Weak Exogeneity, Cointegration and Stability Tests*

Annamaria Bianchi[†], Lynda Khalaf[‡], and Giovanni Urga[§]

January 11, 2025

Abstract

This paper contributes to the literature on time-varying cointegration, identification and exogeneity. We document the implications on stability testing of restrictions resulting from a triangular DGP which entails classifying observables in two groups so that a specific subset does not cointegrate. We show that: (i) such restrictions can substantially affect break tests; (ii) weak exogeneity throughout the sample alleviates such problems; (iii) methodological implications may be avoided by an invariant treatment and bootstraps, using generalized reduced rank regression, with multiple or uncertain break dates. Analytical, Monte Carlo and empirical analyses illustrate the usefulness of the results.

JEL classification number: C12, C15, C32.

Keywords: Weak Exogeneity, Identification, Structural Stability, Time Varying Cointegration, Reduced Rank Regression.

*We extend our gratitude to Michele Bergamelli for his invaluable contributions. Special thanks go to Zhuangyan Li for his support and assistance with the Ox code, as well as for his patience in addressing our numerous and ongoing requests. Lynda Khalaf acknowledges financial support from the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada.

[†]Department of Economics, Università degli Studi di Bergamo, Via dei Caniana, 2 24127, Bergamo, Italy. annamaria.bianchi@unibg.it

[‡]Lynda Khalaf, Economics Department, Carleton University, Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada. LyndaKhalaf@cunet.carleton.ca

[§]Centre for Econometric Analysis and Faculty of Finance, Bayes Business School (formerly Cass), City St George's University of London, 106 Bunhill Row, London (UK). g.urga@city.ac.uk

1 Introduction

A central and enduring premise of Sir David Hendry's work is that exogeneity and parameter stability assumptions should not be taken for granted. In this paper, we illustrate such concerns with focus on a framework that interconnects several features of Hendry's contributions: cointegration, structural breaks and exogeneity.

The literature on time-evolving cointegrating relationships has recently gained momentum, in response to structural stability concerns; see e.g. Phillips et al. (2017) and Banerjee and Carrion-i Silvestre (2025) in the panel context, and Bierens and Martins (2010), Martins (2018), Kapetanios et al. (2020), Eroğlu et al. (2022) in time series. This paper adds to this body of work by revisiting the standard trade-off between a triangular representation (TR) versus a Vector Error Correction Model (VECM), when parameter stability is under test.

The motivation for our focus is two fold. First, while the trade-off in question is long-standing, it has not been analyzed from the perspective of stability testing. Second, we believe - and actually show - that associated statistical problems can be tackled by giving up prior weak exogeneity assumptions. Such an approach has long been a main driver of Hendry's contributions.

There are several ways of specifying cointegration systems (Gonzalo, 1994; Watson, 1994; Johansen, 2009; Gomez-Biscarri and Hualde, 2015), each determining the statistical tools required for inference. In particular, the TR classifies considered observables in two groups so that a set of variables do not cointegrate. The VECM does not depend on such classifications, yet a popular normalization can define a long-run parameter that closely reproduces the TR one. Such normalizations place no constraints on the cointegration space, whereas TRs explicitly restrict this space for identification purposes (Phillips, 1994). While an economic relation can thus be statistically analyzed by interpreting the TR cointegration coefficient or its normalized VECM counterpart, the statistical underpinnings of these alternative specifications are not innocuous.

When the structure is stable, restrictions linking the TR to a VECM form can be formulated. These restrictions resemble linkages between the structural and reduced forms of simultaneous equations (Phillips, 1994), whereby exogeneity intervenes consequentially. In the presence of breaks, the mapping between a TR and its VECM form is less straightforward. Motivated by the above, we revisit the fundamental question raised by Phillips (1991, p.284): "What if the parameters of the transient dynamics themselves rely on the cointegrating coefficient?" Extending a prototypical model proposed by Gonzalo (1994), we address this question for stability testing in a multi-equation systems perspective with known and uncertain break dates.

Available stability tests for cointegration systems are rather scarce. [1] Early work on TRs may be traced back to Hansen (1992). The proposed and well known LM test which is based on a fully-modified or dynamic OLS estimator is designed to assess breaks only in the long-run regression coefficient. One consequence is that the system's residual process, treated non-parametrically, is not assessed for breaks. In contrast, the parametrization given by VECMs can serve to test adjustment as well as long-run coefficients. VECM based stability tests identified via specific normalizations include Seo (1998) and Hansen and Johansen (1999), who also revisit the test of Quintos (1997). Hansen and Johansen (1999) propose a normalization-free eigenvalue based method. Hansen (2003) proposes the generalized reduced rank regression (GRR-reg) framework where parameters are piecewise constant. Bierens and Martins (2010) introduce a VECM with a smoothly varying long run parameter, which nests Hansen (2003)'s specification. Bergamelli et al. (2019) extend GRR-reg tests to account for unknown break dates. GRR-reg seems particularly well suited to detect parameter discontinuities in a flexible way, for all model parameters, that is, long run, short run, determinist trend terms, and error variances/covariances. It is this framework that is our research focus.

Available Monte Carlo designs on all the above-cited stability tests impose weak exogeneity, which unduly restricts high-order dynamics. Furthermore, there is practically no guidance on the main question posed by Phillips (1991) which we rephrase here as it applies to stability assessments: is it necessary that the transient dynamics be jointly assessed, and are there costs otherwise if residuals are treated non-parametrically?.

With this background, this paper has three main contributions. *First*, we ask whether one can reliably test the stability of a TR via GRR-reg based methods. Pursuing the parallel between TRs and simultaneous equations, the usefulness of reduced forms for specification testing in the latter context can be traced back to the exact tests of Harvey and Phillips (1980, 1981a,b, 1989). The TR to VECM mapping is much more intricate

[1] See *e.g.* Kejriwal and Perron (2008, 2010) for single equation tests.

than with linear simultaneous equations, which motivates our work. Formally, we consider alternative TRs underlying Hansen (2003)'s break test, that yield different insights into its properties under weak exogeneity or lack there-off. *Second*, we revisit the bootstrap validity results from Bergamelli et al. (2019) with a TR as a data generating process (DGP) and uncertain break dates. Bergamelli et al. (2019) did not consider TRs and Gonzalo (1994) did not consider breaks. In this respect, this paper is the first that reconciles and unifies both contexts. *Third*, we conduct simulations imposing and relaxing exogeneity, with known and uncertain break dates. An empirical exercise on an interest rate rule for U.S. data further illustrates the concrete consequences of our analysis.

Results confirm the usefulness of GRR-reg. Specifically, we show that it is possible to obtain useful GRR-reg based break tests even when the underlying data generating process (DGP) is a TR. The rationale behind our finding stems from the following fact: the restrictions linking the TR to its VECM form imply that breaks are imminent in *all* VECM parameters when the coefficient of the TR breaks, unless one takes an ex-ante stand on weak exogeneity in which case breaks may only affect the long run VECM coefficient. Serious distortions may thus result from the outstanding (non-tested) instabilities when weak exogeneity fails. By relying on the implications of the TR restrictions, thereby assessing short and long run VECM parameters for breaks and thus avoiding weak exogeneity assumptions, GRR-reg based break tests can reliably inform on the stability of the considered cointegrating relation. Said differently, giving up weak exogeneity makes stability testing possible without the identification requirements of TRs. Our results further underscore the importance of accounting for variance instabilities, which can go a long way towards substantiating cointegrating relationships in empirical work.

The paper is organized as follows. Section 2 discusses the role of weak exogeneity in VECM and TRs. Section 3 shows illustrative simulations, Section 4 presents an empirical application, and Section 5 concludes.

2 Weak exogeneity and triangular forms

Consider a p-dimensional process $\{X_t\}_{t=1}^T$ with *cointegration rank* p^*. Its VECM form can be written as the reduced rank regression

$$\Delta X_t = \alpha\beta^\top X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \epsilon_t, \quad t = 1, \dots, T, \tag{1}$$

where $\epsilon_t \overset{i.i.d}{\sim} N(0, \Omega)$, the *adjustment matrix* α and *cointegration matrix* β are $p \times p^*$ with rank p^*. GRR-reg tests are likelihood ratio procedures to assess (e.g. nested) alternative VECMs with piece-wise constant time varying parameters that take a similar reduced-rank regression form.

Stack $\Delta X_{t-1}, \dots, \Delta X_{t-k+1}$ into the vector $\dot{Y}_t$. Maximizing the system's likelihood over β concentrating all other parameters (Johansen, 1995, Chapter 6) amounts to minimizing

$$\mathcal{S} = \frac{|S_{00}| \left|\beta^\top \left(S_{11} - S_{10}S_{00}^{-1}S_{01}\right)\beta\right|}{|\beta^\top S_{11}\beta|}, \quad S_{ij} = T^{-1}\sum\nolimits_{t=1}^T R_{it}R_{jt}^\top, \quad i, j = 0, 1, \tag{2}$$

where R_{0t} is the least squares (**LS**) residual from the regression of $\dot{X}_t$ on $\dot{Y}_t$ and R_{1t} is the residual from the regression of Y_t on $\dot{Y}_t$, with $\dot{X}_t = \Delta X_t$, $Y_t = X_{t-1}$.

Now consider the normalization $\beta = (I_{p^*}, \tilde{\beta}^\top)^\top$. A TR conformable with $X_t = (X_{1,t}^\top, X_{2,t}^\top)^\top$ where $X_{1,t}$ is $(p^* \times 1)$, $X_{2,t}$ is $((p - p^*) \times 1)$ takes the form

$$X_{1,t} - \tilde{\beta}^\top X_{2,t} = Z_t, \tag{3}$$

$$\Delta X_{2,t} = V_t \tag{4}$$

$$(Z_t^\top, V_t^\top)^\top = B(L)e_t \tag{5}$$

where $B(L)$ is some lag polynomial ensuring stationarity, e_t is *i.i.d.* and contemporaneously correlated, and Z_t and V_t are stationary such that $(Z_t^\top, V_t^\top)^\top$ takes form (5). One aim of this paper is to provide an analytical basis for applying GRR-reg break tests when a triangular DGP is subsumed.

Early references linking and comparing these representations include Phillips (1991), Phillips (1994), Phillips and Loretan (1991), Watson (1994), and Gonzalo (1994). In particular, Watson (1994, Section 3) derives (3)-(5) from the finite order VAR underlying (1). The mapping in question produces a special structure for $B(L)$, which is typically not maintained when DGPs are considered as TRs. Conversely, the above cited works confirm that specific parametric assumptions on $(Z_t^\top, V_t^\top)^\top$ are required to include (1) in (3)-(5), with parameters depending on $\tilde{\beta}$.

To concretize the consequences of these dependencies, we consider the following TR:

$$X_{1,t} - \tilde{\beta}^\top X_{2,t} = Z_t; \qquad \left(I_{p^*} - \textstyle\sum_{j=1}^{m} \rho_j L^j\right) Z_t = u_{1,t} \tag{6}$$

$$A_1 X_{1,t} - A_2 X_{2,t} = W_t; \qquad \left(I_{p-p^*} - \textstyle\sum_{j=1}^{m} \theta_j L^j\right) \Delta W_t = u_{2,t}, \tag{7}$$

where $u_{1,t}$ is $p^* \times 1$, $u_{2,t}$ is $(p - p^*) \times 1$, $\left(u_{1t}^T, u_{2t}^T\right)^T$ is a sequence of independent Gaussian random variables with mean zero and variance matrix Ω. Usual stability assumptions are maintained, unit roots are imposed, and m refers to the maximum length (some of the lags may have zero coefficients). This system takes the (3)-(5) form with, $e_t = \left(u_{1,t}^T, u_{2,t}^T\right)^T$, $V_t = D^{-1}\Delta W_t - D^{-1}A_1\Delta Z_t$, and

$$B(L) = \begin{bmatrix} \left(I_{p^*} - \sum_{j=1}^{m} \rho_j L^j\right)^{-1} & 0 \\ -D^{-1}A_1(I_{p^*} - L)\left(I_{p^*} - \sum_{j=1}^{m} \rho_j L^j\right)^{-1} & D^{-1}\left(I_{p-p^*} - \sum_{j=1}^{m} \theta_j L^j\right)^{-1} \end{bmatrix}, \tag{8}$$

and $D = A_1\tilde{\beta}^\top - A_2$.[2] Weak exogeneity of $X_{2,t}$ with respect to $\tilde{\beta}$ can be imposed in this context setting $A_1 = 0$.

Gonzalo (1994) considered a special case of this process with $m = 1$, $\theta_1 = 0$ and $p = 2$, in order to compare the properties of then available estimators of $\tilde{\beta}$, concluding broadly in favour of VECM reduced rank-based MLE. Khalaf and Urga (2014) conclude along these same lines for inference purposes, allowing for weak identification. Nevertheless, these comparisons were not designed for the analysis of structural breaks. Thus, our goal is a reassessment of this view, taking into account the effects of instability.

The starting point for our analysis is the mapping from the considered TR into a VECM transform, as derived next. Our objective is to characterize the dependencies between short and long run VECM parameters. These relationships motivate and explain our simulation designs and findings.

[2] All the invertability conditions on the polynomials and on D are conformable with Gonzalo (1994), although his model is a special case of our model with $m = 1$, $\theta_1 = 0$, $p = 2$, $p^* = 1$.

Using standard algebra (Harville, 1997, p.99), write (6)-(7) in the VECM form:

$$\Delta X_t = \mathcal{B}\mathcal{D} \begin{bmatrix} \sum_{j=1}^{m} \rho_j - I_{p^*} \\ 0 \end{bmatrix} \begin{bmatrix} I_{p^*} & -\tilde{\beta}^\top \end{bmatrix} X_{t-1} \tag{9}$$

$$+ \mathcal{B}\mathcal{D} \begin{bmatrix} -\sum_{j=2}^{m} \rho_j & 0 \\ 0 & \theta_1 \end{bmatrix} \mathcal{D}^{-1}\mathcal{B}^{-1}\Delta X_{t-1} + ...$$

$$+ \mathcal{B}\mathcal{D} \begin{bmatrix} -\rho_m & 0 \\ 0 & \theta_{m-1} \end{bmatrix} \mathcal{D}^{-1}\mathcal{B}^{-1}\Delta X_{t-(m-1)}$$

$$+ \mathcal{B}\mathcal{D} \begin{bmatrix} 0 & 0 \\ 0 & \theta_m \end{bmatrix} \mathcal{D}^{-1}\mathcal{B}^{-1}\Delta X_{t-m} + \mathcal{B}\mathcal{D} \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix},$$

$$\begin{bmatrix} I_{p^*} & -\tilde{\beta}^\top \\ A_1 & -A_2 \end{bmatrix}^{-1} = \mathcal{B}\mathcal{D}, \quad \mathcal{B} = \begin{bmatrix} I_{p^*} & \tilde{\beta}^\top \\ 0 & I_{p-p^*} \end{bmatrix}, \quad \mathcal{D} = \begin{bmatrix} I_{p^*} & 0 \\ -D^{-1}A_1 & D^{-1} \end{bmatrix}. \tag{10}$$

Recalling that $|\mathcal{B}| = 1$, the associated concentrated likelihood can be optimized through the objective function

$$\tilde{\mathcal{S}} = \left|\mathcal{D}^{-1}\right| \mathcal{S} \left|\left(\mathcal{D}^{-1}\right)^\top\right| \tag{11}$$

where $\mathcal{S}$ is as in (2). Indeed, the residual from the regression of Y_t on $\mathcal{D}^{-1}\mathcal{B}^{-1}\dot{Y}_t$ is unaffected by this transformation and the residual $\tilde{R}_{0t}$ from the regression of $\mathcal{D}^{-1}\mathcal{B}^{-1}\dot{X}_t$ on $\mathcal{D}^{-1}\mathcal{B}^{-1}\dot{Y}_t$ is transformed into $\mathcal{D}^{-1}\mathcal{B}^{-1}R_{0t}$, so

$$\left|T^{-1}\sum_{t=1}^{T} \tilde{R}_{0t}\tilde{R}_{0t}^\top\right| = \left|\mathcal{D}^{-1}\right| |S_{00}| \left|\left(\mathcal{D}^{-1}\right)^\top\right|, \quad \left(T^{-1}\sum_{t=1}^{T} \tilde{R}_{1t}R_{0t}^\top\right) = S_{10}\left(\mathcal{D}^{-1}\right)^\top.$$

When $A_1 = 0$, that is when $X_{2,t}$ is weakly exogenous with respect to $\tilde{\beta}$, $\mathcal{D}$ no longer involves $\tilde{\beta}$, so minimizing $\tilde{\mathcal{S}}$ coincides with minimizing $\mathcal{S}$, which reduces to the standard VECM approach. When $A_1 \neq 0$, in addition to the adjustment term, the coefficients of lags also depend on $\tilde{\beta}$ via $\mathcal{D}$.[3]

The above analysis concretely quantifies how TRs restrict the cointegration space. Yet reduced rank regression can be applied to (9) ignoring the dependence of the adjustment and lag coefficients on $\tilde{\beta}$. Following the same argument with time varying parameters calls

[3]The dependence of the error on $\tilde{\beta}$ can be ignored, as pointed out in Phillips (1994), Phillips (1991).

for caution. Our analysis suggests that a time varying long-run parameter in a TR yields a VECM where all parameters are time varying including lag coefficients when relevant, even when all other remaining parameters which characterize the transient error dynamics are constant. This will be the case, *unless* weak exogeneity holds across the sample. The fact remains that weak exogeneity is rarely granted with cointegration models in economics. A well specified GRR-reg framework allowing for breaks in all VECM coefficients thus holds promise empirically, even if the economic question is formulated via a TR. We provide supportive simulations in the next section.

To conclude this section, we analyze the consequences of breaks in the TR coefficient on the deterministic components of the VECM. Such terms have so far been left out of our discussion for presentation ease, yet their relevance has long been demonstrated; see *e.g.* Johansen et al. (2000) who introduce piece-wise linear trends in VECMs. Importantly, the GRR-reg approach can accommodate breaks in deterministic terms. We build upon (6)-(7) through the following DGP, affected by one break at date T_1:

$$x_{1t} - \tilde{\beta}(t)x_{2t} = z_t, \quad z_t = \rho z_{t-1} + u_{1,t} \tag{12}$$

$$a_1 x_{1t} - a_2 x_{2t} = w_t, \quad w_t = w_{t-1} + u_{2,t}, \quad \begin{bmatrix} u_{1,t}, & u_{2,t} \end{bmatrix}^\top \overset{iid}{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I}_2), \tag{13}$$

$$\tilde{\beta}(t) = \tilde{\beta}_1 \mathbf{1}_{1t} + \tilde{\beta}_2 \mathbf{1}_{2t} \tag{14}$$

with $\mathbf{1}_{1t} = \mathbf{1}(T_0 + 1 \leq t \leq T_1)$ and $\mathbf{1}_{2t} = \mathbf{1}(T_1 + 1 \leq t \leq T)$. This DGP constitutes our baseline simulation design below, where we set $a_1 = 1$ and $a_2 = -1$. Straightforward algebraic manipulations yield the following VECM representation of (12)-(13):

$$\Delta X_t = \Theta(t) \begin{bmatrix} \tilde{\beta}_2 - \tilde{\beta}_1 \\ 0 \end{bmatrix} x_{2,t-1} \mathrm{d}_t + \alpha(t)\beta(t)^\top X_{t-1} + \Theta(t) \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}, \tag{15}$$

$$\Theta(t) = \frac{1}{-a_2 + a_1 \tilde{\beta}(t)} \begin{bmatrix} -a_2 & \tilde{\beta}(t) \\ -a_1 & 1 \end{bmatrix}, \quad \mathrm{d}_t = \mathbf{\Delta 1}_{2t} = -\mathbf{\Delta 1}_{1t}, \tag{16}$$

$$\alpha(t) = \Theta(t) \begin{bmatrix} \rho - 1 \\ 0 \end{bmatrix}, \quad \beta(t) = \begin{bmatrix} 1 \\ -\tilde{\beta}(t-1) \end{bmatrix}, \tag{17}$$

where d_t is a point dummy variable at time $T_1 + 1$. Here $\Delta X_t = [\Delta x_{1,t}, \Delta x_{2,t}]^\top$ and

$X_{t-1} = [x_{1,t-1}, x_{2,t-1}]^\top$. Following Johansen et al. (2000), an unrestricted dummy can be modelled into the conditional likelihood to account for such an outlier at the connection point. Furthermore, breaks in the variance/covariance matrix of disturbance clearly emerge through (15). While $\alpha(t)$ is no longer time varying when $a_1 = 0$, that is when weak exogeneity holds, the disconnect at the junction point and breaks in the error variances and covariances will remain when $\tilde{\beta}_2 \neq \tilde{\beta}_1$.

Despite the simplicity of the considered illustrative DGP, the key take away is that a break in the TR coefficient can induce breaks in all components of the VECM. Yet sample size restrictions may limit ones' ability to apply a GRR-reg based test allowing for such a possibility. Our simulations below emphasize the importance of accounting for a time varying $\alpha(t)$ unless the DGP satisfies weak exogeneity, and of variance instabilities.

3 Numerical evidence

Hansen (1992), Hansen and Johansen (1999) and Hansen (2003) contained no simulation analysis; the bulk of simulations with systems-based methods (Quintos, 1997; Seo, 1998; Bierens and Martins, 2010), imposed weak exogeneity, and most available designs consider a stable model as the null hypothesis. Martins (2018) and Bergamelli et al. (2019) simulated the VECM directly. Instead, we consider null models with or without breaks in the long run coefficient of a TR where weak identification of the cointegration parameter in question can be parametrized and weak exogeneity can be imposed and relaxed.

Two sets of experiments are considered. The first set implements Hansen (2003)'s procedure exactly as proposed, with pre-specified break dates. The null hypothesis is not restricted to parameter constancy though we formulate a single specific alternative. Next, we design more general experiments with uncertain break dates.

3.1 Baseline design

We consider null models with one break in the long run coefficient of a TR where exogeneity does not hold. Formally, we simulate (12)-(13) with $T_1 = \lfloor T/2 \rfloor$, $\tilde{\beta}_1 = 1$, $\tilde{\beta}_2 = \{1.5, 2, 3, 4, 5, 8, 10, 15, 20, 30, 40, 50\}$, $\rho = \{0, 0.8\}$, $a_1 = 1$, $a_2 = -1$, $\sigma = 1$ and $T = \{100, 300, 500\}$. It is worth noting that ρ controls the identification of the cointegrating relation, which deteriorates as persistence nears the unit boundary.

We study two different likelihood ratio (LR) tests of one versus two breaks, based on the GRR-reg

$$\Delta X_t = \alpha(t)\beta(t)^\top X_{t-1} + \epsilon_t, \tag{18}$$

where

$$\beta(t) = \begin{bmatrix} 1 & -\bar{\beta}(t) \end{bmatrix}^\top, \quad \bar{\beta}(t) = \sum\nolimits_{j=1}^{3} \beta_j \mathbf{1}_{jt}, \tag{19}$$

$\alpha(t)$ is conformably piece-wise constant with adjustment parameters denoted by α_1, α_2 and α_3, and the disturbance $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})^\top$ is a sequence of independent zero mean Gaussian variables with a variance matrix that may also break conformably. In this context, the LR tests are based on the following two sets of hypotheses

$$\text{A:} \begin{cases} \mathcal{H}_0^a: & (\beta_1 \neq \beta_2,\, \beta_2 = \beta_3,\, \alpha_1 = \alpha_2 = \alpha_3) \; |T_1 \\ \mathcal{H}_1^a: & (\beta_1 \neq \beta_2 \neq \beta_3,\, \alpha_1 = \alpha_2 = \alpha_3) \; |T_1, T_2 \end{cases},$$

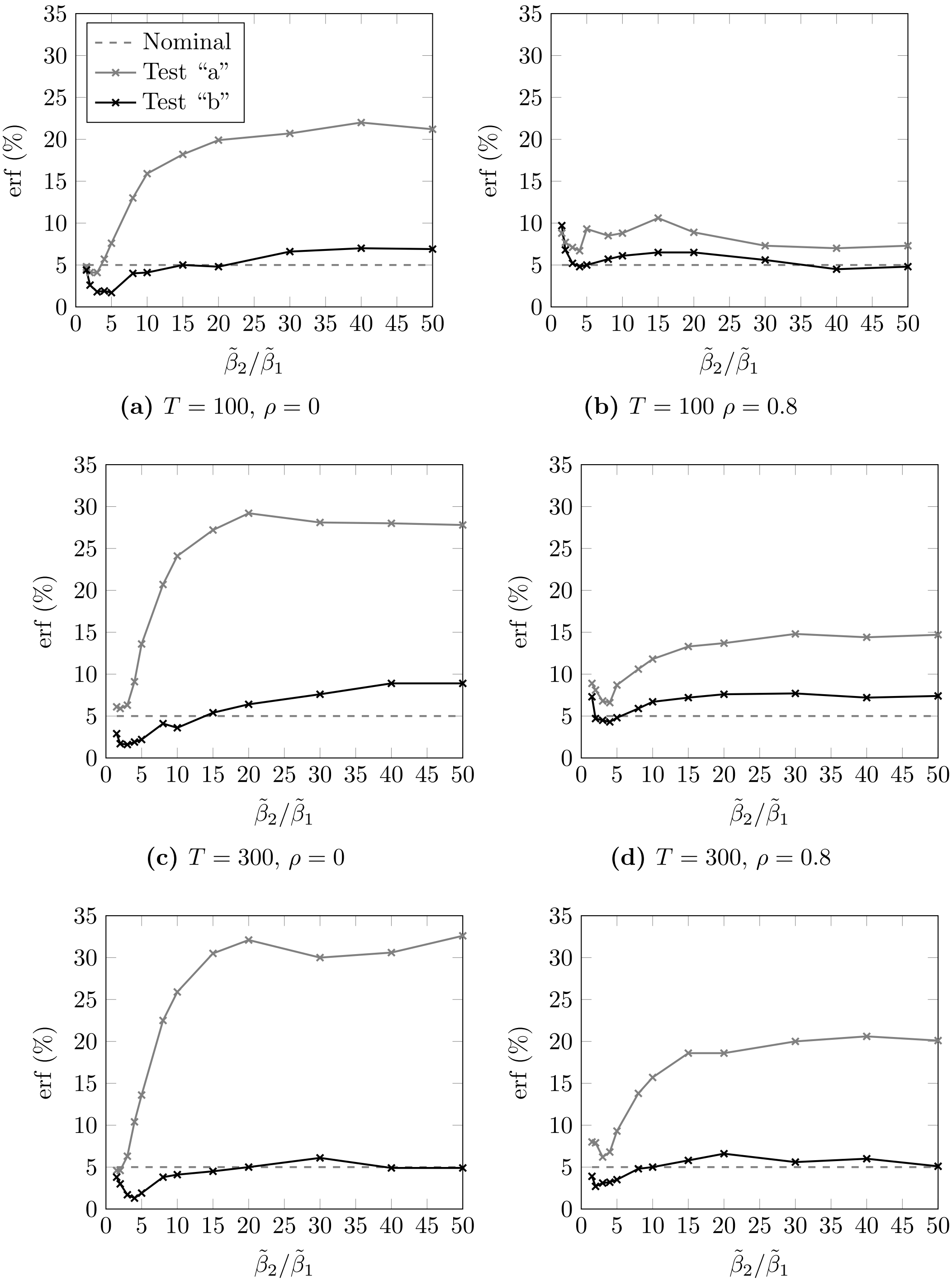
where the error variance is time invariant under the null and the alternative hypotheses, and

$$\text{B:} \begin{cases} \mathcal{H}_0^b: & (\beta_1 \neq \beta_2,\, \beta_2 = \beta_3,\, \alpha_1 \neq \alpha_2,\, \alpha_2 = \alpha_3) \; |T_1 \\ \mathcal{H}_1^b: & (\beta_1 \neq \beta_2 \neq \beta_3,\, \alpha_1 \neq \alpha_2 \neq \alpha_3) \; |T_1, T_2 \end{cases}$$

where we maintain breaks in variance. "|" indicates conditioning on the break date T_i, $i = 1, 2$, and we set $T_1 = T/2$, $T_2 = 2T/3$.

In both cases, (19) ignores the above-discussed breaks at the connection points, so both $\mathcal{H}_0^a$ and $\mathcal{H}_0^b$ do not fully match the considered DGP. Realistically, it is rarely possible to achieve such a match beyond simulation studies. Our design is thus motivated by the following consideration. The restriction $\alpha_1 = \alpha_2 = \alpha_3$ that is embedded in $\mathcal{H}_0^a$ does not conflict with (12)-(13) when $a_1 = 0$ [*i.e.* when weak exogeneity holds], yet forgoes a key source of instability in this DGP when, as in our design, $a_1 = 1$ [*i.e.* when weak exogeneity fails]. In contrast, $\mathcal{H}_0^b$ reflects the linkages between $\alpha(t)$ and $\beta(t)$ when weak exogeneity fails. The breaks in error variance that arise as $\tilde{\beta}_1 \neq \tilde{\beta}_2$ are not directly related to weak exogeneity assumptions, yet it is important to assess the consequences of variance-driven misspecifications. One of the key advantages of the GRR-reg is that it allows for variance instabilities.

Figure 1 reports the empirical rejection frequencies for cases A and B as the magnitude of the break increases, which is measured via the ratio of $\tilde{\beta}_2$ to $\tilde{\beta}_1$ and 1000 replications.



(a) $T = 100$, $\rho = 0$

(b) $T = 100$ $\rho = 0.8$

(c) $T = 300$, $\rho = 0$

(d) $T = 300$, $\rho = 0.8$

(e) $T = 500$ $\rho = 0$

(f) $T = 500$ $\rho = 0.8$

Figure 1: Empirical rejection frequencies of the LR tests. The graph reports plots with on the "x" axis the magnitude of the break, i.e. the ratio between the long-run coefficient post-break ($\tilde{\beta}_2$) and pre-break ($\tilde{\beta}_1$), while on the "y" axis the empirical rejection frequencies of $\mathcal{H}_0$ for Case A and for Case B. Different sample sizes T and values of ρ are explored.

As discussed above, interpreting these rejections as size is strictly incorrect. Yet (15) under $\mathcal{H}_0^b$ can be viewed as close enough to the DGP, assuming that breaks at connection points intervene mildly. If so, then rejections of $\mathcal{H}_0^b$ should be close to 5%.

This is exactly what we find: rejection frequencies for Case B conform closely to the nominal level of 5% for the different sample sizes as well as the different values of ρ, the parameter which controls the identification of the cointegrating relationship. In contrast, the observed rejection frequencies of the LR test based on Case A increase with the magnitude of the break, sizably departing from 5%. These findings, specifically the way rejections differ between Case A and Case B, suggest that the former is grossly and consequentially incompatible with the considered DGP. Interestingly, rejections increase with the sample size and with error serial correlation.

Formally, our experiment assesses power since both $\mathcal{H}_0^a$ and $\mathcal{H}_0^b$ depart from the DGP. The considered LR test does not check for outliers at the connection points. We thus do not expect power in this direction, so the observed rejections with $\mathcal{H}_0^b$ are not disconcerting. Conversely, rejections of $\mathcal{H}_0^a$ underscore the information content of adjustment parameters and variance instabilities.

To conclude, note that even in larger samples, higher values of ρ affect power negatively which - although not surprising - is noteworthy. As ρ approaches unity, identification is compromised in every regime of this design; see Khalaf and Urga (2014) for evidence in this regard.

3.2 Uncertain alternatives

We next design experiments that allow uncertain information on break dates. Concretely, we presume that possible breaks can be broadly characterized so that a certain finite number of GRR-regs can adequately express uncertainty about their number and location. So instead of looking at a single statistic, we are lead to a finite number of LRs. First we describe the way these are used for (here, nested) testing purposes. We next describe the experiments we designed to assess the role of weak exogeneity and identification on the properties of such tests when the DGPs are TRs.

In the cases reported below, we draw from the counterparts of the TR form (12)-(13)

allowing for more breaks. In its most general form, the DGP corresponds to

$$\tilde{\beta}(t) = \sum_{j=1}^{5} \tilde{\beta}_j \mathbf{1}_{jt} \qquad (20)$$

where $\mathbf{1}_{jt}$ are the break indicators associated with various choices of break dates, reported below. So to be clear, we refer to (20) to describe the DGP.

For each DGP, we define a GRR-reg of the (15) form, in which case we formulate a null hypothesis, denoted $\mathcal{M}_{0,\mathrm{i}}$ where $\mathrm{i} = 1$, 2 and 3 refers to the case in question, against a given number n_{i}, of alternative GRR-regs denoted $\mathcal{M}_{j,\mathrm{i}}$, $j = 1, ..., n_{\mathrm{i}}$. $\mathcal{M}_{0,\mathrm{i}}$ may include breaks, and is nested within $\mathcal{M}_{1,\mathrm{i}}, ..., \mathcal{M}_{n,\mathrm{i}}$. Breaks are treated as known, and are specified in each case. As above, $\alpha(t)$ and $\beta(t)$ are piece-wise constant, as $\beta(t)$ takes the (19) where $\bar{\beta}(t)$ conforms to the considered breaks. For each of these VECMs, the breaks in question are associated with: (i) the long-term coefficients only [Table 1] which presumes weak exogeneity, or (ii) with all VECM coefficients including the variance/covariance [Table 2], in which case we view this test as relaxing weak exogeneity. Uncertainty around breaks is incorporated by combining the LR tests that are associated with $\mathcal{M}_{1,\mathrm{i}}, ..., \mathcal{M}_{n,\mathrm{i}}$, which we implement as follows [see Bergamelli et al. (2019)].

n_{i} LR statistics, to assess $\mathcal{M}_{0,\mathrm{i}}$ against each of $\mathcal{M}_{1,\mathrm{i}}, ..., \mathcal{M}_{n,\mathrm{i}}$, are constructed and the p-value for each is calculated using the χ^2 limiting theory in Hansen (2003). The minimum of these p-values is retained as the combined test statistic, which we denote Q^*. We next implement three bootstrap-type procedures to approximate the p-value of Q^*. See Martins (2018) for bootstrap time-varying cointegration tests and references therein, and Bergamelli et al. (2019) for a general parametric theory extending GRR-reg tests beyond known break points. The following algorithm describes the considered methods.

Step 1. Save the estimated parameters under the null hypothesis of m_0 regimes (i.e., model $\mathcal{M}_0$), along with centered residuals, denoted $E = (\hat{\epsilon}_1, \ldots, \hat{\epsilon}_T)^\top$, and corresponding variance/covariance matrices.

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

(a) Re-sample residuals under the null hypothesis leading to the series $\{\epsilon^*_{t,b}\}_{t=1}^{T}$. Three different procedures are considered for this purpose: (1) A wild bootstrap approach: multiply each component of E with a realization from a Rademacher variable η_t (Davidson and Flachaire, 2008) which follows a two-point distribution taking values 1 and -1 with probability 0.5; (2) Separate E into m_0 blocks and draw with replacement within each block, mirroring each regime under the null hypothesis; (3) using the saved variance/covariance matrices from Step 1 corresponding to each regime and a zero-mean Gaussian distribution, draw $\{\epsilon^*_{t,b}\}$ parametrically within each block. The resulting procedures will be denoted the *wild bootstrap (W), semi-parametric [B-SP]* and *parametric [B-P] block-regime bootstrap*, respectively.

(b) Build recursively the bootstrap counterpart of the observed data, using the estimated coefficients in Step 1, and the residuals from Step 2a.

(c) Use the bootstrap data from Step 2b to estimate the null and each of the alternative models, as was done with the observed data. Compute the associated likelihood ratios and their respective p-values, again as was done with the observed data, retaining the minimum of these p-values as the bootstrap test statistic denoted Q^*_b.

STEP 3. Compute the empirical p-value $\hat{p}(Q^*) = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}(Q^*_b \geq Q^*)$.

Our simulation study relies on three cases, as follows.

Test $\mathcal{T}_{0,1}$: $\mathcal{M}_{0,1}$ corresponds to the stable GRR-reg, and one alternative GRR-reg is considered that specifies one break at location $T/2$ or $T/2-20$ or $T/2+20$. Data for size are simulated under the stable TR, that is (20) with $\tilde{\beta}_1 = \tilde{\beta}_2 = \tilde{\beta}_3 = \tilde{\beta}_4 = \tilde{\beta}_5 = 1$. Data for power are simulated given (19) with $\tilde{\beta}_1 = 1$, $\tilde{\beta}_2 = \tilde{\beta}_3 = \tilde{\beta}_4 = \tilde{\beta}_5 = \tilde{\beta}_1 + h$, that is with one a break, and that break occurs at $T_1 = T/2$.

Test $\mathcal{T}_{0,2}$: $\mathcal{M}_{0,2}$ is the GRR-reg affected by one break at location T_1, and the alternative GRR-reg specifies two breaks at locations (T_1, T_1+20), (T_1, T_1-30) or (T_1, T_1-25, T_1+20). Data for size are simulated under the TR with one break, specifically (20) with break date T_1 and $\tilde{\beta}_1 = 1$, $\tilde{\beta}_2 = \tilde{\beta}_3 = \tilde{\beta}_4 = \tilde{\beta}_5 = 2$. Data for power are simulated from the TR with two breaks, specifically at locations $T/2$ and $2T/3$ where, in the context of (20), we set $\tilde{\beta}_1 = 1, \tilde{\beta}_2 = 2$, $\tilde{\beta}_3 = \tilde{\beta}_4 = \tilde{\beta}_5 = 1.5 + h$, and the step h is gradually increased.

Test $\mathcal{T}_{0,3}$: $\mathcal{M}_{0,3}$ is the GRR-reg with breaks at locations T_1 and T_2, and the alternative GRR-regs specify three breaks at locations $(T_1, T_2, T_1 + 20)$, $(T_1, T_2, T_2 - 5, T_1 + 15)$ or $(T_1, T_2, T_1 + 10, T_1 - 10, T_2 - 5)$. Data for size are simulated under the TR (20) with break dates (T_1, T_2), and $\tilde{\beta}_1 = 1$, $\tilde{\beta}_2 = 2$, $\tilde{\beta}_3 = \tilde{\beta}_4 = \tilde{\beta}_5 = 1.5$. Data for power are simulated from the TR (20), with breaks at locations $T/2$, $2T/3$, and $5T/6$, $\tilde{\beta}_1 = 1, \tilde{\beta}_2 = 2, \tilde{\beta}_3 = 1.5$, $\tilde{\beta}_4 = \tilde{\beta}_5 = 1.5 + h$, and the step h is gradually increased.

In all designs, *i.i.d.* standard normal disturbances are used as above, and ρ, a_1 and a_2 are kept constant although various choices are considered reflecting different assumptions about weak identification and exogeneity. We use 1000 replications and $B = 199$. As above, the size interpretation abstracts from breaks in deterministic components.

Results can be summarized as follows. When weak exogeneity is imposed, empirical rejection frequencies are generally close to the nominal level, and the B-P bootstrap seems to perform best. The wild bootstrap seems to suffer as more breaks are included in the model. Basically, we observe oversize in Table 1a just when weak identification is provoked with $\rho = .99$.

Turning to Table 1b, results illustrate good power. Rejections increase, as expected, as h increases away from the null values, and as the sample size increases. The power of tests is roughly equal to their size when weak identification is provoked; power is not expected in this case and one may suspect spurious rejections otherwise.

Size results from Table 2a suggest that accounting for breaks in all coefficients is promising, despite mild oversize with smaller samples with reference to Table 1a. This confirms our findings with our baseline design. It is important to emphasize that ρ was restricted to zero in Table 2a so we can zoom-in on weak exogeneity failures in best case scenarios, since the adverse effects of weak identification are clearly set forth in Table 1a, again in the most favorable scenario. As far as power is concerned, we do not find major differences between Tables 1b and 2b.

Broadly, results confirm the superiority of the B-P method. Note that our notation here differs from the usual block-bootstrap since the null hypothesis clearly defines regimes here. In fact, the method we denote as B-P is a parametric bootstrap, applied to the specific piece-wise constant structure of the null model underlying the applied test. Bergamelli et al. (2019) prove the validity of this bootstrap in GRR-reg. In the examples

Table 1: Empirical Rejection Frequencies with Weak Exogeneity Imposed

(a) Size

T	$a_1 = 0,\ \rho = 0$ w	b-p	b-sp	$a_1 = 0,\ \rho = 0.8$ w	b-p	b-sp	$a_1 = 0,\ \rho = 0.99$ w	b-p	b-sp
				$\mathcal{T}_{0,1}$					
100	5.8	5.4	5.4	6.6	6.4	5.8	8.5	9.0	8.0
200	5.1	4.9	4.8	5.9	4.9	5.0	7.5	7.7	7.2
300	5.5	5.7	5.4	5.4	5.3	5.4	8.7	8.2	7.6
400	6.5	5.8	6.3	6.6	6.1	6.6	9.0	8.4	8.5
500	5.4	5.3	5.0	5.4	5.5	5.1	8.4	7.1	6.4
				$\mathcal{T}_{0,2}$					
				$T_1 = T/2$					
100	4.5	4.2	4.1	6.4	4.7	4.5	6.3	7.6	7.6
200	4.9	5.9	5.5	6.3	5.4	5.3	7.9	8.1	8.0
300	4.6	4.3	5.2	5.6	5.7	5.2	7.2	7.1	6.1
400	6.4	6.6	6.7	6.7	6.5	6.2	6.7	7.0	7.3
500	4.4	4.5	5.0	5.0	4.1	4.7	6.4	6.7	6.9
				$T_1 = 20$					
100	3.2	4.2	4.3	4.8	5.0	4.7	7.3	8.6	8.1
200	5.6	6.2	6.4	7.1	6.7	7.1	6.6	6.1	5.1
300	4.5	4.5	4.7	4.9	5.9	5.6	6.3	5.0	4.8
400	5.7	4.7	5.1	5.0	4.9	4.5	7.5	6.3	5.8
500	4.8	4.6	3.8	4.8	4.1	4.3	6.6	5.2	5.1
				$T_1 = T - 20$					
100	4.6	5.4	5.0	5.1	5.2	5.2	6.9	8.6	8.6
200	4.0	5.5	5.3	5.3	5.8	5.8	4.9	6.3	6.9
300	4.5	5.5	5.5	5.0	4.9	5.1	5.8	6.3	6.1
400	5.1	5.0	5.0	5.3	5.0	5.2	4.5	6.7	6.1
500	4.1	5.1	5.2	4.5	5.2	4.9	4.0	4.8	5.2
				$\mathcal{T}_{0,1}$					
				$T_1 = T/2$ $T_2 = 2T/3$					
100	4.0	4.4	4.3	4.0	4.1	4.5	8.2	8.7	8.3
200	5.3	4.6	5.0	6.6	4.8	4.7	6.8	6.9	7.4
300	5.8	4.5	4.5	8.0	5.6	5.8	7.1	7.4	6.9
400	6.7	5.6	5.4	7.4	5.3	5.2	7.5	8.5	7.6
500	6.0	5.3	4.9	5.8	4.2	4.6	6.4	6.3	6.5
				$T_1 = 20$ $T_2 = 2T/3$					
100	5.3	5.8	5.6	6.4	5.6	5.7	10.1	9.5	9.6
200	6.0	6.7	5.8	7.6	6.5	6.3	7.3	7.4	7.2
300	5.1	5.7	5.4	5.4	5.3	5.3	8.3	8.0	7.3
400	5.2	5.7	5.1	6.7	5.6	5.8	6.5	7.5	6.7
500	3.6	4.3	4.8	5.0	4.5	4.4	6.5	5.9	6.0
				$T_1 = T - 20$ $T_2 = T/2$					
100	5.5	5.0	5.4	5.8	4.5	5.5	9.6	9.3	9.5
200	4.7	4.8	4.7	6.2	5.7	5.0	6.5	7.6	8.0
300	5.4	5.0	4.6	7.6	5.8	6.7	8.0	7.3	7.6
400	6.1	6.3	6.1	6.7	5.9	5.6	5.5	5.9	5.3
500	5.0	4.7	4.2	6.0	4.3	4.5	4.6	5.2	4.7

(b) Power

	$a_1 = 0,\ \rho = 0$ w	b-p	b-sp	$a_1 = 0,\ \rho = 0.8$ w	b-p	b-sp	$a_1 = 0,\ \rho = 0.99$ w	b-p	b-sp
				$\mathcal{T}_{0,1}$					
β_2				$T = 100$					
1.1	56.0	57.3	56.2	10.3	9.9	8.7	10.5	8.9	8.4
1.3	94.6	94.6	94.4	25.2	23.7	24.2	10.3	8.0	8.4
1.5	99.0	99.1	99.0	40.7	38.4	38.3	10.8	8.3	9.2
2.0	99.1	99.6	99.7	56.1	56.0	55.8	14.1	11.0	11.4
β_2				$T = 300$					
1.1	95.7	95.8	95.7	25.9	25.2	24.1	8.7	8.2	7.5
1.3	100.0	100.0	100.0	73.2	72.4	72.3	9.3	7.4	7.3
1.5	100.0	100.0	100.0	89.0	88.7	88.8	10.7	8.8	8.9
2.0	100.0	100.0	100.0	96.0	96.0	95.9	13.6	11.6	11.2
β_2				$T = 500$					
1.1	100.0	99.9	99.8	47.3	46.1	47.1	9.6	6.7	5.9
1.3	100.0	100.0	100.0	91.6	91.2	91.0	11.2	8.4	8.0
1.5	100.0	100.0	100.0	98.9	98.8	98.8	11.1	9.3	8.6
2.0	99.9	99.9	99.9	99.5	99.6	99.8	17.1	14.3	13.1
				$\mathcal{T}_{0,2}$					
β_3				$T = 100$					
2.1	46.0	43.2	41.9	7.8	6.7	6.7	7.5	8.7	7.9
2.3	83.7	81.1	79.8	18.0	15.4	15.6	8.2	8.8	8.5
2.5	93.6	91.7	91.6	30.2	25.4	26.7	8.0	8.4	9.4
3.0	98.5	97.9	97.9	49.1	44.2	45.1	10.0	10.5	9.9
β_3				$T = 300$					
2.1	86.6	86.3	85.1	22.0	20.8	21.2	8.8	7.5	7.6
2.3	99.8	99.8	99.6	62.3	60.7	59.8	8.6	7.4	7.6
2.5	99.9	99.9	99.9	76.5	75.8	76.6	9.6	7.5	7.5
3.0	99.5	99.5	99.5	87.1	86.6	86.0	12.8	9.3	9.6
β_3				$T = 500$					
2.1	95.2	95.6	95.0	40.0	38.2	37.6	8.0	6.9	7.6
2.3	100.0	100.0	100.0	80.7	80.4	80.3	9.7	7.0	7.8
2.5	99.9	99.9	99.9	91.9	91.4	91.1	10.3	7.7	8.7
3.0	99.9	99.9	99.9	97.8	97.9	97.3	13.1	11.1	11.1
				$\mathcal{T}_{0,3}$					
β_4				$T = 100$					
1.6	45.9	46.1	45.9	7.1	6.2	6.7	7.2	9.0	8.3
1.8	80.1	80.4	79.9	24.9	24.2	24.5	7.9	8.8	8.5
2.0	91.0	90.3	90.1	42.6	40.2	39.8	8.8	8.8	8.6
2.5	98.4	98.4	98.3	61.1	59.5	58.8	10.5	9.1	9.7
β_4				$T = 300$					
1.6	78.9	80.1	78.9	21.5	21.3	20.4	5.6	7.0	7.6
1.8	98.1	98.5	98.2	61.0	61.2	60.6	6.4	7.6	8.1
2.0	99.8	99.7	99.7	75.2	74.5	74.3	7.4	7.8	8.6
2.5	100.0	100.0	100.0	87.6	87.5	86.9	10.7	9.9	9.9
β_4				$T = 500$					
1.6	91.3	91.5	91.0	40.0	40.9	40.6	4.8	5.7	6.1
1.8	99.9	99.9	99.9	77.5	78.1	77.0	5.4	6.6	6.4
2.0	99.9	99.9	99.9	88.8	88.6	88.1	6.8	8.0	7.8
2.5	100.0	100.0	100.0	96.3	96.2	96.3	16.2	15.2	14.9

Note: The table reports nominal level used to compute the empirical rejection frequencies is 5%. "w" denotes wild bootstrap, "b-p" block-regime bootstrap with parametric resampling and "b-sp" block-regime bootstrap with semi-parametric resampling.

Table 2: Empirical Rejection Frequencies with Weak Exogeneity Relaxed

(a) Size

	$a_1 = 0,\ \rho = 0$			$a_1 = 0,\ \rho = 0$			$a_1 = 0,\ \rho = 0$		
T	*w*	*b-p*	*b-sp*	*w*	*b-p*	*b-sp*	*w*	*b-p*	*b-sp*
	$\mathcal{T}_{0,1}$			$\mathcal{T}_{0,2}$			$\mathcal{T}_{0,3}$		
				$T_1 = T/2$			$T_1 = T/2$ $T_2 = 2T/3$		
100	5.8	5.0	6.2	6.9	4.8	5.4	9.0	6.6	4.8
200	4.4	4.4	4.1	7.8	6.8	6.1	6.7	5.2	3.8
300	5.7	5.6	4.9	6.1	5.0	4.7	7.8	5.8	4.0
400	5.8	5.3	5.9	7.3	5.3	5.9	9.7	7.0	4.8
500	5.2	5.5	5.2	5.7	4.2	4.3	7.9	5.5	3.2
				$T_1 = 20$			$T_1 = 20$ $T_2 = 2T/3$		
100				2.8	3.7	3.3	2.7	4.0	3.2
200				4.4	4.5	4.9	5.5	5.3	4.8
300				3.3	3.3	3.4	4.2	4.9	4.0
400				5.3	4.6	4.8	4.3	3.9	3.3
500				3.9	3.3	2.9	4.7	4.8	3.8
				$T_1 = T - 20$			$T_1 = T - 20$ $T_2 = T/2$		
100				8.9	6.6	5.7	9.5	5.8	4.8
200				6.3	5.5	5.0	9.5	5.6	5.0
300				6.5	5.9	6.3	7.0	4.1	2.7
400				6.1	5.2	5.3	9.9	5.7	4.7
500				5.8	5.9	5.5	9.6	6.1	5.3

(b) Power

	$a_1 = 0,\ \rho = 0$				$a_1 = 0,\ \rho = 0$				$a_1 = 0,\ \rho = 0$		
	w	*b-p*	*b-sp*		*w*	*b-p*	*b-sp*		*w*	*b-p*	*b-sp*
	$\mathcal{T}_{0,1}$				$\mathcal{T}_{0,2}$				$\mathcal{T}_{0,3}$		
β_2	$T = 100$			β_3	$T = 100$			β_4	$T = 100$		
1.1	20.8	20.3	20.2	2.1	36.1	33.5	32.4	1.6	10.2	8.8	7.3
1.3	61.0	61.4	60.5	2.3	73.8	68.9	69.5	1.8	35.6	32.4	30.5
1.5	80.9	80.8	81.1	2.5	86.2	82.6	81.6	2.0	50.4	47.9	44.5
2.0	95.3	95.4	95.3	3.0	93.9	92.6	91.3	2.5	66.4	63.4	61.7
β_2	$T = 300$			β_3	$T = 300$			β_4	$T = 300$		
1.1	62.5	62.8	62.7	2.1	78.2	77.5	77.2	1.6	36.7	35.0	31.9
1.3	98.2	98.1	98.1	2.3	98.5	98.6	98.6	1.8	74.3	73.3	71.2
1.5	99.8	99.9	99.7	2.5	98.9	99.8	99.6	2.0	84.4	83.2	80.8
2.0	100.0	100.0	100.0	3.0	99.2	99.9	99.9	2.5	90.9	86.9	85.7
β_2	$T = 500$			β_3	$T = 500$			β_4	$T = 500$		
1.1	85.7	85.0	84.8	2.1	91.9	91.7	91.6	1.6	58.8	57.4	53.6
1.3	100.0	100.0	100.0	2.3	99.7	99.7	99.7	1.8	88.6	87.5	85.8
1.5	100.0	100.0	100.0	2.5	97.1	99.7	99.7	2.0	91.8	90.0	89.1
2.0	100.0	100.0	100.0	3.0	99.8	100.0	100.0	2.5	94.3	92.7	91.4

Note: The table reports nominal level used to compute the empirical rejection frequencies is 5%. "w" denotes wild bootstrap, "b-p" block-regime bootstrap with parametric resampling and "b-sp" block-regime bootstrap with semi-parametric resampling..

considered here, the true null model is not a GRR-reg but a TR. However, GRR-reg disregards the TR's identification restrictions which practically implies that the tested null hypothesis is a VECM as analyzed by Bergamelli et al. (2019). This preserves the equi-continuity conditions proved by Bergamelli et al. (2019) and consistency of estimators [see also Hansen (2003, proof of Theorem 8)], which establishes the validity of the B-P method as applied here.

3.3 Discussion

The above simulations have novel implications for stability testing in the cointegration framework. Our main finding is positive suggesting that GRR-reg tests can be applied with success, without knowledge about the precise nature of the DGP. The idea is to give up weak exogeneity whether it holds or not in the unknown DGP, and to account for variance instabilities. When weak exogeneity is relaxed, the possibility that long run and adjustment terms can jointly break needs to be allowed for. This avoids distortions resulting from weak exogeneity failures, and delivers good power even when weak exogeneity holds. Such a framework provides an interesting integration of several fundamental concepts at the heart of Sir David Hendry's contributions.

Our results can be positioned relative to Martins (2018), Khalaf and Urga (2014) and Bergamelli et al. (2019). Martins (2018) recommends (with proof) the wild bootstrap for time varying cointegration, except that the piece-wise constant specification of Hansen (2003) departs from the smoothly evolving assumptions in this case. In contrast, our B-P scheme that mirrors regimes under the null hypothesis recovers validity as shown in Bergamelli et al. (2019). Finally, our B-P scheme with stable samples differs from the Monte Carlo method in Khalaf and Urga (2014) who impose the null explicitly as a TR. This may explain the outstanding oversize we still observe here, relative to Khalaf and Urga (2014). Nevertheless, distortions barely exceed 10% in Table 1, which is noteworthy. Our findings are unique through our analysis of: (i) the cost of ignoring the TR explicit restrictions, (ii) known and uncertain breaks, (iii) various bootstraps, and (iv) the interplay between weak-exogeneity, breaks and identification.

Results with weak identification are still telling. As argued by Khalaf and Urga (2014) and references therein, inference is adversely affected in this case even in stable and standard cointegration models. Identification-robust break tests have attracted recent interest in simultaneous equations yet remain scarce; see e.g. Mavroeidis and Magnusson (2014) and the few references therein. Overall, despite a vast literature in econometrics, identification-robust work on cointegration is scarce. Extending Hansen (2003)'s test in this direction is a worthy research objective.

4 Empirical analysis

We study an interest rate equation for the US with monthly data over the period September 2007 to August 2024. The GRR-reg of interest pertains to the vector

$$X_t = (i_t, p_t, y_t, vix_t)^\top \tag{21}$$

where i_t is the 1-month American dollar LIBOR interest rate, p_t is the US Personal Consumption Expenditure Core (year over year inflation minus the FED's 2% target), y_t is the US year over year industrial production, and the vix_t (Volatility Index) as a measure of stock-market volatility (See Donati et al, 2025). All data is available from Bloomberg. Figure 2 reports the graph of the series. In this context, there is no reason to presume weak exogeneity (see amongst others Clarida et al, 2000 and Carvalho et al, 2021).

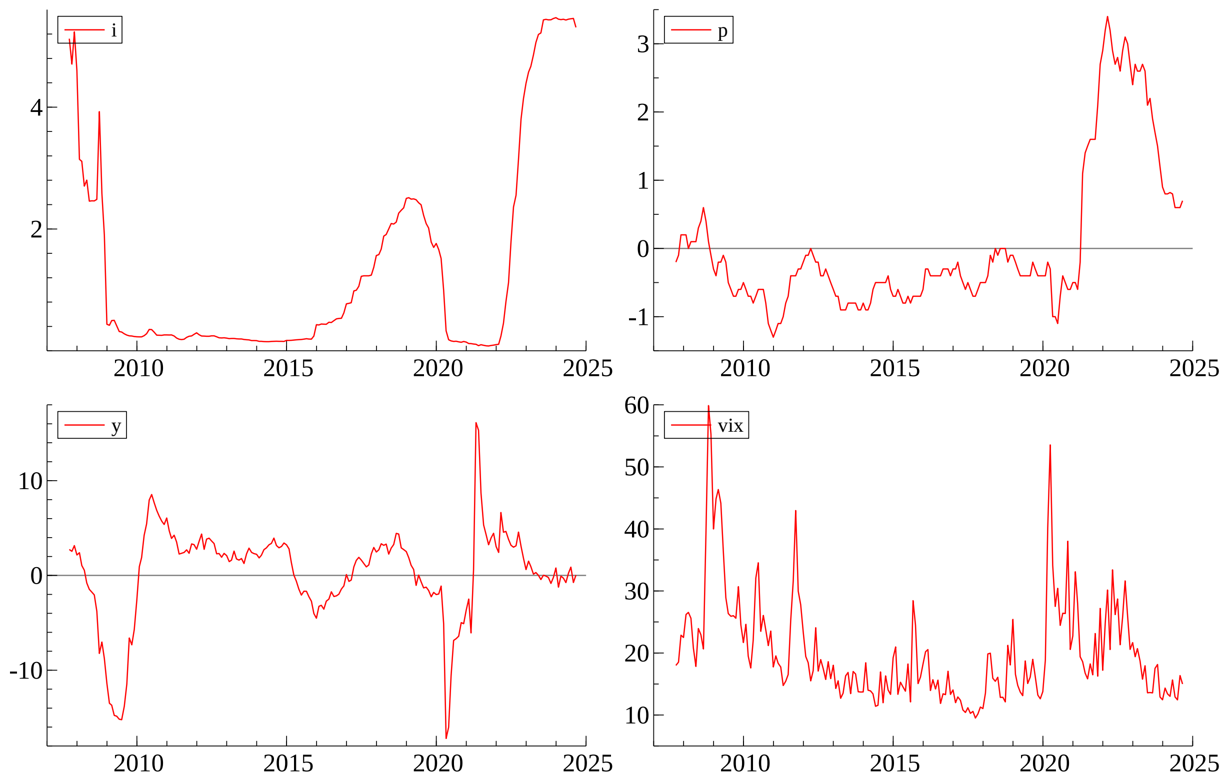


Figure 2: Plot of the series. Source: Bloomberg

From a Taylor (1993, 1999) rule perspective, the focus is on representing how the Federal Reserve adjusts short term interest rates in response to changes in the considered measures of economic conditions. Consequently, all the tests that we conduct adopt a matching normalization. Recall that such a normalization avoids the assumption that is typically required in a TR, namely that inflation, output and stock market volatility do not cointegrate.[4]

We estimate a GRR-reg with break dates corresponding to key historical episodes: (i) the European sovereign-debt crisis of 2011, (ii) the Donald Trump election bump in 2016, (iii) the Covid pandemic in 2020, and (iv) the recent interest rate hikes that started in 2022. The corresponding dates are June 2011, December 2016, March 2020, and March 2022. This framework entails that a piece-wise constant cointegrating relation describes the Federal Reserve reaction function.

In this context, we study the relevance of the European crisis on this function, accounting for the remaining breaks which reflect broadly accepted stylized facts. In contrast to the latter, there seems to be no consensus in the related literature on the relevance of the

[4]As emphasized in Hansen (2003), the normalization affects the degrees-of-freedom in the limiting distribution of the LR statistic (in a known way, which is not problematic).

European crisis to the US economy. The monetary policy report of July 2011[5] discusses some transmission effects on economic growth in the US, and some findings are available documenting some effects on banks and lending. As an example, Allegret et al. (2017) find that US banks have been sheltered from this crisis. In their analysis of sovereign credit default swap spreads, Broto and Pérez-Quirós (2015) treat the US as a "safe country". In contrast, De Marco (2019) reports a reduction in short-term funding of banks. See also Ang and Longstaff (2013) for a comparative analysis of the source of sovereign credit risk in Europe and the US. These studies are not meant to provide an exhaustive perspective on this important contagion question, since the related literature is too broad to be usefully reviewed here. Our aim is to document the existing mixed evidence, which contextualizes our tested hypothesis. For a general perspective on the European crisis, see Lane (2012).

We report the results from four LR tests with asymptotic cut-off points. Under the null hypothesis, the breaks occur in December 2016, March 2020, and March 2022, while under the alternative hypothesis, the breaks occur in June 2011, December 2016, March 2020, and March 2022.

The first test focuses on breaks in the long run parameters only. This test is inspired by the hypothesis that is most commonly assessed within a TR, since the transient dynamics and error variances interfere through the TR residuals. The LR statistic in this case is 4.327 with a p-value of 0.228.

The second test focuses on breaks in the long term and adjustment coefficients, maintaining the stability of the error variance covariance matrix. Abstracting from breaks in variance is also inspired by several existing break tests (in cointegration and other contexts). The LR statistic in this case is 6.428 with a p-value of 0.491.

The third test considers breaks in the long term parameter and in the error variance covariance matrix, maintaining the adjustment coefficients stable. This test may be justified through some weak exogeneity assumption. While there is no formal basis for such an assumption here, this check is motivated by our general objective in this paper. The LR statistic in this case is 85.039 with a p-value of 0.000.

The fourth test broadens our analysis to the case where short and long parameters break, allowing for breaks in error variances and covariances. This test conforms to the

[5] Refer to https://www.federalreserve.gov/monetarypolicy/mpr_20110713_part1.htm.

most general and preferred specification (abstracting from connection points) as discussed in the previous sections. The LR statistic in this case is 107.500, with a p-value of 0.000.

Methodologically, our simulations underscore the information content of: (i) breaks in the adjustment coefficient when there is no reason to presume weak exogeneity, and (ii) variance instabilities. This is clearly reflected in our findings that would have dismissed the effect of the European sovereign debt crisis had we restricted our investigation to the long run cointegration components. Empirically, our results suggest a flexible GRR-reg with breaks in all parameters at these critical historical episodes, which is a valuable addition to the literature.

Interestingly, in addition to adjustment terms, we find that breaks in variance are clear and evident. This reflects the advantages of the GRR-reg approach. From a general modeling perspective, allowing for variance instabilities may enhance the performance of cointegration models in empirical applications.

5 Conclusion

This paper contributes to the literature on time varying cointegration, normalization, identification and exogeneity. Building on the VECM-based GRR-reg framework of Hansen (2003), we study the implications of a DGP in triangular form that entails classifying observables in two groups so that a specific subset does not cointegrate. Results confirm that weak exogeneity throughout the sample alleviates specification problems. Yet it is often unlikely one can take an *ex-ante* stand regarding weak exogeneity in economics. Preferably, we show that time varying cointegration can be statistically analyzed from the VECM reduced form, notwithstanding the DGP. TRs can be linked to parametric VECMs by restricting the cointegration space, in the same way identifying restrictions intervene in simultaneous equations. We find that (unrestricted) GRR-reg tests are remarkably robust to these and other deviations from standard assumptions, provided the tested null hypotheses are carefully formulated. Like simultaneous equations, TRs provide natural representations in economics. However, their identification has methodological implications that may be avoided by an invariant treatment as with GRR-reg stability tests.

References

Allegret, J.-P., H. Raymond, and H. Rharrabti (2017). The impact of the european sovereign debt crisis on banks stocks. some evidence of shift contagion in Europe. *Journal of Banking & Finance 74*, 24–37.

Ang, A. and F. A. Longstaff (2013). Systemic sovereign credit risk: Lessons from the US and Europe. *Journal of Monetary Economics 60*(5), 493–510.

Banerjee, A. and J. L. Carrion-i Silvestre (2025). Panel data cointegration testing with structural instabilities. *Journal of Business & Economic Statistics 43*, 122–133.

Bergamelli, M., A. Bianchi, L. Khalaf, and G. Urga (2019). Combining p-values to test for multiple structural breaks in cointegrated regressions. *Journal of Econometrics 211*, 461–482.

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

Broto, C. and G. Pérez-Quirós (2015). Disentangling contagion among sovereign CDS spreads during the European debt crisis. *Journal of Empirical Finance 32*, 165–179.

Carvalho, C., F. Nechio, and T. Tristão (2021). Taylor rule estimation by ols. *Journal of Monetary Economics 124*, 140–154.

Clarida, R., J. Gali, and M. Gertler (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *The Quarterly journal of economics 115*(1), 147–180.

Davidson, R. and E. Flachaire (2008). Wild bootstrap, tamed at last. *Journal of Econometrics 146*, 162–169.

De Marco, F. (2019). Bank lending and the European sovereign debt crisis. *Journal of Financial and Quantitative Analysis 54*(1), 155–182.

Donati, S., G. Urga, and A. Varaldo. Augmented Taylor rule estimation. *Centre for Econometric Analysis, WP Series*.

Eroğlu, B. A., J. I. Miller, and T. Yiğit (2022). Time-varying cointegration and the kalman filter. *Econometric Reviews 41*(1), 1–21.

Gomez-Biscarri, J. and J. Hualde (2015). Regression-based analysis of cointegration systems. *Journal of Econometrics 186*, 32–50.

Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics 60*, 203–233.

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 S. Johansen (1999). Some tests for parameter constancy in cointegrated VAR-models. *Econometrics Journal 2*, 306–333.

Hansen, P. R. (2003). Structural changes in the cointegrated vector autoregressive model. *Journal of Econometrics 114*, 261–295.

Harvey, A. C. and G. D. A. Phillips (1980). Testing for serial correlation in simultaneous equation models. *Econometrica*, 747–759.

Harvey, A. C. and G. D. A. Phillips (1981a). Testing for heteroscedasticity in simultaneous equation models. *Journal of Econometrics 15*, 311–340.

Harvey, A. C. and G. D. A. Phillips (1981b). Testing for serial correlation in simultaneous equation models: Some further results. *Journal of Econometrics 17*, 99–105.

Harvey, A. C. and G. D. A. Phillips (1989). Testing for structural change in simultaneous equation models. In P. Hackl (Ed.), *Statistical Analysis and Forecasting of Economic Structural Change*, pp. 25–36. Berlin: Springer-Verlag.

Harville, D. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer-Verlag.

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

Johansen, S. (2009). Cointegration: Overview and development. In T. Andersen, R. Davis, J. KreiB, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 671–694. Springer-Verlag.

Johansen, S., R. Mosconi, and B. Nielsen (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal 3*, 216–249.

Kapetanios, G., S. Millard, K. Petrova, and S. Price (2020). Time-varying cointegration with an application to the uk great ratios. *Economics Letters 193*, 109213.

Kejriwal, M. and P. Perron (2008). The limit distribution of the estimates in cointegrated regression models with multiple structural changes. *Journal of Econometrics 146*, 59–73.

Kejriwal, M. and P. Perron (2010). Testing for multiple structural changes in cointegrated regression models. *Journal of Business & Economic Statistics 28*, 503–522.

Khalaf, L. and G. Urga (2014). Identification robust inference in cointegrating regressions. *Journal of Econometrics 182*, 385–396.

Lane, P. R. (2012). The European sovereign debt crisis. *Journal of economic perspectives 26*(3), 49–68.

Martins, L. (2018). Bootstrap tests for time varying cointegration. *Econometric Reviews 37*, 466–483.

Mavroeidis, S. and L. Magnusson (2014). Identification using stability restrictions. *Econometrica 82*, 1799–1851.

Phillips, P. (1994). Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models. *Econometrica 62*, 73–93.

Phillips, P., D. Li, and J. Gao (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics 196*, 180–195.

Phillips, P. and M. Loretan (1991). Estimating long-run economic equilibria. *The Review of Economic Studies 58*, 407–436.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica 59*, 283–306.

Quintos, C. E. (1997). Stability tests in error correction models. *Journal of Econometrics 82*, 289–315.

Seo, B. (1998). Tests for structural change in cointegrated systems. *Econometric Theory 14*, 222–259.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy 39*, 195–214.

Taylor, J. B. (1999). An historical analysis of monetary policy rules. *Monetary Policy Rules/University of Chicago Press*.

Watson, M. (1994). Vector autoregressions and cointegration. In R. F. Engle and D. L. MacFadden (Eds.), *Handbook of Econometrics Vol. IV*, Chapter 47, pp. 2844–2915. Elsevier Science.