nonparametric correction, and the minimum regime size should be set to $15 - 20\%$ of the sample size, according to BP-s suggestions. We let $\epsilon_R = 0.20$. For the rest of the strategies the minimum regime size should not have a large effect on the outcome, so we set $\epsilon_R = 0.05$. It is clear from the table that prewhitening and using the HAC estimator is about sufficient in this specific case if one sets the minimum regime size equal to 20%. Failing to do so will affect the results considerably. For regime lengths $\epsilon_R = 0.05, 0.10, 0.15$ the correct decisions are made in $42.45\%, 55.10\%$ and 63.20% of the time, respectively, instead of 72% that is reported in Table 8. The frequency of choosing the correct number of breaks in the column corresponding to the second strategy is very low. Comparing the results in columns (2) and (3), it is obvious why Bai and Perron strongly recommend one allows for different variances of regressors across segments, or in this case, a different variance for y in each regime. It appears that this assumption has a large positive effect on power. It is rather striking how much the results in columns (2) and (3) can differ from each other. Both strategies also gain power when one increases the minimum required regime length. With $\epsilon_R = 0.20$ the correct decision is made in about 96% of cases. ST-procedures perform as expected, LM_3 is more powerful than LM_1 because there are two breaks in the simulated model. The power is robust to the regime length selected. This small experiment indicates that model uncertainty is a serious issue and that results depend on the modelling approach used in the study.

5   Empirical example

We consider an applications of the procedures presented in this paper. The US ex-post real interest rate series that has been studied in Bai and Perron (2003a) and González and Teräsvirta (2006), among others. The conclusions from these studies are that there are two to three breaks (or shifts) in the mean of the series.

The US ex-post real interest rate is defined as the three-month treasury bill rate deflated by the consumer price index (CPI) inflation rate. Figure 2 presents the quarterly series containing 103 observations for the period 1961:1-1986:3. Looking at the series one may, indeed, suspect the presence of two or three breaks.
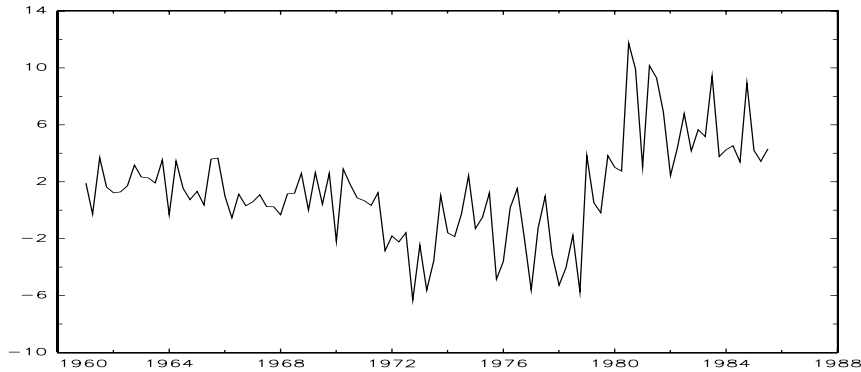
Figure 2: US ex-post real interest rate 1961:1 - 1986:3

Below we present the results from testing sequences at the 5% level of significance. If one is interested in finding out the number of changes only in the mean of the series, one should consider the test specification in Bai and Perron (2003a), that is identical to the Strategy (1) in Section 4.2.3. The only parameter to be tested for instability in this case is the intercept. Following BP, we set the minimum regime size to 15% of the total sample size, take care of serial correlation through pre-whitening and allow the residual variances be different across segments. As a result, we find three breaks, estimated at 1966:4, 1972:3 and 1980:3.

To use any of the other strategies or the sequential procedure proposed in this paper, one has to choose the lag length k, initially assumed equal for every regime. It is selected from the linear autoregressive model using the Schwarz Bayesian information criterion and the Breusch-Godfrey LM test of no error autocorrelation. This results in the estimated lag length $\hat{k} = 3$. If we then follow Strategy (2) we detect three breaks at: 1967:2, 1974:3 and 1979:4. The number of breaks coincides with the results form Strategy (1), but the location of them does not. Following

17

Strategy (3) by relaxing the homogeneity restriction renders two breaks at 1974:3 and 1979:4.

Given the nature of the series (changes in opposite directions) we would expect the testing sequence based on the third-order Taylor expansion have more power than the one based on the first-order expansion. For LM_1-based sequence the first p-value is equal to 0.485 and thus linearity is not rejected. LM_3-based test sequence is able to detect the nonlinearity. The p-value for the presence of a structural break is 0.0053 and the most prominent break found at 1979:4. At the second step, the linearity is again rejected, with $p = 0.0134$ and the second break is found at 1974:3. The third test in our sequence renders a p-value 0.048 and the presence of an additional break is rejected. If we would decide not to halve the significance level on each step the third p-value would signal the presence of another break. Summary of the results for all approaches considered can be found in Table 9.

6   Conclusions

In this paper we show how a smooth transition regression approximation to a piecewise linear structural break model is useful in determining the number of breaks in the latter model when it is not known in advance. The approach proposed and simulated in the paper is based on sequential hypothesis testing and is simple to apply in practice. The whole procedure is based on standard inference and the user can control the overall significance level of the tests in the sequence. In addition, no restrictions are imposed on the number of changing parameters. The simulations show that our procedure is well-sized and works well in comparison with the sequential procedure suggested by Bai and Perron. Neither of the alternatives dominates the other in small and moderate samples.

The examples discussed above show that the results of both our and Bai and Perron's approaches depend on the way the error autocorrelation is being taken care of. Adding lags to the model is a common practice, but sometimes small breaks can get absorbed by the extra dynamics. Then again, in practice a firm knowledge of the presence of a break is rather an exception than rule and a casual modeller would add lags to the model. One has to be careful when applying Bai and Perron's technique as well. The results can depend heavily on the assumptions one is or is not willing to make about the error term and series at hand. Allowing for different distributions of covariates in different segments helps a great deal, but an unnecessary prewhitening

can considerably weaken the procedure's ability to detect breaks.

Overall, our ST-method can be considered a complement to the classical approach of Bai and Perron. Our procedure may be extended to accommodate heteroskedasticity by making the error variance change over time at the same points as the mean. This extension is, however, left for further research.

References

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

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory 13*, 315–352.

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

Bai, J., R. L. Lumsdaine, and J. H. Stock (1998). Testing for and dating common breaks in multivariate time series. *Review of Economic Studies 65*, 395–432.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*, 47–78.

Bai, J. and P. Perron (2003a). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*, 1–22.

Bai, J. and P. Perron (2003b). Critical values for multiple structural change tests. *Econometrics Journal 6*, 72–78.

Bai, J. and P. Perron (2006). Multiple structural change models: A simulation analysis. Econometric Theory and Practice: Frontiers of Analysis and Applied Research, D. Corbea, S. Durlauf and B. E. Hansen (eds.), Cambridge University Press.

Ben Aïssa, M. S., M. Boutahar, and J. Jouini (2004). Bai and Perron's and spectral density methods for structural change detection in the US inflation process. *Applied Economics Letters 11*, 109–115.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica 28*, 591–605.

Eitrheim, Ø. and T. Teräsvirta (1996). Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics 74*, 59–75.

Goldfeld, S. M. and R. E. Quandt (1972). *Nonlinear Methods in Econometrics.* Amsterdam: North-Holland.

González, A. and T. Teräsvirta (2006). Modelling autoregressive processes with a shifting mean. SSE/EFI Working Paper Series in Economics and Finance No. 637, Stockholm School of Economics.

Lin, C.-F. and T. Teräsvirta (1994). Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics 62*, 211–228.

Liu, J., S. Wu, and J. V. Zidek (1997). On segmented multivariate regressions. *Statistica Sinica 7*, 497–525.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988). Testing linearity against smooth transition autoregressive models. *Biometrika 75*, 491–499.

Prodan, R. (2006). Potential pitfalls in determining multiple structural changes with an application to purchasing power parity. *Journal of Business and Economics Statistics. Forthcoming.*

Strikholm, B. and T. Teräsvirta (2005). Determining the number of regimes in a threshold autoregressive model using smooth transition autoregressions. SSE/EFI Working Paper Series in Economics and Finance No. 578, Stockholm School of Economics.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association 89*, 208–218.

Teräsvirta, T. (1998). Modeling economic relationships with smooth transition regression. In A. Ullah and D. E. Giles (Eds.), *Handbook of Applied Economic Statistics*, pp. 507–552. Dekker, New York.

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

Tables

Table 1: Selection frequencies of Bai and Perron sequential procedure (BP) and ST-procedures (LM_1 and LM_3). Data are generated with no breaks, i.e. $m = 0$.

Model	Choice	Uncorrected			Corrected			
		BP	LM_1	LM_3	BP_{PW}	BP	LM_1	LM_3
	$m = 0$	**95.15**	**94.15**	**95.55**	**93.60**	**94.35**	**94.25**	**95.55**
(a)	$m = 1$	4.70	5.85	4.15	6.15	5.55	5.75	3.95
	$m = 2$	0.15	0.00	0.30	0.25	0.10	0.00	0.45
	$m = 0$	**95.95**	**94.45**	**95.55**	**84.40**	**94.00**	**94.55**	**95.50**
(b)	$m = 1$	3.90	5.55	4.35	13.50	5.80	5.45	4.00
	$m = 2$	0.15	0.00	0.10	1.95	0.20	0.00	0.50
	$m = 0$	**94.75**	**94.75**	**94.95**				
(c)	$m = 1$	5.05	5.10	4.65				
	$m = 2$	0.15	0.15	0.35				
	$m = 0$	**46.90**	**73.30**	**53.90**	**91.90**	**86.20**	**93.75**	**93.15**
(d)	$m = 1$	28.15	24.85	30.35	7.45	12.70	6.25	5.95
	$m = 2$	16.85	1.85	12.85	0.65	1.05	0.00	0.80
	$m = 0$	**73.75**	**84.35**	**77.15**	**97.70**	**90.55**	**94.60**	**93.95**
(e)	$m = 1$	20.70	15.50	18.85	2.20	9.15	5.40	5.10
	$m = 2$	4.95	0.15	3.85	0.10	0.30	0.00	0.85
	$m = 0$	**99.85**	**99.70**	**99.90**	**97.35**	**97.40**	**97.20**	**97.70**
(f)	$m = 1$	0.15	0.30	0.10	2.55	2.45	2.80	2.10
	$m = 2$	0.00	0.00	0.00	0.10	0.15	0.00	0.20

Notes: The table contains selection frequencies in per cent based on 2000 Monte Carlo replications. Columns labelled "Uncorrected" contain the results when not correcting for serial correlation, and columns labelled "Corrected" correspond to tests that correct for serial correlation either non-parametrically or parametrically. The column labelled "BP_{PW}" contains the results of the BP sequential test when prewhitening is applied before estimating the long-run covariance matrix. The columns labelled "LM_1" and "LM_3" correspond to the linearity tests that make use of the first-order and third-order Taylor expansions, respectively.

Table 2: Model selection frequencies of the uncorrected and corrected versions of the BP and ST-procedures. Data are generated from equation (10), i.e. errors are uncorrelated and $m = 1$.

Model	Choice	Uncorrected BP	Uncorrected LM_1	Uncorrected LM_3	Corrected BP_{PW}	Corrected BP	Corrected LM_1	Corrected LM_3
Change in the intercept only: $\nu_1 = \nu_2 = 1$								
(a) $\mu_1 = 0$	$m = 0$	55.75	46.50	58.80	36.75	44.25	47.00	59.65
$\mu_2 = 0.5$	$m = 1$	**42.60**	**52.90**	**39.90**	**54.65**	**53.95**	**52.35**	**38.80**
$T = 120$	$m = 2$	1.60	0.55	1.20	8.00	1.80	0.65	1.20
(b) $\mu_1 = 0$	$m = 0$	20.20	14.90	25.25	14.25	13.60	15.35	25.75
$\mu_2 = 0.5$	$m = 1$	**77.85**	**83.40**	**73.40**	**78.00**	**84.60**	**83.25**	**72.90**
$T = 240$	$m = 2$	1.95	1.65	1.35	7.70	1.75	1.40	1.30
(c) $\mu_1 = 0$	$m = 0$	0.90	0.85	2.95	1.35	0.45	2.25	5.25
$\mu_2 = 1$	$m = 1$	**95.00**	**96.75**	**94.65**	**80.90**	**95.20**	**95.85**	**92.65**
$T = 120$	$m = 2$	4.10	2.30	2.20	16.95	4.30	1.90	1.75
Change in the slope only: $\mu_1 = \mu_2 = 0$								
(d) $\nu_1 = 1$	$m = 0$	21.85	15.40	26.75	11.55	14.75	15.85	27.10
$\nu_2 = 1.5$	$m = 1$	**75.45**	**83.15**	**71.25**	**73.50**	**82.30**	**82.80**	**70.80**
$T = 120$	$m = 2$	2.65	1.45	1.95	14.35	2.95	1.30	1.80
(e) $\nu_1 = 1$	$m = 0$	0.90	0.70	2.30	0.45	0.50	0.70	2.25
$\nu_2 = 1.5$	$m = 1$	**96.25**	**97.60**	**95.60**	**88.95**	**97.15**	**97.60**	**95.85**
$T = 240$	$m = 2$	2.75	1.70	2.00	10.30	2.20	1.65	1.90
Change in all parameters: $\nu_1 = 1$, $\mu_1 = 0$								
(f) $\nu_2 = 1.5$	$m = 0$	0.30	0.30	1.10	0.10	0.15	0.45	1.55
$\mu_2 = 0.5$	$m = 1$	**95.75**	**97.15**	**96.40**	**82.15**	**95.20**	**97.30**	**96.20**
$T = 120$	$m = 2$	3.90	2.40	2.30	17.00	4.60	2.25	1.95
(g) $\nu_2 = 2$	$m = 0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\mu_2 = 1$	$m = 1$	**95.40**	**97.10**	**97.00**	**81.65**	**95.20**	**97.45**	**96.95**
$T = 120$	$m = 2$	4.55	2.70	2.85	17.55	4.75	2.55	2.65

Notes: The table contains selection frequencies in per cent based on 2000 Monte Carlo replications. Columns labelled "Uncorrected" contain the results when not correcting for serial correlation, and columns labelled "Corrected" correspond to tests that correct for serial correlation either nonparametrically or parametrically. The column labelled "BP_{PW}" contains the results of the BP sequential test when prewhitening is applied before estimating the long-run covariance matrix. The columns labelled "LM_1" and "LM_3" correspond to the linearity tests that make use of the first-order and third-order Taylor expansions, respectively.

Table 3: Model selection frequencies of the uncorrected and corrected versions of the BP and ST-procedures. Data are generated from equation (11), i.e. errors are serially correlated and $m = 1$.

	Model	Choice	Uncorrected			Corrected			
			BP	LM_1	LM_3	BP_{PW}	BP	LM_1	LM_3
	$\mu_1 = 0$	$m = 0$	27.15	46.55	34.35	75.35	64.80	78.35	82.25
(h)	$\mu_2 = 0.5$	$m = 1$	**38.20**	**48.35**	**45.40**	**24.10**	**33.55**	**21.60**	**16.25**
	$T = 120$	$m = 2$	24.15	4.95	16.60	0.55	1.65	0.05	1.30
	$\mu_1 = 0$	$m = 0$	14.85	30.00	23.05	62.30	52.05	62.60	71.60
(i)	$\mu_2 = 0.5$	$m = 1$	**41.25**	**60.50**	**51.20**	**36.60**	**46.05**	**37.20**	**26.30**
	$T = 240$	$m = 2$	26.60	9.15	20.25	1.05	1.85	0.20	1.70
	$\mu_1 = 0$	$m = 0$	4.95	10.80	7.80	32.85	23.60	41.00	52.30
(j)	$\mu_2 = 1$	$m = 1$	**49.35**	**75.20**	**62.90**	**64.20**	**70.80**	**58.85**	**45.35**
	$T = 120$	$m = 2$	30.90	13.65	23.85	2.85	5.40	0.15	1.90
	$\mu_1 = 0$	$m = 0$	0.40	1.40	0.75	8.15	5.00	11.15	19.10
(k)	$\mu_2 = 1$	$m = 1$	**43.85**	**80.20**	**66.60**	**89.05**	**89.80**	**86.50**	**77.75**
	$T = 240$	$m = 2$	32.75	17.45	25.85	2.75	5.15	2.35	2.65

Notes: The table contains selection frequencies in per cent based on 2000 Monte Carlo replications. Columns labelled "Uncorrected" contain the results when not correcting for serial correlation, and columns labelled "Corrected" correspond to tests that correct for serial correlation either nonparametrically or parametrically. The column labelled "BP_{PW}" contains the results of the BP sequential test when prewhitening is applied before estimating the long-run covariance matrix. The columns labelled "LM_1" and "LM_3" correspond to the linearity tests that make use of the first-order and third-order Taylor expansions, respectively.

Table 4: Model selection frequencies of the uncorrected and corrected versions of the BP and ST-procedures. Data are generated from equation (12), $m = 2$.

	Model	Choice	Uncorrected BP	LM_1	LM_3	Corrected BP_{PW}	BP	LM_1	LM_3
Change in the intercept only: $\nu_1 = \nu_2 = \nu_3 = 1$									
	$\mu_1 = 0$	$m = 0$	88.90	95.10	79.90	70.10	82.70	95.30	80.70
(a)	$\mu_2 = 0.5$	$m = 1$	9.60	4.70	16.90	17.65	13.45	4.55	14.95
	$\mu_3 = 0$	$m = 2$	**1.50**	**0.20**	**3.10**	**11.55**	**3.85**	**0.15**	**4.30**
		$m = 3$	0.00	0.00	0.10	0.70	0.00	0.00	0.05
	$\mu_1 = 0$	$m = 0$	57.60	97.15	29.60	52.00	57.25	97.55	39.25
(b)	$\mu_2 = 1$	$m = 1$	8.90	1.40	27.40	6.45	6.55	1.55	24.40
	$\mu_3 = 0$	$m = 2$	**31.30**	**1.40**	**42.60**	**40.10**	**36.10**	**0.90**	**35.95**
		$m = 3$	2.15	0.05	0.40	1.45	0.10	0.00	0.40
	$\mu_1 = 0$	$m = 0$	0.00	0.00	0.00	0.00	0.00	0.00	0.40
(c)	$\mu_2 = 1$	$m = 1$	41.50	47.50	64.95	25.05	31.40	67.25	79.70
	$\mu_3 = 2$	$m = 2$	**55.10**	**52.05**	**34.30**	**72.85**	**68.15**	**32.50**	**19.20**
		$m = 3$	3.20	0.45	0.75	2.10	0.45	0.25	0.70
	$\mu_1 = 0$	$m = 0$	0.00	10.40	0.05	0.45	0.00	57.50	17.45
(d)	$\mu_2 = 1$	$m = 1$	12.45	15.65	28.85	10.45	9.25	17.45	34.10
	$\mu_3 = -1$	$m = 2$	**83.45**	**73.10**	**70.25**	**88.55**	**90.70**	**24.85**	**47.80**
		$m = 3$	4.00	0.85	0.85	0.55	0.05	0.20	0.65
	$\mu_1 = 0$	$m = 0$	0.00	0.00	0.00	0.00	0.00	35.50	15.65
(e)	$\mu_2 = -1$	$m = 1$	14.80	18.55	30.15	11.50	10.90	23.20	38.10
	$\mu_3 = 2$	$m = 2$	**81.80**	**80.75**	**69.15**	**87.90**	**88.95**	**40.80**	**45.60**
		$m = 3$	3.30	0.70	0.70	0.60	0.15	0.50	0.65
	$\mu_1 = 0$	$m = 0$	0.00	0.00	0.00	0.00	0.00	0.05	0.45
(f)	$\mu_2 = 1$	$m = 1$	14.25	18.70	31.65	11.05	9.65	43.00	58.85
	$\mu_3 = 3$	$m = 2$	**80.95**	**80.50**	**67.50**	**88.40**	**90.20**	**56.35**	**39.45**
		$m = 3$	4.60	0.80	0.85	0.55	0.15	0.60	1.25
	$\mu_1 = 0$	$m = 0$	0.00	29.25	0.00	0.55	0.55	88.90	37.95
(g)	$\mu_2 = 2$	$m = 1$	0.00	0.05	0.05	0.00	0.00	0.90	2.75
	$\mu_3 = -1$	$m = 2$	**93.95**	**69.65**	**98.65**	**99.35**	**99.35**	**10.05**	**58.45**
		$m = 3$	5.85	1.05	1.30	0.10	0.10	0.15	0.85

Notes: See the *Notes* of Table 3.

Table 4: Model selection frequencies of the uncorrected and corrected versions of the BP and ST-procedures. Data are generated from equation (12), $m = 2$, (cont.).

Model	Choice	Uncorrected			Corrected			
		BP	LM_1	LM_3	BP_{PW}	BP	LM_1	LM_3
Change in the slope only: $\mu_1 = \mu_2 = \mu_3 = 0$								
	$m = 0$	78.05	95.30	60.65	58.60	74.45	95.35	61.35
(h) $\nu_1 = 1$	$m = 1$	14.35	4.05	25.85	13.90	14.00	3.85	23.30
$\nu_2 = 1.5$	$m = 2$	**7.25**	**0.60**	**13.15**	**26.15**	**11.45**	**0.80**	**15.20**
$\nu_3 = 1$	$m = 3$	0.35	0.05	0.35	1.35	0.10	0.00	0.15
	$m = 0$	20.05	96.20	6.55	31.40	53.60	96.30	7.60
(i) $\nu_1 = 1$	$m = 1$	1.25	0.65	7.90	1.25	1.10	0.75	7.65
$\nu_2 = 2$	$m = 2$	**72.20**	**3.15**	**84.85**	**65.40**	**44.95**	**2.95**	**84.20**
$\nu_3 = 1$	$m = 3$	6.35	0.00	0.70	1.95	0.35	0.00	0.55
	$m = 0$	0.25	0.10	0.50	0.30	0.50	0.10	0.55
(j) $\nu_1 = 1$	$m = 1$	86.70	86.45	90.70	56.30	75.85	87.20	91.35
$\nu_2 = 1.5$	$m = 2$	**12.55**	**13.45**	**8.55**	**42.15**	**23.55**	**12.65**	**7.85**
$\nu_3 = 2$	$m = 3$	0.50	0.00	0.25	1.25	0.10	0.05	0.25
	$m = 0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(k) $\nu_1 = 1$	$m = 1$	2.35	5.50	10.90	2.05	3.40	7.35	13.40
$\nu_2 = 2$	$m = 2$	**89.50**	**93.60**	**88.30**	**94.10**	**95.45**	**91.90**	**85.70**
$\nu_3 = 3$	$m = 3$	7.75	0.90	0.80	3.85	1.15	0.75	0.90
	$m = 0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(l) $\nu_1 = 1$	$m = 1$	60.95	62.10	74.00	35.70	50.15	66.70	78.85
$\nu_2 = 0.5$	$m = 2$	**37.50**	**37.45**	**25.55**	**62.50**	**49.55**	**33.05**	**20.65**
$\nu_3 = -0.5$	$m = 3$	1.55	0.45	0.45	1.80	0.30	0.25	0.50
Change in all parameters								
$\mu_1 = 0,\quad \nu_1 = 1$	$m = 0$	87.85	94.40	79.40	76.20	82.05	94.60	79.55
(m) $\mu_2 = 0.5, \nu_2 = 0.5$	$m = 1$	10.55	5.40	17.25	15.85	14.70	5.20	15.95
$\mu_3 = 0,\quad \nu_3 = 1$	$m = 2$	**1.55**	**0.20**	**3.10**	**7.60**	**3.20**	**0.20**	**4.30**
	$m = 3$	0.05	0.00	0.25	0.35	0.05	0.00	0.20
$\mu_1 = 0, \nu_1 = 1$	$m = 0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(n) $\mu_2 = 1, \nu_2 = 1.5$	$m = 1$	13.75	15.90	27.55	13.50	9.65	17.80	30.75
$\mu_3 = 2, \nu_3 = 2$	$m = 2$	**83.70**	**83.15**	**71.65**	**86.15**	**90.35**	**81.50**	**68.55**
	$m = 3$	2.55	0.95	0.80	0.35	0.00	0.70	0.70
$\mu_1 = 0, \nu_1 = 1$	$m = 0$	0.00	0.00	0.00	0.00	0.00	3.75	0.30
(o) $\mu_2 = 1, \nu_2 = 2$	$m = 1$	16.20	20.40	30.45	20.15	12.20	21.90	32.50
$\mu_3 = 2, \nu_3 = 1$	$m = 2$	**80.10**	**78.85**	**68.50**	**78.90**	**87.70**	**73.95**	**66.30**
	$m = 3$	3.55	0.75	1.05	0.95	0.10	0.40	0.90

Notes: See the *Notes* of Table 3.

Table 5: Model selection frequencies of the uncorrected and corrected versions of the BP and ST-procedures. Data are generated from equation (13), i.e. errors are serially correlated and $m = 2$.

Model	Choice	Uncorrected			Corrected			
		BP	LM_1	LM_3	BP_{PW}	BP	LM_1	LM_3
(a) $\mu_2 = 0.5$ $T = 120$	$m = 0$	37.65	74.30	43.10	88.75	82.50	94.95	88.40
	$m = 1$	25.20	22.90	31.55	9.90	14.30	5.05	8.90
	$m = 2$	**22.45**	**2.70**	**21.20**	**1.35**	**3.20**	**0.00**	**2.40**
	$m = 3$	10.75	0.10	4.15	0.00	0.00	0.00	0.30
(b) $\mu_2 = 0.5$ $T = 240$	$m = 0$	27.45	75.05	35.05	87.75	82.10	95.45	84.70
	$m = 1$	22.25	19.50	31.35	9.35	12.55	4.55	11.05
	$m = 2$	**27.00**	**5.20**	**28.25**	**2.90**	**5.35**	**0.00**	**4.15**
	$m = 3$	16.45	0.25	5.35	0.00	0.00	0.00	0.10
(c) $\mu_2 = 1$ $T = 120$	$m = 0$	18.75	77.15	22.05	81.80	74.75	97.45	79.05
	$m = 1$	14.15	14.20	25.15	11.95	14.85	2.55	13.15
	$m = 2$	**32.80**	**7.95**	**43.90**	**6.15**	**10.30**	**0.00**	**7.45**
	$m = 3$	24.45	0.70	8.90	0.10	0.10	0.00	0.35
(d) $\mu_2 = 1$ $T = 240$	$m = 0$	4.20	77.95	7.40	67.85	64.90	97.80	57.80
	$m = 1$	5.20	6.95	12.35	12.40	11.35	2.05	19.25
	$m = 2$	**39.55**	**14.05**	**64.95**	**19.55**	**23.40**	**0.15**	**22.80**
	$m = 3$	33.10	1.05	15.30	0.20	0.35	0.00	0.15
(e) $\mu_2 = 2$ $T = 120$	$m = 0$	0.35	85.35	1.30	75.85	78.05	99.80	63.15
	$m = 1$	0.65	1.45	2.00	2.90	2.15	0.20	9.35
	$m = 2$	**39.05**	**11.50**	**75.35**	**20.90**	**19.55**	**0.00**	**27.25**
	$m = 3$	38.50	1.70	21.35	0.35	0.25	0.00	0.25
(f) $\mu_2 = 2$ $T = 240$	$m = 0$	0.00	85.55	0.00	34.30	59.80	99.80	21.05
	$m = 1$	0.00	0.00	0.05	0.05	0.00	0.10	4.75
	$m = 2$	**35.60**	**12.15**	**75.25**	**64.90**	**39.95**	**0.10**	**73.45**
	$m = 3$	37.65	2.30	24.70	0.75	0.25	0.00	0.75
(g) $\mu_2 = 4$ $T = 120$	$m = 0$	0.00	96.95	0.00	92.70	99.25	100.00	74.00
	$m = 1$	0.00	0.00	0.00	0.00	0.00	0.00	1.20
	$m = 2$	**37.50**	**2.25**	**72.75**	**7.30**	**0.75**	**0.00**	**24.05**
	$m = 3$	38.90	0.80	27.25	0.00	0.00	0.00	0.75
(h) $\mu_2 = 4$ $T = 240$	$m = 0$	0.00	96.80	0.00	48.30	98.10	100.00	45.35
	$m = 1$	0.00	0.00	0.00	0.00	0.00	0.00	0.10
	$m = 2$	**36.35**	**2.35**	**72.40**	**51.70**	**1.90**	**0.00**	**53.15**
	$m = 3$	37.55	0.85	27.60	0.00	0.00	0.00	1.40

Notes: See the *Notes* of Table 3.

Table 6: Model selection frequencies of the uncorrected and corrected versions of the BP and ST-procedures. Uncorrelated but heterogenous data and errors across segments, $m = 2$, $\epsilon_R = 0.15$ for the corrected versions.

Model	Choice	Uncorrected			AC-Correct				$\sigma_i^2(u)$
		BP	LM_1	LM_3	BP_{PW}	BP	LM_1	LM_3	BP
	$m = 0$	0.00	20.85	0.00	0.70	1.00	42.90	5.35	0.00
(a) $\sigma_2^2 = 2$, $\varsigma_2 = 2$, $T = 120$ $m = 1$	$m = 1$	11.45	13.35	23.80	23.75	30.70	32.40	55.10	6.20
	$m = 2$	**83.90**	**65.30**	**75.60**	**56.80**	**65.65**	**24.45**	**39.25**	**91.95**
	$m = 0$	0.00	1.10	0.00	0.00	0.00	10.05	0.10	0.00
(b) $\sigma_2^2 = 2$, $\varsigma_2 = 2$, $T = 240$	$m = 1$	0.15	0.35	0.85	3.45	3.55	14.65	24.35	0.05
	$m = 2$	**96.60**	**97.55**	**98.30**	**72.95**	**94.35**	**74.75**	**74.75**	**98.95**
	$m = 0$	0.00	24.30	0.00	12.60	7.90	30.20	1.65	0.00
(c) $\sigma_2^2 = 2$, $\varsigma_2 = 4$, $T = 120$	$m = 1$	25.85	39.40	41.50	63.60	87.20	57.45	71.50	9.45
	$m = 2$	**70.35**	**36.10**	**57.75**	**19.45**	**4.85**	**12.20**	**26.30**	**88.75**
	$m = 0$	0.00	2.20	0.00	0.15	0.15	4.55	0.00	0.00
(d) $\sigma_2^2 = 2$, $\varsigma_2 = 4$, $T = 240$	$m = 1$	0.75	12.40	5.55	73.50	94.80	54.60	45.05	0.10
	$m = 2$	**95.90**	**84.35**	**93.55**	**21.70**	**5.00**	**40.15**	**54.10**	**98.55**
	$m = 0$	0.00	34.70	0.30	1.95	3.10	57.15	14.80	0.00
(e) $\sigma_2^2 = 4$, $\varsigma_2 = 2$, $T = 120$	$m = 1$	33.10	24.80	44.40	35.45	50.30	31.10	59.50	19.80
	$m = 2$	**62.70**	**40.35**	**54.65**	**47.70**	**45.10**	**11.75**	**25.35**	**78.80**
	$m = 0$	0.00	3.40	0.00	0.00	0.00	19.15	0.65	0.00
(f) $\sigma_2^2 = 4$, $\varsigma_2 = 2$, $T = 240$	$m = 1$	4.80	4.30	8.45	17.90	23.05	23.20	40.35	1.75
	$m = 2$	**91.00**	**91.25**	**90.40**	**63.45**	**75.15**	**57.45**	**58.30**	**97.35**
	$m = 0$	0.00	36.05	0.15	31.00	19.05	46.95	7.60	0.00
(g) $\sigma_2^2 = 4$, $\varsigma_2 = 4$, $T = 120$	$m = 1$	51.55	45.35	62.30	47.35	76.70	47.45	76.95	17.80
	$m = 2$	**44.75**	**18.50**	**36.90**	**18.25**	**4.25**	**5.60**	**15.25**	**79.80**
	$m = 0$	0.00	4.80	0.00	4.00	0.45	12.10	0.35	0.00
(h) $\sigma_2^2 = 4$, $\varsigma_2 = 4$, $T = 240$	$m = 1$	10.45	27.80	20.55	70.90	96.15	62.40	64.20	1.05
	$m = 2$	**85.30**	**66.95**	**78.35**	**21.35**	**3.40**	**25.30**	**35.00**	**96.75**

Notes: The table contains selection frequencies in per cent based on 2000 Monte Carlo replications. Columns labelled "Uncorrected" contain the results when not correcting for serial correlation, and columns labelled "AC-Correct" correspond to tests that correct for serial correlation either nonparametrically or parametrically. The column labelled "BP_{PW}" contains the results of the BP sequential test when prewhitening is applied before estimating the long-run covariance matrix. The column labelled "$\sigma_i^2(u)$" contains the results of the BP sequence when the correct specification of the test is applied. The columns labelled "LM_1" and "LM_3" correspond to the linearity tests that make use of the first-order and third-order Taylor expansions, respectively.

Table 7: Model selection frequencies of the uncorrected and corrected versions of the BP and ST-procedures. Uncorrelated but heterogenous data and errors across segments, $m = 2$.

		Uncorrected			AC-Correct				$\sigma_i^2(u)$
Model	Choice	BP	LM_1	LM_3	BP_{PW}	BP	LM_1	LM_3	BP
(a) $\sigma_2^2 = 2$ $\varsigma_2 = 2$ $T = 120$	$m = 0$	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00
	$m = 1$	35.75	38.90	56.95	41.15	59.25	51.80	67.65	21.80
	$m = 2$	**60.35**	**60.50**	**42.55**	**56.70**	**40.65**	**48.05**	**32.00**	**78.00**
(b) $\sigma_2^2 = 2$ $\varsigma_2 = 2$ $T = 240$	$m = 0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$m = 1$	0.80	1.90	3.85	6.80	8.65	11.95	17.70	0.35
	$m = 2$	**90.85**	**97.10**	**95.25**	**90.90**	**90.45**	**87.35**	**81.50**	**97.70**
(c) $\sigma_2^2 = 2$ $\varsigma_2 = 4$ $T = 120$	$m = 0$	0.00	0.00	0.00	5.00	0.00	0.00	0.00	0.00
	$m = 1$	46.50	65.40	65.25	81.00	98.75	72.85	74.80	20.40
	$m = 2$	**50.45**	**34.30**	**34.05**	**13.65**	**1.25**	**27.05**	**24.80**	**78.75**
(d) $\sigma_2^2 = 2$ $\varsigma_2 = 4$ $T = 240$	$m = 0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$m = 1$	2.15	20.75	10.45	83.25	99.00	42.05	31.60	0.15
	$m = 2$	**93.35**	**78.70**	**88.70**	**16.60**	**1.00**	**57.45**	**67.85**	**99.20**
(e) $\sigma_2^2 = 4$ $\varsigma_2 = 2$ $T = 120$	$m = 0$	0.00	0.00	0.10	0.85	0.20	0.00	0.25	0.00
	$m = 1$	75.00	75.00	86.50	61.05	86.80	79.55	88.75	58.90
	$m = 2$	**22.55**	**24.10**	**12.90**	**37.00**	**12.95**	**20.40**	**10.65**	**41.10**
(f) $\sigma_2^2 = 4$ $\varsigma_2 = 2$ $T = 240$	$m = 0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$m = 1$	17.35	17.70	34.10	35.95	51.40	29.95	49.65	8.75
	$m = 2$	**75.80**	**81.35**	**64.45**	**62.30**	**48.55**	**69.80**	**49.45**	**90.50**
(g) $\sigma_2^2 = 4$ $\varsigma_2 = 4$ $T = 120$	$m = 0$	0.00	0.05	0.00	15.60	0.00	0.05	0.00	0.00
	$m = 1$	76.80	84.55	84.70	71.95	99.35	87.40	87.85	36.10
	$m = 2$	**20.25**	**14.95**	**14.25**	**12.20**	**0.65**	**12.55**	**11.70**	**63.15**
(h) $\sigma_2^2 = 4$ $\varsigma_2 = 4$ $T = 240$	$m = 0$	0.00	0.00	0.00	1.75	0.00	0.00	0.00	0.00
	$m = 1$	23.75	49.30	40.50	81.05	99.55	60.55	54.80	2.00
	$m = 2$	**71.00**	**50.45**	**57.90**	**16.85**	**0.45**	**39.35**	**44.20**	**95.70**

Notes: See the *Notes* of Table 6.

Table 8: Model selection frequencies at 5% nominal level, data generated from (16).

Choice	BP(1)	BP(2)	BP(3)	LM_1	LM_3
$m = 0$	0.00	0.00	0.00	29.75	0.65
$m = 1$	27.55	3.40	5.85	10.60	21.35
$m = 2$	**72.00**	**23.95**	**86.05**	**59.10**	**76.80**
$m = 3$	00.45	24.75	7.45	1.40	1.20

Notes: The columns labelled "BP(i)", $i = 1, 2, 3$, refer to the Strategy (i) on page 15. The columns labelled "LM_1" and "LM_3" correspond to the linearity tests that make use of the first-order and third-order Taylor expansions, respectively.

Table 9: Selected break dates for the US ex-post real interest rate series.

BP(1)	BP(2)	BP(3)	LM_1	LM_3
1966:4	1967:2			
1972:3	1974:3	1974:3		1974:3
1980:3	1979:4	1979:4		1979:4

Notes: The columns labelled "BP(i)", $i = 1, 2, 3$, refer to the Strategy (i) on page 15. The columns labelled "LM_1" and "LM_3" correspond to the linearity tests that make use of the first-order and third-order Taylor expansions, respectively.