Combining p-values to Test for Multiple Structural Breaks in Cointegrated Regressions

Michele Bergamelli*
Cass Business School (UK)

Annamaria Bianchi†
Bergamo University (Italy)

Lynda Khalaf‡
Carleton University (Canada)

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

23rd December 2018

Abstract

We propose a multiple hypothesis testing approach to assess structural stability in cointegrating regressions. Underlying tests are constructed via a Vector Error Correction Model and generalize the reduced rank regression procedures of Hansen (2003). We generalize the likelihood ratio test proposed in Hansen (2003) to accommodate unknown break dates through the specification of several scenarios regarding the number and the location of the breaks. We define a combined p-value adjustment, which proceeds by simulating the entire dataset imposing the relevant null hypothesis. This framework accounts for both correlation of underlying tests and the fact that empirically, parameters of interest often pertain to a limited even though uncertain stylized-fact based change points. We prove asymptotic validity of the proposed procedure. Monte Carlo simulations show that proposed tests perform well in finite samples and circumvent Bonferroni-type adjustments. An application to the S&P 500 prices and dividends series illustrates the empirical validity of the proposed procedure.

Keywords: Structural Stability; Vector Error Correction Model; Multiple hypotheses test; Simulation Based Test.

JEL Classification: C12, C15, C32

*Cass Business School, City University of London, 106 Bunhill Row EC1Y 8TZ London (UK). E-mail: Michele.Bergamelli.2@cass.city.ac.uk.

†Department of Management, Economics and Quantitative Methods, Università degli Studi di Bergamo, Via dei Caniana, 2 -24127 BergamoUniversity of Bergamo (Italy). E-mail: annamaria.bianchi@unibg.it

‡Groupe de recherche en économie de l'énergie, de l'énvironnement et des ressources naturelles (GREEN) Université Laval, Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Economics Department, Carleton University. Mailing address: Economics Department, Carleton University, Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada. E-mail: Lynda_Khalaf@carleton.ca.

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

1 Introduction

Time series methods often assume some form of stationarity. When strictly stationary processes lack empirical support, less restrictive definitions including *e.g.* piecewise stationarity and cointegration provide viable alternatives. The former implies stationarity within time *windows*, which may be stated via change points and time-varying parameters.[1] In contrast, cointegration typically describes a time-invariant linear combination of non-stationary series. A cornerstone of the latter literature is Johansen (1991) which introduced the Vector Error Correction Model (VECM).[2]

In this paper, we develop more general procedures in which stationarity within time segments is allowed in the definition of cointegration. In VECMs, cointegration is defined via a reduced rank regression as follows. Given a p-dimensional vector X_t, consider the regression of ΔX_t on X_{t-1}, and *e.g.* a constant, possibly further deterministic terms and lags of ΔX_t. Let Π refer to the coefficient of X_{t-1} in the latter regression. Then cointegration implies $\Pi = \alpha\beta^\top$, where α and β are $p \times r$ matrices and r is the cointegration rank. Hansen (2003) generalizes this model to allow for a time varying coefficient of the form $\Pi(t) = \alpha(t)\beta(t)^\top$ where $\alpha(.)$ and $\beta(.)$ are piecewise constant. With reference to available relevant procedures (reviewed below), this framework has several advantages: (i) all parameters including variance/covariance terms are allowed to change; (ii) both null and alternative hypotheses are well defined reduced rank regressions; (iii) null hypotheses imposing ex-ante breaks tested against different breaking schemes are covered.

Hansen (2003) introduces the so called *generalized reduced rank regression* (G3R) method for estimation and inference, when break dates are given. This paper extends G3R-based tests to accommodate uncertain prior information on break dates. With unknown break dates, the most widely used approach assumes no information is available on change points. Sample-wide change point searches thus often imply combining a number of tests that grows with the sample size. On the other hand, researchers are frequently confronted with decisions on historical events, in which case assuming no prior information is counterfactual and thus statistically inefficient.[3] With

[1] For recent references, see for instance Chan et al. (2014); Matteson and James (2014); Preuss et al. (2015); Schnurr and Dehling (2017).

[2] For a recent perspective, see e.g. Liao and Phillips (2015).

[3] Examples include unexpected geopolitical disruptions, regulatory changes, major market crashes, announcements, etc.

empirical cointegration-based work, such events can be dated though not exactly, allowing for some uncertainty on effective impact and relevance.

In general, formal break tests when some albeit uncertain information is available on dates is scarce, perhaps because empirical process asymptotic theories do not (typically) apply. Dufour et al. (2004) propose finite sample breaks-in-variance tests restricting the search set to a prior window, using the Monte Carlo (MC) test method of Dufour (2006). As it applies to the present paper, the method can be summarized as follows. An induced test is defined to assess stability against a number of alternative hypotheses, each corresponding to several possible break dates, within a pre-specified search window. A simulation-based method[4] is implemented to obtain the combined p-value, resampling from the stability null hypothesis which is shown to be nuisance parameter free. Bernard et al. (2007) follow a similar strategy in multivariate regression. In this paper, we extend these approaches to the nuisance-parameter dependent and non-linear G3R context.

Our procedure can be summarized as follows. Given a null G3R model denoted $\mathcal{M}_0$, we define a number (say n) of alternative models denoted $\mathcal{M}_j$, $j = 1, ..., n$, each of which can fit plausible historical regime changes. $\mathcal{M}_0$ may include breaks, and is nested in each of $\mathcal{M}_1, ..., \mathcal{M}_n$, leading to n nested likelihood ratios. These are next combined into a minimum p-value statistic, for which a p-value is obtained using the MC method and consistent estimates of intervening parameters. In this way, the correlation structure within the data set as well as between the considered tests is replicated and thus accounted for. The number of combined statistics, n, can be large. We prove the validity of this procedure in the nuisance parameter dependent VECM context.

Our method of proof proceeds as follows.[5] We first modify the framework of Dufour (2006) to accommodate sequences of parameters converging to true values. This modification may prove useful beyond the specific problem at hand. Next, we prove that the null distribution of the statistic converges, though not necessarily to a known nor pivotal limiting distribution. To emphasize the usefulness of this result, we discuss a special case for which the distribution can be shown to be

[4] For further references on simulation-based multiple testing see Westfall and Young (1993), Ge et al. (2003), Dufour and Khalaf (2002), Beaulieu et al. (2007, 2013), and Dufour et al. (2015).

[5] Thereafter, we refer to the large n as the many priors case, the term large is not intended in its asymptotic interpretation.

asymptotically pivotal given specific nesting restrictions on the priors. In this case, validity of the bootstrap follows with no further requirements. The nesting scheme is not uninteresting, and may be viewed as an extension of the well known Cochran Theorem (Cochran, 1934) to the G3R context. Yet its specificity underscores the usefulness of our general proof. Another important contribution is a bound on the null distribution of the minimum p-value statistic, which we derive analytically, to validate the required general conditions.

Our approach provides useful contributions to the induced tests literature beyond our specific context. In particular, because size adjustments such as Bonferroni bounds are avoided altogether, power is no longer expected to deteriorate with large n. Power depends on the proximity of priors to the truth. But the MC approach allows to increase the number of priors to reflect uncertainty with no power losses. This increases the likelihood to cover the truth which addresses one of the major hurdles of induced tests.

These properties are investigated in finite-samples through an extensive Monte Carlo study. The study is conducted under the no-break null hypothesis as well as under more challenging multiple-break scenarios. We also compare the performance of our combined procedure relative to a joint test that embeds all alternatives, when feasible.[6] We show that in the small n case, the joint test does not outperform our minimum p-value alternative. As n is increased, both size and power of the joint test deteriorate. The MC method can stabilize its size, a fact we verify. However, the power advantage of our minimum p-value method remains evident.

An empirical application documents the usefulness of our proposed test. We analyze the relationship between the S&P500 price and dividend series, which caught the attention of the profession following Campbell and Shiller (1987, 1989). From their viewpoint, a pair of integrated variables related through a Present Value Model must cointegrate. Since then, related research has evolved in line with econometric work on cointegration.[7] As an alternative to the bubble motivated temporary explosiveness approach in e.g. Phillips et al. (2011), Phillips et al. (2015a), Phillips et al. (2015b) our results suggest a breaking cointegration relation.

To conclude, we note that the test by Hansen (2003) which we extended in this paper seems to

[6] We thank the editor and an anonymous referee who suggested to analyze the large n case as well as the comparison to a joint test.

[7] See e.g. Advanced Information Nobel Prize (2013).

be the only available multiple breaks test in VECM since the bulk of related works deals mainly with one shift. These include *e.g.* the fully modified OLS based method analyzed by Hansen (1992), the sup test of Bai et al. (1998), the fluctuation tests of Quintos (1997) or the LM tests of Seo (1998). Hansen and Johansen (1999) propose a useful specification test, with application to the US term structure of interest rates; see Hansen (2003, Section 5.2). Bierens and Martins (2010) propose a likelihood-ratio test for time-varying cointegration. Oka and Perron (2011) allow for more than a break, but they focus exclusively on the statistical properties of the break dates estimator and the model considered is not the VECM. The literature is more established within a single cointegrating equation framework; see Kejriwal and Perron (2008, 2010) or Bergamelli and Urga (2015) and the references therein.

The remainder of the paper is organized as follows: Section 2 summarizes the VECM formulation in presence of breaks based on Hansen (2003). In Section 3, we introduce the new test based on a minimum p-value approach, the related bootstrap procedures, and proofs of their asymptotic validity. Section 4 reports our simulation exercise, whereas Section 5 reports our empirical analysis. Section 6 concludes.

2 Formulation of the VECM in Presence of Multiple Breaks

Let $X = \{X_t\}_{t=1}^{T}$ be a p-dimensional data generating process (DGP) undergoing m regimes and thus affected by $m-1$ breaks at locations $T_0 = 0 < T_1 < \cdots < T_{m-1} < T_m = T$. A general VECM as in Hansen (2003) where all parameters may break takes the form

$$\Delta X_t = \alpha(t)\beta(t)^\top \ddot{X}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i(t)\Delta X_{t-i} + \Phi(t)D_t + \epsilon_t, \quad t = 1, \ldots, T \tag{1}$$

where $\{\epsilon_t\}$ are *i.i.d.* Gaussian with zero mean and variance matrix $\Omega(t)$, $\alpha(t)$ is the adjustment matrix, $\beta(t)$ the cointegrating matrix, $\ddot{X}_{t-1}$ consists of X_{t-1} and restricted deterministic variables while D_t is a vector of unrestricted deterministic variables. The time-varying parameters are piecewise constant given by

$$\alpha(t)\beta(t)^\top = \alpha_1\beta_1^\top \mathbf{1}_{1t} + \cdots + \alpha_m\beta_m^\top \mathbf{1}_{mt}, \tag{2}$$

$$\Gamma_i(t) = \Gamma_{1,i}\mathbf{1}_{1t} + \cdots + \Gamma_{m,i}\mathbf{1}_{mt}, \quad i = 1, \ldots, k-1, \tag{3}$$

$$\Phi(t) = \Phi_1\mathbf{1}_{1t} + \cdots + \Phi_m\mathbf{1}_{mt}, \quad \Omega(t) = \Omega_1\mathbf{1}_{1t} + \cdots + \Omega_m\mathbf{1}_{mt}, \tag{4}$$

where $\mathbf{1}_{jt} \equiv \mathbf{1}(T_{j-1}+1 \leq t \leq T_j)$, $j = 1, \ldots, m$, $\mathbf{1}\,(C)$ being the indicator function associated with the condition C, taking value 1 if C is true and 0 otherwise. Defining r_j, for $j = 1, \ldots, m$, the cointegrating rank of the j^{th} regime, the dimensions of the parameter matrices are the following: α_j is $(p \times r_j)$, β_j is $(p_1 \times r_j)$ where p_1 is the dimension of $\ddot{X}_t$, $\Gamma_i(t)$ is $(p \times p)$, $\Phi(t)$ is $(p \times q)$, and $\Omega(t)$ is $(p \times p)$. Let $Z_{0t} = \Delta X_t, Z_{1t} = (\mathbf{1}_{1t}\ddot{X}_{t-1}^\top, \ldots, \mathbf{1}_{mt}\ddot{X}_{t-1}^\top)^\top, Z_{2t} = (\mathbf{1}_{1t}\tilde{Z}_{2t}^\top, \ldots, \mathbf{1}_{mt}\tilde{Z}_{2t}^\top)^\top$, where $\tilde{Z}_{2t} = (\Delta X_{t-1}^\top, \ldots, \Delta X_{t-k+1}^\top, D_t^\top)^\top$. Z_{1t} is $(mp_1 \times 1)$ while, denoting with p_2 the number of variables in $\tilde{Z}_{2t}$, Z_{2t} is $(mp_2 \times 1)$. Then (1) can be rewritten as

$$Z_{0t} = AB^\top Z_{1t} + CZ_{2t} + \epsilon_t, \quad t = 1, \ldots, T, \tag{5}$$

where $A = (\alpha_1, \ldots, \alpha_m)$, $B = \text{diag}(\beta_1, \ldots, \beta_m)$, $C = (\Psi_1, \ldots, \Psi_m)$, $\Psi_j = (\Gamma_{j,1}, \ldots, \Gamma_{j,k-1}, \Phi_j)$, $j = 1, \ldots, m$, and diag(.,.) refers to a block-diagonal matrix with blocks as specified.

The required parameter structure, their identification as well as hypotheses about the presence of breaks can be obtained through restrictions of the form

$$\text{vec}(B) = \boldsymbol{H}\phi + h \tag{6}$$

$$\text{vec}(A, C) = \boldsymbol{G}\psi \tag{7}$$

where vec$(\cdot)$ is the vectorization operator, $\boldsymbol{H}$ is a known $[mp_1(r_1 + \cdots + r_m) \times p_\phi]$ matrix, h is a known $[mp_1(r_1 + \cdots + r_m) \times 1]$ vector, ϕ is a vector with p_ϕ free parameters and similarly $\boldsymbol{G}$ is a known $[p(r_1 + \cdots + r_m + mp_2) \times p_\psi]$ matrix, while ψ is a vector with p_ψ free parameters.

Stack the free parameters in A, B, C subject to (6) and (7), and the free parameters in the variance-covariance matrix $\Omega(t)$ of ϵ_t in θ and let Θ denote the relevant parameter space.

Let θ_0 denote the "true" parameter vector. Hansen (2003) introduces the G3R technique, which is cast in the likelihood framework and produces maximum likelihood estimates (**MLE**) for θ. For the purposes of this paper, given θ, an initial value for the level, and Gaussian errors, it is straightforward to draw from (5) imposing what is relevant from (6) and (7).

Let $\mathcal{M}_0$ and $\mathcal{M}_1$ be two models defined by restrictions of the (6) and (7) form, with m_0 and m_1 regimes, respectively. Assume that $\mathcal{M}_0$ is a submodel of $\mathcal{M}_1$ with q fewer parameters, and that $\mathcal{M}_0$ and $\mathcal{M}_1$ have the same cointegration rank in each subsample. To test the model $\mathcal{M}_0$ (null hypothesis $\mathcal{H}_0 : \theta_0 \in \Theta_0$, where Θ_0 is a non-empty subset of Θ) against model $\mathcal{M}_1$ (alternative hypothesis $\mathcal{H}_1$), the framework outlined above suggests likelihood ratio (**LR**) testing. Theorem 10 in Hansen (2003) proves that if breaks dates are known, under suitable conditions on the rank of the restriction matrices the test can be performed using

$$LR = T\left[\sum_{j_0=1}^{m_0} \rho_{j_0} \log|\hat{\Omega}_{j_0}| - \sum_{j_1=1}^{m_1} \rho_{j_1} \log|\hat{\Omega}_{j_1}|\right] \longrightarrow^{d} \chi^2(q), \tag{8}$$

where $\hat{\Omega}_{ji} = (T_j - T_{j-1})^{-1} \sum_{t=T_{j-1}+1}^{T_j} \hat{\epsilon}_{ti}\hat{\epsilon}_{ti}^T$, $\hat{\epsilon}_{ti} = Z_{0t} - \hat{A}_i\hat{B}_i^T Z_{1t} - \hat{C}_i Z_{2t}$, $i = 0, 1$, $j = 1, \ldots, m_i$, and $\rho_j = (T_j - T_{j-1})/T$.

3 Multiple Testing with Uncertain Break Dates

In the above, break dates are taken as given. Here we assume instead that some albeit uncertain information is available on dates: possible breaks can be broadly characterized so that n plausible scenarios can adequately express uncertainty about their number and location. These scenarios are treated as possible priors which seem realistic for various empirical purposes.

Our procedure can be summarized as follows. We test a null model $\mathcal{M}_0$ against a set of prior fixed alternatives $\mathcal{M}_1, \ldots, \mathcal{M}_n$, having $m_1, \ldots, m_n$ regimes and the same cointegration rank as $\mathcal{M}_0$ in each subsample. $\mathcal{M}_0$ is nested in each of $\mathcal{M}_1, \ldots, \mathcal{M}_n$ that are otherwise unrestricted as for the number and location of breaks. Each model is associated with restrictions of the (6) and (7) form. We denote $\mathcal{H}_1, \ldots, \mathcal{H}_n$ the corresponding hypotheses, where $\mathcal{H}_i : \theta_0 \in \Theta_i$, $i = 1, \ldots, n$, Θ_i is a non-empty subset of Θ and $\Theta_0 \subset \Theta_i$ for each i. The intuition is that prior knowledge of stylized facts or policies may suggest a number of n plausible scenarios, where uncertainty translates into different numbers of regimes m_i $(i = 1, \ldots, n)$ or different break dates $T_{i,1} < \cdots < T_{i,m_i-1} < T_{i,m_i}$. Formally, we test $\mathcal{H}_0$ against

$$\mathcal{H}_A = \bigcup_i \mathcal{H}_i : \theta_0 \in \bigcup_i \Theta_i. \tag{9}$$

With reference to Section 2, $\Theta = \bigcup_i \Theta_i$. Let $\mathcal{M}_i$ denote the model estimated under the i^{th} scenario and $LR_{T,i}$, for $i = 1, \ldots, n$, a LR test of $\mathcal{M}_0$ against $\mathcal{M}_i$ as defined in equation (8). To test $\mathcal{H}_0$ versus $\mathcal{H}_A$ in (9), we propose the following minimum p-value statistic

$$Q_T = 1 - \min_{1 \leq i \leq n} [1 - F_i(LR_{T,i})] = \max_{1 \leq i \leq n} F_i(LR_{T,i}). \tag{10}$$

where $F_i(\cdot)$ is the CDF of a $\chi^2(q_i)$ random variable with q_i being the difference between the number of parameters in $\mathcal{M}_i$ with respect to $\mathcal{M}_0$.

While it is possible to extend results to $n \to \infty$, Q_T is defined assuming n is finite yet can be large. This is inspired by the large scale multiple testing literature, where large n does not necessarily imply asymptotics on n itself; see e.g. Cao and Wu (2015) and references therein.

3.1 Simulation-based Multiple Test Correction

To analytically derive the exact distribution of Q_T under the null, we consider a simulation-based procedure. The intended correction requires simulating the relevant structural model under the null at estimated parameter values. The following algorithm describes our proposed parametric simulation-based multiple test adjustment, for any $\theta \in \Theta$.[8]

Step 1. Estimate by G3R a VECM under $\mathcal{H}_0$ (i.e., model $\mathcal{M}_0$) for X_t to obtain a matrix of residuals $(\hat{\epsilon}_1, \ldots, \hat{\epsilon}_T)^\top$ where $\hat{\epsilon}_t = Z_{0t} - \hat{A}_0 \hat{B}_0^\top Z_{1t} - \hat{C}_0 Z_{2t}$. Denote $\hat{\theta}_T$ the MLE for θ, computed under $\mathcal{H}_0$. Estimate the model under each alternative and compute the test statistic Q_{T0}.

Step 2. For any $\theta \in \Theta$, for $b = 1, \ldots, B$ independently repeat:

[8] Recall that θ stacks all parameters that describe the model, including the variance-covariance matrix.

(a). Draw $\epsilon_b = \{\epsilon_{t,b}\}_{t=1}^{T}$ from $\mathcal{N}(0, \Omega(t))$.

(b). Using ϵ_b, build recursively the bootstrap counterpart of X, denoted with $X_b(\theta)$.

(c). Use $X_b(\theta)$ to estimate the null and the alternative models under the various scenarios and compute the associated LR statistics $LR_{b,i}(\theta)$ for $i = 1, \ldots, n$.

(d). Compute the bootstrap statistic $Q_{Tb}(\theta) = 1 - \min_{1 \leq i \leq n}[1 - F_i(LR_{b,i}(\theta))]$ and record its value. Compute the bootstrap p-value for Q, $\hat{p}_{TB}\left[Q|\theta\right]$, where

$$\hat{p}_{TB}\left[x|\theta\right] = \frac{\sum_{b=1}^{B} \mathbf{1}(Q_{Tb}(\theta) \geq x) + 1}{B + 1}. \tag{11}$$

Step 3. Decide on the acceptance/rejection of the null hypothesis by comparing the associated bootstrap p-value with the chosen level of significance.

The parametric bootstrap corresponds to replacing θ by $\hat{\theta}_T$ in Steps 2 to 3, leading to $\hat{p}_{TB}(Q_{T0}|\hat{\theta}_T)$. One important advantage of simulating the underlying structural model is that all n candidate statistics can be replicated, explicitly controlling for their implicit correlation. In this way, the global level of the test can be adequately adjusted.

3.2 Asymptotic validity of the bootstrap procedure

Our proof of validity begins by a modification of Dufour (2006)'s local condition (Section 3.2.1). This is a general condition and can be applied in other situations as well. Next, using this condition, we prove the asymptotic validity of the proposed procedure (Section 3.2.2).

3.2.1 Asymptotic Monte Carlo tests' validity condition

Consider a family of probability spaces $\{(\Omega, \mathcal{F}, P_\theta) : \theta \in \Theta\}$, where Ω is a sample space, $\mathcal{F}$ a σ-algebra of subsets of Ω, and Θ a parameter space in $\mathbb{R}^k$. Let $Q_T = Q_T(\omega)$, $\omega \in \Omega$, be a real-valued $\mathcal{F}$-measurable function whose distribution is determined by P_{θ_0}, i.e. θ_0 is the "true" parameter vector. We test the hypothesis $\mathcal{H}_0 : \theta_0 \in \Theta_0$ where Θ_0 is a non-empty subset of Θ, using a critical region of the form $\{Q_T \geq c\}$. We denote by $G_T(x|\theta) = P_\theta\left[Q_T \leq x\right]$, $x \in \bar{\mathbb{R}}$, the distribution function of Q_T, where $\theta \in \Theta$. Consider

$$Q_{T0}, Q_{T1}(\theta), ..., Q_{TB}(\theta), \theta \in \Theta, T \geq I_0, I_0 \in \mathbb{R}, \tag{12}$$

real random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$. Here Q_{T0} normally refers to a test statistic with distribution function $G_T(x|\theta_0)$ based on a sample of size T, while $Q_{T1}(\theta), ..., Q_{TB}(\theta)$ are i.i.d. replications of the same test statistic obtained independently under the assumption that the parameter vector is θ. Let also

$$\hat{S}_{TB}(x|\theta) = \frac{1}{B}\sum_{j=1}^{B} \mathbf{1}(Q_{Tj}(\theta) \geq x) \tag{13}$$

be the sample tail (or survival) function and the MC p-value function be defined as in (11). We establish an asymptotic validity result under the following assumptions.

Assumption 1 $Q_{T1}(\theta), \ldots, Q_{TB}(\theta)$ are i.i.d. according to the distribution $G_T(x|\theta) = P[Q_T(\theta) \leq x]$, $\forall \theta \in \Theta$.

Assumption 2 Θ is a nonempty subset of $\mathbb{R}^k$.

Assumption 3 $\forall T \geq I_0$, Q_{T0} is a real random variable and $\hat{\theta}_T$ an estimator for θ, both measurable, and $G_T\left(Q_{T0}|\hat{\theta}_T\right)$ is a random variable.

Assumption 4 $\hat{\theta}_T$ is consistent in probability for θ_0 given specific regularity-identification conditions.

Assumption 5

$$Q_{T0} \xrightarrow{p} Q_0, \tag{14}$$

$$D_0 \text{ is a subset of } \mathbb{R} \text{ s.t. } P[Q_0 \in D_0 \text{ and } Q_{T0} \in D_0 \text{ for all } T \geq I_0] = 1, \tag{15}$$

and $\forall x \in D_0$, $\forall \eta > 0$, and given any sequence $\theta_T \xrightarrow{T\to\infty} \theta_0$ under the same regularity-identification conditions as $\hat{\theta}_T$ in Assumption 4, there exists an open neighborhood $B(x, \eta)$ of x such that

$$\limsup_{T\to\infty} \left\{ \sup_{y \in B(x,\eta)\cap D_0} |G_T(y|\theta_T) - G_T(y|\theta_0)| \right\} \leq \eta. \tag{16}$$

Assumptions 4 and 5 require $\hat{\theta}_T$ and the sequence θ_T to satisfy the same regularity-identification properties. These are general properties and are unspecified on purpose. Examples of such properties include a certain rate of convergence $T^{-\alpha}$ to θ_0. The advantage is that it is enough to prove (16) for sequences of parameters θ_T having the same properties as the estimator that is used to implement the MC test.

Theorem 1. Given (12), (13), (11) and Assumptions 1-5, suppose the random variable Q_{T0} and the estimator $\hat{\theta}_T$ are both independent of $Q_{T1}(\theta), ..., Q_{TB}(\theta)$. Then for $0 \leq \alpha \leq 1$ and $0 \leq \alpha_1 \leq 1$

$$\lim_{T\to\infty} \left\{ P\left[\hat{S}_{TB}\left[Q_{T0}|\hat{\theta}_T \right] \leq \alpha_1 \right] - P\left[\hat{S}_{TB}\left[Q_{T0}|\theta_0 \right] \leq \alpha_1 \right] \right\} = 0$$

and

$$\lim_{T\to\infty} \left\{ P\left[\hat{p}_{TB}\left[Q_{T0}|\hat{\theta}_T \right] \leq \alpha \right] - P\left[\hat{p}_{TB}\left[Q_{T0}|\theta_0 \right] \leq \alpha \right] \right\} = 0. \tag{17}$$

The proof of Theorem 1 is provided in Appendix A.

3.2.2 Asymptotic validity of the proposed bootstrap procedure

This section establishes the asymptotic validity of our combined test procedure. Formally, we show that for large T the limit of the rejection probability referring the bootstrap p-value to an α cut-off, is equal to α. This is formalized in the following Theorem.

Theorem 2. Consider hypotheses $\mathcal{H}_0$ and $\mathcal{H}_A$ in (9) and the test statistic Q_T defined in (10). Assume that all the models $\mathcal{M}_0$, $\mathcal{M}_1$,..., $\mathcal{M}_n$ satisfy standard regularity conditions (e.g. Assumptions 1-4 in Hansen (2003)). Then, under $\mathcal{H}_0$, for $0 < \alpha < 1$

$$\lim_{T\to\infty} P\left[\hat{p}_{TB}(Q_{T0}|\hat{\theta}_T) \leq \alpha \right] = \frac{I[\alpha(B+1)]}{B+1}, \tag{18}$$

where $\hat{p}_{TB}[x|\theta]$ is defined in (11) and $I[x]$ is the integer part of a number x. When B is such that $\alpha(B+1)$ is an integer, the right-hand side of equation (18) is exactly α.

The proof of Theorem 2 is provided in Appendix B.[9] Our methodology can be summarized

[9] Note that our proof accommodates lags and deterministic terms automatically in the context of Theorem 10

as follows. The combined statistic is the *minimum* p-value over the n considered tests.[10] We analytically derive a bound on the null distribution of the combined statistic. We also show that its distribution converges. Though we combine asymptotically χ^2 statistics, joint convergence is not granted. The statistics we combine in this paper in fact converge in probability which in contrast to convergence in distribution does imply joint convergence. We emphasize that the limiting distribution in question does not need to be asymptotically pivotal. For this purpose, in Theorem 3 we present a special case where a pivotal limiting distribution follows from the configuration of the priors. Although noteworthy theoretically, this configuration remains restrictive. The special case allows us to emphasize the usefulness of our general result.

Once convergence is shown, validity follows from our generalization of the limiting equicontinuity conditions of Dufour (2006), the bound we derive and the fact that the null model can be estimated consistently. The MCT method does not literally supply an estimate of the joint distribution of combined statistics; in fact, it circumvents the need to do so.

It is worth noticing that our procedure can accommodate the (consistent) estimation of break locations under the null as this is a discrete parameter and thus satisfies (16) automatically.[11]

One consequence of the above is that validity is shown with practically no restrictions on n which can in fact be large, reflecting what we discuss in our simulation section as the "many priors" case. This stands in sharp contrast with the main-stream literature on combining non-independant tests, and for that matter, with its alternative empirical process theory based sample wide agnostic searches. This leads us to discuss the power of our combined test, before turning to simulations aimed to supplement our asymptotic analysis.

Consider the collection of the n statistics $LR_{T,i}$ associated with H_{0i} respectively (these are not necessarily different). We are interested in H_0 which is the intersection of all the hypotheses H_{0i}, $i = 1, ...n$. Let H_{0i}^c denote the complement of H_{0i}, and conformably let H_0^c denote the complement of H_0. Then $H_0^c = \bigcup_{i=1}^{n} H_{0i}^c$, which implies that if H_0 is false then at least one of the H_{0i}s is false. The rationale underlying testing the intersection of null hypotheses against the union of their

in Hansen (2003). For presentation clarity, Appendix B focuses on the constant variance case with minor loss of generality. Breaks in variance can be accommodated using similar arguments by allowing the size of the subsample in each regime to grow as in Hansen (2003, page 274).

[10] We use the (1-*minimum*) transformation for presentation ease just to obtain a right tailed test.

[11] We thank Jean-Marie Dufour for pointing out this property for discrete parameters.

respective alternative hypotheses typically presumes that the individual statistics considered are suitably chosen, that is, can be presumed to have desirable power properties against their respective alternative. Indeed here, we are combining LR statistics, which can generally be presumed to be consistent against their respective alternative.

Because of our reliance on the *min* p-value to test H_0, if at least one of the combined tests is consistent against some alternative H_1, then the combined test will be consistent against this same alternative. The domain of consistency of the combined test is the union of the consistency domains of all individual tests. Our test thus shares the usual consistency properties of induced Bonferroni tests, with the added important advantage that it avoids bounds. Consistency of Bonferroni-type combined induced tests is a (rather) well known property. Intuitively, this property suggests that users should cast their nets wide when defining a set of statistics to combine, unless some prior knowledge is available which may or may not be the case. However, when only bounds are available, for example when the minimum p-value should be referred to α/n, the power of the resulting procedure would deteriorate as n increases. The MCT combined method circumvents this problem. Our simulations show that power does not deteriorate in any way as n increases. Our proposed MCT combinations thus allow us to accommodate tight or wide priors on the possible alternatives, with no power costs. Because Bonferroni adjustments are avoided altogether here, we do not observe (other things constant) a decrease in power as the number of alternatives increases. Because our validity proof does not restrict n, when the number of considered alternatives increases, we have a better chance of covering the alternatives against which the test is consistent.

Finally, a question that is often raised in the combined tests literature is the performance of proposed methods relative to a joint test that embeds all alternatives. Whether such a test is feasible as n increases is beside the point. The relative power when n is manageably small is worth assessing, here particularly because the G3R framework is rich enough to more or less permit embedding n alternatives. This test is denoted in what follows as the H procedure. We report extensive simulations to analyze these questions in finite samples.

To conclude, we report a special case where in fact the limiting distribution of the test statistic is asymptotically pivotal. This result is interesting in itself but it is clear that our underlying priors are restricted with respect to Theorem 2. If the priors are not restricted, the joint distribution of the test statistics to be combined is not guaranteed to be asymptotically pivotal. The following Theorem 3 provides a generalization of the Cochran Theorem to the cointegration asymptotic framework. The proof of Theorem 3 is provided in Appendix C.

Theorem 3. Let $\mathcal{H}_i$ $(i = 1, ..., n)$ be a sequence of nested alternative hypotheses, i.e. $\mathcal{H}_1 \subset \mathcal{H}_2 \subset ... \subset \mathcal{H}_n$ and consider hypotheses $\mathcal{H}_0$ and $\mathcal{H}_A$ as defined in (9). Assume that all the models $\mathcal{M}_0$, $\mathcal{M}_1$,..., $\mathcal{M}_n$ satisfy standard regularity conditions (e.g. Assumptions 1-4 in Hansen (2003)). Consider the test statistic Q_T defined in (10). Then, the test statistic is asymptotically pivotal and hence, under $\mathcal{H}_0$, for $0 < \alpha < 1, \lim_{T\to\infty} P\left[\hat{p}_{TB}(Q_{T0}|\hat{\theta}_T) \leq \alpha\right] = \frac{I[\alpha(B+1)]}{B+1}$, where $\hat{p}_{TB}\left[x|\theta\right]$ is defined in (11).

4 Finite Sample Properties: a Monte Carlo Study

In this section, we report the results of an extensive Monte Carlo simulation study by considering the case of no breaks (Experiment I), one break (Experiment II) and two breaks (Experiment III) under the null, and three alternatives; we also consider two cases (Experiments IV and V) of a blind search when the set of possible locations is very large.

Experiment I is a baseline case, in the following sense. The stability null hypothesis is tested against three prior scenarios for breaks in β, centered around $T/2$. In addition to the mid-sample, we allow for a fixed window around it or for a window that depends on T. In all subsequent Tables, our proposed test is denoted Q_T and the embedding test H. The dimensionality of this problem is our baseline specification, as in further designs we allow for more scenarios, for breaks under the null, and for breaks in both β and α.

Recall that Q_T is defined to test the union of priors as an alternative whereas on face value, H embeds them all which suggests it is an intersection test. Yet a subtle property of the G3R framework allows us, in simulations, to compare both tests: via adequate restrictions the parameters of the models are assumed freely varying so even if the breaks appear to be imposed, the associated parameters are not restricted to be away from zero. This underscores the χ^2 results we exploit. The main requirement is that the union of priors case remains identifiable, and the sample should be large enough to allow adequate maximization in practice.

As the number of priors increases, it is harder to ensure that their union preserves identification; it is thus harder to ensure perfect maximization of the likelihoods. Our baseline case has been calibrated to avoid this problem.[12] We however allow one important exception to stress-test our baseline design: we force weak identification via one of the considered values for α. Our point is that identification problems are not necessarily due to dimensionality in this model. It is thus important to study the size of proposed tests with small to moderate dimensions as identification is challenged by design.

The design of our power analysis also sets the standard for the remaining experiments. In particular, we define two cases depending on whether the prior is covered by the considered DGP or not. We do not claim that good power is to be anticipated in the latter case. The sensitivity of results to mis-specifying the prior is an important matter to document, particularly in our baseline design.

Experiment II increases dimensionality as follows. The null hypothesis assumes a break in β at a known date. We maintain the same number of prior scenarios as in Experiment I, and use a nested set-up where in addition to the break in the null model, we allow for up to two extra breaks. We explore various sources of uncertainty around that break, moving away via fixed steps or with steps the number of which increases with the sample size. This is to stress-test regularity assumptions in each sub-sample. Again we consider DGPs that fit or miss our priors.

Experiment III tightens dimensionality further as we allow for two breaks in β under the null, at T_1 and T_2. In addition to these two, the priors incorporate up to three more breaks. The uncertainty window around the known breaks is modelled via fixed *before and after* steps. We also assess for mis-specification of priors in various forms and extents.

Experiments IV and V are designed to test for stability against one break such that the set of possible locations is very large: every fifth observation or at any possible location after some trimming. Experiment IV considers breaks in β only, while Experi ment V considers breaks in both α and β. This aims to provide a severe stress-test for our asymptotic assumptions, also replicating

[12] This includes in particular automated robustness checks on starting values. In the results shown below, we used two different starting values for ϕ and ψ, and considered the maximum value of the likelihood between the two results.

the completely agnostic search over all possible locations. Despite the fact that we allow for a wide range of priors, we also consider several cases with mis-specified priors.

As the number of priors increases, the H test is no longer feasible. We document these cases as well, particularly via Experiment V. In all cases we report results for sample sizes ranging from 100 to 300. For space considerations, Experiments I-III report results for size for all considered sample sizes $T = \{100, 200, 300, 400, 500\}$. The tables pertaining to the remaining experiments report only a range of sample sizes that suffice to illustrate our main findings. Tests are applied at the 5% nominal significance levels, the number of Monte Carlo simulations is set to $M = 1000$, while the number of bootstrap replications $B = 199$.

4.1 Monte Carlo design.

Experiment I: No breaks under the null

The first experiment considers no breaks under the null and three alternatives.

Hypotheses. We have considered the following hypotheses

1. H_0 : no breaks; $\quad H_1 : T/2 \cup (T/2 - T/5) \cup (T/2 + T/5)$
2. H_0 : no breaks; $\quad H_1 : T/2 \cup (T/2 - 20) \cup (T/2 + 20)$

DGP for size. Data are generated with no breaks:

$$\textbf{DGP}: \quad \begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & -\tilde{\beta} \end{bmatrix} X_{t-1} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \Rightarrow X_t = (\mathbf{I}_2 + \alpha\beta^\top)X_{t-1} + \epsilon_t. \tag{19}$$

$$\text{with } \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \mathbf{\Omega}),\ \mathbf{\Omega} = \mathbf{I}_2, \tag{20}$$

where $\tilde{\beta} = 1$ and $\mathbf{I}_2$ denotes the identity matrix of order 2, and:

$$\alpha = (\alpha_1, \alpha_2)' \in \left\{(-1, 0)', (-0.2, 0)', (-0.01, 0)', (-0.5, 0.5)'\right\} \tag{21}$$

$$T \in \{100, 200, 300, 400, 500\}. \tag{22}$$

The process was initialized as follows: 50 draws from the model starting from zero were disregarded and we retained observation 51 for X_0.

DGP for power. Data are generated with one break via (19)-(22), where now $\tilde{\beta}(t) = \tilde{\beta}_1 \mathbf{1}_{1t} + \tilde{\beta}_2 \mathbf{1}_{2t}$, $\tilde{\beta}_1 = 1$, $\mathbf{1}_{1t} = \mathbf{1}(1 \leq t \leq T_1), \mathbf{1}_{2t} = \mathbf{1}(T_1 + 1 \leq t \leq T)$, and $\tilde{\beta}_2 \in \{1.1, 1.3, 1.5, 2.0\}$.

We study power only for case 1. above, $H_1 : T/2 \cup (T/2 - T/5) \cup (T/2 + T/5)$. Further, we consider two possible situations: a) DGP covered by one of the alternatives: break at $(T_1 = T/2)$; and b) DGP not covered by any of the alternative hypotheses, with a break at $2T/3$.

Experiment II: One break at T_1 under the null

The second experiment considers one break at T_1 under the null and three alternatives.

Hypotheses. We explore several options for the break location. Some are closer to the regularity conditions in Hansen (2003) (the proportion of observations in each subsample remains constant), others mimic realistic situations where observations are added at the sample end.

1. H_0 : break at $T_1 = T/2$; $H_1 : (T_1, T_1 + T/5) \cup (T_1, T_1 - 3T/10) \cup (T_1, T_1 - T/4, T_1 + T/5)$
2. H_0 : break at $T_1 = T/2$; $H_1 : (T_1, T_1 + 20) \cup (T_1, T_1 - 30) \cup (T_1, T_1 - 25, T_1 + 20)$
3. H_0 : break at $T_1 = 20$; $H_1 : (T_1, T_1 + 20) \cup (T_1, T_1 - 30) \cup (T_1, T_1 - 25, T_1 + 20)$
4. H_0 : break at $T_1 = T/5$; $H_1 : (T_1, T_1 + T/5) \cup (T_1, T_1 - 3T/10) \cup (T_1, T_1 - T/4, T_1 + T/5)$

DGP for size. Data are generated via (19)-(22) under the null model of one break at T_1, where $\tilde{\beta}(t) = \tilde{\beta}_1 \mathbf{1}_{1t} + \tilde{\beta}_2 \mathbf{1}_{2t}$, $\tilde{\beta}_1 = 1$, $\tilde{\beta}_2 = 2$, $\mathbf{1}_{1t} = \mathbf{1}(1 \leq t \leq T_1), \mathbf{1}_{2t} = \mathbf{1}(T_1 + 1 \leq t \leq T)$.

DGP for power. Data are generated via (19)-(22) with two breaks, where now $\tilde{\beta}(t) = \tilde{\beta}_1 \mathbf{1}_{1t} + \tilde{\beta}_2 \mathbf{1}_{2t} + \tilde{\beta}_3 \mathbf{1}_{3t}$, $\tilde{\beta}_1 = 1$, $\tilde{\beta}_2 = 2$, $\mathbf{1}_{1t} = \mathbf{1}(1 \leq t \leq T_1), \mathbf{1}_{2t} = \mathbf{1}(T_1 + 1 \leq t \leq T_2), \mathbf{1}_{3t} = \mathbf{1}(T_2 + 1 \leq t \leq T)$, and different values for $\tilde{\beta}_3$ are explored, $\tilde{\beta}_3 \in \{2.1, 2.3, 2.5, 3.0\}$.

We study power for the cases 1. and 3. above, with two designs: a) DGP covered by one of the alternatives, that is, breaks at $(T_1, T_1 + T/5)$ for case 1. and $(T_1, T_1 + 20)$ for case 3.; b) DGP not covered by any of the alternative hypotheses. In both cases 1. and 3., the breaks are at $(T_1, 2T/3)$.

Experiment III: Two breaks under the null

The third experiment considers two breaks under the null and three alternatives.

Hypotheses. We consider the following hypotheses:

1. H_0 : breaks at $T_1 = T/2$ & $T_2 = 2T/3$; H_1 : $(T_1, T_2, T_1 + 20) \cup (T_1, T_2, T_2 - 5, T_1 + 15) \cup (T_1, T_2, T_1 + 10, T_1 - 10, T_2 - 5)$

2. H_0 : breaks at $T_1 = 20$ & $T_2 = 2T/3$; H_1 : $(T_1, T_2, T_1 + 20) \cup (T_1, T_2, T_2 - 5, T_1 + 15) \cup (T_1, T_2, T_1 + 10, T_1 - 10, T_2 - 5)$

3. H_0 : breaks at $T_1 = T/2$ & $T_2 = T - 20$; H_1 : $(T_1, T_2, T_1 + 20) \cup (T_1, T_2, T_2 - 5, T_1 + 15) \cup (T_1, T_2, T_1 + 10, T_1 - 10, T_2 - 5)$

DGP for size. Data are generated via (19)-(22) with two breaks,where now $\tilde{\beta}(t) = \tilde{\beta}_1 \mathbf{1}_{1t} + \tilde{\beta}_2 \mathbf{1}_{2t} + \tilde{\beta}_3 \mathbf{1}_{3t}$, $\tilde{\beta}_1 = 1$, $\tilde{\beta}_2 = 2$, $\tilde{\beta}_2 = 1.5, \mathbf{1}_{1t} = \mathbf{1}(1 \leq t \leq T_1), \mathbf{1}_{2t} = \mathbf{1}(T_1 + 1 \leq t \leq T_2), \mathbf{1}_{3t} = \mathbf{1}(T_2 + 1 \leq t \leq T)$.

DGP for power. Data are generated via (19)-(22) with three breaks, where $\tilde{\beta}(t) = \tilde{\beta}_1 \mathbf{1}_{1t} + \tilde{\beta}_2 \mathbf{1}_{2t} + \tilde{\beta}_3 \mathbf{1}_{3t} + \tilde{\beta}_4 \mathbf{1}_{4t}$, $\tilde{\beta}_1 = 1$, $\tilde{\beta}_2 = 2, \tilde{\beta}_3 = 1.5$, $\mathbf{1}_{1t} = \mathbf{1}(1 \leq t \leq T_1), \mathbf{1}_{2t} = \mathbf{1}(T_1 + 1 \leq t \leq T_2), \mathbf{1}_{3t} = \mathbf{1}(T_2 + 1 \leq t \leq T_3), \mathbf{1}_{4t} = \mathbf{1}(T_3 + 1 \leq t \leq T)$, and $\tilde{\beta}_3 \in \{1.6, 1.8, 2.0, 2.5\}$.

We study power for 1. and 2. above, with two designs: a) DGP covered by one of the alternatives, that is, breaks at $(T_1, T_2, T_1 + 20)$ for case 1. and $(T_1, T_2, T_2 - 5)$ for case 2.; b) DGP not covered by any of the alternative hypotheses. In both cases 1. and 2., the breaks are at $(T_1, T_2, 5T/6)$.

[Tables 1-4 about here]

Experiment IV: Many Priors Search, breaks in β

The experiments reported here mimic more agnostic searches, where one has much more uncertain priors about possible break locations.

Hypotheses. We consider the following experiments:

1. H_0 : no break; H_1 : at (location) $5 \cup 10 \cup 15 \cup ... \cup 40 \cup 45 \cup 50$

2. H_0 : no break; H_1 : at (location) $5 \cup 10 \cup 15 \cup ... \cup 85 \cup 90 \cup 95$

3. H_0 : no break; H_1 : at (location) $5 \cup 6 \cup 7 \cup ... \cup 93 \cup 94 \cup 95$

DGP for size. Data are generated via (19)-(22) with no breaks, where $\tilde{\beta} = 1$.

DGP for power. Data are generated via (19)-(22) with one break, where $\tilde{\beta}(t) = \tilde{\beta}_1 \mathbf{1}_{1t} + \tilde{\beta}_2 \mathbf{1}_{2t}$, $\tilde{\beta}_1 = 1$, $\mathbf{1}_{1t} = \mathbf{1}(1 \leq t \leq T_1)$, $\mathbf{1}_{2t} = \mathbf{1}(T_1 + 1 \leq t \leq T)$, and $\tilde{\beta}_2 \in \{1.1, 1.3, 1.5, 2.0\}$.

We consider two possible situations: a) DGP covered by one of the alternatives, with break at $T_1 = 50$; b) DGP not covered by any of the alternative hypotheses for case 1., with a break at location 55.

For case 3., the H test is of course infeasible, while there was no issue with our proposed test. Table 5 reports size and power results.

[Table 5 about here]

Experiment V: Many Priors Search, breaks in α_1 and β

Finally, we consider the case of more agnostic searches, where the breaks affect both α and β.

Hypotheses. We consider the following hypotheses:

1. H_0 : no break; H_1 : breaks in α_1 and β at (location) $5 \cup 10 \cup 15 \cup ... \cup 40 \cup 45 \cup 50$

2. H_0 : no break; H_1 : breaks in α_1 and β at (location) $5 \cup 10 \cup 15 \cup ... \cup 85 \cup 90 \cup 95$

DGP for size. Data are generated via (19)-(22) with no breaks, where $\tilde{\beta} = 1$. Two cases for the adjustment coefficient are considered: $\tilde{\alpha}_1 = -1$, $\alpha_2 = 0$, and $\tilde{\alpha}_1 = -0.01$, $\alpha_2 = 0$. We only report results with $T = 100$.

DGP for power. Data are generated according to

$$\textbf{DGP}: \quad \begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{bmatrix} = \begin{bmatrix} \tilde{\alpha}_1(t) \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & -\tilde{\beta}(t) \end{bmatrix} X_{t-1} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \tag{23}$$

$$\text{with } \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \boldsymbol{\Omega}),\ \boldsymbol{\Omega} = \mathbf{I}_2, \tag{24}$$

where $\tilde{\beta}(t) = \tilde{\beta}_1 \mathbf{1}_{1t} + \tilde{\beta}_2 \mathbf{1}_{2t}$, $\tilde{\beta}_1 = 1$, $\tilde{\alpha}_1(t) = \tilde{\alpha}_{11} \mathbf{1}_{1t} + \tilde{\alpha}_{12} \mathbf{1}_{2t}$, $\tilde{\alpha}_{11} = -1$, $\mathbf{1}_{1t} = \mathbf{1}(1 \leq t \leq T_1), \mathbf{1}_{2t} = \mathbf{1}(T_1 + 1 \leq t \leq T)$, and $T = 100$. We examine changes in both long run and/or adjustment parameters, since breaks in one may be masked by breaks in the other. For this reason, in Table 6 we consider changes in all directions, namely, $\left(\tilde{\alpha}_{12}, \tilde{\beta}_2\right) \in \{-0.02, -1.1, -1.3, -1.5, -2.0\} \times \{1.1, 1.3, 1.5, 2.0\}$. Further, we consider the following two designs: a) DGP covered by one of the alternatives, that is, break at $T_1 = 50$ for cases 1. and 2.; b) DGP not covered by any of the alternative hypotheses for case 1. with a break at location 55.

[Table 6 about here]

4.2 Discussion of Simulation Results

Our findings can be summarized as follows. In all analyzed cases, except when we provoke poor identification namely the $\alpha_1 = -0.01$, $\alpha_2 = 0$ case, the combined Q_T test is adequately sized, even with $T = 100$. Interestingly, even in the weakly identified case where we do not anticipate adequate size based on existing theory, the distortions are rather minor, not exceeding 11% for a few cases we show in Table 1. In contrast, the asymptotic H test is severely oversized. Empirical rejections exceeding 30% occur in the weakly identified case even with small n, as reported in Table 1; with large n, rejection exceed 50% even in the identified case, as can be seen in Tables 5 and 6. Some of the over-rejections in the large n (the "many priors") experiments can be attributed to imperfect maximization. In the context of Table 6, and although we still report the H test, we have observed that the estimated parameters are unstable. Of course, this test is infeasible for a completely agnostic search over all possible break locations, as reported in Table 5 for breaks in β only. In this case, we only verify that the size of the combined test is controlled, and no further results are reported.

Numerical problems were not encountered in Experiments I-III. These were carefully designed to allow fair comparisons between the Q_T and H: we considered a small number of priors and the size of the samples was acceptable in each subsample. Nevertheless, the asymptotic H test did not perform well size wise. We thus applied the MC method to the H test, with dramatic success: the size of the MC H was brought back quite close to 5% in the identified designs, and close enough although slightly worse than that of the Q test in the weakly-identified case. While the χ^2 limiting distribution provides the asymptotic validity for the MC H, the extent of the correction it achieves is noteworthy, particularly though not exclusively when identification weakens. The most dramatic size corrections are observed in the many priors case [Tables 5 and 6]. The fact that the H test does not completely break down in some of these experiments is by itself noteworthy. Recall that the G3R algorithm in this case admits a concentrated solution optimizing over the β parameters only, which requires less data than a global search. In addition, the null is the stable model here which is well estimated. This is a key ingredient for the validity of the MC method size-wise; power is another story, as we show below.

So despite the severe size distortions of the asymptotic test and the (related) instability of parameter estimates under the alternative, the MC H test remains feasible for the "many priors" designs we report. Particularly, in the case of breaks in α_1 and β (Experiment V) it turns out that the estimates are notably *unstable* for the H test; this result is important to document for empirical practice. The Q_T test by construction is immune to these problems. We nevertheless proceed and study the relative power of these procedures mainly for stress-testing purposes, reporting some results in Table 6.

Table 1 also shows that the asymptotic H test can under-reject. We did verify that under-rejections are accompanied by lower power. The MC correction solves the problem in these cases as well, adjusting the empirical rejections under the null closer to 5%, and improving power.

Consequently, our power comparisons are restricted to the Q_T and MC H tests, since it does not make sense to study the power of over-sized tests. Note that we have also reported - in some though not all of our simulations, which we intend as reference check - a locally size corrected H test, by replacing the χ^2 critical point in the power study with its approximated counterpart we obtained in each of the size studies. Because it is feasible, we base our analysis on the MC correction, yet both corrections convey similar conclusions qualitatively.

How can we reconcile the finite sample behavior of the asymptotic H test with the fact that our asymptotic analysis for the MC combined test relied on the asymptotic behavior of the individual tests we combine, which take the same form as the H test? On face value, this suggests that the curse of dimensionality is the usual culprit. But deeper analysis particularly of Experiments I-III that are designed in favour of the embedding test underscores the following. While it is clear that the asymptotic validity of the MC method was proved in standard identified frameworks, our proofs do not require χ^2 limiting distributions. This result is clearly illustrated via the dramatic improvements we find, in all of our experiments.

On balance, three key results emerge from our size study: (1) dimensionality does matter and often importantly, yet the MC method corrects its size implications; (2) Asymptotic pivotality is not necessary for the MC method, which is revealed here via our size analysis of Q_T relative to the H test; (3) Identification also matters, but the MC method seems to decrease its over-size effects. Weakly identified setting are beyond the formal scope of this paper, yet the point is worth raising. As pointed out in Dufour (1997), likelihood ratios have the potential to be salvaged.

Turning to power, let us first consider the cases that favour the embedding test by design, namely Experiments I and II. The question we ask in this case is whether the combined approach is dominated by the joint test. Here, the MC approach allows us to provide useful insights into this important problem, since no Bonferroni corrections are needed.

A close inspection of Table 2 and focusing on the MC H test (disregarding its asymptotic counterpart) shows that the performance of Q_T and H is roughly comparable. The combined approach dominates though not dramatically when the sample is small and the magnitude of the break is also small, that is, it provides improvements when it is most needed. Table 3 broadly conveys the same message. The major distinction between Tables 2 and 3 is that the former DGP corresponds to a well specified alternative, whereas the latter mispecifies the priors. Comparing these tables will shed light on the following two questions: (i) whether one of the two tests [Q_T versus H] is more or less robust to mispecification, (ii) what is the general cost on power of missing the truth when formulating priors. On (i), we find that both tests react to mispecification in the same way, at least in Experiments I-III. This is worth verifying, even if there is no real reason for one approach to dominate the other in this respect. Turning to question (ii), in the context of Experiment I, we do not find significant power losses in comparing the relevant panels in Tables 2 and 3.

In contrast, Experiment II provides a case that illustrates such losses, which is reported in the last panel of Table 3. The salient feature of that DGP is that the location of breaks is fixed regardless of the sample size. Our intention is to study a design where deviations from the truth are somewhat adhoc, in which case we should expect power losses. Indeed, even when the prior does not cover the truth but remains close enough, which is roughly achieved in our simulations, power does not suffer. But of course when the DGP deviates arbitrarily from the truth, then as it should be, power does drop.

The usual trade-offs between restricted and unrestricted econometric methods apply here. The interesting result is that mispecification costs are not drastic and conform to realistic expectations.

Experiment III tightens the dimensionality in various aspects, moving particularly to three regimes under the null. We intentionally maintain the number of priors to three, as in Experiments I and II. In the multiple test contexts, most discussions focus on the number of combinations. The point we aim to show here in that in the nuisance-parameter dependent case, this is not the whole story. Table 4 confirms that the combined test Q_T now dominates the embedding (joint) test relative H for all the alternatives we studied. This suggests that in addition to the number of prior alternatives, dimensionality of the null model also importantly affects the performance of Q_T relative to H. Pursuing a similar argument, Table 4 also illustrates the interaction of mispecification and dimensionality problems. As the number of breaks under the null increases, we find that mispecification costs are more pronounced. One result that may seem disconcerting at face value: power costs seem less sizable in the small sample case. The fact here is that when the sample is small, the extent of the mispecification as we introduce it is also less important since the priors are by construction closer to the true DGP. The mispecification window is de-facto smaller in smaller samples.

This argument is further concretized via our "many priors" experiments [Table 5]: power does not decrease much when the priors are mispecified. Recall however that the experiments examine breaks in β when these occur at regular short intervals within the sample. Here again the mispecification window is not large, which explains our finding. The salient result in Tables 5 and 6 is to confirm the superiority of our combined approach. While it is (arguably) possible to circumvent the size problems resulting from degrees-of-freedom crunches with the embedding test, power does bear the cost. Combination methods thus emerge as a valuable alternative, particularly in the many-priors case.

To conclude, it is worth noting that power is low for all size-correct methods we consider in the weak identified case, as it should be. On balance, three key results emerge from our power study. (1) As with size, dimensionality does matter and again often importantly, in the comparison between the embedding H test and the combined Q_T: even when the former is well behaved, that is when the number of priors is small, unless the number of regimes is small under both the null or alternatives and/or a small number of parameters break, it pays off to use MC combination methods. (2) In the nuisance parameter dependent case, the number of combined statistics is not the only concern for finite sample performance. This is important since a large chunk of the literature on multiple testing examines dimensionality mostly from the number of tests angle. We show that two factors interact, namely dimensionality of the null model, and the number of combinations. (3) While the MC methods can and does correct the severe over-rejections of the embedding test, this comes at a serious power disadvantage relative to the combined test. In the combined dependent tests literature, the main-stream consensus is that power deteriorates with the number of combinations. The MC method circumvents this problem altogether, despite the nuisance parameter problem.

Our results interpreted collectively suggest that further work on induced MC tests holds concrete promise, even when joint tests are available or when the number of prior alternatives is not small. Joint tests may perform poorly because one of the alternatives lacks fit or is weakly identified. The MC combined approach seems much less prone to this problem.

5 Empirical Application

In this section, we illustrate how multiple break testing can be conducted using the Q_T-statistic by investigating the present value theory for asset prices (see e.g. Campbell and Shiller, 1987, 1989). Let P_t refer to the asset price series and D_t the dividend paid for owning the asset during the period $[t, t+1]$. Then the present value theory implies

$$P_t = \mathbb{E}_t\left(\frac{P_{t+1} + D_{t+1}}{R_{t+1}}\right), \tag{25}$$

where R_{t+1} is the discount factor for the period $[t, t+1]$ and $\mathbb{E}_t(\cdot)$ is a short-hand for $\mathbb{E}(\cdot|\mathcal{I}_t)$, with $\mathcal{I}_t$ denoting the information set available at t. After log-linearisation (see Campbell and Shiller, 1989 for details), (25) can be rearranged as $p_t = \kappa - \mathbb{E}_t(r_{t+1}) + \rho\mathbb{E}_t(p_{t+1}) + (1-\rho)\mathbb{E}_t(d_{t+1})$, where $\rho = 1/(1 + e^{\overline{d-p}})$, $\kappa = (\rho - 1)\log(\rho^{-1} - 1) - \log(\rho)$, and $\overline{d-p}$ is the average log dividend-price ratio. Solving by recursive substitution leads to

$$p_t = \frac{\kappa}{1-\rho} + (1-\rho)\sum_{i=0}^{\infty}\rho^i\mathbb{E}_t(d_{t+i+1}) - \sum_{i=0}^{\infty}\rho^i\mathbb{E}_t(r_{t+i+1}) + \lim_{i\to\infty}\rho^i\mathbb{E}_t(p_{t+i}). \tag{26}$$

If we assume that prices do not follow an explosive process, then $\lim_{i\to\infty}\rho^i\mathbb{E}_t(p_{t+i})$ in (26) converges to zero, and rearranging remaining terms, the dividend-price ratio is given by[13]

$$d_t - p_t = -\frac{\kappa}{1-\rho} + \sum_{i=0}^{\infty}\rho^i\mathbb{E}_t(r_{t+i+1} - \Delta d_{t+1+i}). \tag{27}$$

The empirical validity of (27) corresponds to cointegration between the log-dividend process $\{d_t\}$ and the log-price process $\{p_t\}$.

Our empirical analysis of the above is based on the standard S&P500 prices and associated dividend series measured at quarterly frequency over the period 1960(1) - 2014(2). The dataset is taken from Robert Shiller's web page (`http://www.econ.yale.edu/~shiller/`) while the computations are executed using *OxMetrics 7* (Doornik and Hendry, 2013). The dividend-price ratio for this sample is depicted in Figure 1.

[Figure 1 about here]

We first fit an unrestricted VAR to the vector $X_t = [d_t,\ p_t]^\top$ and assess cointegration using the eigenvalue based tests of Johansen (1991). This test leads to the rejection of (time-invariant)

[13] Without loss of generality, we assume that the discount factor $r_t = r_{t+1} = \cdots = r$ is constant over time, such that (27) simplifies to $d_t - p_t = -\frac{\kappa - r}{1-\rho} - \sum_{i=0}^{\infty}\rho^i\mathbb{E}_t(\Delta d_{t+1+i})$.

cointegration, a result certainly not at odds with the dynamics of the dividend-price ratio in Figure 1.

Our analysis spans through three historical stylized-fact shocks: the Black Monday in 1987, the dot-com burst at the turn of the century, and the latest sub-prime and sovereign debt crises (global financial crises). We thus analyze the relevance of these shocks, in the context of a G3R regression with four lags and unrestricted constant. Table 7 presents four combined tests reported in panels labelled A - D. The underlying alternatives and corresponding individual p-values are also reported in each panel. Each of the four combined tests is size-correct in its own right whereas all four are not combined. Also note that test D is itself a single test; the reported p-value under the Q-heading is the MC p-value, which we deem more reliable than its χ^2 counterpart in view of our simulation results.

Tests A and B assess stability against three alternatives each representing a single break at each of the above locations. In contrast to B, test A imposes stability of the adjustment coefficients. Comparing the outcome of both tests suggests breaks in both cointegration and adjustment coefficients. Indeed, recall that time-invariant cointegration was rejected; we thus proceed to interpret tests B - D. Test C assesses stability against two priors: (i) three breaks [1987(3), 1999(3), and 2007(4)]; (ii) two breaks [1999(3) and 2007(4)]. Test D confronts the latter two priors, assessing (i) as a null against (ii). Interpreted individually or combined via an equal-split Bonferroni adjustment[14] at 10%, the conjunction of tests B, C and D suggests a breaking relation, with breaks at 1999(3) (dot-com boom) and 2007(4) (global financial crisis). This is in line with the conventional wisdom regarding the short-lived impact of the 1987(3) Black Monday.

[Table 7 about here]

On balance, our findings suggest an alternative perspective to the bubbles motivated approach as in Phillips et al. (2011), Phillips et al. (2015a), and Phillips et al. (2015b). These authors test for temporary explosiveness of the price process due to financial bubbles, *i.e.* whether $b_t = \lim_{i\to\infty} \rho^i \mathbb{E}_t(p_{t+i})$ explodes. Instead, our results suggest a breaking cointegrating relationship between the price and dividend time series.

[14] We interpret decisions via this adjustment because the null hypothesis of test D differs from that of the other tests.

These results should however be considered acknowledging the substantial literature on the present value theory, that has not led to a "consensus" view empirically. We bring in evidence of breaks within a bivariate system, which may raise interpretation issues. A changing relation may challenge the present-value model for asset prices, which typically presumes a stable cointegrating vector. Whether a stable adjustment coefficient is required for a structural (e.g. in the sense of Lucas) model interpretation is beyond the scope of our analysis. We do not aim to take a firm view towards the specification retained nor any underpinning financial theories. Our empirical exercise would not do justice to this research topic, in view of the sizable related literature. Instead, our results may be viewed as a motivation for experts in macro/finance to envisage a changing relation. Econometric tool kits such as those we propose here may in fact assist towards its empirical validation.

6 Conclusion

In this paper we proposed a multiple testing technique to assess stability of cointegrating relations, building on the VECM G3R framework of Hansen (2003). We extended the test of Hansen (2003) to accommodate various prior alternatives via a simulation-based approach, and proved its asymptotic validity analytically.

The finite sample properties of proposed procedures are analyzed through an extensive Monte Carlo simulation. Results suggest that a large number of priors can be effectively combined, circumventing Bonferroni-type corrections. An application to the S&P500 prices and dividends series suggests a breaking cointegration relation, in the context of an enduring present-value motivated empirical asset pricing problem.

A Proof of Theorem 1

To prove Theorem 1, it is convenient to first demonstrate two lemmas.

Lemma A.1 (Continuity of p-value function)**.** Under notation (12), (13), (11) and Assumption

1, set $\bar{Q}_{TB}(\theta, x, \alpha_1) = P\left[\hat{S}_{TB}(x|\theta) \leq \alpha_1\right]$, $0 \leq \alpha_1 \leq 1$. For any $\theta, \theta_0 \in \Theta$, $x \in \mathbb{R}$, the inequality

$$|G_T[y|\theta] - G_T[y|\theta_0]| \leq \eta, \quad \forall y \in (x-\delta, x+\delta),\ \delta > 0, \tag{28}$$

entails the inequality $|\bar{Q}_{TB}(\theta, x, \alpha_1) - \bar{Q}_{TB}(\theta_0, x, \alpha_1)| \leq 3C(N, \alpha_1)\eta$, where $C(B, \alpha) = B\sum_{k=0}^{[\alpha_1 B]}\binom{B}{k}$.

Proof. Using Assumption 1 (i.i.d.),

$$\bar{Q}_{TB}(\theta, x, \alpha_1) = P\left[\sum_{i=1}^{B}\mathbf{1}\left(Q_{Ti}(\theta) \geq x\right) \leq \alpha_1 N\right] = \sum_{k=0}^{[\alpha_1 B]}\binom{B}{k}\bar{G}_T(x|\theta)^k[1-\bar{G}_T(x|\theta)]^{B-k},$$

where $\bar{G}_T(x|\theta) = P\left[\mathbf{1}\left(Q_{Ti}(\theta) \geq x\right) = 1\right] = P\left(Q_{Ti}(\theta) \geq x\right) = 1 - G_T(x|\theta)$. Since (using (28)) $|\bar{G}_T(x|\theta) - \bar{G}_T(x|\theta_0)| = |G_T(x|\theta) - G_T(x|\theta_0)| \leq \eta$,

$$\begin{aligned}
&|\bar{Q}_{TB}(\theta, x, \alpha_1) - \bar{Q}_{TB}(\theta_0, x, \alpha_1)| = \\
&\quad = \left|\sum_{k=0}^{[\alpha_1 B]}\binom{B}{k}\left\{\bar{G}_T(x|\theta)^k[1-\bar{G}_T(x|\theta)]^{B-k} - \bar{G}_T(x|\theta_0)^k[1-\bar{G}_T(x|\theta_0)]^{B-k}\right\}\right| \\
&\quad \leq \sum_{k=0}^{[\alpha_1 B]}\binom{B}{k}\left[|\bar{G}_T(x|\theta)^k - \bar{G}_T(x|\theta_0)^k| + |[1-\bar{G}_T(x|\theta)]^{B-k} - [1-\bar{G}_T(x|\theta_0)]^{B-k}|\right] \\
&\quad \leq \sum_{k=0}^{[\alpha_1 B]}\binom{B}{k}\left[k\left|\bar{G}_T(x|\theta) - \bar{G}_T(x|\theta_0)\right| + (B-k)|[1-\bar{G}_T(x|\theta)] - [1-\bar{G}_T(x|\theta_0)]|\right] \leq C(B, \alpha_1)\eta.
\end{aligned}$$

Lemma A.2 (Convergence of Bootstrap p-values)**.** Under the notation (12), (13), (11) and Assumptions 1-5, as $T \to +\infty$, $|\bar{Q}_{TB}(\hat{\theta}_T, Q_{T0}, \alpha_1) - \bar{Q}_{TB}(\theta_0, Q_{T0}, \alpha_1)| \xrightarrow{p} 0$.

Proof. Since $Q_{T0} \xrightarrow{p} Q_0$ (Assumption 5), $(Q_{T0}, \hat{\theta}_T) \xrightarrow{p} (Q_0, \theta_0)$ and the same holds for any subsequence $\{(Q_{\bar{T}_k 0}, \hat{\theta}_{\bar{T}_k}) : k = 1, 2, \ldots\}$ of $\{(Q_{T0}, \hat{\theta}_T) : T \geq I_0\}$: $(Q_{\bar{T}_k 0}, \hat{\theta}_{\bar{T}_k}) \xrightarrow{p} (Q_0, \theta_0)$, as $k \to +\infty$. Since Q_{T0} and $\hat{\theta}_T$, $T \geq I_0$, are random variables (or vectors) on the sample space Ω, we can write $Q_{T0} = Q_{T0}(\omega)$, $\hat{\theta}_T = \hat{\theta}_T(\omega)$, and $Q_0 = Q_0(\omega)$, $\omega \in \Omega$. By Assumption 5 (equation (15)), the event $A_0 = \{\omega : Q_0(\omega) \in D_0 \textit{ and } Q_{T0}(\omega) \in D_0, T \geq I_0\}$ has probability one. Moreover, convergence in probability of $(Q_{T0}, \hat{\theta}_T)$ implies that the subsequence $(Q_{\bar{T}_k 0}, \hat{\theta}_{\bar{T}_k}')'$

contains a further subsequence $(Q_{T_k0}, \hat{\theta}'_{T_k})'$, $k \geq 1$ such that $(Q_{T_k0}, \hat{\theta}'_{T_k})' \overset{T\to+\infty}{\longrightarrow} (Q_0, \theta_0)$ *a.s.*. It follows that the set $C_0 = \{\omega : Q_0(\omega) \in D_0, \lim_{k\to+\infty} Q_{T_k0}(\omega) = Q_0(\omega)\ and\ \lim_{k\to+\infty} \hat{\theta}_{T_k}(\omega) = \theta_0\}$ has probability one. Now let $\eta > 0$. By Assumption 5 (equation (16)), for any $x \in D_0$, given a sequence $\theta_T \overset{T\to\infty}{\longrightarrow} \theta_0$, we can find $T(x,\eta) > 0$ and an open neighborhood $B(x,\eta)$ of x such that, for $T > T(x,\eta)$, $|G_T(y|\theta_T) - G_T(y|\theta_0)| \leq \eta$, $\forall y \in B(x,\eta) \cap D_0$.Furthermore, for $\omega \in C_0$ (by the definition of C_0), $\hat{\theta}_{T_k}(\omega) \overset{k\to+\infty}{\longrightarrow} \theta_0$ and we can find k_0 s.t. $Q_{T_k0}(\omega) \in B(Q_0(\omega),\eta) \cap D_0$ for $k \geq k_0$, so that $T_k > \max\{T(Q_0(\omega),\eta), T_{k_0}\}$ implies $|G_{T_k}[Q_{T_k0}(\omega)|\hat{\theta}_{T_k}(\omega)] - G_{T_k}[Q_{T_k0}(\omega)|\theta_0]| \leq \eta$. Thus $\lim_{k\to+\infty}\{G_{T_k}[Q_{T_k0}(\omega)|\hat{\theta}_{T_k}(\omega)] - G_{T_k}[Q_{T_k0}(\omega)|\theta_0]\} = 0$ for $\omega \in C_0$. By Lemma A.1, $\lim_{k\to+\infty} |\bar{Q}_{T_kB}(\hat{\theta}_{T_k}(\omega), Q_{T_k0}(\omega), \alpha_1) - \bar{Q}_{TB}(\theta_0, Q_{T0}(\omega), \alpha_1)| = 0$, i.e.

$$\lim_{k\to+\infty} |\bar{Q}_{T_kB}(\hat{\theta}_{T_k}, Q_{T_k0}, \alpha_1) - \bar{Q}_{TB}(\theta_0, Q_{T0}, \alpha_1)| = 0\ a.s..$$

This shows that any subsequence of the sequence $|\bar{Q}_{TB}(\hat{\theta}_T, Q_{T0}, \alpha_1) - \bar{Q}_{TB}(\theta_0, Q_{T0}, \alpha_1)|$, $T \geq I_0$, contains a further subsequence which converges a.s. to zero. This is equivalent to $|\bar{Q}_{TB}(\hat{\theta}_T, Q_{T0}, \alpha_1) - \bar{Q}_{TB}(\theta_0, Q_{T0}, \alpha_1)| \overset{p}{\longrightarrow} 0$.

Proof of Theorem 1 Thanks to independence of $\hat{\theta}_T$ and Q_{T0} from $Q_{T1}(\theta), \ldots, Q_{TB}(\theta)$,

$$\begin{aligned} &|P\left[\hat{S}_{TB}(Q_{T0}|\hat{\theta}_T) \leq \alpha_1\right] - P\left[\hat{S}_{TB}(Q_{T0}|\theta_0) \leq \alpha_1\right]| = \\ &= |E\left\{P\left[\hat{S}_{TB}(Q_{T0}|\hat{\theta}_T) \leq \alpha_1|(\hat{\theta}_T, Q_{T0})\right] - P\left[\hat{S}_{TB}(Q_{T0}|\theta_0) \leq \alpha_1|(\hat{\theta}_T, Q_{T0})\right]\right\}| \\ &\leq E\left|\bar{Q}_{TB}(\hat{\theta}_T, Q_{T0}, \alpha_1) - \bar{Q}_{TB}(\theta_0, Q_{T0}, \alpha_1)\right|. \end{aligned}$$

So, applying the dominated convergence theorem and thanks to Lemma A.2, the result follows. Equation (17) follows from the definition of $\hat{p}_{TB}(x|\theta)$.

B Proof of Theorem 2

Let $Q_{T0} = Q_T(\theta_0) = 1 - \min_{1\leq i\leq n}[1 - F_i(LR_{T,i}(\theta_0))]$, where $F_i(\cdot)$ is the CDF of a $\chi^2(q_i)$. Q_{T0} and $LR_{T,i}(\theta_0)$ refer to the test statistics in (10) computed from the observed data when the true parameter is θ_0. Let $G_T(y|\theta) = P(Q_T(\theta) \leq y)$ denote the distribution of $Q_{Tb}(\theta)$, $b = 1, \ldots, B$, where $Q_{Tb}(\theta)$ are the simulation-based counterparts of the test statistics as in Steps 2 to 3 above. Along the same lines, $G_T(y|\theta_0)$ will refer to the distribution of $Q_T(\theta_0)$. To extend Theorem 1, which pertains to the level of a test based on $\hat{p}_{TB}(Q_{T0}|\hat{\theta}_T)$, Assumptions 1-5 need to hold. Assumptions 1-4 require that: (i) the above defined $Q_{T1}(\theta), \ldots, Q_{Tb}(\theta), \ldots, Q_{TB}(\theta)$ are i.i.d. according to $G_T(y|\theta)$, $\forall\theta \in \Theta$; (ii) $\hat{\theta}_T$ and Q_{T0} are measurable; (iii) $G_T(Q_{T0}|\hat{\theta}_T)$ is a random variable; and (iv) $\hat{\theta}_T$ is consistent for θ_0 with convergence rate $T^{-1/2}$. Assumptions 1-4 are verified by definition. To prove Assumption 5, we need to show that $Q_{T0} \xrightarrow[T\to\infty]{p} Q_0$ and, given any sequence of parameters $\theta_T \xrightarrow[T\to\infty]{} \theta_0$ with convergence rate $T^{-1/2}$, D_0 a subset of $\mathbb{R}$ s.t. $P[Q_0 \in D_0 \; and \; Q_{T0} \in D_0 \; for \; all \; T \geq I_0] = 1$ and $\forall x \in D_0$, $\forall \eta > 0$, there exists an open neighborhood $B(x, \eta)$ of x such that

$$\limsup_{T\to\infty} \left\{ \sup_{y \in B(x,\eta)\cap D_0} |G_T\left(y|\theta_T\right) - G_T\left(y|\theta_0\right)| \right\} \leq \eta. \tag{29}$$

First, Theorem 10 in Hansen proves that (under Assumptions 1-4 in Hansen (2003)) $LR_{T,i}(\theta) \xrightarrow[T\to\infty]{d} X_i \sim \chi^2_{q_i}$, where q_i corresponds to the reduction of free parameters between the model under the null and the model under the considered alternative. From the proof of Theorem 10 in Hansen (2003), it is apparent that the convergence of the likelihood ratios is not only in distribution but also in probability, namely, $LR_{T,i}(\theta) \xrightarrow[T\to\infty]{p} X_i \sim \chi^2_{q_i}$.Differently from convergence in distribution, convergence of two (or more) sequences in probability implies joint convergence in probability (see, e.g., Lemma 3.4 in Kallenberg (1997)). This implies that the vector of likelihood ratios $(LR_{T,1}(\theta), \ldots, LR_{T,n}(\theta))$ jointly converges in probability to the vector $(X_1, \ldots, X_n)$. It thus follows that the joint distribution of $(LR_{T,1}(\theta), \ldots, LR_{T,n}(\theta))$ converges to the joint distribution of $(X_1, \ldots, X_n)$. Convergence of the test statistic

$$Q_T\left(\theta\right) = 1 - \min_{1\leq i\leq n} \left[1 - F_i\left(LR_{T,i}\left(\theta\right)\right)\right] = h\left(LR_{T,1}(\theta), \ldots, LR_{T,n}(\theta)\right), \tag{30}$$

follows from the continuous mapping theorem. So, denoting $Q_0 = h\left(X_1, \ldots, X_n\right)$, we have that $Q_{T0} \xrightarrow[T\to\infty]{p} Q_0$. Next, considering $D_0 = (0, 1)$, it follows immediately that $P[Q_0 \in D_0 \; and \; Q_{T0} \in D_0 \; for \; all \; T \geq 1] = 1$. Finally, in order to prove (29), it is convenient to rewrite the test statistic as $Q_T(\theta) = g[X_T(\theta)]$, where the function $g(\cdot)$ is the composition of the functions underlying the definition of the test statistic (10), as summarized in Step 2 of the algorithm. The distribution of $X_T(\theta)$ is the joint likelihood $L_\theta(x) = L(x, \theta)$.

Now consider $P(Q_T(\theta) \leq y) = P(g[X_T(\theta)] \leq y) = P(X_T(\theta) \in R_y)$, where $R_y = \{x \in \mathbb{R}^{Tp} : g(x) \leq y\}$ is the pre-image of the set $(-\infty, y]$, $y \in \mathbb{R}$. We can always write

$$
\begin{aligned}
|G_T(y|\theta_T) - G_T(y|\theta_0)| &= |P(Q_T(\theta_T) \leq y) - P(Q_T(\theta_0) \leq y)| \\
&= |\int_{R_y} L_{\theta_T}(x)dx - \int_{R_y} L_{\theta_0}(x)dx| \\
&\leq E_{\theta_0}\left[\left|\frac{L_{\theta_T}(X_T)}{L_{\theta_0}(X_T)} - 1\right| I(X_T \in R_y)\right] \\
&\leq \sqrt{E_{\theta_0}\left[\left(\frac{L_{\theta_T}(X_T)}{L_{\theta_0}(X_T)} - 1\right)^2\right]},
\end{aligned}
$$

where we denote by $E_{\theta_0}[\cdot]$ the expected value with respect to the true distribution P_{θ_0}. The last quantity does not depend on y so we have that

$$
\sup_{y\in(0,1)} |G_T(y|\theta_T) - G_T(y|\theta_0)| \leq \sqrt{E_{\theta_0}\left[\left(\frac{L_{\theta_T}(X_T)}{L_{\theta_0}(X_T)} - 1\right)^2\right]}. \tag{31}
$$

Recall now that the likelihood is given by

$$
L_\theta(X_T) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{p/2}|\Omega|^{1/2}} \exp\left[-\frac{1}{2}(Z_{0t} - AB'Z_{1t} - CZ_{2t})'\Omega^{-1}(Z_{0t} - AB'Z_{1t} - CZ_{2t})\right],
$$

with $\theta = (A, B, C, \Omega)$ where correct specification implies zero mean variates. So we consider the sequence θ_T to satisfy the same property. Let us now focus on the right-hand side of (31). $L_{\theta_T}(X_T)$ corresponds to the likelihood for a $N_p(0, \Omega_T)$ and $L_{\theta_0}(X_T)$ to the likelihood for a $N_p(0, \Omega_0)$. In the following, we denote these likelihoods L_A and L_0, respectively. Now, $L_A = \prod_{t=1}^{T} \frac{1}{(2\pi)^{p/2}|\Omega_T|^{1/2}} \exp\left[-\frac{1}{2}y_t'\Omega_T^{-1}y_t\right]$, which may be written in the canonical exponential form as $L_A = \frac{1}{(2\pi)^{Tp/2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{T} tr(y_t y_t' K_T) + \frac{T}{2}\ln|K_T|\right)$, with $K_T = \Omega_T^{-1}$. The same applies to L_0.

Thanks to Lemma 1 in Nielsen and Nock (2013) we thus can write

$$\begin{aligned} E_{\theta_0}\left[\left(\frac{L_{\theta_T}(X_T)}{L_{\theta_0}(X_T)}-1\right)^2\right] &= e^{-(T/2)\ln|2K_T-K_0|+T\ln|K_T|-(T/2)\ln|K_0|}-1 \\ &= |2K_T-K_0|^{-T/2}|K_T|^T|K_0|^{-T/2}-1, \end{aligned}$$

where we denoted $K_T=\Omega_T^{-1}$. In terms of the original parameter, this corresponds to

$$E\left[\left(\frac{L_{\theta}(X_T)}{L_{\theta_0}(X_T)}-1\right)^2\right] = |2\Omega_T^{-1}-\Omega_0^{-1}|^{-T/2}|\Omega_T^{-1}|^T|\Omega_0^{-1}|^{-T/2}-1,$$

where Ω_T is a sequence of variance-covariance matrices s.t. $\Omega_T\to\Omega_0$. Now consider the term

$$\begin{aligned} |2\Omega_T^{-1}-\Omega_0^{-1}|^{-T/2}|\Omega_T^{-1}|^T|\Omega_0^{-1}|^{-T/2} &= |2\Omega_T^{-1}-\Omega_0^{-1}|^{-T/2}|\Omega_T^{2}|^{-T/2}|\Omega_0^{-1}|^{-T/2} \\ &= \frac{1}{(|2I-\Omega_T\Omega_0^{-1}||\Omega_T\Omega_0^{-1}|)^{T/2}} \\ &= \frac{1}{|2M_T-M_T^2|^{T/2}}, \end{aligned}$$

where $M_T=\Omega_T\Omega_0^{-1}$. It follows that

$$E\left[\left(\frac{L_{\theta}(X_T)}{L_{\theta_0}(X_T)}-1\right)^2\right] = \frac{1}{|2M_T-M_T^2|^{T/2}}-1=\frac{1}{(|M_T||2I-M_T|)^{T/2}}-1.$$

We thus need to prove that

$$|M_T||2I-M_T|\geq\frac{1}{(1+\eta^2)^{2/T}}. \tag{32}$$

Now, denote by $\lambda_{i,T}$ the eigenvalue of M_T, i.e. $\lambda_{i,T}$ is a solution to $|M_T-\lambda_{i,T}I|=0$, so that $|M_T|=\prod_{i=1}^p\lambda_{i,T}$. Notice that $\lambda_{i,T}-2$ is the eigenvalue of M_T-2I so (32) becomes $\prod_{i=1}^p\lambda_{i,T}\prod_{i=1}^p(2-\lambda_{i,T})\geq\frac{1}{(1+\eta^2)^{2/T}}$, which is the same as $\prod_{i=1}^p\lambda^*_{i,T}\geq\frac{1}{(1+\eta^2)^{2/T}},\quad \lambda^*_{i,T}=\lambda_{i,T}(2-\lambda_{i,T})$. Notice now that $\lambda^*_{i,T}=-(\lambda_{i,T}-1)^2+1$ and thus $\prod_{i=1}^p\left[-(\lambda_{i,T}-1)^2+1\right]\geq1-\sum_{i=1}^p(\lambda_{i,T}-1)^2$, so we are left to prove that $\sum_{i=1}^p(\lambda_{i,T}-1)^2\leq1-\frac{1}{(1+\eta^2)^{2/T}}$. Now $1-\frac{1}{(1+\eta^2)^{2/T}}=1-e^{-\frac{2}{T}\ln(1+\eta^2)}\simeq+\frac{2}{T}\ln(1+\eta^2)$, so that for condition (29) to hold, it is requested that $\Omega_T-\Omega_0=O(T^{-1/2})$. This is exactly the rate of convergence of the estimator $\hat{\Omega}_T$ (see Johansen (1995)).

Finally, $Q_T(\theta_0)$ and $\hat{\theta}_T$ are both independent of $Q_{T1}(\theta),\ldots,Q_{TB}(\theta)$, by construction. Theorem

2 then entails that for $0 \leq \alpha \leq 1$

$$\lim_{T\to\infty} \left\{ P\left[\hat{p}_{TB}\left(Q_{T0}|\hat{\theta}_T\right) \leq \alpha\right] - P\left[\hat{p}_{TB}\left(Q_{T0}|\theta_0\right) \leq \alpha\right] \right\} = 0. \tag{33}$$

We can therefore apply Proposition 2.4 in Dufour (2006) as follows. This proposition proves that $P\left[\hat{p}_{TB}\left(Q_{T0}|\theta_0\right) \leq \alpha\right] = \frac{I[\alpha(B+1)]}{B+1}$, because of exchangeability of the observed and simulated statistics. Substituting this expression in (33) proves the theorem.

C Proof of Theorem 3

Proof. First, we show that the limiting distribution exists and is pivotal. The proof of the existence of the limiting distribution follows along the lines of the proof of Theorem 2. To prove asymptotic pivotality, it is useful to note that the test statistic can be represented as (see proof of Theorem 10 in Hansen (2003)) $LR_{T,i}(\theta) = \eta^T M_{T,i}(\theta)\eta + o_p(1)$, $i = 1, 2, ..., n$, where $\eta = \Sigma^{-1/2}\epsilon \sim N(0, I_{Tp})$, $\epsilon = vec(\epsilon_1, ..., \epsilon_T)$, $\Sigma = var(\epsilon)$, and $M_{T,i} = P_{AC,i} - P_{AC,0} + P_{B,i} - P_{B,0}$, with $P_{AC,i}, P_{AC,0}, P_{B,i}, P_{B,0}$ projection matrices such that $P_{AC}P_B = o_p(1)$, so that each alternative is tested against the null hypothesis by $LR_{T,i}(\theta) = \eta'\left(P_{AC,i} - P_{AC,0} + P_{B,i} - P_{B,0}\right)\eta + o_p(1)$.

Now consider $LR_{T,i}(\theta) - LR_{T,i-1}(\theta) = \eta'\left(P_{AC,i} - P_{AC,i-1} + P_{B,i} - P_{B,i-1}\right)\eta + o_p(1)$. Since $\mathcal{H}_{i-1}$ is nested in $\mathcal{H}_i$ ($i = 1, ..., n$), $P_{AC,i} - P_{AC,i-1}$ and $P_{B,i} - P_{B,i-1}$ are p.s.d.. It thus follows that $LR_{T,i}(\theta) - LR_{T,i-1}(\theta) \to^d Y_i \sim \chi^2_{q_i - q_{i-1}}$, where we have denoted by Y_i the limit r.v.. Now, for $j < i$, $LR_{T,j}(\theta) - LR_{T,j-1}(\theta)$ and $LR_{T,i}(\theta) - LR_{T,i-1}(\theta)$ are asymptotically independent as $\left(P_{AC,j} - P_{AC,j-1} + P_{B,j} - P_{B,j-1}\right)\left(P_{AC,i} - P_{AC,i-1} + P_{B,i} - P_{B,i-1}\right) = o_p(1)$.

The joint vector of likelihood ratios $(LR_{T,1}(\theta), ..., LR_{T,n}(\theta))$ can thus be written as a linear transformation of $(LR_{T,1}(\theta), LR_{T,2}(\theta) - LR_{T,1}(\theta), ..., LR_{T,n}(\theta) - LR_{T,n-1}(\theta))$, which has asymptotically independent components and hence it is asymptotically pivotal. Thus, it follows that $(LR_{T,1}(\theta), ..., LR_{T,n}(\theta))$ is asymptotically pivotal, i.e. $(LR_{T,1}(\theta), ..., LR_{T,n}(\theta)) \to^d (X_1, ..., X_n)$, where $X_i \sim \chi^2_{q_i}$ (q_i denoting the reduction of free parameters between the alternative and the null model) and $(X_1, ..., X_n)$ jointly pivotal. Hence, also the test statistic $Q_T(\theta)$ is asymptotically pivotal. We denote by $Q_0 = h(X_1, ..., X_n)$ the limit (pivotal) random variable, so $Q_T(\theta) \to^d Q_0$.

Next, we prove the validity of the asymptotic MCT directly showing that Assumptions (6.1)-(6.4) in Dufour (2006) hold in this case. Assumptions (6.1)-(6.3) require that: (i) the statistics $Q_{T1}(\theta), ..., Q_{Tb}(\theta), ..., Q_{TB}(\theta)$ are i.i.d. according to $G_T(y|\theta)$, $\forall \theta \in \Theta$; (ii) Θ is a nonempty subset of $\mathbb{R}^k$; and (iii) $\hat{\theta}_T$ and Q_{T0} are measurable, and $G_T(Q_{T0}|\hat{\theta}_T)$ is a random variable. These assumptions are verified by definition. Assumption (6.4) states that $\forall \eta_0 > 0$, $\forall \eta_1 > 0$, $\exists \delta > 0$ and a sequence of open subsets $D_{T0}(\eta_0)$ in $\mathbb{R}$ such that $\liminf_{T\to\infty} P\left[Q_{T0} \in D_{T0}(\eta_0)\right] \geq 1 - \eta_0$ and

$$||\theta - \theta_0|| \leq \delta \Longrightarrow \limsup_{T\to\infty} \sup_{y \in D_{T0}(\eta_0)} |G_T(y|\theta) - G_T(y|\theta_0)| \leq \eta_1.$$

Noting that we can take $D_{T0}(\eta_0) = (0, 1)$ as $Q_{T0} \in [0, 1]$, the above condition is equivalent to: $\forall \eta_1 > 0$, $\exists \delta > 0$ and $\exists T_1 > 0$ such that

$$\forall y \in (0, 1), \forall \theta : ||\theta - \theta_0|| \leq \delta, \forall T > T_1 = T_1(\theta) \quad |G_T(y|\theta) - G_T(y|\theta_0)| \leq \eta_1. \tag{34}$$

Notice that $T_1 = T_1(\theta)$ as the condition is not uniform in θ. That is $\lim_{T\to\infty} |G_T(y|\theta) - G(y)| = 0$, at all continuity points y of G, i.e., given θ, $\forall \eta > 0$, $\forall y \in (0, 1)$ $\exists T_1 : T > T_1$, $|G_T(y|\theta) - G(y)| < \eta$, where we denoted by $G(y) = P(Q_0 \leq y)$ the asymptotic distribution. Since the function $h(\cdot)$ is a composition of monotonic functions, it is (piecewise) monotonic, so that Q_0 is an absolutely continuous random variable. Thus, Polya's Theorem ensures that, given θ, the convergence is uniform in y, i.e. $\sup_{y\in(0,1)} |G_T(y|\theta) - G(y)| \to 0$, which means that $\forall \epsilon > 0$, $\exists T_1 : T > T_1 = T_1(\theta)$: $\sup_{y\in(0,1)} |G_T(y|\theta) - G(y)| \leq \eta$. So, finally, the inequality

$$|G_T(y|\theta) - G_T(y|\theta_0)| \leq |G_T(y|\theta) - G(y)| + |G_T(y|\theta_0) - G(y)|,$$

shows that condition (34) is satisfied. The final result then follows along the same lines as in the proof of Theorem 2.

Acknowledgements

We are deeply indebted to Jushan Bai, Jean-Marie Dufour, Søren Johansen, Kajal Lahiri, and Paolo Paruolo for useful comments and suggestions. We wish to thank the Editor, Oliver Linton, an Associate Editor, and four referees for very useful comments and suggestions which have helped to improve and develop the paper further. We thank Ryan Taylor for his support in running the Monte Carlo simulations. The usual disclaimer applies. Michele Bergamelli acknowledges financial support from the "1st PhD Scholarship in Memory of Ana Timberlake" and the Centre for Econometric Analysis of Cass Business School. Lynda Khalaf acknowledges financial support from the "2015-2017 Progetto ITALY" at the Department of Management, Economics and Quantitative Methods of University of Bergamo (Italy), and the "Leverhulme Trust Visiting Professorship (VP2-2013-003)" at the Centre for Econometric Analysis, Cass Business School, London (UK).

References

Advanced Information Nobel Prize. (2013). Understanding Asset Prices. *Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2013*.

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

Beaulieu, M.-C. and Dufour, J.-M. and Khalaf, L. (2007). Multivariate Tests of Mean–Variance Efficiency With Possibly Non-Gaussian Errors. *Journal of Business & Economic Statistics 25*, 398–410.

——— (2013). Identification-Robust Estimation and Testing of the Zero-Beta CAPM. *The Review of Economic Studies 80*, 892–924.

Bergamelli, M. and Urga, G. (2015). Detecting Multiple Structural Breaks: Dummy Saturation vs Sequential Bootstrapping. *Discussion Paper, Cass Business School, London*.

Bernard, J.T. and Idoudi, N. and Khalaf, L. and Clement, Y. (2007). Finite Sample Inference

Methods for Dynamic Energy Demand Models. *Journal of Applied Econometrics 22*, 1211–1226.

Bierens, H.J. and Martins, L.F. (2010). Time-varying Cointegration. *Econometric Theory 26*, 1453–1490.

Campbell, J. Y. and Shiller, R. (1987). Cointegration and Tests of Present Value Models. *Journal of Political Economy 95*, 1062–1088.

——— (1989). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies 1*, 195–228.

Cao, H. and Wu, W.B. (2015). Changepoint estimation: another look at multiple testing problems. *Biometrika 102*, 974âĂŞ980.

Chan, N.H. and Yau, C.Y. and Zhang, R.M. (2014). Group LASSO for Structural Break Time Series. *Journal of the American Statistical Association 109*, 590–599.

Cochran, W.G. (1934). The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. *Mathematical Proceedings of the Cambridge Philosophical Society 30*, 178–191.

Dhrymes, P. J. (2000). *Mathematics for Econometrics* (3rd ed.). New York: Springer-Verlag.

Doornik, J. and D. F. Hendry (Eds.) (2013). *Empirical Econometric Modelling - PcGive 14*, Volume I. London: Timberlake Consultants.

Dufour, J.M. (2006). Monte Carlo Tests with Nuisance Parameters: a General Approach to Finite-Sample Inference and Nonstandard Asymptotics. *Journal of Econometrics 133*, 443–478.

Dufour, J.M. and Khalaf, L. and Voia, M. (2015). Finite-sample Resampling-based Combined Hypothesis Tests, with Applications to Serial Correlation and Predictability. *Communications in Statistics - Simulation and Computation 44*, 2329–2347.

Dufour, J.-M. (1997). Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models. *Econometrica 65*, 1365–1387.

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

Dufour, J.-M. and Khalaf, L. and Bernard, J.-T. and Genest, I. (2004). Simulation Based Finite-sample Tests for Heteroskedasticity and ARCH Effects. *Journal of Econometrics 122*, 317–347.

Ge, Y. and Dudoit, S. and Speed, T. P. (2003). Resampling-based multiple testing for microarray data analysis. *Test 12*, 1–77.

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

Hansen, H. and Johansen, S. (1999). Some Tests for Parameter Constancy in Cointegrated VAR-models. *Econometrics Journal 2*, 306–303.

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

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica 56*, 1551–1580.

——— (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* New York: Oxford University Press.

Kallenberg, O. (1997). *Foundations of Modern Probability.* New York: Springer.

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.

Liao, Z. and Phillips, P.C.B. (2015). Automated estimation of vector error correction models. *Econometric Theory 31*, 581–646.

Matteson, D.S. and James, N.A. (2014). A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data. *Journal of the American Statistical Association 109*, 334–345.

Nielsen, F. and Nock, R. (2013). On the Chi Square and Higher-Order Chi Distances for Approximating f-Divergences. *IEEE Signal Processing Letters 0*, 1–4.

Oka, T. and Perron, P. (2011). Testing for Common Breaks in a Multiple Equations System. *Working Paper*.

Phillips, P.C.B. and Shi, S. and Yu, J. (2015a). Testing for multiple bubbles: historical episodes of exuberance and collapse in the S&P 500. *International Economic Review 56*, 1043–1077.

——— (2015b). Testing for multiple bubbles: limit theory of real-time detectors. *International Economic Review 56*, 1079–1133.

Phillips, P.C.B. and Wu, Y. and Yu, J. (2011). Explosive Behavior in the 1990s NASDAQ: When Did Exuberance Escalate Asset Values? *International Economic Review 52*, 201–226.

Preuss, P. and Puchstein, R. and Dette, H. (2015). Detection of Multiple Structural Breaks in Multivariate Time Series. *Journal of the American Statistical Association 110*, 654–668.

Quintos, C. E. (1997). Stability Tests in Error Correction Models. *Journal of Econometrics 82*, 289–315.

Schnurr, A. and Dehling, H. (2017). Testing for Structural Breaks via Ordinal Pattern Dependence. *Journal of the American Statistical Association 112*, 706–720.

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

Westfall, P.H. and Young, S.S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Values Adjustment.* New York: John Wiley & Sons.

Figure 1: Log Dividend-Price Ratio 1871-2014

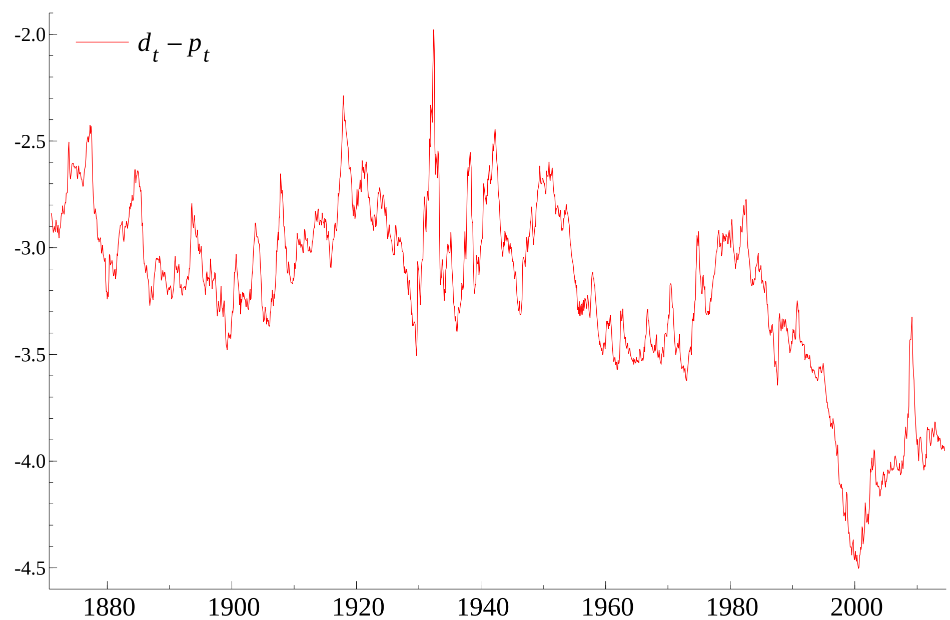

Table 1: Empirical size for Experiments I, II, and III.

	$\alpha_1=-1$	$\alpha_2=0$	$\alpha_1=-0.2$	$\alpha_2=0$	$\alpha_1=-0.01$	$\alpha_2=0$	$\alpha_1=-0.5$	$\alpha_2=0.5$
T	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*
Experiment I								
$H_0: no\ breaks,\ H_1: (T/2) \cup (T/2-T/5) \cup (T/2+T/5)$								
100	5.1	5.6 (4.8)	5.4	11 (5.5)	7.4	30.3 (7.4)	5.2	7 (4.9)
200	5.3	6.3 (5.6)	5.7	9.3 (6.1)	7.3	26.8 (7.5)	5	6.1 (5.7)
300	5.1	6.6 (6)	5.7	7.5 (6.4)	6	27.1 (8.5)	5.7	5.9 (6.1)
400	6.2	5.3 (5.4)	5.4	6.4 (5.2)	6.5	21.7 (6.1)	4.4	5.6 (5.1)
500	5.4	4.7 (4.2)	5.7	5.8 (5.5)	4.9	22.5 (6.1)	5.3	4.9 (4.6)
$H_0: no\ breaks,\ H_1: (T/2) \cup (T/2-20) \cup (T/2+20)$								
100	5.1	5.6 (4.8)	5.4	11 (5.5)	7.4	30.3 (7.4)	5.2	7 (4.9)
200	5.7	5.5 (5.5)	5.2	8 (5.7)	7.8	23 (8.2)	4.6	4.9 (4.4)
300	5.2	5.6 (5.6)	5.3	7.8 (5.8)	7	20 (5.6)	6.4	4.8 (4.9)
400	5.6	6.2 (6.2)	5.4	7.6 (6.6)	6	18.1 (6.8)	5	5.6 (7.2)
500	5.3	4.4 (4.6)	5.5	4.8 (4.4)	4.8	12.7 (5.6)	5.9	5 (5.2)
Experiment II								
$H_0: T_1=T/2,\ H_1: (T_1, T_1+T/5) \cup (T_1, T_1-3T/10) \cup (T_1, T_1-T/4, T_1+T/5)$								
100	6.4	5.9 (6.2)	6	10.1 (6.2)	9.1	23.9 (9.9)	5.3	6.6 (5.5)
200	6.3	6.4 (6.8)	5.7	7.1 (5.2)	10.1	22 (9.8)	4.5	5.1 (4.6)
300	7.2	6.8 (7.4)	8.1	7.6 (9.9)	7.7	19.1 (8.5)	5.2	5 (5.2)
400	7	5.9 (7.9)	5.8	6.9 (6.7)	9.5	17 (10)	6.6	5.8 (6.4)
500	6.8	4.5 (6.3)	5.4	4.9 (4.7)	7.9	16.1 (8.1)	7	4.9 (5.5)
$H_0: T_1=T/2,\ H_1: (T_1, T_1+20) \cup (T_1, T_1-30) \cup (T_1, T_1-25, T_1+20)$								
100	6.4	5.9 (6.2)	6	10.1 (6.2)	9.1	23.9 (9.9)	5.3	6.6 (5.5)
200	7.2	7.2 (7.6)	6.7	8.2 (6.7)	10.4	17.4 (9.8)	5.9	4.9 (5)
300	5.7	5 (5.6)	8.1	5.8 (7.3)	6.4	12.3 (7.1)	5.3	5.5 (6.2)
400	8.3	6.8 (8.4)	6.5	7 (6.9)	10	12.9 (10.1)	6.4	6.3 (6.8)
500	7.1	4.6 (7.1)	4.4	5.1 (3.9)	6.2	9 (6.8)	5.8	5.2 (5.6)
$H_0: T_1=20,\ H_1: (T_1, T_1+20) \cup (T_1, T_1-30) \cup (T_1, T_1-25, T_1+20)$								
100	7.5	0.7 (7.4)	6.6	1 (6.7)	8.1	3.4 (8.3)	4.4	0.3 (4.4)
200	9.7	1.6 (9.7)	8.2	1.6 (8.2)	9.2	2.6 (10.4)	5.7	0.4 (5.8)
300	6.4	0.9 (6.4)	6.9	1.4 (7)	6.4	1.9 (6.8)	6.9	0.5 (6.9)
400	8.6	0.9 (8.4)	6.1	1 (6)	10.2	1.7 (10.1)	6.5	0.5 (6.5)
500	6.4	0.6 (6.5)	4.2	0.9 (4.2)	6	1.2 (6.9)	6.7	0.1 (6.7)
$H_0: T_1=T/5,\ H_1: (T_1, T_1+T/5) \cup (T_1, T_1-3T/10) \cup (T_1, T_1-T/4, T_1+T/5)$								
100	7.5	0.7 (7.4)	6.6	1 (6.7)	8.1	3.4 (8.3)	4.4	0.3 (4.4)
200	6.3	0.7 (6.4)	6.1	0.7 (5.9)	9.5	3.4 (9.5)	5.6	0.7 (5.5)
300	6.8	0.7 (6.7)	6.9	0.5 (6.6)	6.5	2.3 (7)	8.1	0.9 (8.3)
400	8.9	1 (8.9)	6.5	0.5 (6.6)	9.9	2.5 (9.5)	6.2	0.5 (6.3)
500	7.1	0.3 (6.8)	4.9	0.6 (5.2)	6.8	2.2 (7)	7	0.3 (6.9)
Experiment III								
$H_0: T_1=T/2, T_2=2T/3,\ H_1: (T_1, T_2, T_1+20) \cup (T_1, T_2, T_2-5, T_1+15) \cup (T_1, T_2, T_1+10, T_1-10, T_2-5)$								
100	6.7	7.9 (6.2)	6.5	10.1 (5.7)	7	16.5 (7.1)	6.4	7.8 (5.7)
200	7.1	6.5 (6.3)	7.2	8.6 (5.6)	7.7	14.8 (8.5)	5.3	5.2 (4.9)
300	7	5.8 (6.3)	9	7.4 (8.4)	7.3	11.7 (7)	6.8	5.3 (5.2)
400	7.7	7.1 (7.9)	8.4	7.9 (7.8)	8.9	11.4 (9.3)	6.9	5.1 (6)
500	5.9	4.5 (7)	4.3	4.8 (5.4)	4.6	9.1 (6.2)	4.6	3.6 (4.4)
$H_0: T_1=20, T_2=2T/3,\ H_1: (T_1, T_2, T_1+20) \cup (T_1, T_2, T_2-5, T_1+15) \cup (T_1, T_2, T_1+10, T_1-10, T_2-5)$								
100	5	8 (6.6)	5.9	11.8 (5.4)	9.1	25.2 (10.2)	4.8	6.4 (4.9)
200	8.2	6.2 (6.4)	7.7	8 (5.9)	10.1	13.8 (11)	4.3	5 (5.1)
300	7.1	6.2 (7.2)	7.4	7.3 (7.2)	6.7	10.6 (7.2)	6.1	5.3 (5.9)
400	7.8	5.8 (7.2)	7.9	6.5 (7.2)	11.4	11.7 (12.3)	8.2	5.9 (7.1)
500	7.1	5.9 (7.8)	4.9	6.1 (6.2)	6.2	8.6 (8.5)	5.4	3.3 (4.9)
$H_0: T_1=T/2, T_2=T-20,\ H_1: (T_1, T_2, T_1+20) \cup (T_1, T_2, T_2-5, T_1+15) \cup (T_1, T_2, T_1+10, T_1-10, T_2-5)$								
100	5.9	7.2 (6.3)	5.3	10 (4.7)	9.2	23.8 (9.7)	5.4	6.5 (5)
200	7.1	6.9 (6.3)	6.4	8 (6.3)	10.2	18.8 (9.8)	6.2	6.2 (5.6)
300	5.2	4.7 (5.8)	5	6.7 (6.1)	7.6	12.4 (7.3)	5.2	4.6 (4.4)
400	7.2	6.4 (7.3)	6.4	6.9 (6.9)	11.1	12.5 (11)	6	5.6 (6.8)
500	6.3	5.5 (7.4)	4.7	6.1 (5.1)	5.4	10 (7)	4.9	5 (5.3)

Note: Nominal level used to compute the empirical rejection frequencies is 5%. "Q_T" indicates Q_T-statistic, "H" Hansen's asymptotic test statistic, and "H-BT" bootstrapped Hansen's asymptotic test.

Table 2: Empirical power for Experiments I and II (DGP corresponds to one of the alternatives)

	$\alpha_1 = -1$ $\alpha_2 = 0$		$\alpha_1 = -0.2$ $\alpha_2 = 0$		$\alpha_1 = -0.01$ $\alpha_2 = 0$		$\alpha_1 = -0.5$ $\alpha_2 = 0.5$	
Experiment I								
	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*
DGP:$(T_1 = T/2)$, $H_0 : no\ breaks$, $H_1 : (T_1) \cup (T_1 - T/5) \cup (T_1 + T/5)$								
β_2	$T = 100$							
1.1	66	63.8 (61; 60.7)	9.3	14.8 (7.2; 7.5)	6.7	29.9 (5; 7.4)	35.4	33.4 (31.6; 30.4)
1.3	98.5	98.2 (97.8; 97.5)	35.7	41.8 (29.7; 30.4)	7.1	30 (5.1; 7.8)	84.1	82.2 (80.9; 79.8)
1.5	100.0	100 (99.9; 99.8)	58.8	64.6 (52.7; 52.6)	6.5	29.7 (5; 7.5)	95.7	95 (94.4; 94.1)
2.0	100	100 (100; 100)	89.8	90.3 (85.3; 85.9)	7.4	30.5 (4.9; 8.1)	99.8	99.7 (99.7; 99.7)
β_2	$T = 300$							
1.1	98.1	97.3 (97.1; 97.1)	34.5	32.4 (27.3; 28.9)	6.0	26.5 (4.8; 8.4)	76.3	74 (71.9; 73.6)
1.3	100.0	100 (100; 100)	85.0	83.1 (79.9; 80.2)	5.7	27.1 (4.5; 7.9)	99.9	99.8 (99.8; 99.8)
1.5	100.0	100 (100; 100)	97.8	97.4 (96.5; 96.8)	6.7	27.3 (5.1; 8.3)	100.0	100 (100; 100)
2.0	100.0	99.9 (99.9; 99.9)	99.0	99.7 (99.7; 99.7)	10.3	30.9 (6.4; 10.8)	100.0	100 (100; 100)
Experiment II								
	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*
DGP:$(T_1 = T/2, T_1 + T/5)$, $H_0 : T_1$, $H_1 : (T_1, T_1 + T/5) \cup (T_1, T_1 - 3T/10) \cup (T_1, T_1 - T/4, T_1 + T/5)$								
β_2	$T = 100$							
2.1	52.3	52.6 (50.5)	9.0	14.1 (8.0)	6.8	23.9 (5.1)	15.6	17.2 (14.3)
2.3	89.1	87.8 (87.1)	29.2	33.2 (25.4)	6.5	23.9 (5)	54.5	54.3 (52)
2.5	97.0	97 (96.8)	46.6	51.3 (43.0)	6.5	24.3 (5.2)	74.6	73.3 (72.1)
3.0	99.1	100 (100.0)	73.7	75.8 (69.5)	7.1	24.9 (5.3)	96.5	96.2 (95.9)
β_2	$T = 300$							
2.1	84.6	87.3 (84.5)	24.1	26.6 (20.4)	7.1	19.3 (5.4)	50.3	47.5 (48.1)
2.3	95.2	99.3 (99.3)	67.1	69 (64.1)	7.3	19.8 (5.1)	90.2	89.2 (89.5)
2.5	97.3	99.9 (99.9)	80.7	86 (83.2)	8.5	21 (5.0)	97.6	97.7 (97.7)
3.0	96.9	99.6 (99.6)	88.6	97.5 (96.7)	10.8	24.4 (6.9)	98.2	100 (100.0)
DGP:$(T_1 = 20, T_1 + 20)$, $H_0 : T_1$, $H_1 : (T_1, T_1 + 20) \cup (T_1, T_1 - 30) \cup (T_1, T_1 - 25, T_1 + 20)$								
β_3	$T = 100$							
2.1	58.1	41.5 (59.0)	10.5	2.2 (10.2)	5.8	3.5 (5.1)	19.2	5.2 (21.7)
2.3	89.8	86.1 (92.7)	33.3	17.9 (32.9)	5.9	3.7 (4.9)	64.0	44.5 (65.3)
2.5	98.2	96.1 (98.6)	54.4	38.2 (54.0)	5.9	3.6 (5.1)	85.3	69.9 (86.6)
3.0	99.1	99.6 (100.0)	79.9	70.6 (80.2)	6.7	3.4 (5.3)	98.5	96.8 (99.2)
β_3	$T = 300$							
2.1	64.3	48.7 (64.6)	10.6	2.7 (10.4)	4.7	2 (5.4)	25.2	9.7 (22.7)
2.3	88.3	87.3 (93.4)	37.5	23.4 (39)	4.8	1.7 (5)	70.4	55.9 (70)
2.5	97.5	97.5 (99.3)	59.0	47.3 (62.4)	5.5	1.6 (4.9)	87.5	78.3 (87.9)
3.0	95.9	99.6 (99.8)	77.6	76.8 (84.2)	6.3	2.4 (7)	98.5	96.8 (98.9)

Note: Nominal level used to compute the empirical rejection frequencies is 5%. "Q_T" indicates Q_T-statistic, "H" Hansen's asymptotic test statistic, "H-c" Hansen's asymptotic test with size correction, and "H-BT" bootstrapped Hansen's asymptotic test.

Table 3: Empirical power for Experiments I and II (DGP does not correspond to any of the alternatives)

	$\alpha_1=-1$ $\alpha_2=0$		$\alpha_1=-0.2$ $\alpha_2=0$		$\alpha_1=-0.01$ $\alpha_2=0$		$\alpha_1=-0.5$ $\alpha_2=0.5$	
Experiment I								
	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*
DGP:$(2T/3)$, H_0 : *no breaks*, $H_1:(T/2)\cup(T/2-T/5)\cup(T/2+T/5)$								
β_2				$T=100$				
1.1	62.1	59.5 (57.2; 56.8)	8.9	15.2 (7.7; 8.5)	7.4	30 (5.2; 7.7)	32.2	31.5 (30.1; 29.6)
1.3	96.2	95.8 (94.7; 95.2)	30.8	37.3 (25.9; 24.3)	7.3	30.2 (5; 7.6)	77.6	75.9 (74.1; 73.7)
1.5	98.8	99.1 (99.0; 98.5)	47.8	54.9 (41.8; 41.2)	7.0	30.6 (5.7; 8.3)	92.3	91 (90.5; 90.1)
2.0	92.6	95.6 (95.3; 92.8)	72	78.4 (68.7; 66.8)	8.4	30.9 (5.6; 7.9)	99.3	99.3 (99.2; 99.3)
β_2				$T=300$				
1.1	95.8	94.7 (94.3; 95.1)	30.3	30.3 (24; 26.5)	6.0	26.4 (5; 8.7)	74.9	71.8 (70.2; 71.9)
1.3	100.0	100 (100; 100)	79.6	77.3 (73.1; 74.4)	6.4	27.2 (4.8; 8.4)	98.5	98.5 (98.2; 98.4)
1.5	100.0	100 (100; 100)	91.5	91.7 (89.9; 89.9)	7	27.1 (4.9; 8.1)	100	99.9 (99.9; 99.9)
2.0	95.4	97.9 (97.5; 95.6)	87.6	91.6 (90.4; 87.7)	9.1	30.2 (5.4; 9.7)	100.0	100 (100; 100)
Experiment II								
	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*
DGP:$(T_1=T/2,2T/3)$, $H_0:T_1$, $H_1:(T_1,T_1+T/5)\cup(T_1,T_1-3T/10)\cup(T_1,T_1-T/4,T_1+T/5)$								
β_2				$T=100$				
2.1	43.1	43.4 (41.2)	7.5	12.1 (6.6)	6.7	23.9 (5.1)	12.3	13.9 (11)
2.3	82.5	82.2 (81.5)	19.1	22.8 (15.6)	6.3	23.5 (5)	46.3	46.5 (43.7)
2.5	94.1	92.9 (92.2)	31.2	34.9 (25.8)	6.8	23.6 (5)	67.9	66.5 (64.7)
3.0	97.8	98.1 (97.8)	52.5	55.9 (46.2)	7.3	23.4 (5)	92.8	92.1 (90.9)
β_2				$T=300$				
2.1	79.7	79.5 (76.1)	16.3	19.1 (13.3)	7.2	19.1 (5)	40.6	37.7 (38.4)
2.3	94.4	97.4 (96.8)	53.4	57 (49.8)	8.0	20.1 (5)	83.5	80.5 (81)
2.5	97.2	99.5 (99.4)	71.5	76.9 (71)	7.8	20.1 (4.8)	92.7	94.2 (94.3)
3.0	97.7	99.7 (99.5)	79.2	86.8 (83.5)	9.5	20.6 (5.8)	98.6	99.5 (99.6)
DGP:$(T_1=20,2T/3)$, $H_0:T_1$, $H_1:(T_1,T_1+20)\cup(T_1,T_1-30)\cup(T_1,T_1-25,T_1+20)$								
β_3				$T=100$				
2.1	28.4	10.8 (28.6)	7.0	1.2 (5.9)	5.5	3.6 (5.1)	8.4	0.8 (9.6)
2.3	52.8	31.2 (53.4)	10.0	2.2 (9.0)	6.1	3.5 (4.8)	28.7	9.4 (30.1)
2.5	53.4	35.8 (55.3)	12.5	4.7 (12.3)	5.0	3.7 (5.1)	44.7	21.2 (46.7)
3.0	46.2	32.6 (49.3)	18.0	8.8 (17.8)	4.9	3.2 (4.7)	80.5	59.8 (82)
β_3				$T=300$				
2.1	15.9	4.6 (15.9)	4.9	1.3 (4.6)	5.3	2.2 (5.5)	8.9	0.9 (7.6)
2.3	23.8	12.4 (24.5)	5.3	1.5 (5.7)	4.7	1.8 (4.7)	18.4	5 (15.8)
2.5	20.9	11.5 (22)	6.8	1.9 (6.5)	4.6	1.6 (4.8)	26.1	9.8 (24.1)
3.0	14.6	7.1 (15.8)	8.3	3.3 (11.3)	4.2	1.6 (4.8)	61.1	37.6 (58.2)

Note: Nominal level used to compute the empirical rejection frequencies is 5%. "Q_T" indicates Q_T-statistic, "H" Hansen's asymptotic test statistic, "H-c" Hansen's asymptotic test with size correction, and "H-BT" bootstrapped Hansen's asymptotic test.

Table 4: Empirical power for Experiment III

	$\alpha_1=-1$ $\alpha_2=0$		$\alpha_1=-0.2$ $\alpha_2=0$		$\alpha_1=-0.01$ $\alpha_2=0$		$\alpha_1=-0.5$ $\alpha_2=0.5$	
Experiment III								
	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*
DGP covered by one of the alternatives								
DGP:$(T_1=T/2, T_2=2T/3, T_3=\mathbf{T_1+20})$, $H_0:(T_1,T_2)$, $H_1:(T_1,T_2,T_1+20)\cup(T_1,T_2,T_2-5,T_1+15)\cup(T_1,T_2,T_1+10,T_1-10,T_2-5)$								
β_2	$T=100$							
1.6	**26.7**	26.9 (21.5; 22.3)	**6.2**	11.3 (5.4; 5.3)	5.2	16.5 (4.7; **5.8**)	**9.5**	10.8 (8; 8.5)
1.8	**67.1**	66.7 (62.7; 62.9)	**13.4**	17.6 (10.0; 10.6)	5.2	16.9 (5.1; **5.9**)	**31.2**	30 (26.8; 26.8)
2.0	**81.3**	81.2 (78.9; 79.0)	**26.4**	28.9 (20.6; 21.0)	5	17 (5.5; **6.5**)	**48.6**	47.2 (44.5; 44.2)
2.5	**91.4**	91.5 (90.8; 90.7)	**52.7**	55.2 (45.4; 44.7)	4.9	16.3 (5.5; **6.1**)	**70.9**	69.7 (67.2; 67.0)
β_2	$T=300$							
1.6	**64.5**	60.4 (59.7; 58.5)	**11.1**	11.8 (10.1; 9.6)	**5.8**	11.7 (5.1; 5.4)	**27.7**	23.7 (23.4; 22.9)
1.8	**91.5**	90.2 (89.7; 87.9)	**38.6**	37.3 (33.9; 32.0)	**6.2**	11.4 (5; 5.9)	**68.5**	61.0 (60.8; 61.0)
2.0	**96.2**	97.5 (97.3; 94.7)	**56.9**	57.1 (53.8; 51.5)	5.7	11.2 (5.4; **5.8**)	**83.7**	78.2 (78; 78.2)
2.5	**96.2**	99.7 (99.7; 96.1)	**75.0**	80.3 (78.3; 71.9)	**7.2**	13 (6.3; 7.1)	**97.1**	96.6 (96.6; 94.9)
DGP:$(T_1=20, T_2=2T/3, T_3=T_2-5)$, $H_0:(T_1,T_2)$, $H_1:(T_1,T_2,T_1+20)\cup(T_1,T_2,T_2-5,T_1+15)\cup(T_1,T_2,T_1+10,T_1-10,T_2-5)$								
β_2	$T=100$							
1.6	**30.5**	32.4 (26.4; 28.5)	**7.2**	13 (6.2; 5.9)	8.1	25.9 (4.9; **8.5**)	**7.4**	9.0 (7.2; 7.2)
1.8	**71**	71.7 (68.0; 69.0)	**17**	22.6 (14.6; 15.6)	7.5	24.4 (4.5; **8.6**)	**27.4**	28.2 (25.7; 25.6)
2.0	**83.4**	83.4 (81.3; 81.7)	**28.9**	35.7 (25.9; 26.0)	7.1	24.3 (4.2; **8.0**)	**48.6**	47.6 (44.6; 44.0)
2.5	**94**	94.0 (92.5; 92.0)	**55.7**	59.2 (51.8; 52.1)	6.8	23 (4.5; **8.7**)	**77.6**	76.5 (75.1; 75.9)
β_2	$T=300$							
1.6	**45.9**	45.9 (43.5; 41.6)	7	8.3 (6.5; **7.1**)	4.6	11.1 (5.7; **5.7**)	**15.3**	12.1 (11.8; 12.4)
1.8	**77.2**	79.8 (78.9; 75.9)	**23.0**	25.9 (23.0; 20.7)	4.6	11.2 (5.9; **6.0**)	**47.7**	46.1 (45.7; 46.2)
2.0	**84.6**	87.2 (86.7; 83.7)	**43.3**	45.1 (42.1; 40.0)	4.3	11.5 (5.8; **5.9**)	**67.1**	64.4 (64.0; 63.4)
2.5	**92.1**	95.4 (95.1; 91.6)	**65.9**	71.0 (69.7; 65.0)	5.2	11.2 (5.3; **6.4**)	**85.1**	84.5 (84.4; 83.5)
DGP not covered by any of the alternatives (misspecified case)								
DGP:$(T_1=T/2, T_2=2T/3, T_3=5T/6)$, $H_0:(T_1,T_2)$, $H_1:(T_1,T_2,T_1+20)\cup(T_1,T_2,T_2-5,T_1+15)\cup(T_1,T_2,T_1+10,T_1-10,T_2-5)$								
β_2	$T=100$							
1.6	11.9	10.5 (6.6; 8.1)	5.7	10.7 (4.7; 4.4)	5.4	16.9 (5.1; 6.1)	6.9	8.2 (5; 6.3)
1.8	32.5	24.3 (18.2; 17.5)	6.4	10.6 (5.1; 5.1)	5.5	17.3 (5.4; 7.1)	9.1	8.4 (6.9; 6.7)
2.0	42.4	30.5 (24.4; 23.7)	9.2	12.4 (6.7; 7.0)	5.1	16.8 (5.3; 6.6)	11	7.7 (6.1; 6.1)
2.5	43.7	30.6 (23.1; 20.1)	15.9	20.3 (13.1; 12.3)	5.2	15.8 (5.2; 6.2)	15.6	8.7 (6.3; 6.4)
β_2	$T=300$							
1.6	3.5	2.7 (2.3; 2.9)	5.5	6.9 (4.4; 5.2)	5.9	11.4 (4.7; 5.3)	5.3	4.2 (4.2; 4.3)
1.8	2.1	1.3 (1.3; 1.6)	6.4	8.1 (5.5; 5.9)	6.2	11.3 (5; 5.6)	3.7	2.5 (2.4; 2.3)
2.0	2.8	3.2 (2.5; 2.1)	10.5	13.9 (11.9; 10.1)	5.4	11.2 (4.8; 5.2)	3.1	1.9 (1.9; 1.8)
2.5	2.7	3.2 (3.1; 2.3)	15.9	23.5 (20.1; 16.7)	4.6	10.8 (4.5; 5.3)	1.8	1.3 (1.3; 1.2)
DGP:$(T_1=20, T_2=2T/3, T_3=5T/6)$, $H_0:(T_1,T_2)$, $H_1:(T_1,T_2,T_1+20)\cup(T_1,T_2,T_2-5,T_1+15)\cup(T_1,T_2,T_1+10,T_1-10,T_2-5)$								
β_2	$T=100$							
1.6	3.7	6.1 (3.7; 4.3)	5.6	11.8 (5.1; 5.3)	8.3	25.5 (5.1; 8.7)	4.4	5.7 (4.5; 4.4)
1.8	2.4	3.1 (1.5; 2)	5.3	10.2 (3.9; 3.7)	7.3	24.5 (4.3; 7.7)	3.4	4.3 (3.3; 3.3)
2.0	1.5	1.8 (0.7; 1.2)	4.2	8.5 (3.4; 3.6)	7.1	25.2 (4.6; 8.8)	2.7	3.3 (2.7; 2.5)
2.5	0.9	1 (0.6; 0.4)	3.8	8.7 (3.7; 3.4)	6.4	23.1 (4; 8.1)	1.4	1.4 (1; 1.1)
β_2	$T=300$							
1.6	2.7	2.8 (2.6; 2.2)	5.1	6.6 (4.8; 5.4)	4.6	11 (5.4; 5.4)	5.6	4.5 (4.4; 4.4)
1.8	1.1	1.1 (0.8; 0.7)	4.7	5.7 (4.5; 4.3)	4.3	10.6 (5.7; 5.9)	3.1	2.7 (2.4; 2.9)
2.0	1.0	0.5 (0.4; 0.3)	5.7	6.9 (5.3; 4.7)	4.5	10.7 (5.7; 5.6)	2.7	1.6 (1.6; 2.1)
2.5	0.7	0.6 (0.6; 0.3)	7.0	10.9 (8.1; 6.7)	4.1	10.3 (5; 5.2)	2.6	1.2 (1.2; 1.3)

Note: Nominal level used to compute the empirical rejection frequencies is 5%. "Q_T" indicates Q_T-statistic, "H" Hansen's asymptotic test statistic, "H-c" Hansen's asymptotic test with size correction, and "H-BT" bootstrapped Hansen's asymptotic test.

Table 5: Empirical size and empirical power for Experiment IV (breaks in β only).

	$\alpha_1=-1$ $\alpha_2=0$		$\alpha_1=-0.2$ $\alpha_2=0$		$\alpha_1=-0.01$ $\alpha_2=0$		$\alpha_1=-0.5$ $\alpha_2=0.5$	
Experiment IV								
Empirical size								
T	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*
$H_0: no\ breaks,\ H_1: 5\cup10\cup15\cup....\cup45\cup50$								
100	5.7	9.4 (4.9)	5.7	17.4 (5.4)	7.7	32.2 (9)	4.6	9.9 (5.2)
$H_0: no\ breaks,\ H_1: 5\cup10\cup15\cup....\cup90\cup95$								
100	5.2	19.8 (6.2)	5.6	53.1 (6.8)	7.4	69.9 (8.2)	4.4	22.2 (4.4)
$H_0: no\ breaks,\ H_1: 5\cup6\cup7\cup....\cup93\cup94\cup95$								
100	5.6	-	3.1	-	4.7	-	5.8	-
Empirical power								
	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*
DGP covered by one of the alternatives								
β_2	$DGP: T_1=50,\ H_0: no\ breaks,\ H_1: 5\cup10\cup15\cup....\cup45\cup50$							
1.1	**60.7**	55.1 (47.6; 46.6)	**7.3**	19 (7; 6.5)	7.3	32.2 (5.2; **8.5**)	**30.3**	29.8 (19.7; 20.2)
1.3	**97.2**	95.7 (94; 93.8)	**28.6**	38.6 (20; 18.5)	7	31.5 (4.7; **8.3**)	**79.9**	72.3 (64.4; 65.2)
1.5	**100**	99.9 (99.4; 99.4)	**52.4**	57.1 (36.8; 36)	6.7	30.6 (4.2; **7.8**)	**94.2**	89.9 (86.5; 87.5)
2.0	100	100 (100 ; 100)	**86.5**	83.1 (70; 69.5)	7.1	29.7 (4.6; **7.5**)	**99.6**	99.4 (99; 99.2)
β_2	$DGP: T_1=50,\ H_0: no\ breaks,\ H_1: 5\cup10\cup15\cup....\cup90\cup95$							
1.1	**58.3**	57.4 (36.7; 37.5)	7.4	53.7 (5.5; **7.7**)	7.7	70.6 (5; **8.5**)	**28.3**	38.3 (17.3; 13.9)
1.3	**96.8**	94.6 (86.2; 87.4)	**23.3**	63.5 (8.5; 11.2)	7.7	71.1 (5; **8.8**)	**77.4**	72.2 (55.6; 53)
1.5	**99.9**	99.6 (97.9; 97.9)	**46.5**	71.5 (17.2; 19.2)	6.8	70.6 (4.8; **7.8**)	**93.1**	88.5 (78.4; 78.7)
2.0	100	100 (100; 100)	**81.9**	85.4 (40.4; 47.2)	5.9	71.1 (4.7; **8.7**)	**99.6**	98.9 (96.8; 97)
β_2	$DGP: T_1=50,\ H_0: no\ breaks,\ H_1: 5\cup6\cup7\cup....\cup93\cup94\cup95$							
1.1	55.3	-	5.2	-	5.7	-	27.1	-
1.3	96.3	-	10.5	-	5	-	75.6	-
1.5	99.7	-	30.5	-	4	-	92.9	-
2.0	99.8	-	63.7	-	3.3	-	99.5	-
DGP not covered by any of the alternatives (misspecified case)								
β_2	$DGP: T_1=55,\ H_0: no\ breaks,\ H_1: 5\cup10\cup15\cup....\cup45\cup50$							
1.1	**54.1**	47.9 (39.6; 38.6)	**6.6**	18.6 (5.8; 6)	7.5	31.9 (4.7; **8.4**)	**24.6**	23.6 (15.1; 15.4)
1.3	**94.7**	88.2 (84.5; 83.6)	**17.2**	26.9 (11.9; 10.3)	7.4	31.2 (4.1; **8.1**)	**72.4**	61.4 (52; 51.9)
1.5	**99.3**	96.8 (94.5; 92.5)	**33.8**	37.2 (18; 15.6)	7.5	31.4 (4.8; **7.9**)	**89.4**	80.1 (71.4; 70.5)
2.0	**98.9**	93.7 (90.7; 84.6)	**64.7**	59.7 (37.9; 35.5)	6.7	28.9 (4.3; **7.4**)	**98.6**	94.8 (90.3; 89.6)

Note: Nominal level used to compute the empirical rejection frequencies is 5%. "Q_T" indicates Q_T-statistic, "H" Hansen's asymptotic test statistic, and "H-BT" bootstrapped Hansen's asymptotic test.

Table 6: Empirical size and empirical power for Experiment V (breaks in α_1 and β).

Experiment V										
Empirical size										
	$\alpha_1 = -1$, $\alpha_2 = 0$		$\alpha_1 = -0.01$, $\alpha_2 = 0$							
T	Q_T	*H (H-BT)*	Q_T	*H (H-BT)*						
H_0 : *no breaks*, $H_1 : 5 \cup 10 \cup 15 \cup .... \cup 45 \cup 50$										
100	4.3	35.5 (4.1)	6.6	96.4 (6.5)						
H_0 : *no breaks*, $H_1 : 5 \cup 10 \cup 15 \cup .... \cup 90 \cup 95$										
100	4.2	49.3 (4.6)	9.5	99.9 (8.2)						
Empirical power										
	$\alpha_{12} = -0.02$		$\alpha_{12} = -1.1$		$\alpha_{12} = -1.3$		$\alpha_{12} = -1.5$		$\alpha_{12} = -2.0$	
	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*	Q_T	*H (H-c; H-BT)*
DGP covered by one of the alternatives										
β_2	$DGP : T_1 = 50$, H_0 : *no breaks*, $H_1 : 5 \cup 10 \cup 15 \cup .... \cup 45 \cup 50$									
1.1	**7.6**	96.6 (4.6; 6.2)	**63.6**	71.8 (39; 36.3)	**82.2**	83.8 (58; 56.4)	**98**	94.8 (80.4; 78.3)	**100**	100 (99.9; 99.9)
1.3	**7**	96.1 (4.8; 6.5)	**97**	98.2 (92.7; 89.7)	**99.3**	99.6 (97.3; 96.4)	**100**	100 (99.2; 99.2)	100	100 (100; 100)
1.5	**7.3**	95.9 (5 ; 6.4)	**99.9**	99.9 (99.5; 98.8)	**100**	100 (99.9; 99.9)	100	100 (100; 100)	100	100 (100; 100)
2.0	**9.3**	96.1 (5.1; 7.1)	100	100 (100; 100)	100	100 (100; 100)	100	100 (100; 100)	100	100 (100; 100)
β_2	$DGP : T_1 = 50$, H_0 : *no breaks*, $H_1 : 5 \cup 10 \cup 15 \cup .... \cup 90 \cup 95$									
1.1			**59.4**	72.2 (25.4; 22.9)	**80.1**	78.7 (36.2; 33.9)	**97.3**	88 (52.8; 50.8)	**100**	99.3 (97.3; 96.8)
1.3			**96.9**	98.5 (79.9; 75.4)	**99.4**	99.1 (87.7; 84.7)	**100**	99.2 (94.2; 93.1)	100	100 (100; 100)
1.5			**99.7**	99.9 (97; 94.3)	**100**	100 (98.6; 97.7)	**100**	100 (99.7; 99.4)	**100**	100 (100; 99.9)
2.0			**100**	100 (100; 99.7)	**100**	100 (100; 99.8)	100	100 (100; 100)	100	100 (100; 100)
DGP not covered by any of the alternatives (misspecified case)										
β_2	$DGP : T_1 = 55$, H_0 : *no breaks*, $H_1 : 5 \cup 10 \cup 15 \cup .... \cup 45 \cup 50$									
1.1	**7.1**	95.9 (4.9; 5.7)	**54**	65.1 (29.1; 25.5)	**72.4**	76.7 (43; 39.3)	**94**	90.1 (66; 63)	100	100 (100; 99.6)
1.3	**6.9**	96 (4.9; 6.4)	**94.2**	95.6 (79.8; 71)	**97.6**	97.9 (87.1; 81.6)	**99.8**	99.3 (93.7; 90.6)	100	100 (100; 99.9)
1.5	**6.9**	95.7 (4.4; 6.6)	**99.2**	99.2 (93.6; 81.7)	**99.7**	99.7 (95.9; 85.1)	**100**	100 (98.2; 91)	100	100 (100; 100)
2.0	**7.4**	95.7 (4.6; 6.2)	**99.6**	99.9 (94.3; 73.3)	**99.9**	99.9 (94.3; 72.6)	**100**	100 (95.8; 77.8)	**100**	100 (99.9; 99.9)

Note: Nominal level used to compute the empirical rejection frequencies is 5%. "Q_T" indicates Q_T-statistic, "H" Hansen's asymptotic test statistic, "H-c" Hansen's asymptotic test with size correction, and "H-BT" bootstrapped Hansen's asymptotic test.

Table 7: Q_T-statistic results

A.				Breaks in $\boldsymbol{\beta}$				
	# breaks	Break1	Break2	Break 3	log-lik	$LR(\mathcal{H}_0\|\mathcal{H}_1)$	DoF	p-value
$\mathcal{H}_0$	0				1681.3	0	0	0
$\mathcal{H}_1 1$	1	1987(3)			1681.4	0.2319	1	0.6301
$\mathcal{H}_1 2$	1	1999(3)			1684.2	5.8025	1	0.0160
$\mathcal{H}_1 3$	1	2007(4)			1682.2	1.7152	1	0.1903
				$\boldsymbol{Q_T}$ $\boldsymbol{p}$-value = **0.20736**				
B.				Breaks in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$				
	# breaks	Break1	Break2	Break 3	log-lik	$LR(\mathcal{H}_0\|\mathcal{H}_1)$	DoF	p-value
$\mathcal{H}_0$	0				1681.3	0	0	0
$\mathcal{H}_1 1$	1	1987(3)			1681.4	0.2988	3	0.9603
$\mathcal{H}_1 2$	1	1999(3)			1685.3	8.0054	3	0.0459
$\mathcal{H}_1 3$	1	2007(4)			1692.7	22.7650	3	0.0000
				$\boldsymbol{Q_T}$ $\boldsymbol{p}$-value = **0.08027***				
C.				Breaks in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$				
	# breaks	Break1	Break2		log-lik	$LR(\mathcal{H}_0\|\mathcal{H}_1)$	DoF	p-value
$\mathcal{H}_0$	0				1681.3	0	0	0
$\mathcal{H}_1 1$	2	1999(3)	2007(4)		1698.5	34.4850	6	0.0000
$\mathcal{H}_1 2$	3	1987(3)	1999(3)	2007(4)	1699.5	36.4970	9	0.0000
				$\boldsymbol{Q_T}$ $\boldsymbol{p}$-value = **0.03345****				
D.				Breaks in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$				
	# breaks	Break1	Break2		log-lik	$LR(\mathcal{H}_0\|\mathcal{H}_1)$	DoF	p-value
$\mathcal{H}_0$	2	1999(3)	2007(4)		1698.5	0	0	0
$\mathcal{H}_1 1$	3	1987(3)	1999(3)	2007(4)	1699.5	2.012	3	0.5699
				$\boldsymbol{Q_T}$ $\boldsymbol{p}$-value = **0.53846**				

Note: For each test the number of break dates under the null and the alternative hypotheses is reported together with the location of the breaks. Under the column labelled "Log-lik" the value of the maximised log-likelihood function for the corresponding model is reported while $LR(\mathcal{H}_0|\mathcal{H}_1)$ denotes the value of the likelihood-ratio test of the null hypothesis against each scenario. "**" and "*" denotes rejection of the null hypothesis at 5% and 10% level, respectively. The Q_T-statistic p-values are obtained by parametric bootstrap (see Section 3.2) with 299 replications.