Structural Breaks and Outliers Detection in Time-series Econometrics: Methods and Applications

by

Michele Bergamelli

Thesis submitted to Cass Business School in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Faculty of Finance
Cass Business School
City University London

March 2015

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I wish to thank all the people that in different ways have supported me throughout this long and exciting journey. A special mention goes to:

My wonderful wife Michela for constant moral support and understanding during stressful times.

My parents, Maria and Pietro, for allowing me to follow the path I wished.

My PhD supervisor, Prof. Giovanni Urga, for his thorough guidance, profound support and wise advice.

Prof. Lynda Khalaf for instilling in me her passion for Econometrics and for the extreme kindness in passing on her knowledge.

All my PhD fellows with whom I shared the office and had lengthy discussions over the past few years and in particular, Riccardo Pianeti, Andrew Hunt, Cheng Yan, Jan Novotny and Riccardo Borghi.

The Cass faculty and in particular Lorenzo Trapani, Ian Marsh, Richard Payne, and Sonia Falconieri.

The PhD officers – Abdul Momin and Malla Pratt – who contributed considerably to facilitate travelling and attending conferences.

The examiners – Prof. George Kapetanios and Prof. Lucio Sarno – who provided capital feedbacks in light of future publication in peer-reviewed journals of the chapters of this thesis.

Declaration of Authorship

I, Michele Bergamelli, declare that this thesis entitled *"Structural Breaks and Outliers Detection in Time-series Econometrics: Methods and Applications"* together with the work presented in it are my own and that any material contained in this thesis has not been submitted for a degree to any other university. I further declare that whenever the results of this thesis are used as a source for co-authored research papers this is clearly stated in the document. Finally, powers of discretion are hereby granted to the University Librarian to allow this thesis to be copied in whole or in part without further reference to the author. This permission covers only simple copies made for study purpose, subject to the normal conditions of acknowledgement.

Michele Bergamelli
London, March 2015

ABSTRACT

This thesis contributes to the econometric literature on structural breaks analysis and outliers detection in parametric linear models. The focus is on the development of new econometric tools as well as on the analysis of novel but largely unexplored approaches. The econometric methods under analysis are illustrated using macroeconomic and financial relationships. The thesis is organised in three main chapters. In Chapter 2, we consider two novel methods to detect multiple structural breaks affecting the deterministic component of a linear system. The first is an extension of the dummy saturation method whereas the second method deals with a sequential bootstrapping procedure based on the sup-F statistic. Through an extensive Monte Carlo exercise, we explore the ability of the two approaches to detect the correct number and the correct location of the breaks. Additionally, we illustrate how to apply empirically the two procedures by investigating the stability of the Fisher relationship in the United States. In Chapter 3, we consider testing for multiple structural breaks in the vector error correction framework. First, we study the role of weak exogeneity when testing for structural breaks in the cointegrating matrix. Second, we extend the existing likelihood ratio test of Hansen (2003) to the case of unknown break dates through the specification of a minimum p-value statistic with critical values approximated by bootstrapping. Monte Carlo simulations show that the proposed statistic has good finite sample properties whilst three small empirical applications illustrate how the minimum p-value statistic can be used in practice. In Chapter 4, we tackle the purchasing power parity puzzle developing a robust estimator for the half-life of the real exchange rate. Specifically, we propose to identify outlying observations by means of a dummy saturation type algorithm designed for ARMA processes which enables to detect additional and innovative outliers as well as level shifts. An empirical application involving US dollar real exchange rates shows that the estimated half-lives are considerably shorter when outlying observations are correctly modelled, therefore shedding some light on the purchasing power parity puzzle.

Chapter 1

Introduction

Structural stability raises enduring concerns when investigating economic and financial relationships. The main reason lies in the fact that there is plenty of evidence that macroeconomic and financial time series exhibit changes in their tendency. In a very comprehensive study, Stock and Watson (1996) analyse 76 U.S. post-war economic time-series including, amongst the others, price indexes, interest rates, exchange rates, and showing that the series experience widespread instability. The issue gained popularity amongst the econometric community in the eighties, following the article by Perron [Perron, P., 1989. "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis". *Econometrica* 57, 1361–1401], and since it became a central theme in time-series econometrics.

In practice, together with the fact that we want to accommodate empirical evidence, the main reason why we take into account structural breaks is because ignoring their presence leads to severe inferential distortions. As largely documented in the literature (see e.g. Pesaran, Pettenuzzo, and Timmerman, 2006), these distortions have striking implications for forecasting

as well as for policy analysis. Similar consequences arise when the inferential process disregards the presence of outlying observations which contaminate the available data. These occur when an economic/financial shock has a short-lived effect in contrast with the long-lasting effects of structural breaks. In light of these considerations, a joint treatment of structural breaks and outliers seems to be advisable. This is the path followed throughout the thesis, where we focus on the analysis and development of new econometric methods to deal with the presence of structural breaks and outlying observations.

1.1 Unstable Relationships and Outlying Observations

In this section, we overview the frameworks adopted to model relationships that change over time and/or are affected by outliers. The aim is neither to provide a detailed nor a comprehensive literature review of all the available methods, but rather try to settle the approaches later developed in the context of the existing econometric and statistical literature. At the beginning of each chapter, a detailed review of the relevant literature specific to the issues analysed is provided.

1.1.1 Modelling Unstable Relationships

From a modelling perspective, there are two ways to incorporate in a statistical/econometric model the changing behaviour of a time series: either by assuming that shifts happen at discrete times or by using time-varying para-

meter (TVP) models (see e.g. Harvey, 1989), which constitute a limiting case where a change occurs at all the periods. Discrete changes may be captured either in a deterministic or in a stochastic way. In the deterministic case, the discrete shifts are usually referred as "structural breaks" or "change points" and can be incorporated in the model through dummy variables. Structural breaks analysis has a long tradition in the statistical and econometric literature[1] dating back to the seminal works by Chow (1960), Quandt (1960), Gardner (1969) and Brown, Durbin, and Evans (1975). On the other hand, stochastic discrete shifts are usually captured by specifying a finite number of regimes with an associated matrix of transition probabilities. This implies that the model moves from one regime to the other in a stochastic fashion. In the econometric literature, a model class incorporating stochastic discrete shifts which gained considerable popularity is the Markov-switching class of models, following the work of Hamilton (1989). In practice, however, there is no uniform evidence of the superiority of one class of model with respect to the other. Despite their generality, TVP models do not necessarily over-perform a discrete structural change model as shown in Stock and Watson (1996). The choice between deterministic breaks and stochastic regime-switching models depends instead on how likely a series moves between a predetermined number of regimes. Regime-switching models implicitly assume that a relationship breaks in the same way it happened in the past which is clearly not always reasonable. On the other hand, discrete deterministic changes might be easier to justify and to link to particular shocks (policy intervention, financial crisis, etc.) that might have affected an eco-

[1]Stock (1994), Banerjee and Urga (2005), Perron (2006), Andreou and Ghysels (2009), and Aue and Horváth (2013) provide extensive surveys discussing a wide range of problems related to structural breaks analysis.

nomic relationship. In this dissertation, we deal with deterministic structural breaks and the word deterministic will be omitted from here onward.

1.1.2 Outlying Observations

An outlying observation is a data point which is distant from the other observations. Extreme observations can arise either because there are measurement errors in the data or because the data come from fat-tailed distributions. In the former case, we normally discard the outlying observations whereas in the latter case, we like to take them into account by using a modelling framework which allows us to incorporate them in the model. There are essentially two approaches to deal with outliers in regression analysis. The first relies on outliers detection procedures – see for instance the time series procedure described by Tsay (1988) – and it is aimed at identifying outliers based on test statistics or regression diagnostics. Statistically significant outliers can then be corrected or removed by the modeller and the model parameters re-estimated. The second approach relies instead on *robust statistics* in the sense of Rosseauw and Leroy (1986). In this case, the model is fitted directly to the "well" behaved data whilst outlying observations are identified as a by-product. This literature was initiated by the seminal contribution of Huber (1964) where the author introduced the class of M-estimators. Recently, a new outlier-robust method has been proposed in the econometric literature by Hendry (1999), known as *Dummy Saturation*. The resulting estimator has been classified by Johansen and Nielsen (2009) as a member of the class of M-estimators. The method is quite powerful although its statistical properties are largely unexplored. Its main peculiar-

ity is that it allows a joint treatment of outlying observations and structural breaks by saturating the initial model with different kind of dummies. Next section summarises the contribution of each chapter.

1.2 Contribution of the thesis

In this thesis, we contribute to the econometric literature concerned with the problem of detecting and locating structural breaks and outlying observations affecting parametric linear models. Special emphasis is put on the analysis and extension of existing approaches with unknown finite sample properties as well as the development of completely new techniques to make inference on the presence of structural breaks and outliers. The empirical relevance of the proposed econometric methods is illustrated through the empirical analysis of macroeconomic and financial relationships. The next sections describe the specific contributions of each of the three chapters.

1.2.1 Chapter 2: Detecting Breaks by Dummy Saturation and Sequential Bootstrapping

Chapter 2 consists of a simulation based comparison and extension of two largely unexplored procedures to detect multiple structural breaks in univariate systems involving stationary, non-stationary and cointegrated variables. Through an extensive Monte Carlo simulation study, we investigate the ability of the *Dummy Saturation* (Hendry, 1999; Hendry et al., 2008; Johansen and Nielsen, 2009) and the *sequential bootstrapping* procedure (Banerjee, Lazarova, and Urga, 1998; De Peretti and Urga, 2005) of the sup-F test

of Andrews (1993) in detecting the correct number of breaks and correctly locate them. In particular, we consider structural breaks affecting the deterministic component of linear models (level and linear trend). We explore for the first time the behaviour of the *Dummy Saturation* in capturing breaks in trend by means of trend dummies. Finally, we illustrate the practical relevance of the two procedures through an application to investigate the validity of the Fisherian hypothesis in the US economy.

There are two motivations behind the decision to analyse the above mentioned procedures. First, in the existing literature, there is scarce and incomplete evidence on how they behave in finite-sample and especially when considering non-stationary variables and cointegrated relationships. This is due to the fact that both procedures are relatively recent and the underlying theory is still under development. The extensive simulation study in this thesis fills this gap by providing evidence of their performance for a large class of models usually adopted in applied economic research. Second, the two procedures under analysis constitute alternative frameworks to the methodology of Bai and Perron (1998, 2003) and its extension to include non-stationary variables by Kejriwal and Perron (2008, 2010). These methods are theoretically well established though they rely on non-pivotal statistics to decide on the number of breaks which require extensive simulations to generate the appropriate critical values. Conversely, both the dummy saturation and the sequential bootstrapping procedure do not rely on non-pivotal statistics to ascertain the number of structural breaks.

1.2.2 Chapter 3: Testing for Multiple Breaks in the VECM

In Chapter 3, we deal with the issue of testing for multiple structural breaks in the vector error correction modelling (VECM) framework. In particular, being the VECM a heavily parametrised model, it is crucial to correctly specify the subset of parameters of which we want to test the stability. This is the underlying motivation of the results in Chapter 3. Specifically, we contribute to the existing literature in two main directions. First, we show that breaks in the long run matrix β imply breaks in the short run impact matrix α, unless weak exogeneity is imposed, and breaks in β imply also breaks in the covariance matrix of the error term. Hence, this result implies the presence of restrictions on the way the parameters to be tested for breaks are selected. Second, we extend the likelihood ratio test proposed in Hansen (2003) to the case of unknown break dates through the specification of several scenarios regarding the number and the location of the breaks. We define a minimum p-value statistic with critical values approximated by bootstrapping methods. Monte Carlo simulations show that the proposed statistic has optimal finite sample properties when imposing and relaxing weak exogeneity as well as when exploring the impact of weak identification of the cointegrating relationship. Finally, the chapter reports three empirical applications which illustrate how the minimum p-value statistic can be used in practice as well as the relevance of the presence of weak exogeneity on break testing. In particular, we use the minimum p-value statistic to study the dividend-price ratio for the S&P500 and the money demand in the US. To investigate the role of weak exogeneity, we study instead the relationship between three exchange rates.

1.2.3 Chapter 4: Robust Estimation of Real Exchange Rate Process Half-life

Chapter 4 tackles the puzzling behaviour of the real exchange rate (see Rogoff, 1996) developing an estimation framework for ARMA processes which is robust to the presence of outliers and level shifts. Specifically, we argue that the data generating process of the real exchange rate is likely to include outliers that, if not accounted for, lead to unreliable half-lives estimates. In order to obtain robust estimates of the half-life, we propose to identify outlying observations by means of a dummy saturation type algorithm designed for ARMA processes. The proposed procedure allows us to detect additional and innovative outliers as well as level shifts in the real exchange rate process. An empirical application involving US dollar real exchange rates shows that the estimated half-lives are consistently shorter when outlying observations are correctly modelled, thus shedding some light on the purchasing power parity puzzle.

Chapter 2

Detecting Structural Breaks by Dummy Saturation and Sequential Bootstrapping*

2.1 Introduction

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

*A research paper joint with my PhD supervisor, Prof. Giovanni Urga, entitled *"Detecting Structural Breaks by Dummy Saturation and Sequential Bootstrapping"* is based on the results in this chapter and it has been submitted for publication. The paper has been presented at the 12th OxMetrics User Conference (Cass Business School, September 3-4 2012) and at the 1st IAAE Annual Conference (Queen Mary, University of London, June 26-28 2014).

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

In this chapter, we consider two novel approaches to detect multiple structural breaks in a wide range of linear models including the cointegra-

tion set up. The first procedure is an extension of the so called *Dummy Saturation* (DS) introduced by Hendry (1999) and further developed by Hendry et al. (2008) and Johansen and Nielsen (2009), which involves not only impulse dummies but also step and trend dummies, as originally mentioned by Banerjee et al. (1998), De Peretti and Urga (2005) and Ericsson (2011). The DS procedure consists in detecting structural breaks affecting a linear system by saturating a regression with dummies and then removing the non-significant ones through a *general-to-specific* approach. The second methodology consists in a *Sequential Bootstrapping* (SB) procedure based on the sup-F statistic of Andrews (1993). We evaluate the performance of both the extended DS and SB procedures under several DGPs including stationary, nonstationary and cointegrated models. In particular, we assess the two procedures according to the following two criteria: first, the ability to pick up the correct number of structural breaks, and second, the ability to correctly locate them. To this extent, we set up an extensive Monte Carlo exercise in order to compute the empirical retention frequencies of the DS dummies and the empirical rejection frequencies of the SB sup-F tests as well as to measure the goodness of the resulting break date estimates. More precisely, evaluation criteria include empirical size and power as well as the *gauge* and *potency* criteria as described in Castle et al. (2012); the latter assess the considered method from the model selection perspective rather than as formal statistical break tests. For the DS method, evaluation criteria yield mixed results confirming the model selection nature of the procedure. In contrast, the empirical size and power of the SB suggest that the method hold clear and definitive promise. Finally, in order to study the performance of the two novel procedures with a real dataset, we empirically investigate

the stability of the Fisher equation in the United States.

The remainder of the chapter is organised as follows. Section 2.2 presents the extended DS and the SB procedures. Section 2.3 reports the criteria used to assess the Monte Carlo exercise, while the results of the simulations are reported and discussed in Section 2.4, where we also offer some useful guidelines about the implementation of the two procedures. Section 2.5 reports the empirical application involving the analysis of the Fisher relationship in the United States. Section 2.6 concludes

2.2 Two Novel Approaches to Detect Multiple Structural Breaks

In this section, we describe the ideas underlying the DS and the SB procedures to detect breaks in the deterministic component of linear regressions. Specifically, let

$$y_t = \boldsymbol{\gamma}_t^\top \mathbf{x}_t + \varepsilon_t = \begin{bmatrix} \boldsymbol{\alpha}_t^\top & \boldsymbol{\beta}^\top \end{bmatrix} \begin{bmatrix} \mathbf{w}_t \\ \mathbf{z}_t \end{bmatrix} + \varepsilon_t, \quad t = 1, \dots, T$$

where $\mathbf{w}_t$ collects the deterministic components, $\mathbf{z}_t$ is a $k \times 1$ vector collecting lags of the endogenous variable as well as exogeneous regressors and $\varepsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\varepsilon^2)$ satisfies $\mathbb{E}(\mathbf{z}_t \varepsilon_t) = 0$. Given that a linear trend usually suffices for most economic applications, we can restrict $\mathbf{w}_t$ to consist of a constant and a linear trend without losing too much generality. Hereafter, we consider models that are nested in the following specification

$$y_t = \alpha_t^0 + \alpha_t^1 t + \boldsymbol{\beta}^\top \mathbf{z}_t + \varepsilon_t \qquad t = 1, \dots, T. \tag{2.1}$$

Conditionally on m unknown break dates $\{T_1, T_2, \ldots, T_m\}$, we have a piecewise constant model of the form

$$\begin{aligned}
y_t &= \alpha_0^0 + \alpha_0^1 t + \boldsymbol{\beta}^\top \mathbf{z}_t + \varepsilon_t \quad && T_0 = 0 \leq t < T_1 \\
y_t &= \alpha_1^0 + \alpha_1^1 t + \boldsymbol{\beta}^\top \mathbf{z}_t + \varepsilon_t \quad && T_1 \leq t < T_2 \\
&\vdots \\
y_t &= \alpha_m^0 + \alpha_m^1 t + \boldsymbol{\beta}^\top \mathbf{z}_t + \varepsilon_t \quad && T_m \leq t \leq T_{m+1} = T.
\end{aligned}$$

In the following, we want study the performance of DS and SB in making inference about the vector of break dates $\{T_1, T_2, \ldots, T_m\}$.

2.2.1 Dummy Saturation (DS)

The DS approach – as originally proposed by Hendry (1999) and Hendry, Johansen, and Santos (2008) – is a technique to test model constancy by means of a set of dummy variables. The underlying idea is to saturate a linear model involving T observations with T dummy variables (one for each observations) to capture outliers and structural breaks. Following the "*general-to-specific*" approach, the technique starts with an initial model where an outlier/break may happen at all times and then removes the statistically insignificant dummies. The framework is very general allowing to test for the presence of multiple structural breaks in a wide range of systems.

In particular, the original approach as outlined in Hendry (1999) and Hendry, Johansen, and Santos (2008) involves a saturation with 0-1 impulse dummies (*impulse indicator saturation*, IIS). Recently, Doornik, Hendry, and Pretis (2013) develop the theoretical properties of the step indicator saturation, where impulse dummies are replaced by partial sums of impulse

dummies (step dummies). In this chapter, we propose to extend the original IIS approach to include also step dummies and double partial sums of impulse dummies (sums of step dummies or simply trend dummies) to capture structural breaks in the level or in the trend. Following the terminology in Ericsson (2011), we define *super saturation* (SS) the version with also step dummies, and *ultra saturation* (US) the regression saturation involving also trend dummies.

According to the DS principle, inference about the vector of break dates is based on the selection of the parsimonious representation from one of the following *saturated* regressions

$$\text{DS} \begin{cases} \text{IIS:} & y_t = \alpha_0^0 + \alpha_0^1 t + \boldsymbol{\beta}^\top \mathbf{z}_t + \sum_{i=1}^{T} \gamma_i I_{i,t} + \varepsilon_t \\ \text{SS:} & y_t = \alpha_0^0 + \alpha_0^1 t + \boldsymbol{\beta}^\top \mathbf{z}_t + \sum_{i=1}^{T} (\gamma_i I_{i,t} + \psi_i S_{i,t}) + \varepsilon_t \\ \text{US:} & y_t = \alpha_0^0 + \alpha_0^1 t + \boldsymbol{\beta}^\top \mathbf{z}_t + \sum_{i=1}^{T} (\gamma_i I_{i,t} + \psi_i S_{i,t} + \omega_i L_{i,t}) + \varepsilon_t \end{cases}$$

where $I_{i,t} = \mathbf{1}(t = i)$, $S_{i,t} = \mathbf{1}(t \geq i)$ and $L_{i,t} = (t - i + 1)\mathbf{1}(t \geq i)$ for $i = 1, \ldots, T$. After dropping the statistically insignificant dummies from the saturated regression, one is left with a set of dummies which can be interpreted as outliers (impulse dummies) and/or structural breaks affecting the deterministic component of the process (step and trend dummies)[1].Additionally, it is important to note that the coefficients of the dummies capture the magnitude of the changes in the coefficients of (2.1), i.e. $\psi_i = \alpha_i^0 - \alpha_{i-1}^0$ and $\omega_i = \alpha_i^1 - \alpha_{i-1}^1$.

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

There are two problems with the specifications just introduced. For all cases, though especially in SS and US, multicollinearity between some of the dummies themselves and between the dummies and the deterministic component of the process clearly arises. Second, the estimation of the saturated regressions is infeasible because of lack of degrees of freedom, given that N, the total number of regressors, is larger than the number of observations T. In principle, the multicollinearity problem within dummies can be solved quite easily by excluding some dummies (last step dummy or last trend dummy for instance). Typically, one sets $i = \ell_y + 1, \ldots, T-1$, where ℓ_y is the highest order of lagged dependent variables entering the process, and excludes the first step dummy, which is exactly collinear with the set of impulse dummies, and the last trend dummy. As far as the dimensionality problem is concerned, a possible solution is to split the set of dummies in J blocks such that the number of dummies in each block (N_j) plus the number of elements in the deterministic components, the number of lagged dependent variables and exogenous regressors (k) is less than the sample size ($N_j + 2 + k < T$, for $j = 1, \ldots, J$), following the strategy introduced in Hendry, Johansen, and Santos (2008). More explicitly, in the general case of US, assume to form J blocks of about the same size[2] such that $\mathcal{I}_1 = \{I_i, S_i, L_i : i = 1, \ldots, \lceil T/J \rceil\}$, $\mathcal{I}_2 = \{I_i, S_i, L_i : i = \lceil T/J \rceil + 1, \ldots, \lceil 2T/J \rceil\}$,..., $\mathcal{I}_J = \{I_i, S_i, L_i : i = \lceil T(J-1)/J \rceil + 1, \ldots, T\}$. The procedure then runs as follows:

- for $j = 1, \ldots, J$ include the $\mathcal{I}_j$ subset of dummies in the equation of interest (e.g. (2.1)) and estimate the partially saturated regression re-

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

cording the significant dummies from each regression.

- Combine all relevant dummies from the previous iterations and re-estimate the model, assuming that the total number of the retained dummies from each subset $\mathcal{I}_j$ is less than the sample size.

- Retain the significant dummies.

This is the standard way to carry out a structural break analysis using the DS approach. As explained in Castle et al. (2012), under the null of no outliers or breaks, αT impulse indicators are retained on average (α being the level of significance). For this reason, if we fix $\alpha \leq r/T$ we control the false null retention at r dummies. This is rather satisfactory if we think that we are testing the potential relevance of a big number of dummies which is a multiple of the sample size T. This point is investigated in the Monte Carlo exercise where we show that the procedure delivers the correct retention rates only for some specific choice of the target size α.

An alternative and more convenient route to implement the DS is through the algorithm for automated model selection *Autometrics* (see Doornik, 2009a). This is accessible through the software OxMetrics© which can handle $N > T$ as well as non-orthogonal candidate regressors. Using *Autometrics* , it is possible to specify a *general unrestricted model* (GUM), in this case a regression saturated with dummies, and a tree search algorithm is able to eliminate statistically insignificant regressors. The entire procedure of block creation is carried out by the algorithm as well as the management of non-orthogonal regressors. An important aspect is that *Autometrics* provides different ways to create blocks in case of $N > T$ (i.e. sequential blocks, random blocks, cross blocks, etc.). Despite in theory, different ways of selecting

the candidate regressors which enter in each block should not affect the outcome radically, Doornik (2009b) reports that the ordering of the variables as they enter the GUM may influence the final results. We investigate also this issue through simulations in the Monte Carlo exercise.

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

2.2.2 Sequential Bootstrapping (SB) Procedure

The second methodology we consider is based on the sequential bootstrapping of the sup-F test for breaks (SB). This method is specifically developed to estimate multiple structural breaks in systems with a conditional process

and several marginal (the regressors) processes of the form:

$$y_t = \alpha_0 + \delta_0 t + \sum_{\ell_y=1}^{q} \rho_{y,\ell_y} y_{t-\ell_y} + \sum_{i=1}^{K} (\psi_i S_{i,t} + \omega_i L_{i,t}) + \\ + \sum_{n=1}^{N} (\tilde{\rho}_n x_{n,t-1} + \sum_{j=1}^{K_n} (\tilde{\upsilon}_{n,j} S_{n,j,t} + \tilde{\phi}_{n,j} L_{n,j,t})) + u_t \quad (2.2)$$

$$x_{n,t} = \gamma_{n,0} + \zeta_{n,0} t + \sum_{\ell=1}^{p_n} \rho_{x,\ell_x} x_{n,t-\ell_x} + \sum_{j=1}^{K_n} (\upsilon_{n,j} S_{n,j,t} + \phi_{n,j} L_{n,j,t}) + e_{n,t} \quad (2.3)$$

where (2.2) is the conditional process, (2.3) is a set of $n = 1, \ldots, N$ marginal processes experiencing their own independent breaks, $S_{i,t} = \mathbf{1}(t \geq T_i)$ and $L_{i,t} = (t - T_i + 1)\mathbf{1}(t \geq T_i)$ indicate the dummies capturing breaks in the level and the trend respectively and T_i, $i = 1, \ldots, K$ denotes the i^{th} break date of the conditional process; $S_{n,j,t}$ and $L_{n,j,t}$ indicate dummies capturing breaks in the level and the trend respectively in the marginal process and T_j, $j = 1, \ldots, K_n$ are not in general equal to T_i.

In a first instance, the algorithm estimates breaks in the marginal processes then these enter into the conditional process before the search for breaks in the conditional process starts. The estimation of the breaks is sequential in that break dates are added one by one until a stopping criterion is reached[3]. Thus, contrary to the DS approach, here the philosophy is from *"specific-to-general"*. Both stationary and nonstationary processes can be handled.

Conceptually, the sequential procedure has two building blocks: the first deals with the estimation of the break dates, whereas the second with the statistical testing framework required for stopping the sequential search.

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

The next two sections summarise the main features of the SB procedure.

2.2.2.1 Structural Breaks Estimation

Consider the case of a marginal process as in (2.3) and assume we are interested in estimating possible breaks affecting this process at unknown times. The procedure starts by searching for break which may occur at each date $t = 3, \ldots, T-1$ by means of a Wald type test. Thus, for each t, compute the F-statistic, F_t, for testing the null hypothesis of no more breaks at $T_{j=t}$, $H_0 : \upsilon_j = \phi_j = 0 | T_j = t$. When looking for the second break and so on for the others apart from the first one, the statistics is not computed for the already estimated break dates (more precisely for a small neighbourhood around them). In particular, F_t takes the usual form:

$$F_t = \frac{T-k-2(j+1)}{2} \frac{RSS_{T_{j-1}} - RSS_{T_j=t}}{RSS_{T_j=t}} \tag{2.4}$$

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

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

$$\hat{T}_j = \text{argmax}\{F_t\} \quad \text{for all admissable } t \tag{2.5}$$

This amounts to define the break date estimator as the argument which maximises the usual sup-F statistic introduced initially by Quandt (1960) [4] with non-standard asymptotic distribution derived in Andrews (1993) for the stationary case. Bai (1994) proved that the estimation of the break date T_j obtained by least squares in a linear model leads to asymptotically biased – and hence inconsistent – estimates. However, the bias is small when the magnitude of the break is big and in particular $\hat{T}_j - T_j = O_p(\lambda^{-2})$, with λ the magnitude of the break. It is evident from (2.4) that

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

hence, the proposed estimator is equivalent to the least squares estimator.

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

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

2.2.2.2 Stopping Rule for the Sequential Search

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

$$\hat{\tau} = \max\{F_t\}. \tag{2.6}$$

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

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

$$x_t = \gamma_0 + \zeta_0 t + \sum_{\ell=1}^{p} \rho_{x,\ell_x} x_{t-\ell_x} + \sum_{k=1}^{j-1} (\upsilon_k S_{k,t} + \phi_k L_{k,t}) + e_t \tag{2.7}$$

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

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

(a) Draw with replacement T values from the centred residuals[5]

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

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

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

$$x^*_{b,t} = \hat{\gamma}_0 + \hat{\zeta}_0 t + \sum_{\ell=1}^{p}\hat{\rho}_{x,\ell_x} x^*_{b,t-\ell_x} + \sum_{k=1}^{j-1}(\hat{\upsilon}_k S_{k,t} + \hat{\phi}_k L_{k,t}) + e^*_{b,t}. \quad (2.8)$$

(c) Compute the bootstrapped counterpart of $\hat{\tau}$, namely, $\hat{\tau}^*_b = \max\{F^*_{b,t}\}$ where $F^*_{b,t}$ is computed as in (2.4) but using the bootstrapped sample $x^*_{b,t}$.

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

$$\frac{1}{B}\sum_{b=1}^{B}\mathbf{1}\left((\hat{\tau})^2 < (\tau^*_b)^2\right). \quad (2.9)$$

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

[5]We assume the residuals do not exhibit autocorrelation or heteroskedasticity. Conversely, the bootstrap procedure can easily be modified either using blocking techniques or multiplying the centred residuals by a symmetric random variable with mean zero and variance one (see e.g. Davidson and Flachaire, 2008).

2.3 Monte Carlo Design

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

2.3.1 Evaluating the Detection of the Correct Number of Breaks

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

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

As far as the DS is concerned, however, the computation of the empirical rejection frequency it is not always straightforward. This is true in particular to the case of the IIS. More specifically, it is difficult in the case of the IIS to

test for instance the hypothesis of the form $H_0 : K = K_0$ versus $H_1 : K = K_1$ level breaks given the nature of the impulse dummies. When instead we saturate the model with step and/or trend dummies, it is easier to cast the procedure in terms of testing $H_0 : K = K_0$, given that a step or a trend dummy is directly related to a break in the level or the trend component. Alternatively, we may assess the performance of the procedure considering how frequently a dummy entering the saturated regression is retained across the Monte Carlo replications, thus treating the DS as a model selection procedure rather than a test for breaks.

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

$$\hat{r}_j = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1} \left(\hat{\beta}_{j,m} \neq 0 \right), \quad j = 1, \ldots, N \tag{2.10}$$

where $\mathbf{1}(\cdot)$ is the indicator function and $\hat{\beta}_{j,m}$ is the coefficient of the j^{th} dummy computed at the m^{th} iteration. If the dummy is statistically significant ($\hat{\beta}_{j,m} \neq 0$)[6], then the indicator function equals unity. Given (2.10), it is possible to define the *gauge* and the *potency* the procedure as

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

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

[6]Note that this is a short-hand for $\left| \text{t}_{\hat{\beta}_{j,m}} \right| \geq c_{\frac{\alpha}{2}}$, with $\text{t}_{\hat{\beta}_{j,m}}$ denoting the t-ratio and $c_{\frac{\alpha}{2}}$ the associated critical value at a significance level α.

where the gauge is the average retention rate of the non-significant dummies (according to the DGP) and the potency is the average retention rate of the significant dummies[7]. Note that, although gauge and potency are related to the concepts of size and power, they are actually distinct: while size and power refer to a well specified null hypothesis in terms of K_0 versus K_1 breaks, the gauge and the potency are useful criteria for the assessment of the procedure but are not clearly linked to any statistical hypothesis about the number of the breaks. Finally, note that according to the definitions above, a dummy contributes to the potency if it enters the DGP. This is straightforward when saturating with impulse dummies. On the other side, if saturating the GUM with step and/or trend dummies (SS and US models respectively), we have to be less restrictive in that we have to define an interval around the break date such that a step and/or trend dummies falling in this interval contributes to the potency. Assume for instance that a break occurs at $T_1 = 50$, it would be unreasonable to count $S_{52,t} = \mathbf{1}(t \geq 52)$ as contributing to the gauge of the SS. For this reason, when computing the gauge and the potency of the SS or US, we consider a dummy as contributing to the potency if it captures a break occurring at $T_i \pm 3$ observations. If more then one dummy falling in that interval is retained, we choose the one closest to the break date, while in the unlikely case that two dummies are equally-distant from T_i we count randomly one as contributing to the potency and one contributing to the gauge.

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

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

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

sup-F statistic and given a significance level α, we then compute

$$\frac{1}{M}\sum_{m=1}^{M}\mathbf{1}\left(p\text{-value}_{m,K1}<\alpha\right) \tag{2.11}$$

where p-value$_{m,K_1}$ represents the p-value of the K_1^{th} break date computed according to (2.9) at the m^{th} iteration. If the null is true ($K_0 = K^*$), (2.11) gives the empirical size of the procedure since it computes the number of significant breaks that do not enter the DGP. Then, for the procedure to have the correct size, we expect that the number of times that the K_1^{th} break date is significant (its p-value is small) and of course should not exceed α. On the contrary, if the null is false, (2.11) represents the empirical power. To summarise, assume that the DGP is formulated with $K^* = 1$, then testing $H_0 : K = 0$ versus $H_1 : K = 1$ gives the power of the procedure while a test of $H_0 : K = 1$ versus $H_1 : K = 2$ gives the size. So, we have

$$\begin{aligned}\text{size} &= \frac{1}{M}\sum_{m=1}^{M}\mathbf{1}\left(p\text{-value}_{m,T_2}<\alpha\right)\\ \text{power} &= \frac{1}{M}\sum_{m=1}^{M}\mathbf{1}\left(p\text{-value}_{m,T_1}<\alpha\right).\end{aligned}$$

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

2.3.2 Evaluating the Dating of Breaks

In addition to the desiderable property of selecting the correct number of breaks, a structural breaks procedure has to able to pick up the correct tim-

ing of when the breaks occur. This translates in assessing the goodness of the resulting break date estimates.

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

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

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

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

For sake of simplicity, assume we have a process with one break at T_1. The first issue concerns with the DS procedure. When working with the IIS, we need to define a rule such that at each iteration of a Monte Carlo experiment we can consider a spike dummy ($I_{i,t} = \mathbf{1}(i = t)$) as the reference (i.e. $\hat{T}_{m,1}$) to compute the bias. In this case, we use the dummy belonging to the relevant set of dummies which is closest to the break date[8]. If there are multiple breaks, first we have to group the spike dummies according to their magnitude and then select the first one from each set. To reduce the

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

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

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

2.4 Simulation Results

In this section, we present the results from our simulations. In each experiment, the number of simulations is set to $M = 1,000$ and the sample size

to $T = 100$. We fix the seed of the random number generator and we build recursively $T + n$ (where n depends on the characteristics of the process) observations M times, starting from $y_{-n} = 0$ and then discarding the first n in order to create independence from the initial conditions.

2.4.1 Breaks in Level

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

$$y_t = \alpha + \delta t + \beta x_t + \sum_{i=1}^{K} \psi_i S_{i,t} + u_t \tag{2.12}$$

$$x_t = \gamma + \zeta t + \rho x_{t-1} + \sum_{j=1}^{K_x} \upsilon_j S_{j,t} + e_t \qquad t = 1, \ldots, T \tag{2.13}$$

where $S_{i,t} \equiv \mathbf{1}(t > T_i)$ denotes a step dummy (the same applies to $S_{j,t}$). Moving from this simple benchmark to the alternative more general model including the breaks in trend allows us to understand the impact of the different ways to create blocks by *Autometrics* . It is reasonable to expect that, if different mechanisms to create blocks have some impact on the selection of the final model, this becomes even more relevant as we move from the simplest IIS to the US.

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

For the DS, all the GUMs correspond to the DGPs saturated with dummies and we fix all the variables that enter the DGP except the dummies. This entails that the rejection frequencies are computed only with respect to the dummies (i.e. we do not test the ability of *Autometrics* to reject non-significant regressors apart from dummies). On the other hand, the SB size and power properties are assessed by computing both parametric and non-parametric p-values through 99 bootstrap replications. The nominal significance levels used in the simulations are 1% and 5%. All DGPs are investigated assuming that the number of breaks is $K = \{0, 1, 2\}$. When conditional models with broken marginals are considered, the number of breaks in the marginal model is fixed to one and the inference focuses then on the number of breaks in the conditional model. The break dates and the coefficients of the related dummies (ψ_i and υ_j) are reported in the tables summarizing the results of the simulations. For the IIS, we restrict the experiments to the empirical retention frequency of the dummies entering the DGP, i.e. the gauge and the potency. For the SS, we assess also the size and power by computing the empirical rejection frequencies of the following hypothesis: $H_0 : K = 0$ for the DGP with no breaks (size only); $H_0 : K = 0$ (power) and $H_0 : K = 1$ (size) for the DGP with one breaks; and $H_0 : K = 1$ (power) and $H_0 : K = 2$ (size) for the DGP with two breaks. As far as the SB is concerned, we set the maximum number of breaks, to be sequentially estimated in each experiment, equal to the true number of breaks plus one. Specifically, we have that for the case with no breaks a test of $H_0 : K = 0$ vs $H_1 : K = 1$ (size only); for the case with one break a test of $H_0 : K = 0$ vs $H_1 : K = 1$ (power), and $H_0 : K = 1$ vs $H_1 : K = 2$ (size); and for the case with 2 breaks a test of $H_0 : K = 1$ vs $H_1 : K = 2$ (power), and $H_0 : K = 2$ vs $H_1 : K = 3$ (size).

2.4.1.1 Marginal Processes

The first set of experiments involves modelling the exogenous regressors one at the time. These include location-scale models, stationary or non-stationary autoregressive processes and autoregressive processes with or without trends which are all nested in (2.13). Table 2.1 reports the model specifications used in the Monte Carlo experiments for the DGPs.

[Table 2.1 about here.]

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

Number of Breaks Detection

Table 2.2 reports gauge and potency as well as the empirical size and power for the processes described in Table 2.1. In order, we consider the IIS, the SS and the SB based on both parametric and nonparametric resampling schemes. For the IIS and the SS, we set the block method creation in the *Autometrics* options both *standard* (i.e. sequential) and *random*. Since both gauge/potency and empirical size/power are practically identical either using standard block method or the random block method creation, we report the results for the standard option only.

Table 2.2 reports the gauge/potency and the size/power of the two procedures.

[Table 2.2 about here.]

Dummy Saturation. Focusing on the DS, it emerges that both the IIS and the SS have a good gauge at 1% target size, meaning that the average retention rate of the insignificant dummies is satisfactorily under control. On the opposite, when working with 5% target size the gauge is generally between 20% and 30%. Thus, in line with Castle et al. (2012), this suggest for the DS to set a 1% target size in the *Autometrics* options. The average retention rate of the non null dummies is instead mixed across the different DGPs changing from the IIS to the SS. While the potency level for the SS are rarely below 80%, the IIS is affected in the cases of the autoregressive processes and the location of the breaks. In particular, when working with IIS and the AR models, there is lower potency because *Autometrics* tends to select just the dummies near the break dates, without retaining dummies within the regimes. This is due probably to the fact that after the shocks have been correctly captured by the impulse dummies, the subsequent (if 1 break) or intermediate instability (if 2 breaks) is absorbed by the autoregressive component. This may explain the loss in potency with respect to the LS model. From Table 2.2, we can see that the loss in potency decreases if we restrict the interval between the two break dates. Furthermore, the simulation exercise allows also to highlight that the location of the break(s) matter for *Autometrics* when applying the IIS. If the level shift is close to the end of the sample, in general is no problem. However, as we move the break in the middle of the sample, the gauge increases in particular in the case of multiple level shifts as the distance between two break dates increases. The first effect might be explained in that *Autometrics* selects the minimum number of impulse dummies that provide a good approximation of the underlying DGP. If the level shift is at the beginning of the sample, *Autometrics* selects

the impulses from the first observation up to that point.

As far as the SS is concerned, Table 2.2 also reports the empirical size and power about the number of structural breaks as described in Section 2.4.1. There is evidence of very high oversize both at 1% and 5% nominal levels. When considering the 1% nominal level, in almost 50% of the cases at least one insignificant step dummy is retained. As a consequence, the number of times that the procedure detects the correct number of breaks is around 50% as well. However, as already anticipated, this result is expected in that when applying the DS we are testing the significance of a huge number of dummies and this comes at the cost that some insignificant dummies are retained. For this reason, the information given by the retention rates of the "null" and "non-null" dummies is in general a better criterion to evaluate the performances of the DS. For the DS, in the next sets of simulations, we focus only on the computation of the gauge and potency, assessing the approach as a model selection procedure.

Sequential Bootstrapping. Table 2.2 show that the procedure has an empirical size almost identical to the nominal level at both 1% and 5% nominal levels when applied to stationary and trend-stationary models, while is slightly oversized for the non-stationary autoregressive model. The most dramatic oversize is for the multiple breaks case when we observe a 13% with a nominal of 5% and a 3% for a nominal of 1%. This implies that the SB detects the correct number of structural breaks in almost all cases. As far as the power is concerned, the empirical rejection frequencies of false null hypotheses are close to 100% for all DGPs considered in the simulations.

Break Dates Estimates

To evaluate the goodness of the break estimates, in Table 2.3 we report the sample mean, the mean bias, the root mean squared error and the median for each estimate.

[Table 2.3 about here.]

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

The main conclusions from the Monte Carlo exercises can be summarised as follows. First, as expected the SS outperforms the IIS when considering multiple breaks affecting the level component. The IIS seems to work well for capturing outliers and single level shifts but when dealing with multiple breaks in level, the SS and the SB prove to be more appropriate. Second, using empirical rejection frequencies to assess the DS can be misleading while it is more appropriate to work with the empirical retention rates of the dummies. Third, the SB provides empirical size close to the

nominal one and very good power across all simulations with the exception of the nonstationary autoregression model. Finally, both the IIS and SS as well as the SB provide good estimates of the break dates when considering structural breaks affecting the level component. A final important point is about that to discriminate between alternative methodologies it is better to use the gauge/potency and size/power properties rather than the goodness of the break dates estimates.

2.4.1.2 Conditional Processes

We now turn to the case of conditional models with stationary and non-stationary regressors. This setting allows us to explore how breaks in the marginal processes (regressors) affect the conditional equation[9]. Table 2.4 reports the DGPs considered in this second set of Monte Carlo simulations.

[Table 2.4 about here.]

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

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

Number of Breaks Detection

Table 2.5 reports the gauge and the potency as well as the empirical size and the power for the conditional processes introduced above. For the IIS and the SS, as in the previous set of simulations involving location-scale and autoregressive processes, we do not find significant differences among the size and power recorded in changing the block method and thus we report the results for the standard block method option only. Thus, the assessment of the SS is undertaken on the basis of the gauge and potency.

[Table 2.5 about here.]

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

Sequential Bootstrapping: As far as the SB is concerned, the procedure delivers empirical size close to nominal both at 1% and 5% nominal levels and the power is close to 100% for all DGPs. Moving from stationary to nonstationary processes does not seem to affect the performance of the IIS, the SS and the SB, the only exception being the IIS with one break case ("Cci" and "CMci"). This is of particular interest for the SB procedure in support of the validity of the bootstrapping in a nonstationary cointegrating framework. Another important result is that both the DS and the SB are robust to the presence of a break in the marginal process. The gauge and the empirical size of the procedures are not affected by letting the marginal process

experience an independent break from the ones in the conditional. As far as the DS is concerned, the Monte Carlo exercise extends the results of the simulations contained in Castle et al. (2012) for the IIS and shows the good performances of the SS. Turning to the SB, the robustness of the procedure to broken marginal processes is even more important in light of that Hansen (2000) shows that the distribution of the sup-F statistic to test for breaks in the conditional process is affected when a break is observed in the marginal process. To avoid rejecting stability in conditional models in absence of breaks, Hansen (2000) proposes a bootstrapping approach denoted *fixed regressor bootstrap*. However, the author does not take into account the cointegrating framework while presenting an autoregressive case with a moderate level of persistence (0.5). Though the SB is based on a different bootstrapping approach[10], the results confirm once again that the adoption of the bootstrap in cointegrating equations seem to perform well (see e.g. Li and Maddala, 1997).

Break Dates Estimates

Table 2.6 reports the results of break date estimates. All the three procedures perform quite well in terms of goodness of the resulting estimates. The estimates of the break dates are median unbiased and the mean bias is very controlled and less than one except for the IIS in some cases. Overall, the best results in terms of unbiasedness and smaller RMSE are those obtained applying the SB.

[Table 2.6 about here.]

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

2.4.2 Breaks in Trend

In this section, we consider processes with breaks also affecting the deterministic linear time trend thus extending (2.12)-(2.13) as follows:

$$y_t = \alpha + \delta t + \beta x_t + \sum_{i=1}^{K} (\psi_i S_{i,t} + \omega_i L_{i,t}) + u_t \tag{2.14}$$

$$x_t = \gamma + \zeta t + \rho x_{t-1} + \sum_{j=1}^{K_x} (\upsilon_j S_{j,t} + \phi_j L_{j,t}) + e_t \qquad t = 1, \ldots, T \tag{2.15}$$

where $L_{i,t} \equiv (t - T_i + 1)\mathbf{1}(t > T_i)$ denotes a trend dummy. We consider both autoregressive and conditional processes and we let also the marginal process experience independent breaks. We restrict our analysis to the US and the SB. In Table 2.7, we report the list of the DGPs considered.

[Table 2.7 about here.]

ARsTr" and "ARnsTr" denote the stationary and nonstationary autoregressive processes respectively, "CsTr" and "CciTr" the conditional model with stationary regressors and nonstationary cointegrated regressors respectively, "CMsTr" and "CMciTr" the stationary and cointegrated conditional models where the marginal processes experience independent breaks, respectively. We assume that the structural break(s) affect contemporaneously both the level and the trend components, though we let the shock to be not necessarily of the same signs between the level and the trend. This implies that a negative shock in the level might be present together with a positive shock in the trend. Similarly as for the DGPs with just breaks in level, we carry out a set of experiments for $K = \{0, 1, 2\}$. The null hypotheses are the same as those reported in Section 2.4.1.

Number of Breaks Detection

Table 2.8 reports the gauge/potency and the empirical size/power for Monte Carlo simulations involving the processes described in Table 2.7. As far as the US is concerned, we explore again the standard and the random methods to create blocks as also suggested in Johansen and Nielsen (2009) for the trending and unit root cases. In particular, the GUM is specified alternating step and trend dummies such that the standard block search algorithm of *Autometrics* takes into account the dummies referring to the same potential break date in the same block[11]. As expected, when saturating with step and trend dummies, the role of the different methods for creating blocks impacts on the final selected model. Indeed, when moving from standard to random blocks, although the variations in terms of potency are not dramatic, the gauge is affected. Thus, we also report the results for the random block method. Again, the assessment of the US is based on the gauge and the potency only since the saturation with further dummies would make the empirical rejection frequency even less meaningful than before.

[Table 2.8 about here.]

Dummy Saturation: As far as the US is concerned, the gauge appears to be under control setting the target size to be either 1% or 5% when working with the standard block method. When using the random blocks instead, the gauge of the US tends to increase above the nominal level while the gain in potency is negligible. In particular, for the autoregressive processes, we see high average retention rates of null dummies also when setting 1% as

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

target size. Thus, we may conclude that the use of the random blocks does not improve the performance of the US. On the other side, when working with standard blocks, the controlled gauge at both 1% and 5% is especially surprising given that the number of dummies that enter the initial GUM is $T \times 3$. However, the controlled gauge comes with a considerable drop in potency. The best potency level we find does not go above 70% with an average around 50%. This results is expected since there are two dummies which should be retained for each break date.

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

Break dates estimates

Table 2.9 reports the structural break estimates for the US and the SB procedures. For the US, we report separately the estimates of the break dates in the level from those affecting the trend.

[Table 2.9 about here.]

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

timates. For the SB, the estimates are all median unbiased and the RMSEs smaller than the US counterparts.

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

2.4.3 Some Useful Guidelines to Practitioners

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

With respect to the DS, an important role is played by the *target size*. As shown in the the Monte Carlo simulations, the DS works properly only if the 1% target size is used. A higher target size would lead to the retention of too many dummies. As far as the *block method* is concerned, the standard block method performs best. As also highlighted in Doornik (2009b), we discourage the use of the random block scheme because it leads to the retention of more insignificant dummies. As far as the *choice between the IIS*

and the SS in systems affected only by breaks in mean, what emerges from the simulations is that IIS and SS provide satisfactory performances both in terms of gauge and potency as well as in terms of break date estimates. In particular, the SS provides better results in terms of potency than the IIS, as also highlighted by Doornik et al. (2013). Moreover, saturating with both step and impulse dummies allows the step dummies to capture breaks in level while the impulse dummies capture possible outliers. This helps also to simplify the interpretation of the final output. When using the SS, there is no need to group the impulse dummies according to their sign and magnitude to identify a break in level, as it happens when using the IIS. The procedure of grouping the impulse dummies may be difficult to implement as the sample size grows. Note that when considering systems affected also by *breaks in trend*, the potency of the US tends to be lower than the case of only breaks in level. Despite the loss in potency, the performance of the procedure is similar to that of the SB. In addition, since the *ordering of the variables* matters as the number of dummies grows, we suggest to try alternative ordering schemes for the step and trend dummies. What we learn from the experiments reported in this chapter is that the best approach is to alternate one step and one trend dummy in a chronological order. In this way, *Autometrics* takes into account contemporaneously dummies capturing the same break date when forming the blocks.

As far as the SB is concerned, the procedure works very well in detecting breaks in both level and trend. There are only few cases when the SB tends to be oversized occurring in the nonstationary autoregressions while the procedure works quite well when dealing with cointegrating equations. An additional point is about the *stopping rule in the sequential search*. As

already reported in De Peretti and Urga (2005), the optimal rule is to stop the sequential search when two consecutive break dates are statistically insignificant. This procedure is robust to the the fact that the estimator of the initial break dates is biased when further breaks are neglected.

2.5 Testing the Fisher Effect in the United States Economy

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

$$r_t^{t+1} = r_t^{real} + \mathbb{E}_t \pi_t^{t+1} \tag{2.16}$$

where r_t^{t+1} is the nominal interest rate paid for instance by a bond over the period $[t, t+1]$, r_t^{real} is the real interest rate, and $\mathbb{E}_t \pi_t^{t+1}$ is the expected rate of inflation over the period $[t, t+1]$ given the information set available at t. In a rational expectations framework, the expected inflation is equal to the actual one plus a mean-zero Gaussian forecast error term, i.e. $\mathbb{E}_t(\pi_t^{t+1}) = \pi_t^{t+1} + e_t$ with $e_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$. The Fisher equation (2.16) holds if estimating the linear equation

$$r_t = \alpha + \delta t + \beta \pi_t + \varepsilon_t. \tag{2.17}$$

we find cointegration – assuming r_t and π_t are both $I(1)$ – and β is statistically equal to one. The fact that many empirical studies (see Beyer et al., 2009) conclude that β differs significantly from unity may be due to the presence

of unmodelled level/trend shifts in the Fisher relationship. To this purpose, in this chapter we implement both the US and the SB procedures to equation (2.17).

We employ quarterly data running from 1985-Q1 to 2012-Q3 (111 observations) obtained from the Federal Reserve Bank of Saint Louis database (FRED). We use the 3 months Treasury Bill rate as short term nominal interest rate (r_t) and we compute the inflation rate (π_t) as the annual relative change in the CPI $\left(\frac{CPI_t - CPI_{t-4}}{CPI_{t-4}}\right)$. Figure 2.1 reports the plots of the two series.

[Figure 2.1 about here.]

Both series are tested for the presence of unit roots and consistently with other studies involving United States post-war data (see for instance Crowder and Hoffman, 1996) we cannot reject the hypothesis that both series are I(1). Following Beyer et al. (2009), in order to address potential endogeneity between interest rates and inflation rates, we estimate (2.17) by DOLS including five leads and lags of the first differenced rate of inflation.

We first look for potential breaks in (2.17) by saturating with impulse, step, and trend dummies, i.e. applying the US method. To this extent, we set the target size at 1%, we fix all the variables entering the Fisher equation except the dummies, and we explore different combinations in the ordering of the step and trend dummies as well as different block methods. In particular, we specify GUMs where the step and trend dummies enter sequentially (i.e. all the step dummies and then all the trend dummies) versus GUMs which alternate one step and one trend dummy. We also employ both standard and random block methods. Consistently with what emerges

in the Monte Carlo study, the final model that support the economic theory and passes all misspecification tests is that obtained alternating the dummies and using the standard block method. Table 2.10 reports the final model selected by *Autometrics* when estimating (2.17) by DOLS (to control for endogeneity) and applying the US:

[Table 2.10 about here.]

Nine breaks and one outlier are identified. All misspecification tests are satisfied and the residuals have all the desirable properties as Figure 2.2 shows, where we report the residuals, the ACF up to 20 lags and the QQ-plot against a $\mathcal{N}(0,1)$. Moreover, the residuals are stationary implying the presence of a cointegrated relationship as found by Crowder and Hoffman (1996).

[Figure 2.2 about here.]

Table 2.10 shows that in the Fisher effect holds being $\beta \in [0.857; 1.05]$ with probability 95%.

Next, we repeat the analysis using the SB to search for potential breaks. To this extent, we estimate the same model by DOLS stopping the sequential search when two consecutive break dates are statistically insignificant. In particular, the significance of the break dates is assessed through parametric and non-parametric bootstrapped p-values of the corresponding sup-F statistics. Table 2.11 reports the statistically significant break dates identified by the SB procedure.

[Table 2.11 about here.]

On the basis of the break dates selected by the SB, we estimate the Fisher equation by DOLS and adding the corresponding step and trend dummies. Table 2.12 reports the final selected model after removing the insignificant variables[12].

[Table 2.12 about here.]

Via the implementation of the SB, we identify six significant structural breaks captured by step and trend dummies. Similarly to what we find for the US, the selected model provides a good fit of r_t with no sign of misspecification. Figure 2.3 reports a plot of the residuals together with their ACF and a QQ-plot. Once more, after accounting for structural breaks, the relationship between the short term interest rate and the inflation rate cointegrates and the Fisher effect holds given that $\beta \in [0.829; 1.187]$ with probability of 95%.

[Figure 2.3 about here.]

Finally, to give a visual representation of the similarity of the structural breaks identified by the US and the SB procedures, in Figure 2.4 we report the actual series of the interest rate with the structural breaks identified by vertical solid lines to denote breaks affecting the level component and vertical dashed lines to denote breaks affecting the trend component.

[Figure 2.4 about here.]

The dynamics which emerges from Figure 2.4 shows that there is almost a one-to-one correspondence between the structural breaks identified by the US and the SB. In particular, the differences in the breaks dates are all within

[12]We additionally run *Autometrics* over the final model with the "Large Residuals" option to control for large residuals. This is denominated "I:2000(3-4)" in Table 2.12.

one quarter when comparing a break found by the US with the correspondent closest break found by the SB. Hence, we have that the two procedures validate each other and this is an important result given that they are based on two completely different logics.

2.6 Conclusions

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

linear time trend component (US model), the potency decreased although the gauge remained well controlled. Furthermore, we evaluated the sensibility of US model to the block method options provided by *Autometrics*: the use of a random block method did not improve the overall performance of the DS consistently with what reported in Doornik (2009b). On the other hand, the SB proved to have empirical size close to nominal one both at 1% and 5% levels for all the DGPs considered, the only exception being the case of nonstationary autoregressive processes where it exhibited a small oversize. Opposite to the US case, the SB was not affected by the inclusion of trend dummies in terms of both size and power. Finally, both the DS and the SB procedures showed good performance considering broken marginal processes when regressors are nonstationary.

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

Finally, the findings in this chapter suggest some further developments. First, it may be interesting to extend the framework considered in this chapter to identify breaks affecting the slope coefficients. While this should not complicate the inference for the SB procedure, the computational cost may increase substantially for the DS procedure due to additional T dummies for each regressor. Second, it will be useful to compare the performance of the DS with breaks in mean and trend estimated using *Autometrics* with the robust dummy saturation estimator proposed in Johansen and Nielsen (2009),

based on an M-estimator with a bias-corrected variance term. Third, in the light of our empirical results on the Fisher effect and considering the profound instabilities experienced over the last decades, detecting and modelling structural breaks may also be important to validate other well established economic relationships. The robust procedures considered in this chapter certainly help to undertake correct inference. This is part of an ongoing research agenda.

Appendix 2.A *Autometrics*

The aim of this appendix is to give a sketch of the algorithm constituting *Autometrics* . Very briefly, *Autometrics* allows the empirical modeller to easily apply the *"general-to-specific"* approach or *"LSE approach"* advocated by David Hendry. As a software, *Autometrics* is included in OxMetrics^{TM} and is the evolution of *PcGets* a computer program developed by David Hendry and Hans-Martin Krolzig (Hendry and Krozlig, 1999, 2005) which builds on Hoover and Perez (1999). The main references used in this short overview are Doornik (2009a,b).

The Algorithm

Starting from a *general unrestricted model* (GUM) including all the explanatory variables that the modeller believes to potentially matter, *Autometrics* is able to select a final model through a tree search algorithm and selecting the relevant variables according to a battery of misspecification tests as well as individual significance tests (t-tests) on the candidate regressors.

The selection of the final model is performed using a reduction p-value (p_α) that can be modified by the user. Moreover, *Autometrics* is able to deal with more variables than observations $N > T$ (unidentified GUM) as well as with non-orthogonal candidate regressors. If $N > T$ then a block search algorithm kicks in. When working with the DS, this is the case of interest and so here we analyse the general version of the algorithm incorporating the block search. The standard case with $N < T$ is a special case of the algorithm with the block search.

Following Doornik (2009b), first we introduce the block-splitting component of the algorithm, then we present the overall algorithm. In particular, inside the block-search algorithm all the N variables entering the GUM, $\bar{\mathcal{B}}$, are split in two sets: the *selected* variables at iteration i, denoted $\mathcal{S}_i$, and the *excluded* variables, denoted $\mathcal{B}^0 = \bar{\mathcal{B}} \setminus \mathcal{S}_i$. The excluded set is partitioned in blocks and two steps alternate in succession:

1. *Expansion step*: partition the excluded variables in blocks $(\mathcal{B}_1^0, \ldots, \mathcal{B}_B^0)$ and run all over the blocks $\mathcal{B}_1^0 \cup \mathcal{S}_i, \ldots, \mathcal{B}_{B^0}^0 \cup \mathcal{S}_i$ to look for omitted variables ($\mathcal{O}_i$). This is an iterative step and it stops when the number of regressors selected from the initial set of excluded variables is small enough.

2. *Reduction step*: find a new candidate set $\mathcal{S}_{i+1}$ from model selection on $\mathcal{S}_i \cup \mathcal{O}_i$.

As explained above, the selection of the relevant variables is governed throughout by a p-value that can be set by the user. However, a temporary increase of the p-value in the reduction step is used to improve the sensitivity of *Autometrics* at the cost of slightly increasing the risk of overfitting. For exposure

clarity then define the following stages:

stage A: starts from an empty model, and is run only once.

stage B: is run until convergence, using p_α for the expansion step.

stage C: is run only once, using $2p_\alpha$ for the expansion step.

stage D: starts with $4p_\alpha$ for the expansion step, using p_α thereafter. When stage D converges, the block search terminates.

The *Autometrics* algorithm then runs as follows:

Set $i = 0, \mathcal{S}_{i-1} = \emptyset,$ stage $= A$.

1. *Expansion step* to find $\mathcal{O}_i$;
2. *Reduction step* to find $\mathcal{S}_i$;
3. Change the stage:
   - 3a. **if** stage is A set stage = B and go to *Continuation*;
   - 3b. **if** stage is C set stage = D and go to *Continuation*;
   - 3c. **else** go to *Convergence* step;
4. *Convergence* **if** $\mathcal{S}_0 \cup \ldots \mathcal{S}_i = \mathcal{S}_0 \cup \ldots \mathcal{S}_{i-1}$ **then**
   - 4a. **if** stage is B increment it to C,
   - 4b. **else** terminate block search.
5. *Continuation* increment i and return to Step 1.

To improve on efficiency, the way in which the algorithm searches for significant variables is a tree-search type procedure where redundant branches are skipped. In the case there are multiple terminal models, *Autometrics* forms the final GUM as the union of these. Finally, the overall final model is chosen using the Schwarz criterion.

Additionally, there are different ways in which the blocks may be formed in the *Expansion* step. This option, together with the block size, can be set by the user. In particular, together with the *standard* block method where the blocks are formed sequentially, other options are the *random* blocks and two variants where the algorithms also searches more extensively crossing the standard blocks. However, as also noted in Doornik (2009b), the overall algorithm is sensible to the ordering of the variables so a different way to constitute the blocks may lead to slightly different outputs as we observe in our simulation exercise.

For further details on the performances and the options of *Autometrics* , we invite the interested reader to refer to Doornik (2009a) and Doornik (2009b).

Appendix 2.B SB Algorithm

We report in the box below a pseudo-code version to better summarise each step of the SB and we refer the interested reader to De Peretti and Urga (2005) for further details on the procedure.

a. Set $j = 1$ and $i = 0$:

   1. Estimate the j^{th}-break date as:

$$\hat{T}_j = \text{argmax}\,\{F_t\} \quad t = 3, \ldots, T-1$$

      where F_t is the F-test statistic to test jointly the non-significance of the coefficients of the dummies capturing a potential break occurring at t.

   2. Test the significance of $\hat{T}_j$ ($H_0 : j-1$ breaks) computing the bootstrapped p-value

$$\frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\{(\tau_b^*)^2 > (\hat{\tau}_j)^2\}$$

      where $\hat{\tau}_j = \max\{F_t\}$ and τ_b^* is the bootstrapped counterpart.

   3. If significant, re-estimate the previous $j-1$ breaks otherwise $i = i + 1$.

b. Set $j = j + 1$ and repeat steps 1-3. Stop when two consecutive break dates are not significant ($i = 2$).

c. Repeat for all the marginals and then impose the breaks in the conditional process before looking for its own breaks.

Table 2.1: Marginal DGPs.

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

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

Table 2.2: Correct Number of Breaks Detection for Marginal Processes.

		IIS		SS		SS			SB (param.)		SB(nonparam)		
DGP	T_i	Gauge	Potency	Gauge	Potency	Size	Power	NCB	Size	Power	Size	Power	Dummies' coeff.
LS	-	1.6	-	1.6	-	48.5	-	-	1.4	-	1.5	-	
		24.1	-	18.5	-	100.0	-	-	4.5	-	5.2	-	
	70	1.6	98.7	1.2	100.0	41.4	100.0	58.6	1.2	100.0	1.0	100.0	$\upsilon_1 = 5$
		19.1	99.9	17.6	100.0	99.5	100.0	0.5	5.3	100.0	5.4	100.0	
	20, 40	0.5	49.9	0.9	98.8	35.8	100.0	64.2	1.5	99.9	1.1	100.0	$\upsilon_1 = 3, \upsilon_2 = -3$
		16.4	91.2	20.6	99.2	98.9	100.0	1.1	5.4	100.0	5.0	100.0	
ARs	-	1.6	-	3.4	-	65.3	-	-	1.2	-	0.7	-	-
		25.7	-	24.6	-	100.0	-	-	3.5	-	3.6	-	
	70	1.7	19.2	2.5	100.0	59.1	100.0	40.9	1.3	99.4	0.9	98.4	$\upsilon_1 = 5$
		22.7	35.4	23.0	100.0	99.9	100.0	0.1	5.4	100.0	5.8	100.0	
	20, 40	2.1	14.7	1.1	83.1	39.1	99.4	60.9	1.6	99.7	0.9	99.1	$\upsilon_1 = 5, \upsilon_2 = -5$
		24.3	44.0	19.7	96.7	99.2	100.0	0.8	6.0	99.9	5.3	99.9	
	20, 30	1.2	81.9	1.3	99.8	44.3	100.0	55.7	1.2	99.9	1.2	99.9	$\upsilon_1 = 5, \upsilon_2 = -5$
		18.2	90.8	21.0	99.6	99.8	99.9	0.1	4.9	99.9	5.8	99.9	
ARst	-	1.6	-	3.4	-	68.9	-	-	0.9	-	1.3	-	-
		27.7	-	21.3	-	100.0	-	-	5.7	-	5.3	-	
	70	1.7	17.7	2.4	100.0	60.7	100.0	39.3	0.8	99.1	0.9	98.1	$\upsilon_1 = 5$
		24.2	37.9	20.1	100.0	99.8	100.0	0.2	5.2	100.0	5.4	100.0	
	20, 40	2.0	14.6	1.2	75.1	40.5	98.2	59.5	1.3	96.8	0.8	95.8	$\upsilon_1 = 5, \upsilon_2 = -5$
		26.2	45.3	17.1	95.4	99.3	99.9	0.7	5.5	99.3	5.2	99.2	
	20, 30	1.3	83.3	1.3	99.6	46.1	99.9	53.9	1.0	95.1	1.0	95.1	$\upsilon_1 = 5, \upsilon_2 = -5$
		18.5	91.2	18.2	99.4	99.8	100.0	0.2	5.0	99.4	5.4	99.3	
ARns	-	1.6	-	3.4	-	74.2	-	-	2.0	-	1.6	-	-
		26.1	-	23.8	-	100.0	-	-	7.4	-	7.6	-	
	70	0.7	38.8	1.3	100.0	44.0	100.0	56.0	2.8	100.0	2.5	100.0	$\upsilon_1 = 5$
		17.7	66.3	19.6	99.8	99.7	100.0	0.3	10.7	100.0	11.6	100.0	
	20, 40	13.4	73.1	0.9	98.6	37.0	100.0	63.0	3.0	99.2	3.4	99.3	$\upsilon_1 = 3, \upsilon_2 = -5$
		28.5	82.7	17.7	99.1	98.6	100.0	1.4	13.0	99.9	12.6	100.0	
	20, 30	4.8	83.7	1.1	99.7	44.0	100.0	56.0	3.3	94.5	3.3	93.2	$\upsilon_1 = 3, \upsilon_2 = -5$
		21.9	91.8	19.1	99.4	99.6	100.0	0.4	12.0	99.4	13.5	99.4	

Notes: The specification of the processes is given in Table 2.1. The numbers under the columns "gauge" and "potency" represent the empirical retention frequencies of the null and non-null dummies respectively according to the DGP, while the numbers under the columns labelled "size" and "power" are the empirical rejection frequencies of null hypothesis about the number of breaks. For a detail of the null hypotheses, see the end of Section 2.4.1. Gauge and potency as well as size and power for the SS are computed only with reference to the step dummies. "NCB" denotes the percentage of times a procedure detects the correct number of structural breaks according to the DGP. Finally, "-" indicates the no break case and thus only gauge and size can be calculated.

Table 2.3: Break Dates Estimates for Marginal Models.

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

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

Table 2.4: Conditional DGPs.

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

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

Table 2.5: Correct Number of Breaks Detection for Conditional Models.

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

Notes: The specification of the processes is given in Table 2.4. The numbers in brackets in the column "T_i" represent the break dates affecting the marginal process. The numbers under the columns "gauge" and "potency" represent the empirical retention frequencies of the null and non-null dummies respectively according to the DGP, while "size" and "power" represent the empirical rejection frequencies of null hypothesis about the number of breaks. For a detail of the null hypotheses, see the end of Section 2.4.1. Gauge and potency for the SS are computed only with reference to the step dummies. Finally, "-" indicates the no break case and thus only gauge and size can be calculated.

Table 2.6: Break Date Estimates for Conditional Models.

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

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

Table 2.7: Marginal and Conditional DGPs with Broken Linear Trend.

DGP name	
ARsTr:	$x_t = 0.2 + 0.05t + 0.6x_{t-1} + \sum_{i=1}^{K}(\upsilon_i S_{i,t} + \phi_i L_{i,t}) + u_t$
ARnsTr:	$x_t = 0.2 + 0.01t + x_{t-1} + \sum_{i=1}^{K}(\upsilon_i S_{i,t} + \phi_i L_{i,t}) + u_t$
CsTr:	$y_t = 0.2 + 0.05t + 0.5x_t + \sum_{i=1}^{K}(\psi_i S_{i,t} + \omega_i L_{i,t}) + u_t$
	$x_t = 0.6x_{t-1} + e_t$
CciTr:	$y_t = 0.2 + 0.02t + 0.5x_t + \sum_{i=1}^{K}(\psi_i S_{i,t} + \omega_i L_{i,t}) + u_t$
	$x_t = x_{t-1} + e_t$
CMsTr:	$y_t = 0.2 + 0.05t + 0.5x_t + \sum_{i=1}^{K}(\psi_i S_{i,t} + \omega_i L_{i,t}) + u_t$
	$x_t = 0.5 + 0.02t + 0.6x_{t-1} + \upsilon S_t^x + \phi L_t^x + e_t$
CMciTr:	$y_t = 0.2 + 0.01t + 0.5x_t + \sum_{i=1}^{K}(\psi_i S_{i,t} + \omega_i L_{i,t}) + u_t$
	$x_t = 0.5 + 0.001t + x_{t-1} + \upsilon S_t^x + \phi L_t^x + e_t$

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

Table 2.8: Correct Number of Breaks Detection for Marginal and Conditional Models with Broken Trend.

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

Notes: The specification of the processes is given in Table 2.7. The numbers in brackets in the column "T_i" represent the break dates affecting the marginal process. The numbers under the columns "gauge" and "potency" represent the empirical retention frequencies of the null and non-null dummies respectively according to the DGP, while the numbers under the columns denoted "size" and "power" are the empirical rejection frequencies of null hypothesis about the number of breaks. For a detail of the null hypotheses, see the end of Section 2.4.1. Gauge and potency for the US are computed only with reference to the step and trend dummies. Finally, "-" indicates the no break case and thus only gauge and size can be calculated.

Table 2.9: Break Dates Estimates for Marginal and Conditional Models with Broken Trend.

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

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

Table 2.10: US Procedure: Final Selected Model for the Fisher Relationship.

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

Note: I:YYYY(Q) indicates an impulse dummy ($I_{i,t} = \mathbf{1}(t = \text{YYYY(Q)})$), S:YYYY(Q) a step dummy ($S_{i,t} = \mathbf{1}(t \geq \text{YYYY(Q)})$) and L:YYYY(Q) a trend dummy ($L_{i,t} = (t - \text{YYYY(Q)} + 1)\mathbf{1}(t \geq \text{YYYY(Q)})$).

Table 2.11: Break Dates Selected by the SB Procedure.

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

Notes: The bold dates represent the statistically significant structural breaks. ** indicates significance at 1% while * significance at 5%. Both parametric (P) and nonparametric (NP) p-values are computed according to 99 bootstrap replications. For each significant break date a step and trend dummies are created in order to estimate the final model reported in Table 2.12.

Table 2.12: SB Procedure: Final Selected Model for the Fisher Relationship.

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

Note: I:YYYY(Q) indicates an impulse dummy ($I_{i,t} = \mathbf{1}(t = \text{YYYY(Q)})$), S:YYYY(Q) a step dummy ($S_{i,t} = \mathbf{1}(t \geq \text{YYYY(Q)})$) and L:YYYY(Q) a trend dummy ($L_{i,t} = (t - \text{YYYY(Q)} + 1)\mathbf{1}(t \geq \text{YYYY(Q)})$).

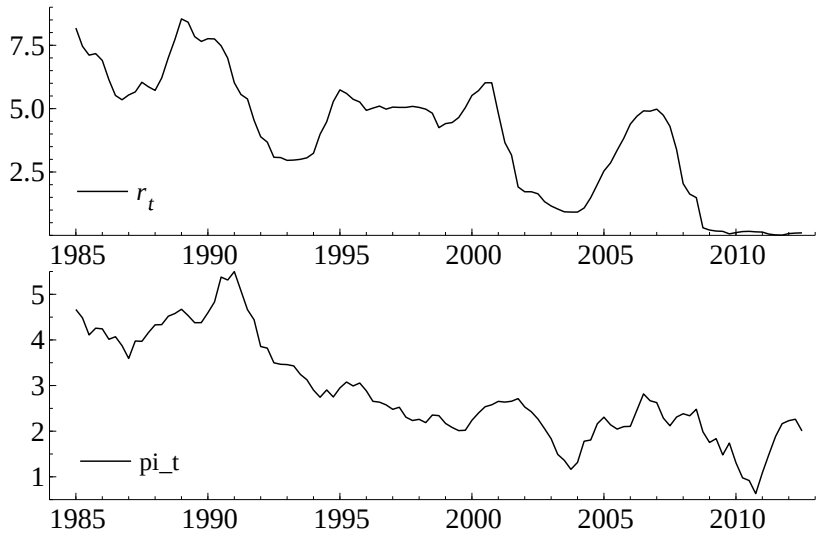


Figure 2.1: Three months Treasury Bill rate (r_t) and inflation rate (π_t).

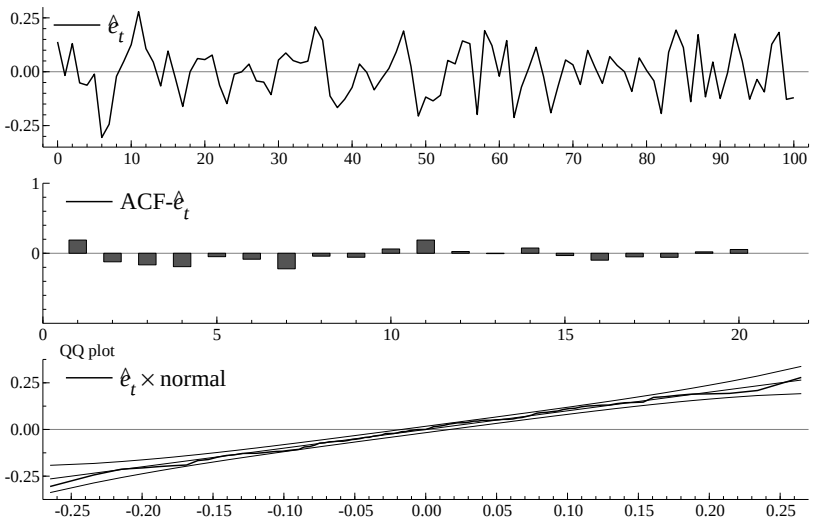


Figure 2.2: US procedure final selected model plots. The first plot reports the residuals $\hat{e}_t$ time series. The second plot reports the autocorrelation function of $\hat{e}_t$ up to the 20^{th} lag. The third plot reports the QQ-plot of $\hat{e}_t$ against a standard normal.

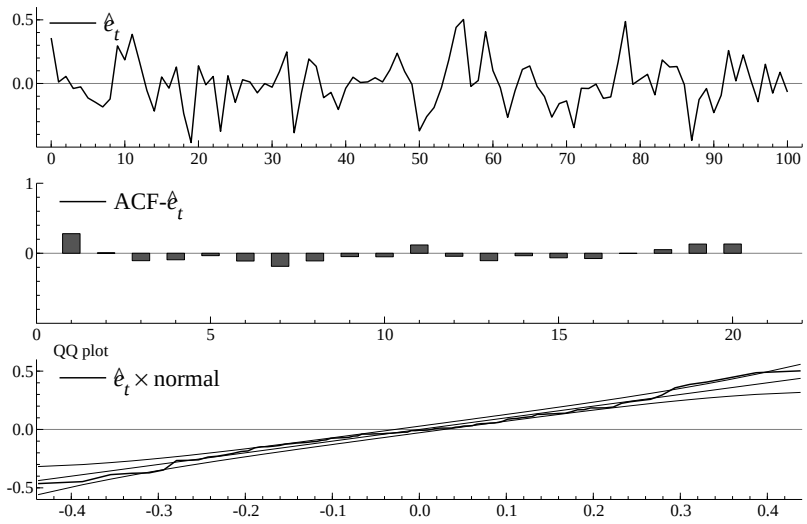


Figure 2.3: SB procedure final selected model plots. The first plot reports the residuals $\hat{e}_t$ time series. The second plot reports the autocorrelation function of $\hat{e}_t$ up to the 20^{th} lag. The third plot reports the QQ-plot of $\hat{e}_t$ against a standard normal.

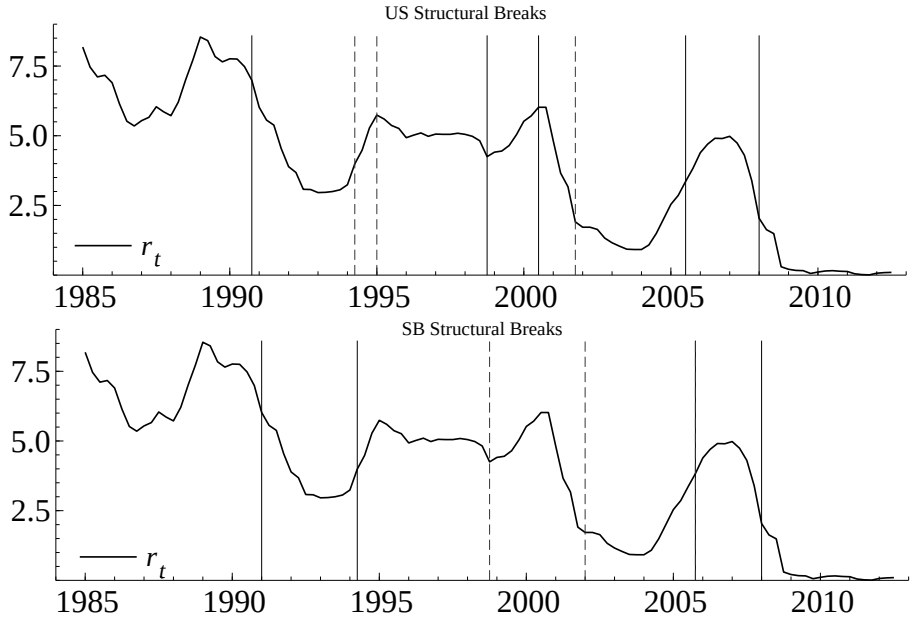


Figure 2.4: Structural Breaks found by the US and the SB in the r_t time series. Solid lines denote breaks in level while dashed lines denote breaks in trend.

Chapter 3

Testing for Multiple Breaks in the VECM Framework*

3.1 Introduction

In the analyses of long run relations, structural stability raises enduring concerns. This chapter focuses on multiple structural breaks in the likelihood based Vector Error Correction Model (VECM). To set focus, given a p-dimensional vector X_t, consider the regression of ΔX_t on X_{t-1}, and e.g. a constant and further lags of ΔX_t. Let Π refer to the coefficient of X_{t-1} in the latter regression. The long run relation is defined in this context via a reduced rank restriction of the form $\Pi = \alpha\beta^\top$, where α and β are the $(p\times r)$ matrices of short and long-run parameters and r refers to the cointegration rank. This chapter focuses on assessing breaks in both in the long-run

*A research paper joint with my PhD supervisor, Prof. Giovanni Urga, and with Prof. Lynda Khalaf entitled *"Multiple Testing and Stability in Reduced Rank Non-Stationary Regressions"* (2014) is based on the results in this chapter and it has been submitted for publication. The paper has been presented at the 13th OxMetrics User Conference (CREATES, Århus, 5-6 September 2013), at the IX New York Econometrics Camp (Watkins Glen, 3-4 April 2014), and at the Oxford Econometrics Conference (Oxford University, 1-2 September, 2014).

matrix β and the adjustment matrix α.

The literature on structural breaks in multivariate time series is relatively scarse. Qu and Perron (2007) provide a fairly general quasi-maximum likelihood (QML) framework to estimate and test for multiple unknown structural breaks in systems with stationary regressors. In the multivariate non-stationary setting, the existing literature deals mainly with one break. Hansen (1992) studies the asymptotic properties of tests for a single shift in a fully modified OLS (FM-OLS) framework while Bai et al. (1998) focus on the QML break date estimate. Bernard et al. (2007) provide a finite-sample corrected version of the tests in Bai et al. (1998). More recently and within a non-stationary system of equations similar to Qu and Perron (2007), Oka and Perron (2011) analyze the QML estimates of common (cross-equation) multiple breaks overlooking questions concerning their number. In a single cointegrating equation framework, Kejriwal and Perron (2008, 2010) provide a comprehensive treatment of multiple unknown changes. Through an extensive simulation analysis, Bergamelli and Urga (2013) explore the ability of the "dummy saturation" and the sequential bootstrapping of the sup-F test to detect the correct number of breaks and to correctly locate them.

In the cointegrated vector autoregressive model (VECM) framework of Johansen (1988), Quintos (1997) and Hansen and Johansen (1999) consider fluctuation tests for parameter stability while Seo (1998) considers Lagrange multipliers (LM) tests (Ave-LM, Exp-LM and Sup-LM as in Andrews and Ploberger, 1994) for a single unknown shift in the cointegrating vector or/and in the adjustment vector. Hansen (2003) proposes the so called generalised reduced rank regression framework where the parameters of the VECM

process are allowed to experience multiple structural changes. The Hansen (2003) set-up is very general in that all the parameters (adjustment matrix, cointegrating matrix, coefficients of the deterministic components, variance terms) are allowed to change, though by imposing adequate restrictions it is also possible to focus on specific components. The main limitation of Hansen (2003) is that the break dates are assumed to be known, which restricts the empirical relevance of this framework.

In this context, the contribution of this chapter is twofold. We first study the interaction of weak exogeneity with assumptions on breaks in each of the components of Π. We also discuss the implication of such breaks on the stability of variance parameters. In particular, we show that structural breaks in β entail, *a fortiori*, breaks in α, unless weak exogeneity is imposed and maintained. If weak exogeneity is not acceptable *a priori*, we suggest to test for breaks in α and β jointly to avoid size distortions arising from misspecifications. Non-separability of breaks in these parameters from breaks-in-variance is also shown for specific parametrizations. The underlying intuition for such interactions stems from the reduced rank restrictions which validate the model under the null and alternative hypotheses. We provide an analytical treatment of these problems, which seem to have escaped notice in available related works. Further, we provide numerical evidence in order to strengthen our theoretical conclusions.

Second, we propose a new test where the multiple breaks detection in the VECM framework is conducted without the assumption of knowledge of the break dates, thus generalising the Hansen (2003) testing procedure. We propose to model instability by letting the researcher suggest a certain number of scenarios (say n) to describe multiple potential breaks. The scenarios

in question allow one to specify several hypotheses about both the number and the location of the breaks. For instance, if we define m_n to be the number of regimes in each scenario (which implies $m_n - 1$ breaks), it is possible to build a set of n likelihood ratio (LR) tests, one for each scenario, to decide on the number of potential regimes. The ensuing simultaneous hypotheses are thus tested jointly via a minimum p-value approach corresponding to the least favourable scenario under the null model. This translates into the definition of a minimum p-value statistic with distribution approximated through various bootstrapping procedures. The advantage of this approach is that the minimum p-value statistic is derived from n LR tests which follow standard limiting distributions while break dates are not restricted to be known. In contrast, we presume that possible breaks can be broadly characterized so that n plausible scenarios can adequately express uncertainties about their number and location. In an economic environment, this seem a realistic specification.

The finite-sample properties of the minimum p-value statistic are investigated through an extensive Monte Carlo study under different combinations of breaking times and sample lengths, imposing and relaxing weak exogeneity and considering different degrees of identification of the cointegrating relationship. Further, in assessing the finite sample properties of the proposed combined statistic, we consider DGPs that experience breaks in the covariance matrix (resulting from breaks in the components of Π) and we provide bootstrapping procedures that allow to restore the correct size nevertheless. The simulation study is conducted under the null hypothesis of no breaks as well as under the more challenging scenario of multiple breaks.

The remainder of the chapter is organised as follows: Section 3.2.1 briefly summarises the formulation and the estimation of the VECM in presence of breaks based on Hansen (2003). Section 3.3 deals with the role of weak exogeneity in testing for structural breaks in the VECM framework and derives the theoretical result. Section 3.4 introduces the new test, based on a minimum p-value approach, and the related bootstraps procedures. Section 3.5 reports the Monte Carlo simulations exercise whereas Section 3.6 reports three empirical illustrations. Section 3.7 concludes.

3.2 Formulation and Estimation of the VECM in Presence of Multiple Breaks

3.2.1 Model Formulation

The modelling framework is the error correction formulation of a cointegrated VECM affected by multiple structural breaks as in Hansen (2003).

Let $\{X_t\}_{t=1}^{T}$ be a p-dimensional process undergoing m regimes and thus affected by $m-1$ break dates, denoted $T_0 = 0 < T_1 < \cdots < T_{m-1} < T_m = T$. A general VECM formulation where all parameters are allowed to change may take the following form

$$\Delta X_t = \alpha(t)\beta(t)^{\top}\ddot{X}_{t-1} + \sum_{i=1}^{k-1}\Gamma_i(t)\Delta X_{t-i} + \Phi(t)D_t + \varepsilon_t, \quad t = 1, \ldots, T \tag{3.1}$$

where $\{\varepsilon_t\}$ is a sequence of independent Gaussian random variables 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 vari-

ables while D_t is a q-dimensional vector of unrestricted deterministic variables. Defining r_j, for $j = 1, \ldots, m$, the cointegrating rank of the j^{th} regime (implying that the cointegrating rank may vary across the different regimes), 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 $(p \times p)$. The time-varying parameters are piecewise constant given by

$$
\begin{aligned}
&\alpha(t)\beta(t)^\top = \alpha_1\beta_1^\top \mathbf{1}_{1t} + \cdots + \alpha_m\beta_m^\top \mathbf{1}_{mt}, \\
&\Gamma_i(t) = \Gamma_{1,i}\mathbf{1}_{1t} + \ldots \Gamma_{m,i}\mathbf{1}_{mt}, \quad i = 1, \ldots, k-1, \\
&\Phi(t) = \Phi_1\mathbf{1}_{1t} + \cdots + \Phi_m\mathbf{1}_{mt} \\
&\Omega(t) = \Omega_1\mathbf{1}_{1t} + \cdots + \Omega_m\mathbf{1}_{mt}
\end{aligned}
$$

where for each of the m subsamples the corresponding indicator function is defined as

$$
\mathbf{1}_{jt} \equiv \mathbf{1}(T_{j-1} + 1 \leq t \leq T_j), \quad j = 1, \ldots, m. \tag{3.2}
$$

In the same spirit of Johansen (1991), it is possible to rewrite (3.1) in compact form by defining the variables $Z_{0t} = \Delta X_t$, $Z_{1t} = (\mathbf{1}_{1t}\ddot{X}_{t-1}^\top, \ldots, \mathbf{1}_{mt}\ddot{X}_{t-1}^\top)^\top$, $\tilde{Z}_{2t} = (\Delta X_{t-1}^\top, \ldots, \Delta X_{t-k+1}^\top, D_t^\top)^\top$, and $Z_{2t} = (\mathbf{1}_{1t}\tilde{Z}_{2t}^\top, \ldots, \mathbf{1}_{mt}\tilde{Z}_{2t}^\top)^\top$. Z_{1t} is of order $(mp_1 \times 1)$ while, if we denote with p_2 the number of variables in $\tilde{Z}_{2t}$, the dimension of Z_{2t} is $(mp_2 \times 1)$. We may group the parameters accordingly

by defining

$$A = (\alpha_1, \ldots, \alpha_m)$$

$$B = \text{diag}(\beta_1, \ldots, \beta_m) = \begin{pmatrix} \beta_1 & 0 & \ldots & 0 & 0 \\ 0 & \beta_2 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & \beta_{m-1} & 0 \\ 0 & 0 & \ldots & 0 & \beta_m \end{pmatrix}$$

$$C = (\Psi_1, \ldots, \Psi_m), \quad \text{where} \quad \Psi_j = (\Gamma_{j,1}, \ldots, \Gamma_{j,k-1}, \Phi_j), \quad j = 1, \ldots, m.$$

In this way, it is possible to cast (3.1) in a reduced rank regression framework to obtain

$$Z_{0t} = AB^\top Z_{1t} + CZ_{2t} + \varepsilon_t, \quad t = 1, \ldots, T \tag{3.3}$$

which has constant parameters.

As explained in Hansen (2003), the required structure of the parameters, their identification as well as hypotheses about the presence of structural breaks can be obtained through restrictions of the form

$$\text{vec}(B) = H\phi + h \tag{3.4}$$

$$\text{vec}(A, C) = G\psi \tag{3.5}$$

where vec$(\cdot)$ is the vectorization operator, 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 parameters and similarly G is a known $[p(r_1 + \cdots + r_m + mp_2) \times p_\psi]$ matrix, while ψ is a vector with p_ψ parameters that for convenience can be partitioned as $\psi = (\psi_A^\top, \psi_C^\top)^\top$. The combination of (3.3) with (3.4) and (3.5)

gives rise to a constrained reduced rank problem.

Note that the above notation is very flexible in that it allows to formulate changes in the parameters (adjustment vector, cointegrating vector, coefficients of the deterministic components, variance terms) and at the same time, by imposing the adequate restrictions, allows to explore changes affecting only a specific component (partial structural changes). Furthermore, the cointegrating rank is not restricted to be the same over the whole sample but it can vary across the different regimes.

3.2.2 Estimation

For the purpose of estimation, Hansen (2003) introduces the so-called *generalized reduced rank regression* (GRRR) technique. GRRR is a likelihood based iterative method which aims at maximising the following log-likelihood function

$$\ell(\phi,\psi,\omega) = -\frac{Tp}{2}\log 2\pi - \frac{T}{2}\sum_{j=1}^{m}\rho_j \log|\Omega_j(\omega)|$$
$$-\frac{1}{2}\sum_{j=1}^{m}\sum_{T_{j-1}+1}^{T_j}(Z_{0t}-A(\psi_A)B(\phi)^\top Z_{1t}-C(\psi_C)Z_{2t})^\top \Omega_j(\omega)^{-1}(Z_{0t}-A(\psi_A)B(\phi)^\top Z_{1t}-C(\psi_C)Z_{2t}) \tag{3.6}$$

where ω collects the parameters which characterise $\Omega_{j=1,\dots,m}$ and $\rho_j = (T_j - T_{j-1})/T$. In practice, the methodology, based on Boswijk (1995), allows to impose restrictions also on the C matrix and it consists in maximizing iteratively the following three equations, starting from $\phi^{(0)}$, $\psi^{(0)}$ and $\omega^{(0)}$:

1. $\hat{\phi}^{(n)} = \arg\max \ell(\phi, \psi^{(n-1)}, \omega^{(n-1)})$
2. $\hat{\psi}^{(n)} = \arg\max \ell(\phi^{(n-1)}, \psi, \omega^{(n-1)})$

3. $\hat{\omega}^{(n)} = \arg\max \ell(\phi^{(n-1)}, \psi^{(n-1)}, \omega)$

for $n \geq 1$. Thus, the algorithm allows to split the maximisation problem in smaller ones which should be easier to solve. Algorithms of this kind are usually referred as *switching algorithms* due to their nature of optimizing with respect to a subset of the parameters and then moving to next subset (see Johansen, 1995).

3.2.2.1 Concentrated Log-likelihood

It is possible to further simplify the maximisation problem concentrating (3.6) with respect to Ω_j. In particular, given that

$$\hat{\Omega}^{ML}_{j=1,\dots,m} = (T_j - T_{j-1})^{-1} \sum_{t=T_{j-1}+1}^{T_j} \hat{\varepsilon}_t \hat{\varepsilon}_t^\top$$

with $\hat{\varepsilon}_t = Z_{0t} - \hat{A}(\hat{\psi}_A)\hat{B}(\hat{\phi})^\top Z_{1t} - \hat{C}(\hat{\psi}_C)Z_{2t}$, we can find numerically the maximum likelihood estimators of the remaining parameters choosing the values of $\hat{\phi}$, and $\hat{\psi} = (\hat{\psi}_A^\top, \hat{\psi}_C^\top)^\top$ which maximise

$$\ell_c(\hat{\phi}, \hat{\psi}) \propto -\frac{T}{2} \sum_{j=1}^{m} \rho_j \log \left| \hat{\Omega}^{ML}_j(\hat{\phi}, \hat{\psi}) \right| \tag{3.7}$$

where the above concentrated log-likelihood function is obtained substituting the expression for $\hat{\Omega}^{ML}_j$ in (3.6) and leaving out the resulting constant term. A derivation of (3.7) is reported in Appendix A. The estimator of the possibly time-varying covariance matrix $\hat{\Omega}(t) = \hat{\Omega}_{j=1,\dots,m}$ is then computed from the properly grouped residuals $\{\hat{\varepsilon}_t\}_{t=T_{j-1}+1}^{T_j}$. Notice that, in the case we want to apply the iterative algorithm described above, it is sufficient to iterate on the first two equations only.

3.3 The Role of Weak Exogeneity in Testing for Breaks

In this section, we explore the role of weak exogeneity in testing for breaks in the cointegrating matrix Π. We also report numerical evidence showing the size distortions that might arise in testing for breaks if ignoring the implications of weak exogenity.

3.3.1 Hypotheses Specification with Known Break Dates

Consider the full set of parameters $\Theta = \{\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m, \Gamma_{1,1}, \ldots, \Gamma_{1,m}, \ldots, \Phi_1, \ldots, \Phi_m, \Omega_1, \ldots, \Omega_m | \tau_m\} = \{\Theta_1, \ldots, \Theta_m | \tau_m\}$ which defines the VECM process (3.1) undergoing m regimes for a given set of break dates $\tau_m = \{T_1, \ldots, T_{m-1}\}$. Further, define $\Theta^* \subseteq \Theta$ the subset of parameters (*partial* structural changes model) we test for structural breaks. This amounts to split Θ into two components, the first one is $\Theta^* = \{\Theta_i^* \subset \Theta : \Theta_i^* \neq \Theta_{i+1}^* | \tau_m$ for $i = 1, \ldots, m-1\}$ while the second $\tilde{\Theta} \equiv \Theta \setminus \Theta^* = \{\tilde{\Theta}_i \subset \Theta : \tilde{\Theta}_i = \tilde{\Theta}_{i+1} | \tau_m$ for $i = 1, \ldots, m-1$ $\}$. In Section 3.2.1, we explain how it is possible to impose the necessary restrictions to select Θ^* from Θ in order to estimate the VECM. Generally, Θ^* can differ under the null and the alternative hypotheses, i.e. $\Theta_{\mathcal{H}_0}^* \neq \Theta_{\mathcal{H}_1}^*$, thus defining another dimension, together with the number and location of the breaks, along which one may be interested in testing for. However, we now focus our attention on the case where $\Theta_{\mathcal{H}_0}^* \equiv \Theta_{\mathcal{H}_1}^* = \Theta^*$ to test for the presence of breaks. Then, to decide on the number of regimes,

it is plausible to specify the hypotheses of the form

$$\begin{aligned} \mathcal{H}_0: &\quad \Theta_1^* \neq \cdots \neq \Theta_{m_0}^* | \tau_{m_0} \\ \mathcal{H}_1: &\quad \Theta_1^* \neq \cdots \neq \Theta_{m_1}^* | \tau_{m_1} \end{aligned} \tag{3.8}$$

with $m_1 > m_0$ and $\tau_{m_0} \subset \tau_{m_1}$ in order to ensure that the null model is a restricted version of the alternative one. The notation adopted in (3.8) implies that Θ_i^* changes at discrete intervals conditional on the set of break dates τ_m. This implies that for each $t \notin \tau_m$, $\Theta^*(t) = \Theta^*(t-1)$ for $t = 1, \ldots, T$.

In order to accept or reject $\mathcal{H}_0$, the natural tool arising from the framework introduced above consists of standard likelihood ratio (LR) tests. In particular, denote $\mathcal{M}_0$ the restricted model under the null hypothesis of m_0 regimes and $\mathcal{M}_1$ the unrestricted model under the alternative of m_1 regimes, Theorem 10 in Hansen (2003) proves that if the dates of the structural breaks are known and under suitable conditions on the rank of the restriction matrices and assuming both $\mathcal{M}_0$ and $\mathcal{M}_1$ have the same cointegrating rank in each subsample, the stability of the subset of VECM parameters Θ^* can be tested through

$$LR = -2\log\frac{\mathcal{L}_{\max}^{\mathcal{M}_0}}{\mathcal{L}_{\max}^{\mathcal{M}_1}} = T\left[\sum_{j_0=1}^{m_0} \rho_{j_0} \log|\hat{\Omega}_{j_0}| - \sum_{j_1=1}^{m_1} \rho_{j_1} \log|\hat{\Omega}_{j_1}|\right] \sim \chi^2(q) \tag{3.9}$$

where $\mathcal{L}_{\max}$ denotes the maximum value of the likelihood function, $\hat{\Omega}_{j_0}$ and $\hat{\Omega}_{j_1}$ are the sample covariance matrices of the null and the alternative model respectively, and q is the difference between the number of parameters of $\mathcal{M}_1$ and those of $\mathcal{M}_0$. Hence, the requirement that the number of regimes under the alternative, m_1, must be bigger than the number of regimes under

the null, m_0, is necessary to build statistics with non-degenerate distributions.

3.3.2 Role of Weak Exogeneity in Specifying Hypotheses about Structural Breaks

The framework introduced to test for multiple breaks is very general and can accommodate structural breaks affecting either all components of the VECM, Θ, or only a subset of them, $\Theta^* \subseteq \Theta$. In particular, focusing on $\Theta^* = \{\Pi_1, \ldots, \Pi_m | \tau_m\}$, with $\Pi = \alpha\beta^\top$, Hansen (2003) deals with the case in which the breaks affect either only α, $\Theta^* = \{\alpha_1, \ldots, \alpha_m | \tau_m\}$, or only β, $\Theta^* = \{\beta_1, \ldots, \beta_m | \tau_m\}$, or both components, $\Theta^* = \{\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m | \tau_m\}$, and the discussion runs as if it is up to the modeller to choose the structure that best fits the data, using for instance LR tests. A similar analysis is carried out by Seo (1998).

However, using a weak exogeneity argument (see Engle et al., 1983), it can be shown that the choice of which component of the Π matrix is to be tested is indeed limited, due to the link between the short-run impact matrix α and the long-run matrix β. Similarly, it can be shown that structural breaks affecting β impact on the covariance matrix of the error term. These two results are valid under the assumption that cointegration continues to hold also after the break dates, i.e. we do not allow for situations where after the break rank$(\Pi) = 0$. Theorem 1 formalises this point:

Theorem 1. Consider the VECM representation of a p-dimensional process such that $X_t = (X_{1,t}^\top, X_{2,t}^\top)^\top$, where $X_{1,t}$ is of dimension p^* and $X_{2,t}$ of dimension $p - p^*$. If the long-run matrix β experiences different regimes, then this

implies:

a. structural breaks in the adjustment matrix α unless $X_{2,t}$ is weakly exogenous with respect to $X_{1,t}$

b. structural breaks in the covariance matrix of the error term Ω.

Proof. Consider the standard cointegrating DGP adopted in Gonzalo (1994), which follows very closely the DGP used in the seminal contribution of Engle and Granger (1987), and write

$$X_{1,t} - \tilde{\beta}^{\top} X_{2,t} = Z_t; \quad Z_t = \rho Z_{t-1} + u_{1t} \tag{3.10}$$

$$A_1 X_{1,t} - A_2 X_{2,t} = W_t; \quad W_t = W_{t-1} + u_{2,t} \tag{3.11}$$

where $X_{1,t}$ is of dimension $p^* \times 1$ while $X_{2,t}$ is of dimension $(p-p^*) \times 1$. Note that after normalisation of the cointegrating matrix β as $\beta = [I_{p^*} \ -\tilde{\beta}^{\top}]^{\top}$, every cointegrating system can be rewritten as (3.10)-(3.11). The formulation above is a direct application of the concept of cointegration. It simply tells us that if X_t is a unit root vector process and it cointegrates then there may be only one linear combination $\beta^{\top} X_t = X_{1,t} - \tilde{\beta} X_{2,t}$ which is stationary (assuming $|\rho - \lambda I| = 0$ only for $|\lambda_i| < 1$, $i = 1, \ldots, p^*$) while every other combination $\beta^{*\top} X_t = A_1 X_{1,t} - A_2 X_{2,t}$ is not.

Using matrix notation, we can rewrite (3.10)-(3.11) as

$$\begin{bmatrix} I_{p^*} & -\tilde{\beta} \\ A_1 & A_2 \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \rho & 0 \\ 0 & I_{p-p^*} \end{bmatrix} \begin{bmatrix} I_{p^*} & -\tilde{\beta} \\ A_1 & A_2 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$

and after some algebra, we get the VECM representation

$$\begin{bmatrix}\Delta X_{1,t}\\ \Delta X_{2,t}\end{bmatrix} = \begin{bmatrix}(I_{p^*} - \tilde{\beta}^\top D^{-1} A_1)(\rho - I_{p^*})\\ D^{-1}A_1(\rho - I_{p^*})\end{bmatrix} \begin{bmatrix} I_{p^*} & -\tilde{\beta}^\top \end{bmatrix} \begin{bmatrix} X_{1,t-1}\\ X_{2,t-1}\end{bmatrix} + \\ + \begin{bmatrix} u_{1t} - \tilde{\beta}^\top D^{-1}(A_1 u_{1t} - u_{2t})\\ -D^{-1}(A_1 u_{1t} - u_{2t})\end{bmatrix} \quad (3.12)$$

with $D = A_1\tilde{\beta}^\top - A_2$. The process in (3.12) can be reparametrised in the usual way, and under the additional assumption of gaussian i.i.d. errors, we have

$$\begin{bmatrix}\Delta X_{1t}\\ \Delta X_{2t}\end{bmatrix} = \begin{bmatrix}\alpha_{p^*}\\ \alpha_{p-p^*}\end{bmatrix} \beta^\top X_{t-1} + \begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{bmatrix} \quad \begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{bmatrix}\right). \quad (3.13)$$

In this more general setting, we have that X_{2t} is weakly exogenous if and only if A_1 is a zero matrix which clearly implies $\alpha_{p-p^*} = 0$ in (3.13). The proof of part *a)* follows by noting that if β changes across m regimes, i.e $\beta(t) = \beta_1^\top \mathbf{1}_{1t} + \cdots + \beta_m^\top \mathbf{1}_{mt}$ then

$$\alpha = \begin{bmatrix}(I_{p^*} - \tilde{\beta}(t)^\top D(t)^{-1} A_1)(\rho - I_{p^*})\\ D(t)^{-1}A_1(\rho - I_{p^*})\end{bmatrix} = \alpha(t) = \alpha_1 \mathbf{1}_{1t} + \cdots + \alpha_m \mathbf{1}_{mt}.$$

Thus a change in β implies a change in α. The only exception is when $X_{2,t}$ is weak exogenous with respect to $X_{1,t}$. As explained above, this happens

when $\alpha_{p-p^*} = 0 \Leftrightarrow A_1 = 0$ and implies

$$\alpha = \begin{bmatrix} \rho - \mathbf{I}_{p^*} \\ 0 \end{bmatrix}$$

which is independent of β. In this case, a potential change in α would occur only if the autoregressive structure of the error correction term Z_t undergoes more than one regime.

For part *b)*, note that

$$\begin{aligned} \Omega &= \mathbb{E}\left(\begin{bmatrix} u_{1t} - \tilde{\beta(t)}^\top D(t)^{-1}(A_1 u_{1t} - u_{2t}) \\ -D(t)^{-1}(A_1 u_{1t} - u_{2t}) \end{bmatrix} \begin{bmatrix} u_{1t} - \tilde{\beta}(t)^\top D(t)^{-1}(A_1 u_{1t} - u_{2t}) \\ -D(t)^{-1}(A_1 u_{1t} - u_{2t}) \end{bmatrix}^\top \right) \\ &= \begin{bmatrix} \Omega_{11}(t) & \Omega_{12}(t) \\ \Omega_{21}(t) & \Omega_{22}(t) \end{bmatrix} \end{aligned}$$

and hence, structural breaks in β imply structural breaks in the covariance matrix of the VECM error term.

■

The results in Theorem 1 are very important for the correct identification of structural breaks. The implications of such results are completely unexplored in the framework proposed in Hansen (2003), where a break in the cointegrating matrix does not imply a break neither in the short term impact matrix nor in the covariance matrix of the VECM. Moreover, we show that ignoring the role of weak exogeneity leads to a misspecification of the subset of parameters Θ^* which is allowed to change and, as we show via the simulation exercise reported in the next section, this has detrimental effects on the size and power of the tests used to decide on the number of structural

breaks affecting the cointegrating matrix β. When weak exogeneity cannot be imposed, a joint test on the stability of both α and β must be performed.

3.3.3 Numerical Evidence

In order to stress the relevance for correct inference of Theorem 1, we undertake a simulation exercise to show the effects of ignoring weak exogenity on the size of a likelihood ratio statistic used for testing m_0 vs m_1 regimes in the cointegrating matrix.

We simulate the following DGP affected by one break at $T_1 = \lfloor T/2 \rfloor$:

$$\begin{aligned} x_{1t} - (\tilde{\beta}_1 \mathbf{1}_{1t} + \tilde{\beta}_2 \mathbf{1}_{2t}) x_{2t} = z_t \quad & z_t = \rho z_{t-1} + e_{z_t} \\ a_1 x_{1t} - a_2 x_{2t} = w_t \quad & w_t = w_{t-1} + e_{w_t} \\ & \begin{bmatrix} e_{z_t} \\ e_{w_t} \end{bmatrix} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I}_2) \end{aligned}$$

with $\mathbf{1}_{1t} = \mathbf{1}(T_0 + 1 \leq t \leq T_1)$ and $\mathbf{1}_{2t} = \mathbf{1}(T_1 + 1 \leq t \leq T)$. We consider the following parameter space: $\tilde{\beta}_1 = 1$, $\tilde{\beta}_2 = \{1.5, 2, 3, 4, 5, 8, 10, 15, 20, 30, 40, 50\}$, $\rho = \{0, 0.8\}$, $a_1 = 1$, $a_2 = -1$, $\sigma = 1$ and $T = \{100, 300, 500\}$. Note that, by setting $a_1 = 1$, we rule out weak exogeneity. Alternatively, the above DGP can be rewritten in VECM form leading to

$$\underbrace{\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix}}_{Z_{0t}} = \underbrace{\begin{bmatrix} \frac{(1-\rho)a_2}{a_1\tilde{\beta}_1 - a_2} & \frac{(1-\rho)a_2}{a_1\tilde{\beta}_2 - a_2} \\ \frac{(1-\rho)a_1}{a_1\tilde{\beta}_1 - a_2} & \frac{(1-\rho)a_1}{a_1\tilde{\beta}_2 - a_2} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 1 & 0 \\ -\tilde{\beta}_1 & 0 \\ 0 & 1 \\ 0 & -\tilde{\beta}_2 \end{bmatrix}^\top}_{B^\top} \underbrace{\begin{bmatrix} \mathbf{1}_{1t} x_{1,t-1} \\ \mathbf{1}_{1t} x_{2,t-1} \\ \mathbf{1}_{2t} x_{1,t-1} \\ \mathbf{1}_{2t} x_{2,t-1} \end{bmatrix}}_{Z_{1t}} + \begin{bmatrix} \frac{-a_2 e_{z_t} + (\mathbf{1}_{1t}\tilde{\beta}_1 + \mathbf{1}_{2t}\tilde{\beta}_2) e_{w_t}}{a_1(\mathbf{1}_{1t}\tilde{\beta}_1 + \mathbf{1}_{2t}\tilde{\beta}_2) - a_2} \\ \frac{-a_1 e_{z_t} + e_{w_t}}{a_1(\mathbf{1}_{1t}\tilde{\beta}_1 + \mathbf{1}_{2t}\tilde{\beta}_2) - a_2} \end{bmatrix}$$

$$\Rightarrow Z_{0t} = \begin{bmatrix} \alpha_1 & \alpha_2 \end{bmatrix} \begin{bmatrix} \beta_1 & \mathbf{0} \\ \mathbf{0} & \beta_2 \end{bmatrix}^{\top} Z_{1t} + \varepsilon_t.$$

We consider two different LR tests of $m = 1$ versus $m = 2$ regimes based on the following two sets of hypotheses

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

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

where "|" indicates conditional on the T_i, $i = 1, 2$ break date and we set $T_1 = \lfloor T/2 \rfloor$ while $T_2 = \lfloor 2T/3 \rfloor$. Both cases involve to estimate under the null hypothesis a model with one break and under the alternative hypothesis a model with two breaks, though for Case A we do not consider possible effects of a break in β on α. This amounts to set $\Theta^*_{A,j} = \{\beta_i \text{ for } i = 1, \ldots, m_j | \tau_{m_j}\}$, $j = 0, 1$ for Case A, and $\Theta^*_{B,j} = \{\alpha_i, \beta_i \text{ for } i = 1, \ldots, m_j | \tau_{m_j}\}$, $j = 0, 1$ for Case B. The hypotheses formulated in Case A, where α is seen as independent of β, can arise implicitly (testing for breaks in β disregarding the effect that these breaks have on α) in the framework of Seo (1998) or Hansen (2003).

[Figure 3.1 about here.]

Figure 3.1 reports the empirical rejection frequencies of the null hypotheses of the two LR tests – Case A and Case B – as function of an increasing magnitude of the break measured as the ratio between the values of $\tilde{\beta}$ post ($\tilde{\beta}_2$) and pre ($\tilde{\beta}_1$) break. The the results are clear-cut. The empirical rejec-

tion frequency of the LR test based on Case B is well controlled and close to the nominal level of 5% for the different sample sizes as well as values of ρ, the parameter which controls the identification of the cointegrtaing relationship. On the other side, the empirical rejection frequencies of the LR test based on Case A increase with the magnitude of the break. This result deserves a careful interpretation. In the Seo (1998) and Hansen (2003) frameworks, this is interpreted as size, thus we would over-reject a "true" null hypothesis concluding that the LR test suffers from size distortions. In empirical applications, this means that we would select a model with more breaks than those actually affecting β. However, in our setting, the rejection of the null $\mathcal{H}_0^a$ has to be interpreted as power since in Case A we are indeed testing a false null hypothesis where the α component is regarded as being not affected by breaks in β while Theorem 1 states that this is not true unless there is weak exogeneity. As often observed, the power is increasing with the sample size and decreasing for higher values of ρ.

In conclusion, if weak exogeneity cannot be imposed, the appropriate testing strategy for deciding on the number of breaks affecting $\Pi = \alpha\beta^\top$ is to jointly test for the presence of breaks in α and β. Focusing exclusively on the long-run component β should be done only if we are sure that weak exogeneity is a reasonable feature of the system. This is quite unlikely to be known in advance especially in the VECM framework.

3.4 The Minimum p-value Statistic (Q-Statistic)

In this section, we extend (3.9) to the case of unknown break dates, given that very rarely the exact number and location of structural breaks are known.

3.4.1 Hypotheses Specification with Unknown Break Dates: The $Q-$Statistic

The knowledge of the breaks under the alternative, τ_{m_1}, is usually uncertain while the assumption of known breaks under the null, τ_{m_0}, is instead reasonable. In order to define a more general testing strategy, we can think of building a test where the alternative hypothesis incorporates the uncertainty about the number and location of the break dates. An extended version of the framework introduced in (3.8) can be written as follows

$$\begin{aligned} \mathcal{H}_0 : \quad & \Theta_1^* \neq \cdots \neq \Theta_{m_0}^* | \tau_{m_0} \\ \mathcal{H}_1 : \quad & \begin{cases} \Theta_{1,1}^* \neq \cdots \neq \Theta_{m_{1,1}}^* | \tau_{m_{1,1}} \\ \vdots \\ \Theta_{1,n}^* \neq \cdots \neq \Theta_{m_{1,n}}^* | \tau_{m_{1,n}} \end{cases} \end{aligned} \tag{3.14}$$

where $\tau_{m_{1,i}}$ for $i = 1, \ldots, n$, contains different breaking dates. The intuition is that *a priori* knowledge of stylised facts and economic policies may suggest a number of n plausible scenarios which may be described by a matrix of the form

$$\Sigma = [\tau_{m_{1,1}}, \ldots, \tau_{m_{1,n}}] = \begin{bmatrix} T_{1,1} & \ldots & T_{1,n} \\ T_{2,1} & \ldots & T_{2,n} \\ \vdots & \ddots & \vdots \\ \vdots & & T_{m_n-1,n} \\ T_{m_1-1,1} & \ldots & 0 \end{bmatrix}$$

where uncertainty about the number of regimes translates in scenarios characterised by different m_i while uncertainty about the dates of the breaks is

expressed through different $T_{i,k}$ for $k = 1, \ldots, m_i - 1$. In order to accept or reject the null hypothesis, we need to map (3.14) into a test statistic. By noting that (3.14) can be decomposed in n LR tests of the form (3.9), we can define the following statistic

Definition 1 (Q-statistic)**.** Let $\mathcal{M}_{1,i}$ denote the model estimated under the i^{th} scenario and $LR_i = -2\log(\mathcal{L}_{\max}^{\mathcal{M}_0}/\mathcal{L}_{\max}^{\mathcal{M}_{1,i}})$ for $i = 1, \ldots, n$ a LR test of $\mathcal{M}_0$ against $\mathcal{M}_{1,i}$ for a given set of break dates $\tau_{m_{1,i}}$ affecting the parameters in $\Theta^* \subseteq \Theta$. To test $\mathcal{H}_0$ versus $\mathcal{H}_1$ in (3.14), we can compute the following minimum p-value statistic

$$Q = 1 - \min_{1 \leq i \leq n} [1 - F(LR_i)] \tag{3.15}$$

where $F(\cdot)$ is the CDF of a $\chi^2(q_i)$ random variable with q_i being the number of more parameters in $\mathcal{M}_{1,i}$ with respect to $\mathcal{M}_0$.

The expression (3.15) defines a new statistic that under the alternative hypothesis allows us to specify different values of m as well as conditioning on different break dates. The Q-statistic we propose makes use of the combination of n individual tests by taking the minimum $p-$value associated to each test. This approach has a long tradition in the statistical literature dating back to Tippett (1931) and recently reconsidered by Dufour et al. (2014). Considering the lowest p-value implies that the decision is determined by the least favourable scenario in terms of the null hypothesis. In addition, (3.15) is based on a general framework in between the case of the exact knowledge of the breaking date and that of a total unawareness of it. For $n = 1$, the standard likelihood ratio test reported in (3.9) holds, whereas by considering all possible combinations of break dates (for some

upper bound on the maximum number of regimes allowed) we can get a test for multiple unknown breaks.

The computation of the Q-statistic is extremely convenient as we know from Hansen (2003) that the individual LR test, being computed conditionally on a set of break dates specified in the i^{th} scenario, is asymptotically distributed as a χ^2. Finally, the asymptotic p-values in (3.15) can be replace by bootstrapped p-values counterpart. For B bootstrap replications, we have that

$$\hat{p}(LR_i) = \frac{\sum_{b=1}^{B} \mathbf{1}(LR_i^* \geq LR_i) + 1}{B+1} \simeq \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}(LR_i^* \geq LR_i) = 1 - \hat{F}^*(LR_i)$$

where we thus replace the CDF $F(\cdot)$ with the empirical distribution function of the bootstrapped samples, $\hat{F}^*(\cdot)$.

3.4.2 Bootstrapping the Distribution of the $Q-$Statistic

The distribution of (3.15) is easy to derive if the n underlying statistics are independent, given that $p(LR_i) = 1 - F(LR_i) \sim \mathcal{U}[0,1]$ under $\mathcal{H}_0$ for all $i = 1, \ldots, n$. In our case, however, the individual likelihood ratios are far from being independent and so to analytically derive the exact distribution of Q is not a trivial task. Thus, in order to generate the critical values of the distribution of $Q-$statistic under the null, we consider a bootstrapping procedure based on resampling from the residuals of the VECM. Given the autoregressive nature of the VECM, a sieve approach as in Chang et al. (2006) to correct for autocorrelation is not needed if we assume a correctly specified lag structure of the VECM. However, our process might be affected by heteroskedasticity induced by the presence of structural breaks,

i.e. $\Omega = \Omega(t)$. To control for heteroskedasticity, we propose to use both a wild bootstrap approach and a 'block-regime' bootstrap procedure where blocks from which the residuals are resampled coincide with the regimes. Of course, it is interesting to evaluate the performance of the bootstrapping procedures when weak exogeneity is either imposed or relaxed.

The procedure runs as follows

Step 1. Fit to the p-dimensional process X_t a VECM model under the null hypothesis of m_0 regimes (i.e., model $\mathcal{M}_0$) to obtain a matrix of residuals $E = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T)^\top$ where $\hat{\varepsilon}_t = Z_{0t} - \hat{A}_0 \hat{B}_0^\top Z_{1t} - \hat{C}_0 Z_{2t}$. The estimation of the model parameters can be carried out using the standard reduced rank regression technique with closed-form solution if no breaks occur under the null.

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

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

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

to get $\{\varepsilon_{t,b}^*\}_{t=1}^T$ (semi-parametric). Alternatively, compute the sample covariance matrix $\hat{\Omega}_c$ of the centred residuals and draw $\{\varepsilon_{t,b}^*\}_{t=1}^T$ from $\mathcal{N}(0, \hat{\Omega}_c)$ (parametric). For the wild bootstrap approach, we proceed as in the semi-parametric case but we further multiply each resampled residual with a realization from a Rademacher variable η_t (see Davidson and Flachaire, 2008) which follows a

two-point distribution defined as

$$F:\ \eta_t = \begin{cases} 1 & \text{with probability } 0.5 \\ -1 & \text{with probability } 0.5. \end{cases}$$

For the block-regime bootstrap instead, we form m_0 blocks, one for each regime under the null hypothesis and we resample either parametrically or non-parametrically from each block.

(b). Build recursively the bootstrapped counterpart of X_t, denoted with $X^*_{t,b}$, noting that $Z^*_{0t,b} = \hat{A}_0 \hat{B}_0^\top Z^*_{1t,b} + \hat{C}_0 Z^*_{2t,b} + \varepsilon^*_{t,b}$ can be conveniently rewritten as

$$X^*_{t,b} = (\hat{A}_0 \hat{B}_0^\top \Xi_1 + \mathbf{I}_{p_1}) X^*_{t-1,b} + \hat{C}_0 \Xi_2 \tilde{Z}^*_{2t,b} + \varepsilon^*_{t,b} \quad t \geq k,$$

where the first k observations are taken from X_t, the parameters estimates come from Step 1 and the matrices Ξ_1 of dimensions $(mp_1) \times p_1$ and Ξ_2 of dimensions $(mp_2) \times p_2$ are used to reconcile the above expression with (3.3). In particular, Ξ_1 is defined as

$$\Xi_1 \equiv \begin{bmatrix} \mathbf{1}_{1,t} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{1}_{1,t} \\ \vdots & \vdots & \\ \mathbf{1}_{m,t} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{1}_{m,t} \\ & & \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{1,t} \\ \vdots \\ \mathbf{1}_{m,t} \end{bmatrix} \otimes \mathbf{I}_{p_1}$$

such that $\Xi X^*_{t-1,b} = Z^*_{1t,b}$ and Ξ_2 is similarly defined in order to get $\Xi \tilde{Z}^*_{2t,b} = Z^*_{2t,b}$.

(c). Use $X^*_{t,b}$ to estimate the null and the broken models under the various scenarios and compute the associated likelihood ratio tests $LR^*_{b,i}$ of $\mathcal{M}^*_{0,b}$ against $\mathcal{M}^*_{1,i,b}$ for $i = 1, \ldots, n$.

(d). Compute the bootstrapped minimum p-value statistic

$$Q^*_b = 1 - \min_{1 \leq i \leq n} [1 - F(LR^*_{b,i})]$$

and record its value. If we bootstrap also the p-values by repeating Step 1 and Step 2(a)-2(c) using $X^*_{t,b}$ instead of X_t, then we obtain the following double bootstrapped statistic

$$Q^{**}_b = 1 - \min_{1 \leq i \leq n} [1 - \hat{F}^*(LR^*_{b,i})]$$

with the bootstrapped p-values defined as

$$\hat{p}(LR^*_{b,i}) = 1 - \hat{F}^*(LR^*_{b,i}) = \frac{1}{B_p} \sum_{b_p=1}^{B_p} \mathbf{1}(LR^{**}_{b_p,b,i} \geq LR^*_{b,i})$$

and where B_p can be different from B.

Step 3. Decide on the acceptance/rejection of the null hypothesis by comparing Q either with the desired quantile of the bootstrapped distribution of Q^* or Q^{**} or by comparing the associated bootstrapped p-value with the chosen level of significance.

According to Theorem 1 in Dufour et al. (2014), the minimum p-value statistic has exactly the nominal size if there are no ties, i.e. $\mathbb{P}[Q^*_i = Q^*_j] = 0$

for $i \neq j$, and $\alpha(B+1)$ is an integer (α being the nominal level). Consequently, at least 19 replications are needed if α is set to 5%[1]. However, as again noted by Dufour et al. (2014), ties have non-zero probability in minimum p-value statistics and to solve this issue they suggest the tie-breaking procedure developed in Dufour (2006). The procedure amounts in pairing the set of statistics $\{Q \equiv Q_0, Q_1^*, \ldots, Q_B^*,\}$ with $B+1$ randomly drawn uniform random variables $\{Z_0, Z_1, \ldots, Z_B\}$ and sorting the pairs according to

$$(Q_i^*, Z_i) \geq (Q_j^*, Z_j) \Leftrightarrow \left[Q_i^* > Q_j^* \vee (Q_i^* = Q_j^* \wedge Z_i \geq Z_j)\right].$$

The bootstrapped p-value can be then computed as

$$\hat{p}(Q) = \frac{B\left[1 - \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}(Q \geq Q_b^*) + \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}(Q_b^* = Q)\mathbf{1}(Z_b \geq Z_0)\right] + 1}{B+1}.$$

The strength of using minimum p-value statistics combined with the above approach is that the global level of the test is controlled even if the p-values of the individual tests (in our case LR_i, $i = 1, \ldots, n$) are not exact or based on asymptotic approximations, provided that the statistics are nuisance-parameter free under H_0 (see Dufour et al., 2014).

[1] See also Davidson and MacKinnon (2005) for a discussion on how to choose the correct number of bootstrap replications

3.5 Finite Sample Properties: Size and Power of the Q-statistic

In this section, we investigate the small sample properties of the Q statistic through a Monte Carlo simulation. We compute the empirical size and power of the minimum p-value statistic in order to assess its ability in detecting the correct number of breaks. Further, we study the effects of imposing and relaxing weak exogeneity as well as the impact of weak identification of the cointegrating relationship.

3.5.1 Monte Carlo Design

The Monte Carlo experiments are based on two alternative formulations of a cointegrating system. The first is a standard VECM (DGP1) while the second is based on the Engle and Granger (1987) representation as formulated in Gonzalo (1994) (DGP2). In applying the test, DGP2 is mapped into a VECM. By omitting breaks, the two DGPs take the following forms

$$\textbf{DGP1}: \quad \begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & -\tilde{\beta} \end{bmatrix} X_{t-1} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \Rightarrow X_t = (\mathbf{I}_2 + \alpha\beta^\top) X_{t-1} + \varepsilon_t. \tag{3.16}$$

with $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \mathbf{I}_2)$

$$\textbf{DGP2}: \quad x_{1t} - \tilde{\beta} x_{2t} = z_t, \; z_t = \rho z_{t-1} + e_{z_t} \tag{3.17}$$

$$a_1 x_{1t} - a_2 x_{2t} = w_t, \; w_t = w_{t-1} + e_{w_t} \tag{3.18}$$

with $[e_{z_t}, e_{w_t}]^\top \overset{iid}{\sim} \mathcal{N}(0, \mathbf{I}_2)$.

Both formulations allow to control for weak exogeneity through α_2 in DGP1 and a_1 in DGP2, as well as to study the impact of weak identification of the cointegrating relationship through α_1 (DGP1) and ρ (DGP2). DGP1 does not uncover the relationship between the VECM parameters (α, β, Ω) allowed to be treated independently. When simulating instead from DGP2, from Theorem 1 we have that when $a_1 \neq 0$ we need to test for breaks both in α and β; in addition, breaks in β cause breaks in the covariance matrix of the error term and this needs to be taken into account when bootstrapping the distribution of the Q-statistic.

The number of Monte Carlo simulations is set to $M = 1,000$ while the number of bootstrap replications to $B = 199$. Each experiment is performed initialising the random number generator with the same seed, and for each simulated vector process $X_{t,m} = [x_{1t,m}, x_{2t,m}]^\top$, $m = 1, \ldots, M$ and $t = 1, \ldots, T$, we estimate $\mathcal{M}_{0,m}$ and $\mathcal{M}_{1,i,m}$ for $i = 1, \ldots n$ and we compute the corresponding likelihood ratios in order to get Q_m. We then compute the empirical rejection frequencies of hypotheses of the form (3.14) by counting how many times Q_m exceed the 95th percentile of the distribution of $Q^*_{b,m}$ obtained through the bootstrapping procedure described before. In particular, define $\Theta^* = \{\Theta^*_1, \ldots, \Theta^*_{m_D} | \tau_D\}$ to be the set which contains the time-varying parameters for a given set of break dates τ_D that actually characterise the DGP. The time-varying parameters defining the DGP are such that $\Theta^*_{m_D} = \Theta^*_{m_D-1} + h$ where h is a vector of known constants. We consider

the empirical rejection frequencies of tests of the form

$$\mathcal{H}_0: \quad \Theta_1^* \neq \cdots \neq \Theta_{m_0=m_D-1}^* | \tau_{m_0}$$

$$\mathcal{H}_1: \begin{cases} \Theta_{1,1}^* \neq \cdots \neq \Theta_{m_{1,1} \geq m_D}^* | \tau_{m_{1,1}} \\ \vdots \\ \Theta_{1,n}^* \neq \cdots \neq \Theta_{m_{1,n} \geq m_D}^* | \tau_{m_{1,n}} \end{cases}$$

such that $\tau_{m_0} \subset \tau_D$ and $\tau_{m_0} \subset \tau_{m_{1,i}}$, $i = 1, \ldots, n$. The two constraints on the location of the breaks are needed to ensure that if the number of regimes under the null hypothesis m_0 is different from 1, then we require that the model under the null incorporates the correct location of the break dates and the alternative scenarios contains the same breaks (plus additional ones) of the null hypothesis to ensure that the null model is nested into the alternative. This is to avoid situations where the unrestricted model has a smaller log-likelihood than the restricted model. The empirical rejection frequencies of $\mathcal{H}_0$ approximate the size of the Q statistic when h equals zero (we are simulating under the null) while for h different from zero (we are simulating under the alternative), we approximate the power. In order to specify the scenarios under $\mathcal{H}_1$, we input a matrix containing n columns corresponding to n different scenarios. Consistently with the spirit of the Q statistic, we do not require the number of the breaks in each scenario to be the same.

Finally, to estimate the models undergoing the various regimes, we maximise the concentrated log-likelihood function (3.7) using the BFGS algorithm[2] and monitoring for convergence, in order to avoid spurious rejections. To

[2]The simple DGP considered here is characterised by few parameters and thus it is possible to maximise directly (3.7) without recurring to the switching algorithm. We notice that if the optimisation over the full set of parameters reaches convergence, then the switching algorithm gives the same estimates though requiring more computing time.

select the correct subset of parameters, we define the appropriate restriction matrices H, h and G according to (3.4) and (3.5). Note that since under the alternative we have different scenarios, in practice we need to define $H_{i,1}$, $h_{i,1}$, and $G_{1,i}$ for each experiments plus the three restriction matrices to estimate the null model H_0, h_0, and G_0. Details are reported in Appendix B.

3.5.2 Simulations Results

We evaluate the performance of the Q-statistic in detecting multiple structural breaks in the normalised cointegrating vector $[1, -\tilde{\beta}]^\top$. In our bivariate setting, this means that we consider versions of DGP1 and DGP2 with structural breaks affecting the long-run coefficient $\tilde{\beta}$. Following the discussion in Section 3.3.2, breaks in $\tilde{\beta}$ also implies breaks in the covariance matrix for DGP2 and, depending on whether weak exogeneity holds, in the short-run coefficients. Breaks in the covariance matrix cause heteroskedasticity in the residuals time series which needs to be taken into account when bootstrapping the distribution of Q. As a consequence, the residuals are resampled using the wild and the regime-block approaches. On the other hand, the structure of DGP1 is such that breaks in β do not affect the covariance matrix of the errors or any other component, thus the standard parametric or non-parametric bootstrap suffices.

3.5.2.1 Results for DGP1

We simulate from the following version of DGP1 with breaks in $\tilde{\beta}$

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & -\tilde{\beta}(t) \end{bmatrix} X_{t-1} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}. \tag{3.19}$$

We consider three sets of experiments where $\tilde{\beta}$ is affected by one, two and three breaks respectively. In this way, we study the size and the power of Q under the null of structural breaks which is more informative than studying the statistic under the null of no breaks only. For explanation purposes, consider the case with one break. In detail, we have that $\tilde{\beta}(t) = \tilde{\beta}_1 \mathbf{1}_{1t} + \tilde{\beta}_2 \mathbf{1}_{2t}$ where $\mathbf{1}_{1t} = \mathbf{1}(1 \leq t \leq T_1)$, $\mathbf{1}_{2t} = \mathbf{1}(T_1 + 1 \leq t \leq T)$ and $\tilde{\beta}_2 = \tilde{\beta}_1 + h$. The simulated process (3.19) is therefore affected by one structural break at T_1 when $h \neq 0$ or zero when $h = 0$.

To evaluate the size and the power of Q-statistic, we consider tests of zero, one and two breaks against more than zero, one and two breaks respectively, specifying different scenarios under the alternative hypothesis. When $h = 0$, the null hypothesis is then specified consistently with the DGP (size) whereas when $h \neq 0$ we count the rejection frequency of a false null (power).

Table 3.1 reports the empirical size of the Q-statistic obtained using both parametric and non-parametric bootstrap. For all three experiments, we set $h = 0$ and we consider the following parameter space: $\tilde{\beta}_1 = 1$ (first experiment), $\tilde{\beta}_1 = 1, \tilde{\beta}_2 = 2$ (second experiment) and $\tilde{\beta}_1 = 1, \tilde{\beta}_2 = 2, \tilde{\beta}_3 = 1.5$ (third experiment), $\alpha_1 = \{-1, -0.2, -0.01, -0.5\}$, $\alpha_2 = \{0, 0.5\}$ (imposing and relaxing weak exogeneity) and $T = \{100, 200, 300, 400, 500\}$. Further, we

explore different locations of the break dates affecting the DGP by letting the shifts to happen at the middle as well as towards the beginning or the end of the sample. Details on the location of the breaks are given in Table 3.1.

[Table 3.1 about here.]

The size of the Q-statistic is close to the nominal one for all cases. As expected, we observe some over-size when there is relevant weak identification ($\alpha_1 = -0.01$). The distortions however are not severe and tend to disappear in samples as large as $T = 500$. The modest size distortion when in presence of weak identification is also encouraging since it means that the behaviour of the statistic is controlled even for extreme cases.

In order to study the power, we consider instead the same null hypotheses but letting the value of h to gradually increase. In particular, we consider the coefficient of the last regime to be augmented with respect to the one of the penultimate regime by $h = \{0.1, 0.3, 0.5, 1\}$ for $T_1 = \lfloor T/2 \rfloor$ (first experiment) , $T_2 = \lfloor 2T/3 \rfloor$ (second experiment), and $T_3 = \lfloor 5T/6 \rfloor$ (third experiment). The value of the other parameters which defines (3.19) is the same as for the size experiments while the sample length is explored for $T = \{100, 300, 500\}$.

Table 3.2 reports the power of the test. We observe that the power increases with T and when gradually moving away from the null hypothesis, i.e. for larger deviations of $\tilde{\beta}_{m_D}$ with respect to $\tilde{\beta}_{m_D-1}$. Further, the test performs well even for short samples such as $T = 100$ with the only exception being, as expected, the case of relevant weak identification ($\alpha_1 = -0.01$).

[Table 3.2 about here.]

3.5.2.2 Results for DGP2

We consider the same three sets of experiments as for DGP1, where $\tilde{\beta}$ is affected by one, two and three breaks respectively, but simulating from the following process instead

$$x_{1t} - \tilde{\beta}(t)x_{2t} = z_t \quad z_t = \rho z_{t-1} + e_{z_t}$$
$$\Delta x_{2t} = e_{w_t}$$

which corresponds to (3.18) when weak exogeneity is imposed. For the case with one break, we define $\tilde{\beta}(t) = \tilde{\beta}_1 \mathbf{1}_{1t} + \tilde{\beta}_2 \mathbf{1}_{2t}$ where $\mathbf{1}_{1t} = \mathbf{1}(1 \leq t \leq T_1)$, $\mathbf{1}_{2t} = \mathbf{1}(T_1 + 1 \leq t \leq T)$ and $\tilde{\beta}_2 = \tilde{\beta}_1 + h$. The DGPs with two and three breaks are defined in a similar fashion.

Adopting a VECM representation, we can write

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} \rho - 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & -\tilde{\beta}(t) \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{z_t} + \tilde{\beta}(t) e_{w_t} \\ e_{w_t} \end{bmatrix} \tag{3.20}$$

which entails structural breaks in the variance term as well. For this reason, the empirical rejection frequencies that follow are obtained using a wild and a 'block-regime' bootstrap approach as discussed in Section 3.4.2.

In order to study the empirical size, we set $h = 0$ and we consider an extended testing framework with respect to the previous one where the parameters of the VECM that change across the regimes are both the long-run coefficient and the covariance matrix of the error term, i.e. $\Theta^* = \{\tilde{\beta}_1, \Omega_1, \ldots, \tilde{\beta}_m, \Omega_m | \tau_m\}$. The scenarios under the alternative are chosen such that they match exactly those used for DGP1. Similarly, the time-varying $\tilde{\beta}$ is such that $\beta_1 = 1$ (first

experiment) , $\beta_1 = 1$ and $\beta_2 = 2$ (second experiment), $\beta_1 = 1$, $\beta_2 = 2$ and $\beta_3 = 1.5$ (third experiment) while $\rho = \{0, 0.8, 0.99\}$ in order to obtain the same values of α_1 as in DGP1. The location of the break dates and the sample sizes explored are the same as well. Table 3.3a reports the results.

[Table 3.3 about here.]

The empirical rejection frequencies are generally very close to the nominal level with a better performance of the 'block-regime' bootstrap. The wild bootstrap works almost equally well for the case with no breaks and one break under the null hypothesis while it seems to suffer as more breaks are included in the model. Indeed, the bigger the number of breaks the more important is the role played by heteroskedasticity. Since the block-regime version of the bootstrap provides exact adjustment for regime-specific heteroskedasticity, it is not surprising that it performs as good as the standard bootstrap for DGP1. Further, as already observed for DGP1, weak identification seems to generally worsen the results.

Table 3.3b summarises the results for the power. Just as in the case of DGP1, the values reported are obtained by letting h to gradually increase. The results show the statistic has good power also when adopting DGP2. As commonly observed, the power increases with larger deviations from the null and when increasing the sample length. There are no notable differences when using either the wild or the 'block-regime' bootstrap.

A final set of experiments refers to the case where we relax the assumption of weak exogeneity according to the following process

$$x_{1t} - \tilde{\beta}(t)x_{2t} = z_t,\ z_t = \rho z_{t-1} + e_{z_t}$$

$$x_{1t} - a_2 x_{2t} = w_t,\ w_t = w_{t-1} + e_{w_t}$$

which corresponds to the most general case with $a_1 = 1$. From Section 4, structural breaks in β cause structural breaks in α in the VECM representation. Therefore, we extend the set of broken parameters Θ^* in order to include also α, i.e. $\Theta^* = \{\alpha_1, \beta_1, \Omega_1, \ldots, \alpha_m, \beta_m, \Omega_m | \tau_m\}$, while all the remaining parameters are kept the same as for the case with weak exogeneity imposed. Table 3.4 reports the empirical rejection frequencies.

[Table 3.4 about here.]

By relaxing weak exogeneity, the maximization of the likelihood function becomes more demanding and thus this framework is the most challenging consider so far. As long as the size is concerned, the results reported in Table 3.4 allow us to conclude that the block-regime bootstrap provides rejection frequencies close to the nominal level in all the cases with a better performance of the parametric version. Similarly to the case when weak exogeneity is imposed, the wild bootstrap does not perform so satisfactorily. As far as the power is concerned, the test shows power of about the same magnitude than that observed in the previous experiments when weak exogeneity is imposed.

3.6 Empirical Illustrations

In this section, we want to illustrate the empirical relevance of taking into account the role of weak exogeneity when selecting the subset of parameters $\Theta^* \subset \Theta$ to be tested for breaks in the VECM framework as well as how to practically implement the Q-statistic.

3.6.1 Dividend-Price Ratio

In this first application, we illustrate how multiple break testing can be conducted using the Q-statistic by investigating the present value theory for asset prices, which states that asset prices can be expressed as the sum of the present discounted values of expected future dividends (see e.g. Campbell and Shiller, 1987, 1989).

According to the present value theory, the relationship between the current price P_t of an asset, the expected price at the next period P_{t+1}, and the the expected dividend D_{t+1} paid for owning the asset during the period $[t, t+1]$ can be then expressed as

$$P_t = \mathbb{E}_t \left(\frac{P_{t+1} + D_{t+1}}{R_{t+1}} \right) \tag{3.21}$$

where R_{t+1} is the discount factor for the period $[t, t+1]$ and $\mathbb{E}_t(\cdot)$ is a shorthand 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), (3.21) 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

$$p_t = \frac{\kappa}{1-\rho} + (1-\rho)\sum_{i=0}^{\infty} \rho^i \mathbb{E}_t(d_{t+i+1}) - \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t(r_{t+i+1}) + \lim_{i\to\infty} \rho^i \mathbb{E}_t(p_{t+i}). \quad (3.22)$$

If we further assume that prices do not follow an explosive process, then $\lim_{i\to\infty} \rho^i \mathbb{E}_t(p_{t+i})$ in (3.22) converges to zero, and by rearranging the remaining terms, the dividend-price ratio is given by[3]

$$d_t - p_t = -\frac{\kappa}{1-\rho} + \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t(r_{t+i+1} - \Delta d_{t+1+i}). \quad (3.23)$$

The empirical validity of the present value relationship in (3.23) corresponds to cointegration between the log-dividend process $\{d_t\}$ and the log-price process $\{p_t\}$. This is because the right-hand side of (3.23) is the sum of a stationary process plus a constant term.

Our empirical analysis of the above is based on the 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[4] while the computations are executed using *OxMetrics 7* (Doornik and Hendry, 2013). The dividend-price ratio for this sample is depicted in Figure 3.2.

[Figure 3.2 about here.]

In order to test (3.23), we 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). These tests lead to the rejection of cointegration, a result certainly not at

[3]Without loss of generality, we assume that the discount factor $r_t = r_{t+1} = \cdots = r$ is constant over time, such that (3.23) simplifies to $d_t - p_t = -\frac{\kappa - r}{1-\rho} - \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t(\Delta d_{t+1+i})$.

[4]`http://www.econ.yale.edu/~shiller/`

odds with the dynamics of the dividend-price ratio in Figure 3.2.

[Figure 3.3 about here.]

To investigate why cointegration fails, it is instructive to plot the one-step ahead residuals. Figure 3.3 reports the one-step ahead residuals, obtained by recursively estimating the VAR with an initial window of 50 observations, with reference to the stylized-fact shocks such as the Black Monday in 1987, the dot-com burst in early 2000, and the more recent sub-prime and sovereign debt crises (global financial crises, GFC).

We consider the above stylized-fact dates as alternative scenarios and estimate the underlying VECM from a VAR(4) with an unrestricted constant term. We aim to assess stability without taking any ex-ante stand regarding weak exogeneity, since there is no theoretical nor empirical guidance in this regard. Table 3.5 reports the results. In the following discussion, rejection refers to the 5% level.

[Table 3.5 about here.]

The individual LR tests show that the break dates that matter are 1999(3) (dot-com boom) and 2007(4) (global financial crisis) while the 1987(3) (Black Monday) it is not significant. This is in line with the conventional wisdom regarding short-lived impact of the Black Monday.

From test A in Table 3.5, the null of no breaks is not rejected when we consider $\Theta^* = \{\beta, \Omega\}$. However, we strongly reject the null of stability when we augment the vector of parameters to be tested for breaks with the short-run impact matrix α (test B). This illustrates the importance of our results in Section 3, as the formulation of a restrictive null may mask important potential breaks. Test C assesses no breaks against two scenarios: (i) one

break at 1987(3), 1999(3), and 2007(4); (ii) only two breaks at 1999(3) and 2007(4). In this case, the Q-statistic rejects the null of no breaks. Finally, test D evaluates a restricted model with only the two last break dates, namely 1999(3) and 2007(4), against a model which incorporates all three breaks at 1987(3), 1999(3) and 2007(4). The restricted model which allows for breaks in both the cointegrating parameters at the time of the dot-com boom and the global financial crisis cannot be rejected.

Jointly interpreted, our analysis provides an alternative perspective to the bubbles motivated approach as in Phillips et al. (2011). 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, as long as a multiple simulation-based adjustment is applied and the specifications considered do not take a stand on exogeneity.

3.6.2 Money Demand Stability in the US

We turn now to another application where we test for the presence of breaks by means of the minimum p-value statistic Q. Specifically, we consider modelling the narrow money demand (M1) in the US in a VECM framework and testing for its stability. For the US economy, the evidence on the stability of the long-run relationship between money demand, real output and short-term interest rate is predominant in the literature. Some of the most influential studies are Baba et al. (1992) and Hoffman et al. (1995) which found indeed that the money demand is stable after accounting for omitted variables and imposing unit long-run income elasticity, respectively. Non-

etheless, the data-span considered by these studies does not cover for obvious reasons the recent financial crisis where the relationship is expected to break down at some point around 2007-2008. In this section, we reconsider the stability of the money demand in the US using an updated sample and testing for breaks by means of the Q-statistic.

In particular, we specify the following vector process

$$X_t = \begin{bmatrix} (m-p)_t \\ y_t \\ r_t \end{bmatrix}$$

where $(m-p)_t$ is the logarithm of the real money demand (where p_t is the GDP deflator), y_t is the logarithm of the real GDP and r_t is the three month T-bill rate representing the external opportunity cost for holding real money. The data are obtained from the FRED database at quarterly frequency over the period 1980(1)-2014(2). If the money demand is theoretically updated every three months, it makes sense to see as alternative investment opportunity on the same time horizon the T-Bill with maturity three months as explained in Hoffman et al. (1995). Figure 3.4 plots the time series and allows us to infer that the relationship between $(m-p)_t$ and y_t is expected to be positive while the relationship between $(m-p)_t$ and r_t is expected to be negative as predicted by theory.

[Figure 3.4 about here.]

We start by modelling the trivariate system using a VAR(4). However, when testing for cointegration using the λ_{trace} test we find no evidence of any cointegrating relationship (see Table 3.6). To investigate the issue fur-

ther, we analyse the stability of the individual equations as well as the whole system over the sample period through the 1-step Chow test[5] using an initial window of 40 observations. Figure 3.5 plots the 1-step Chow test outcomes together with the critical value at 1% level. Evidence of instability is intended as the red line crossing the blue one.

[Figure 3.5 about here.]

As we can see from Figure 3.5, both the real money demand equation and the real GDP one show signs of instability corresponding to the period of the recent financial crisis which transmit also to the system when considered as a whole. To check whether the failure of finding cointegration is indeed due to the 2007-2008 period, we repeat the cointegration analysis by cutting the sample at the end of 2006.

[Table 3.6 about here.]

Table 3.6 clearly shows that when considering only the data before the recent financial crisis, we find cointegration and the presence of instabilities seems to disappear both at individual as well as global level as shown by Figure 3.6. This result is consistent with the findings of the previous studies (Baba et al., 1992; Hoffman et al., 1995).

[Figure 3.6 about here.]

In order to assess statistically the time of the break date, we apply the minimum p-value statistic where the underlying VECM process is obtained

[5]Given a regression $y_t = \beta^\top x_t + \varepsilon_t$, $t = 1, \ldots, T$, with k regressors, the 1-step Chow test (see Nielsen and Whitby, 2008) tests for the stability of the regressors coefficients and it is computed as $\frac{(RSS_t - RSS_{t-1})(t-k-1)}{RSS_{t-1}}$ where RSS_t denotes the residuals sums of squares of a regression involving $t < T$ observations. Starting by selecting an initial window of observations, the test is usually computed recursively to produce plots similar to those reported in Figure 3.5.

transforming a VAR(4) with an unrestricted constant term. Nevertheless, as the system is not stable over the whole sample, it is not obvious how to test for weak exogeneity although theoretically y_t and r_t are exogenous with respect to the money demand. For this reason, all the tests are repeated considering both the case $\Theta^* = \{\beta, \Omega\}$ and the case $\Theta^* = \{\alpha, \beta, \Omega\}$. In particular, we consider three tests where the alternative hypotheses incorporate different scenarios. In the first one, we consider testing the null of no breaks against three scenarios, where the break dates are selected on the basis of the outcome of 1-step Chow test (Figure 3.5). The two break dates, i.e. $2008(3)$ and $2011(4)$, are considered both individually and jointly to give rise to the following test:

$$A: \quad \begin{aligned} &\mathcal{H}_0 : \Theta^*(t) = \Theta^*(t-1), && \forall t = 1, \dots, T \\ &\mathcal{H}_1 : \begin{cases} \Theta_1^* \neq \Theta_2^* \mid \tau_1 = 2008(3) \\ \Theta_1^* \neq \Theta_2^* \mid \tau_2 = 2011(4) \\ \Theta_1^* \neq \Theta_2^* \neq \Theta_3^* \mid \tau_3 = [2008(3), 2011(4)] \end{cases} \end{aligned}$$

Furthermore, we consider also a simplified version of test A, where the joint scenario is dropped:

$$B: \quad \begin{aligned} &\mathcal{H}_0 : \Theta^*(t) = \Theta^*(t-1), && \forall t = 1, \dots, T \\ &\mathcal{H}_1 : \begin{cases} \Theta_1^* \neq \Theta_2^* \mid \tau_1 = 2008(3) \\ \Theta_1^* \neq \Theta_2^* \mid \tau_2 = 2011(4) \end{cases} \end{aligned} .$$

Finally, we also check the robustness of the Q-statistic by contaminating the scenarios with an apparently insignificant break date at $1992(2)$. The null

and alternative hypotheses are given by

$$C: \quad \begin{array}{ll} \mathcal{H}_0 : \Theta^*(t) = \Theta^*(t-1), & \forall t = 1, \ldots, T \\ \mathcal{H}_1 : \begin{cases} \Theta_1^* \neq \Theta_2^* \mid \tau_1 = 2008(3) \\ \Theta_1^* \neq \Theta_2^* \mid \tau_2 = [1992(2), 2008(3)] \end{cases} & \end{array} .$$

[Table 3.7 about here.]

The outcome of the three tests are reported in Table 3.7. The Q-statistic p-values are obtained by semi-parametric block-regime bootstrap (see Section 3.4.2) with $B = 299$ replications whereas the individual LR tests p-values are based on standard asymptotics.

From an inspection of the results, we can infer that the three tests all lead to the rejection of the null of no breaks – at a standard 5% significance level – which is consistent with our initial finding of no cointegration over the full sample. More in detail, in test A the Q-statistic leads to the rejection of the null of no breaks whether the stability of α and β is tested jointly or not. This is strong evidence against the null of stability as well as supportive of the fact that y_t and r_t are exogenous variables. Further, from an inspection of the individual LR tests p-values, it is evident that only $2008(3)$ is a significant break date. A similar conclusion is drawn when considering the outcome of test B. In particular, the test further confirms that a break in the relationship between money demand, real output and short-term investment opportunities occurred in the third quarter of 2008. Again, the same conclusion is reached with test C, where we perturb the testing framework introducing a scenario with an insignificant break date in $1992(2)$.

To summarise, our analysis of the stability of the narrow money demand in the US allows us to draw the following conclusions: first, consistently with the existing literature, the money demand equation is stable up to the end of 2006; second, the global financial crisis of 2008 causes a break down in the relationship which manifests itself in the lack of cointegration when extending the sample to present days. Through the minimum p-value statistic derived in Section 3, we have been able to identify the third quarter of 2008 as the time of the break-down.

3.6.3 Exchange Rate Modelling

In this last application, we want to empirically investigate the role of weak exogeneity in order to show the consequences that we might incur in if we ignore the results in Theorem 1. In particular, we model the joint dynamics of three US dollar denominated exchange rates, namely Euro, Swiss Franc and Japanese Yen using a VECM. The data are at quarterly frequency over the period 1999(1)-2014(1) and are obtained from the FRED database. Figure 3.7 reports the three time series for which we cannot reject the null hypothesis of being unit root processes.

[Figure 3.7 about here.]

As Figure 3.7-(d) shows it seems plausible that the three series share some common stochastic trends. In order to test for cointegration, we first model the log-levels selecting a VAR(4) process with a constant and we then decide on the number of cointegrating vectors (i.e. cointegarting rank) using sequentially the well known eigenvalue-based test for cointegration (Jo-

hansen, 1991),

$$\lambda_{\text{trace}}(\mathcal{H}_0(\text{rank} \leq r)|\mathcal{H}_1(\text{rank} = p)) = -T \sum_{i=r+1}^{p} \log(1-\hat{\lambda}_i) \qquad r = 0, \ldots, p-1,$$

where in this specific case $p = 3$.

[Table 3.8 about here.]

Table 3.8 reports the outcome of the sequential testing procedure from which we can conclude that there is one cointegrating vector and two common stochastic trends. The estimated system of equations with a normalised cointegrating vector and the p-values of the estimated coefficients in brackets is given by

$$\begin{bmatrix} \Delta EUR_t \\ \Delta CHF_t \\ \Delta YEN_t \end{bmatrix} = \mu + \begin{bmatrix} -0.19302 \\ \underset{(0.0293)}{} \\ -0.11224 \\ \underset{(0.0499)}{} \\ -0.12340 \\ \underset{(0.0256)}{} \end{bmatrix} \begin{bmatrix} 1.0000 & -1.5605 & 1.6967 \\ & \underset{(0.0000)}{} & \underset{(0.0000)}{} \end{bmatrix} \begin{bmatrix} EUR_{t-1} \\ CHF_{t-1} \\ YEN_{t-1} \end{bmatrix} + \sum_{i=1}^{3} \hat{\Gamma}_i \Delta X_{t-i}$$

from which we can exclude the presence of weak exogeneity, since all the elements of the short-run impact matrix are statistically different from zero at 5% significance level. Additionally, we check the stability of the cointegrating relationship by means of a recursive graphical analysis of the eigenvalue corresponding to the cointegrating vector as reported in Figure 3.8.

[Figure 3.8 about here.]

The recursive plot of the estimated value of $\hat{\lambda}_1$ shows that the cointegrating vector is stable across the sample period and thus when testing for no breaks in α and β jointly – given the absence of weak exogeneity – against an al-

ternative hypothesis where they are allowed to change, we should not reject the null hypothesis. To investigate this claim statistically, we can compute the following LR test

$$A: \quad \begin{array}{ll} \mathcal{H}_0 : \Theta^*(t) = \{\alpha(t), \beta(t), \Omega(t)\} = \Theta^*(t-1), & \forall t = 1, \ldots, T \\ \mathcal{H}_1 : \Theta_1^* \neq \Theta_2^* \mid \tau_1 = 2003(1) & \end{array}$$

where the break date under the alternative hypothesis corresponds to the least favourable scenario in terms of stability, i.e. the date that minimises the p-value corresponding to the null hypothesis of no breaks. Furthermore, it also interesting to investigate what is the outcome of the test when incorrectly specifying the subset of parameters which we want to test for breaks. This can be achieved changing the subset of parameters Θ^* to exclude α, i.e. ignoring the the role of weak exogeneity, leading to the following LR test

$$B: \quad \begin{array}{ll} \mathcal{H}_0 : \Theta^*(t) = \{\beta(t), \Omega(t)\} = \Theta^*(t-1), & \forall t = 1, \ldots, T \\ \mathcal{H}_1 : \Theta_1^* \neq \Theta_2^* \mid \tau_1 = 2003(1) & \end{array} .$$

Table 3.9 reports the outcomes of the the two LR tests. In particular, we computed the p-values by parametric (assuming normality) and semi-parametric block-regime bootstrap following the steps described in Section 3.4.2 and using 299 replications to approximate the tests distributions.

[Table 3.9 about here.]

The interpretation of the outcome is straightforward: when wrongly specifying the subset of parameters to be tested (Test A), i.e. ignoring the role of weak exogeneity and focusing only on breaks in β, we reject the null of no breaks even if the standard VECM with no time-varying parameters ap-

pears to be correctly specified; however, when testing jointly for breaks in α and β (Test B), consistently with the absence of weak exogeneity, we cannot reject the null of no breaks. The outcome of the testing procedure is thus consistent with both Theorem 1 as well as the simulations results reported in Section 3.5.2.

3.7 Conclusions

In this chapter, we investigate how to specify the subset of parameters when testing for breaks in a VECM framework. First, we showed that the choice of which parameters of the VECM can be tested for breaks is indeed constrained given that breaks in the long run matrix β implies breaks in the short run impact matrix α, unless weak exogeneity can be imposed, and breaks in β imply also breaks in the covariance matrix of the error term. Second, we developed a new test for multiple structural breaks in a VECM framework, by extending the likelihood ratio test proposed in Hansen (2003) to the case of unknown break dates through the specification of several scenarios regarding the number and the location of the breaks. We defined a minimum p-value statistic and a bootstrap procedure to approximate its critical values robust to the presence of breaks in the covariance matrix of the error term.

The finite sample properties of the proposed statistic are analysed through an extensive Monte Carlo simulation where under the null hypothesis we allow for the presence of multiple breaks. The effects of imposing and relaxing weak exogeneity as well as weak identification of the cointegrating matrix are also explored. The statistic we proposed has size close to the

nominal level and good power when the data are simulated either using a standard VECM (DGP1) or the data generating process adopted in Gonzalo (1994) mapped into a VECM (DGP2). We showed that for DGP2, breaks in the cointegrating parameters cause breaks in the covariance matrix as well as in the short-term adjustment matrix if weak exogeneity is not imposed. The breaks-induced heteroskedasticity is addressed in the simulations by using the wild bootstrap and a block-regime specific bootstrap approach which allow to restore the correct size. For both DGPs, relaxing weak exogeneity, if taken into account, does not affect the finite-sample properties of the test statistic. As expected, the Q-statistic shows smaller size distortions when the cointegrating relationship is weakly identified.

Finally, we illustrate how the Q-statistic can be used in applied work to detect multiple structural breaks by means of three empirical applications: the first involving the dividend-price ratio of the S&P 500; the second concerning the money demand stability in the US; and the third considering a vector of exchange rates.

Appendix 3.A Concentrated Log-likelihood

We present the derivation of the concentrated log-likelihood function reported in equation (3.7) used to estimate the parameters of (3.3).

First, a reminder of some standard results of matrix algebra will reveal useful in the derivation. We have that:

- for the trace operator $\text{tr}(\cdot)$ and given two matrices X and Y, we have that

$$\text{tr}(XY) = \text{tr}(YX) = \sum_{i,j} X_{i,j} Y_{j,i} \,, \tag{3.24}$$

- for a square non-singular matrix X,

$$\frac{\partial \log|X|}{\partial X} = (X^{-1})^\top = (X^\top)^{-1} \,, \tag{3.25}$$

- for a square non-singular matrix X, and two conformable matrices, A and B not function of X,

$$\frac{\partial \text{tr}(AX^{-1}B)}{\partial X} = -(X^{-1}BAX^{-1})^\top. \tag{3.26}$$

The log-likelihood function of $Z_{0t} = AB^\top Z_{1t} + CZ_{2t} + \varepsilon_t$ with the error term distributed as $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \Omega(t))$ and piecewise constant covariance matrix $\Omega(t) = \Omega_1 \mathbf{1}_{1t} + \cdots + \Omega_m \mathbf{1}_{mt}$ is

$$\ell(A, B, C, \Omega_{j=1,\ldots,m}) = -\frac{Tp}{2}\log 2\pi - \frac{T}{2}\sum_{j=1}^{m} \rho_j \log|\Omega_j| - \frac{1}{2}\sum_{j=1}^{m}\sum_{t=T_{j-1}+1}^{T_j} \varepsilon_t^\top \Omega_j^{-1} \varepsilon_t.$$

On the basis of (3.24), note that

$$\textstyle\sum_t \varepsilon_t^\top \Omega_j^{-1} \varepsilon_t = \sum_t \sum_{m,n} (\Omega^{-1}_{m,n,j} \varepsilon_{t,m} \varepsilon_{t,n}) = \sum_t \text{tr}(\Omega_j^{-1} \varepsilon_t \varepsilon_t^\top) = \text{tr}(\Omega_j^{-1} \textstyle\sum_t \varepsilon_t \varepsilon_t^\top)$$

and hence the log-likelihood function can be equivalently rewritten as

$$\ell(A, B, C, \Omega_{j=1,\ldots,m}) = -\frac{Tp}{2}\log 2\pi - \frac{T}{2}\sum_{j=1}^{m} \rho_j \log|\Omega_j| - \frac{1}{2}\sum_{j=1}^{m} \text{tr}\left(\Omega_j^{-1} \sum_{t=T_{j-1}+1}^{T_j} \varepsilon_t \varepsilon_t^\top\right). \tag{3.27}$$

We can now derive the maximum likelihood estimator $\hat{\Omega}^{ML}_{j=1,\ldots,m}$ of $\Omega_{j=1,\ldots,m}$ by setting the first derivative of the log-likelihood function (3.27) with respect to Ω_j

equal to zero. In particular, for each $j = 1, \ldots, m$ and reminding that $\rho_j = \frac{T_j - T_{j-1}}{T}$, we can apply (3.25)-(3.26) to get

$$\begin{aligned}\frac{\partial \ell(A, B, C, \Omega_{j=1,\ldots,m})}{\partial \Omega_j} &= -\frac{T}{2}\rho_j \frac{\partial \text{log}|\Omega_j|}{\partial \Omega_j} - \frac{1}{2}\frac{\partial \text{tr}(\Omega_j^{-1}\sum_{t=T_{j-1}+1}^{T_j}\varepsilon_t\varepsilon_t^\top)}{\partial \Omega_j} = 0 \\ &\Rightarrow (T_j - T_{j-1})\Omega_j^{-1} = (\Omega_j^{-1}\left(\sum_{t=T_{j-1}+1}^{T_j}\varepsilon_t\varepsilon_t^\top\right)\Omega_j^{-1}) \\ &\Rightarrow \hat{\Omega}_j^{ML} = \frac{1}{T_j - T_{j-1}}\sum_{t=T_j-1+1}^{T_j}\varepsilon_t\varepsilon_t^\top. \end{aligned} \tag{3.28}$$

Note that (3.28) is entirely determined by the other set of parameters since $\varepsilon_t = Z_{0t} - AB^\top Z_{1t} - CZ_{2t}$. Substituting back (3.28) in the log-likelihood function, we can write

$$\begin{aligned}\ell(A, B, C, \hat{\Omega}_{j=1,\ldots,m}) &= -\frac{Tp}{2}\text{log}2\pi - \frac{T}{2}\sum_{j=1}^{m}\rho_j \text{log}|\hat{\Omega}_j| - \frac{1}{2}\sum_{j=1}^{m}\text{tr}\left(\hat{\Omega}_j^{-1}\sum_{t=T_{j-1}+1}^{T_j}\varepsilon_t\varepsilon_t^\top\right) \\ &= -\frac{Tp}{2}\text{log}2\pi - \frac{T}{2}\sum_{j=1}^{m}\rho_j \text{log}|\hat{\Omega}_j| + \\ &\quad - \frac{1}{2}\sum_{j=1}^{m}(T_j - T_{j-1})\text{tr}\left(\frac{\hat{\Omega}_j^{-1}}{T_j - T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}\varepsilon_t\varepsilon_t^\top\right) \\ &= -\frac{Tp}{2}\text{log}2\pi - \frac{T}{2}\sum_{j=1}^{m}\rho_j \text{log}|\hat{\Omega}_j| - \frac{1}{2}\sum_{j=1}^{m}(T_j - T_{j-1})\text{tr}\left(I_p\right) \\ &= -\frac{Tp}{2}(\text{log}2\pi + 1) - \frac{T}{2}\sum_{j=1}^{m}\rho_j \text{log}|\hat{\Omega}_j|. \end{aligned} \tag{3.29, 3.30}$$

From (3.30), it is evident that the maximum likelihood estimates of the remaining parameters (ϕ and $\psi = (\psi_A^\top, \psi_C^\top)^\top$) that defines the VECM can be found solving the following

$$\max_{\phi,\psi} -\frac{T}{2}\sum_{j=1}^{m}\rho_j \text{log}\left|\hat{\Omega}_j(\phi, \psi)\right|$$

or more explicitly

$$\max_{\phi,\psi} -\frac{T}{2}\sum_{j=1}^{m}\rho_j \log\left|\frac{1}{T_j - T_{j-1}}\sum_{t=T_{j-1}}^{T_j} (Z_{0t} - A(\psi_A)B(\phi)^\top Z_{1t} - C(\psi_C)Z_{2t})(Z_{0t} - A(\psi_A)B(\phi)^\top Z_{1t} - C(\psi_C)Z_{2t})^\top\right|. \tag{3.31}$$

Appendix 3.B Constructing the Restriction Matrices

In this appendix, we show how to construct the restriction matrices as in Hansen (2003) in order to impose the adequate restrictions when estimating the broken VECM.

Consider a simple LR test of $m_0 = 1$ regime against $m_1 = 2$ regimes where, similarly to our simulation study in Section 5, the breaks affect the long-run coefficient $\tilde{\beta}$. By defining $\mathbf{1}_{1t} \equiv \mathbf{1}(0 \leq t < T_1)$, $\mathbf{1}_{2t} \equiv \mathbf{1}(T_1 \leq t < T_2)$, and $\mathbf{1}_{3t} \equiv \mathbf{1}(T_2 \leq t \leq T)$ we can write the VECM form of the restricted and unrestricted models as

$\mathcal{M}_0$:

$$\underbrace{\begin{bmatrix}\Delta x_{1t}\\ \Delta x_{2t}\end{bmatrix}}_{Z_{0t}} = \underbrace{\begin{bmatrix}\alpha_1 & \alpha_1 & \alpha_1\\ \alpha_2 & \alpha_2 & \alpha_2\end{bmatrix}}_{A} \underbrace{\begin{bmatrix}1 & 0 & 0\\ -\beta_1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -\beta_2 & 0\\ 0 & 0 & 1\\ 0 & 0 & -\beta_2\end{bmatrix}^{\top}}_{B^{\top}} \underbrace{\begin{bmatrix}\mathbf{1}_{1t}x_{1,t-1}\\ \mathbf{1}_{1t}x_{2,t-1}\\ \mathbf{1}_{2t}x_{1,t-1}\\ \mathbf{1}_{2t}x_{2t-1}\\ \mathbf{1}_{3t}x_{1,t-1}\\ \mathbf{1}_{3t}x_{2,t-1}\end{bmatrix}}_{Z_{1t}} + \begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{bmatrix} \tag{3.32}$$

$\mathcal{M}_1$:

$$\underbrace{\begin{bmatrix}\Delta x_{1t}\\ \Delta x_{2t}\end{bmatrix}}_{Z_{0t}} = \underbrace{\begin{bmatrix}\alpha_1 & \alpha_1 & \alpha_1\\ \alpha_2 & \alpha_2 & \alpha_2\end{bmatrix}}_{A} \underbrace{\begin{bmatrix}1 & 0 & 0\\ -\beta_1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -\beta_2 & 0\\ 0 & 0 & 1\\ 0 & 0 & -\beta_3\end{bmatrix}^{\top}}_{B^{\top}} \underbrace{\begin{bmatrix}\mathbf{1}_{1t}x_{1,t-1}\\ \mathbf{1}_{1t}x_{2,t-1}\\ \mathbf{1}_{2t}x_{1,t-1}\\ \mathbf{1}_{2t}x_{2t-1}\\ \mathbf{1}_{3t}x_{1,t-1}\\ \mathbf{1}_{3t}x_{2,t-1}\end{bmatrix}}_{Z_{1t}} + \begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{bmatrix} \tag{3.33}$$

Therefore, the null model $\mathcal{M}_0$ assumes the same specification as $\mathcal{M}_1$ with the only difference that $\beta_3 = \beta_2$. So, given $\phi_0 = (\beta_1, \beta_2)^{\top}$, $\phi_1 = (\beta_1, \beta_2, \beta_3)^{\top}$ and

$\psi = (\alpha_1, \alpha_2)^\top$, we can define the following restriction matrices

$$H_1 = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ \underset{(6\times1)}{\mathbf{0}} & \underset{(6\times1)}{\mathbf{0}} & \underset{(6\times1)}{\mathbf{0}} \\ 0 & 0 & 0 \\ 0 & -1 & 0 \\ \underset{(6\times1)}{\mathbf{0}} & \underset{(6\times1)}{\mathbf{0}} & \underset{(6\times1)}{\mathbf{0}} \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad H_0 = \begin{bmatrix} 0 & 0 \\ -1 & 0 \\ \underset{(6\times1)}{\mathbf{0}} & \underset{(6\times1)}{\mathbf{0}} \\ 0 & 0 \\ 0 & -1 \\ \underset{(6\times1)}{\mathbf{0}} & \underset{(6\times1)}{\mathbf{0}} \\ 0 & 0 \\ 0 & -1 \end{bmatrix}, \quad h = \begin{bmatrix} 1 \\ 0 \\ \underset{(6\times1)}{\mathbf{0}} \\ 1 \\ 0 \\ \underset{(6\times1)}{\mathbf{0}} \\ 1 \\ 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

such that $\text{vec}(B_0) = H_0\phi_0 + h$, $\text{vec}(B_1) = H_1\phi_1 + h$, and $\text{vec}(A) = G\psi$.

Table 3.1: Empirical Size for DGP1.

	$\alpha_1 = -1$	$\alpha_2 = 0$	$\alpha_1 = -0.2$	$\alpha_2 = 0$	$\alpha_1 = -0.01$	$\alpha_2 = 0$	$\alpha_1 = -0.5$	$\alpha_2 = 0.5$
T	*p*	*sp*	*p*	*sp*	*p*	*sp*	*p*	*sp*
0 breaks								
100	6.1	6.1	6.4	6.9	8.6	8.2	4.7	5.2
200	5.4	5.9	5.4	5.2	8.8	8.6	4.4	4.4
300	4.4	5.2	4.0	4.2	6.5	6.7	5.9	5.8
400	5.8	5.4	7.1	6.8	6.1	6.0	4.9	5.1
500	4.8	5.3	4.8	4.6	6.0	6.5	5.6	5.1
1 break								
$T_1 = T/2$								
100	4.8	4.4	5.4	6.0	6.1	6.3	4.7	4.9
200	6.0	6.3	5.5	5.5	6.9	6.6	5.4	5.4
300	5.0	4.8	5.2	5.1	6.1	5.1	5.1	4.9
400	7.0	6.1	6.5	5.8	7.2	7.3	5.7	6.2
500	4.9	4.7	4.7	4.6	5.4	4.6	5.5	5.9
$T_1 = 20$								
100	4.5	4.4	5.8	5.1	9.5	9.2	5.2	4.9
200	7.1	6.8	7.3	7.4	6.5	6.3	4.9	4.7
300	4.6	4.8	4.9	5.0	6.1	5.6	6.1	6.7
400	5.6	5.5	5.3	5.5	5.4	5.9	7.0	6.7
500	4.7	4.9	4.9	5.3	5.4	4.6	5.7	5.9
$T_1 = T - 20$								
100	6.0	5.3	4.4	4.6	8.2	7.3	5.2	4.3
200	5.7	5.7	5.6	5.3	6.5	6.5	4.3	4.7
300	5.3	5.2	5.0	5.0	6.4	6.6	6.4	6.1
400	6.0	5.5	5.0	5.3	7.0	7.3	5.1	4.9
500	5.0	4.8	4.3	4.9	6.4	6.7	6.0	6.3
2 breaks								
$T_1 = T/2\, T_2 = 2T/3$								
100	4.7	4.5	5.2	5.6	6.7	6.7	5.3	4.7
200	5.7	6.2	5.7	5.6	6.3	6.5	4.5	4.3
300	5.0	4.8	5.3	5.5	5.6	6.4	5.7	4.9
400	6.0	5.6	5.7	5.7	6.6	7.3	6.0	5.9
500	4.6	4.9	5.4	4.8	4.4	4.4	4.9	5.1
$T_1 = 20\, T_2 = 2T/3$								
100	5.8	5.7	5.6	5.6	8.2	8.2	4.5	3.9
200	6.0	6.7	6.0	5.8	6.8	7.4	4.7	5.4
300	6.5	5.7	6.2	5.9	5.3	5.1	6.4	6.1
400	6.1	6.0	6.0	5.9	8.7	8.1	5.4	5.4
500	4.8	5.3	5.1	5.7	5.6	5.5	5.7	5.9
$T_1 = T - 20\, T_2 = T/2$								
100	4.8	5.1	5.2	5.4	7.0	6.7	6.0	5.5
200	5.7	5.4	5.3	5.1	5.6	6.1	4.3	4.4
300	5.7	5.1	4.9	4.9	5.5	5.4	6.3	5.7
400	5.3	5.4	5.7	5.3	6.3	6.8	6.3	6.6
500	5.4	5.8	5.0	5.7	4.5	4.8	5.7	5.7

Note: Nominal level used to compute the empirical rejection frequencies is 5%. "p" indicates parametric bootstrap while "sp" semi-parametric bootstrap.

Table 3.2: Empirical Power for DGP1

	$\alpha_1 = -1$	$\alpha_2 = 0$	$\alpha_1 = -0.2$	$\alpha_2 = 0$	$\alpha_1 = -0.01$	$\alpha_2 = 0$	$\alpha_1 = -0.5$	$\alpha_2 = 0.5$
	p	*sp*	*p*	*sp*	*p*	*sp*	*p*	*sp*
				$T_1 = T/2$				
β_2				$T = 100$				
1.1	58.0	58.1	10.0	8.9	8.5	8.1	27.4	28.1
1.3	96.7	96.9	29.0	28.5	9.0	8.2	77.8	77.5
1.5	99.9	100.0	52.1	50.9	9.3	8.3	94.6	94.5
2.0	99.9	100.0	84.0	84.4	9.4	9.1	99.5	99.3
β_2				$T = 300$				
1.1	94.5	94.5	26.4	25.9	6.4	6.5	71.2	70.4
1.3	98.3	98.5	73.5	73.4	6.7	6.6	99.6	99.6
1.5	100.0	100.0	93.3	93.7	7.3	6.8	100.0	100.0
2.0	100.0	100.0	96.4	96.3	10.2	9.5	100.0	100.0
β_2				$T = 500$				
1.1	99.6	99.8	44.3	43.8	6.3	6.6	91.4	91.5
1.3	100.0	100.0	95.1	94.8	7.0	7.3	100.0	100.0
1.5	100.0	100.0	99.4	99.4	7.9	9.3	100.0	99.8
2.0	100.0	100.0	99.6	99.6	14.6	14.8	100.0	100.0
				$T_1 = T/2\, T_2 = 2T/3$				
β_3				$T = 100$				
2.1	55.0	54.6	9.4	9.7	7.2	7.1	18.1	17.6
2.3	88.9	88.9	27.6	27.5	7.5	7.8	56.3	56.1
2.5	96.7	96.5	49.3	47.8	8.4	7.9	76.4	76.5
3.0	99.9	99.9	71.2	69.9	9.5	9.2	96.3	96.5
β_3				$T = 300$				
2.1	88.2	88.0	19.5	19.7	6.5	6.3	51.3	51.8
2.3	98.2	98.3	55.7	55.9	7.5	6.7	90.6	90.7
2.5	98.3	98.2	67.9	67.4	8.5	8.1	98.4	98.8
3.0	97.2	97.3	80.6	79.8	10.8	10.4	100.0	100.0
β_3				$T = 500$				
2.1	96.1	96.0	36.1	36.4	6.1	6.0	73.0	72.8
2.3	99.2	99.2	67.0	66.7	8.3	8.0	97.6	97.8
2.5	99.2	99.2	79.4	79.0	8.5	8.1	99.8	99.8
3.0	99.7	99.7	85.9	85.4	9.3	9.3	100.0	100.0
				$T_1 = T/2\, T_2 = 2T/3\, T_3 = 5T/6$				
β_4				$T = 100$				
1.6	50.2	49.6	7.3	7.0	6.0	6.3	19.0	19.2
1.8	84.0	83.7	26.2	26.4	6.4	6.8	56.9	56.2
2.0	94.4	94.1	45.3	45.4	6.7	6.7	74.5	74.3
2.5	99.7	99.7	69.0	69.0	7.3	7.6	92.3	92.1
β_4				$T = 300$				
1.6	81.8	81.4	21.9	21.8	7.0	6.2	52.9	52.8
1.8	99.4	99.3	58.7	57.6	6.8	6.6	88.5	88.4
2.0	100.0	100.0	65.0	64.3	7.0	7.2	97.1	97.0
2.5	100.0	100.0	78.3	77.5	11.0	11.8	100.0	100.0
β_4				$T = 500$				
1.6	92.6	92.8	43.9	43.4	4.8	4.5	71.0	71.4
1.8	100.0	99.9	77.6	76.6	5.3	5.0	97.2	97.2
2.0	100.0	100.0	82.3	82.1	6.1	5.5	99.5	99.5
2.5	100.0	100.0	88.2	87.4	12.5	12.7	100.0	100.0

Note: See notes below Table 3.1.

Table 3.3: Empirical Rejection Frequencies for DGP2 with Weak Exogeneity Imposed

(a) Size

	$a_1=0,\rho=0$			$a_1=0,\rho=0.8$			$a_1=0,\rho=0.99$		
T	*w*	*b-p*	*b-sp*	*w*	*b-p*	*b-sp*	*w*	*b-p*	*b-sp*
0 breaks									
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
1 break									
$T_1=T/2$									
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
2 breaks									
$T_1=T/2\,T_2=2T/3$									
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

	$a_1=0,\rho=0$			$a_1=0,\rho=0.8$			$a_1=0,\rho=0.99$		
	w	*b-p*	*b-sp*	*w*	*b-p*	*b-sp*	*w*	*b-p*	*b-sp*
$T_1=T/2$									
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
$T_1=T/2\,T_2=2T/3$									
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
$T_1=T/2\,T_2=2T/3\,T_3=5T/6$									
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: 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 3.4: Empirical Rejection Frequencies for DGP2 with Weak Exogeneity Relaxed

(a) Size

	$a_1 = 1, \rho = 0$			$a_1 = 1, \rho = 0$			$a_1 = 1, \rho = 0$		
T	*w*	*b-p*	*b-sp*	*w*	*b-p*	*b-sp*	*w*	*b-p*	*b-sp*
	0 breaks			**1 break**			**2 breaks**		
				$T_1 = T/2$			$T_1 = T/2\ T_2 = 2T/3$		
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

	$a_1 = 1, \rho = 0$				$a_1 = 1, \rho = 0$				$a_1 = 1, \rho = 0$			
	w	*b-p*	*b-sp*		*w*	*b-p*	*b-sp*		*w*	*b-p*	*b-sp*	
	$T_1 = T/2$				$T_1 = T/2\ T_2 = 2T/3$				$T_1 = T/2\ T_2 = 2T/3\ T_3 = 5T/6$			
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_4	$T = 500$			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: See notes below Table 3.3.

Table 3.5: Q-statistic Results

A.				$\boldsymbol{\Theta^* = \{\beta, \Omega\}}$				
	# 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
$\boldsymbol{Q}$ $\boldsymbol{p}$-value = $\mathbf{0.21405}$								
B.				$\boldsymbol{\Theta^* = \{\alpha, \beta, \Omega\}}$				
	# 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
$\boldsymbol{Q}$ $\boldsymbol{p}$-value = $\mathbf{0.07692}^*$								
C.				$\boldsymbol{\Theta^* = \{\alpha, \beta, \Omega\}}$				
	# 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
$\boldsymbol{Q}$ $\boldsymbol{p}$-value = $\mathbf{0.04013}^{**}$								
D.				$\boldsymbol{\Theta^* = \{\alpha, \beta, \Omega\}}$				
	# 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
$\boldsymbol{Q}$ $\boldsymbol{p}$-value = $\mathbf{0.59532}$								

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-statistic p-values are obtained by semi-parametric block-regime bootstrap (see Section 4.2) with 299 replications.

Table 3.6: λ_{trace} test for cointegration.

rank $\leq$	λ_{trace}	p-value
Full Sample: 1981(1)-2014(1)		
0	29.012	[0.062]
1	6.3404	[0.660]
2	0.047966	[0.827]
Reduced Sample: 1981(1)-2006(4)		
0	42.999	[0.001] **
1	13.185	[0.108]
2	0.49446	[0.482]

Table 3.7: Q-statistic Results

				Test A			
				$\Theta^* = \{\beta, \Omega\}$			
	# breaks	Break1	Break2	log-lik	$LR(\mathcal{H}_0\|\mathcal{H}_1)$	DoF	p-value
$\mathcal{H}_0$	0			1409.40			0.00000
$\mathcal{H}_1 1$	1	2008(3)		1419.20	19.561	2	0.00006
$\mathcal{H}_1 2$	1	2011(4)		1411.10	3.278	2	0.19417
$\mathcal{H}_1 3$	2	2008(3)	2011(4)	1420.00	21.143	4	0.00030
				Q* p-value = 0.0402			
				$\Theta^* = \{\alpha, \beta, \Omega\}$			
	# breaks	Break1	Break2	log-lik	$LR(\mathcal{H}_0\|\mathcal{H}_1)$	DoF	p-value
$\mathcal{H}_0$	0			1409.40			0.00000
$\mathcal{H}_1 1$	1	2008(3)		1425.40	31.866	5	0.00001
$\mathcal{H}_1 2$	1	2011(4)		1412.00	5.055	5	0.40927
$\mathcal{H}_1 3$	2	2008(3)	2011(4)	1432.50	46.145	10	0.00000
				***Q* p-value = 0.00000**			

				Test B			
				$\Theta^* = \{\beta, \Omega\}$			
	# breaks	Break1		log-lik	$LR(\mathcal{H}_0\|\mathcal{H}_1)$	DoF	p-value
$\mathcal{H}_0$	0			1409.40			0.00000
$\mathcal{H}_1 1$	1	2008(3)		1419.20	19.561	2	0.00006
$\mathcal{H}_1 2$	1	2011(4)		1411.10	3.278	2	0.19417
				Q* p-value = 0.0302			
				$\Theta^* = \{\alpha, \beta, \Omega\}$			
	# breaks	Break1		log-lik	$LR(\mathcal{H}_0\|\mathcal{H}_1)$	DoF	p-value
$\mathcal{H}_0$	0			1409.40			0.00000
$\mathcal{H}_1 1$	1	2008(3)		1425.40	31.866	5	0.00001
$\mathcal{H}_1 2$	1	2011(4)		1412.00	5.055	5	0.40927
				***Q* p-value = 0.00502**			

				Test C			
				$\Theta^* = \{\beta, \Omega\}$			
	# breaks	Break1	Break2	log-lik	$LR(\mathcal{H}_0\|\mathcal{H}_1)$	DoF	p-value
$\mathcal{H}_0$	0			1409.40			0.00000
$\mathcal{H}_1 1$	1	2008(3)		1419.20	19.561	2	0.00006
$\mathcal{H}_1 2$	2	1992(2)	2008(3)	1421.00	23.083	4	0.00012
				***Q* p-value = 0.00503**			
				$\Theta^* = \{\alpha, \beta, \Omega\}$			
	# breaks	Break1	Break2	log-lik	$LR(\mathcal{H}_0\|\mathcal{H}_1)$	DoF	p-value
$\mathcal{H}_0$	0			1409.40			0.00000
$\mathcal{H}_1 1$	1	2008(3)		1425.40	31.866	5	0.00001
$\mathcal{H}_1 2$	2	1992(2)	2008(3)	1436.80	54.691	10	0.00000
				***Q* p-value = 0.00000**			

Note: The Q-statistic p-values are obtained by semi-parametric block-regime bootstrap (see Section 3.4.2) with $B = 299$ replications. * denotes acceptance of $\mathcal{H}_0$ at 1% level.

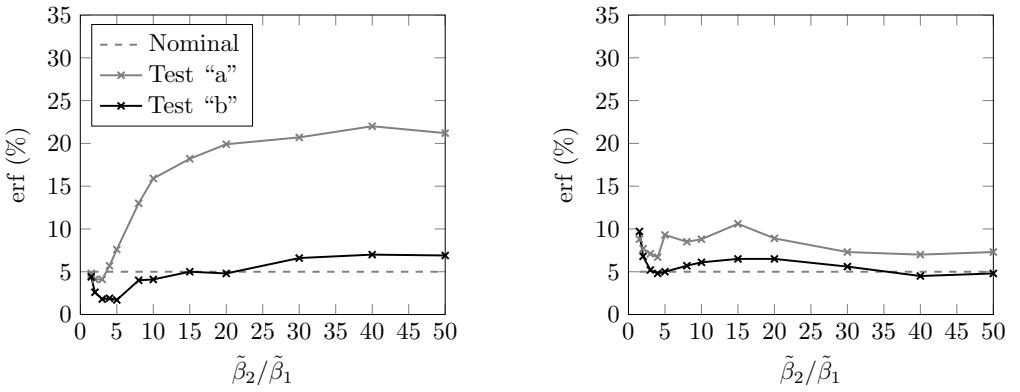
Table 3.8: λ_{trace} test for cointegration.

rank $\leq$	λ_{trace}	p-value
0	34.293	[0.013] *
1	6.6910	[0.619]
2	0.38880	[0.533]

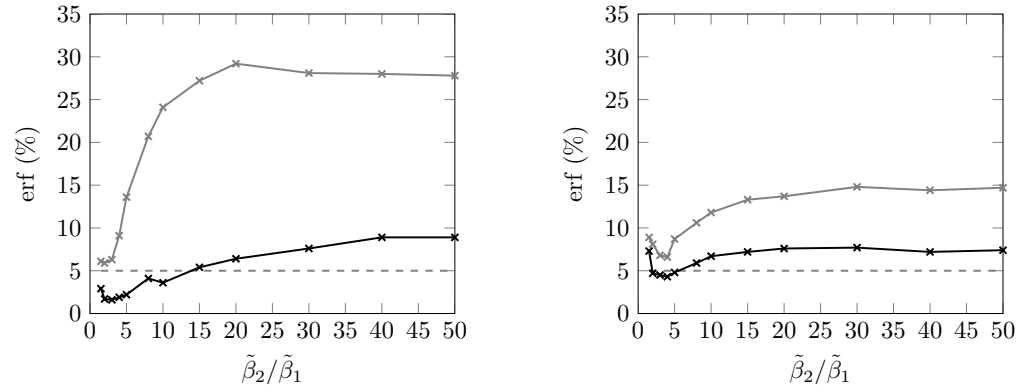
Table 3.9: Bootstrapped LR Tests for Breaks

Break Date(s)		p-value	
$\mathcal{H}_0$	$\mathcal{H}_1$	param	non-param
$\Theta^* = \{\beta, \Omega\}$			
0	2003(1)	0.0635^*	0.0435^{**}
$\Theta^* = \{\alpha, \beta, \Omega\}$			
0	2003(1)	0.2609	0.2241

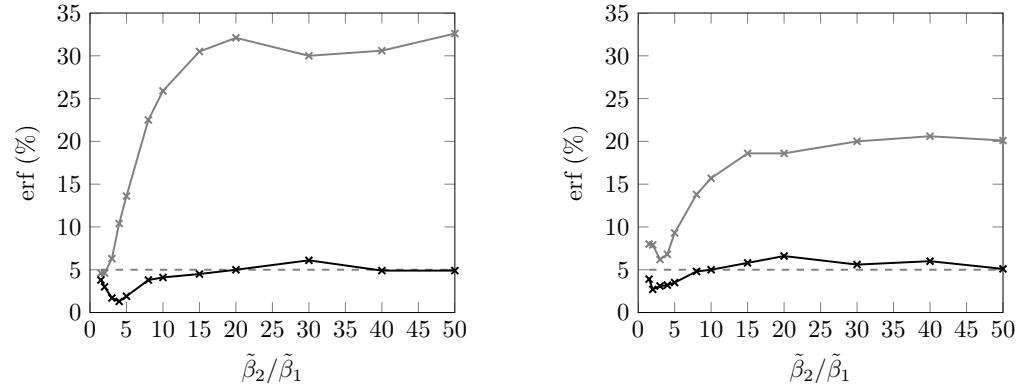
Note: The reported p-values are obtained by block-regime bootstrap (see Section) with $B = 299$ replications.



(a) $T = 100$, $\rho = 0$ **(b)** $T = 100$ $\rho = 0.8$



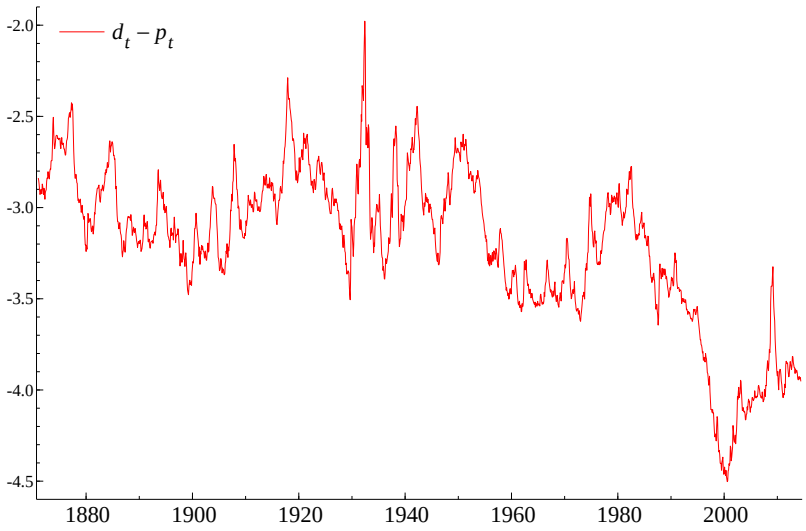
(c) $T = 300$, $\rho = 0$ **(d)** $T = 300$, $\rho = 0.8$

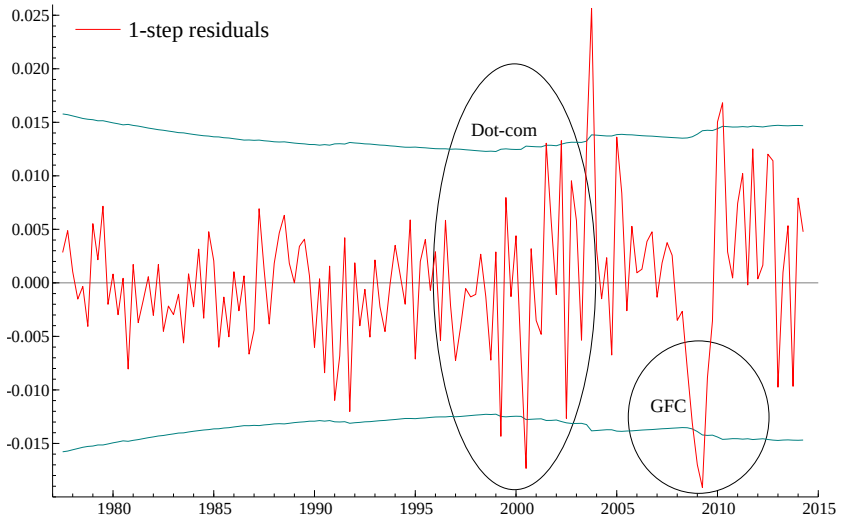


(e) $T = 500$ $\rho = 0$ **(f)** $T = 500$ $\rho = 0.8$

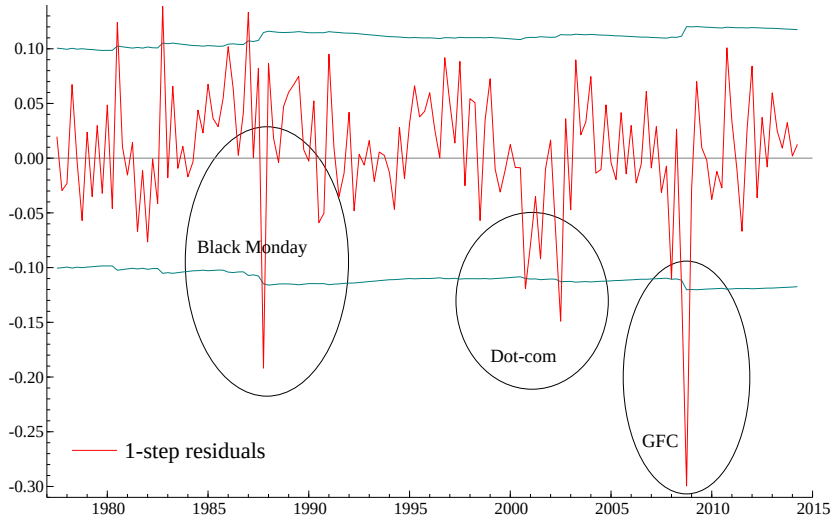
Figure 3.1: The above graphs plot on the "x" axis the magnitude of the break, i.e. the ratio between the long-run coefficient post-break ($\tilde{\beta}_2$) and pre-break ($\tilde{\beta}_1$), while on the "y" axis the empirical rejection frequencies of $\mathcal{H}_0$ for Case"A" and for Case "B". Different sample sizes T and values of ρ are explored.

Figure 3.2: Log Dividend-Price Ratio 1871-2014



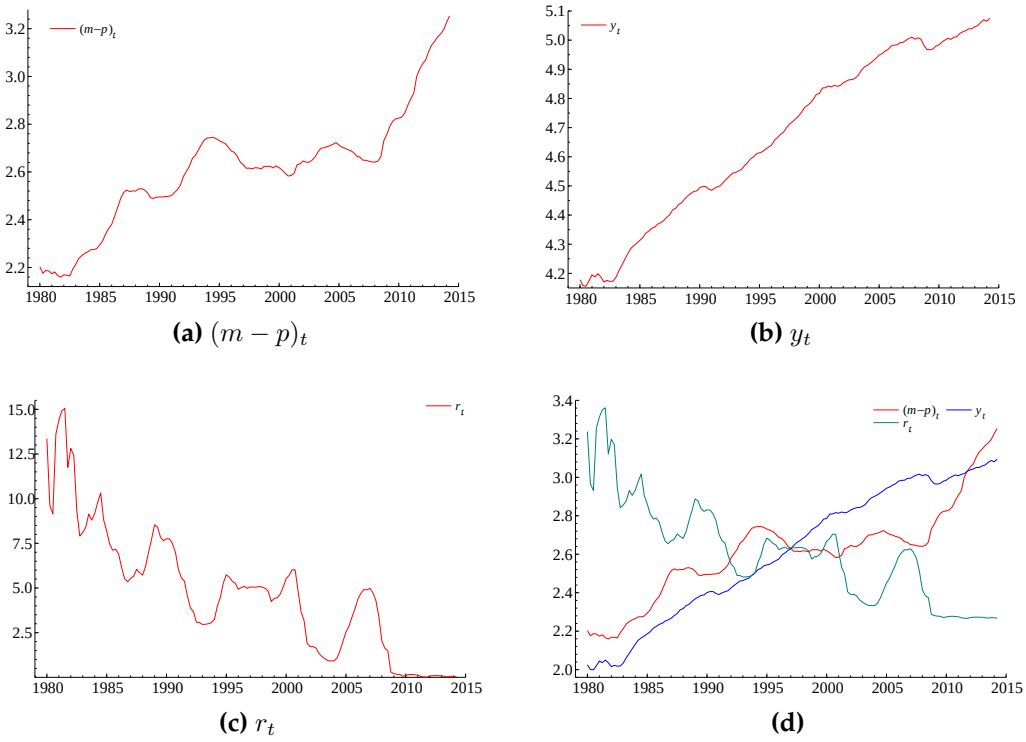


(a) d_t



(b) p_t

Figure 3.3: The graphs plot the one-step ahead residuals together with $\pm 2\hat{\sigma}_{res}$ for both the dividend (d_t) and the price (p_t) series. The residuals are obtained by recursive estimation of the VECM over the period 1960(1)-2014(2) with an initial window of 50 observations.



(a) $(m-p)_t$ **(b)** y_t

(c) r_t **(d)**

Figure 3.4: Money Demand Variables

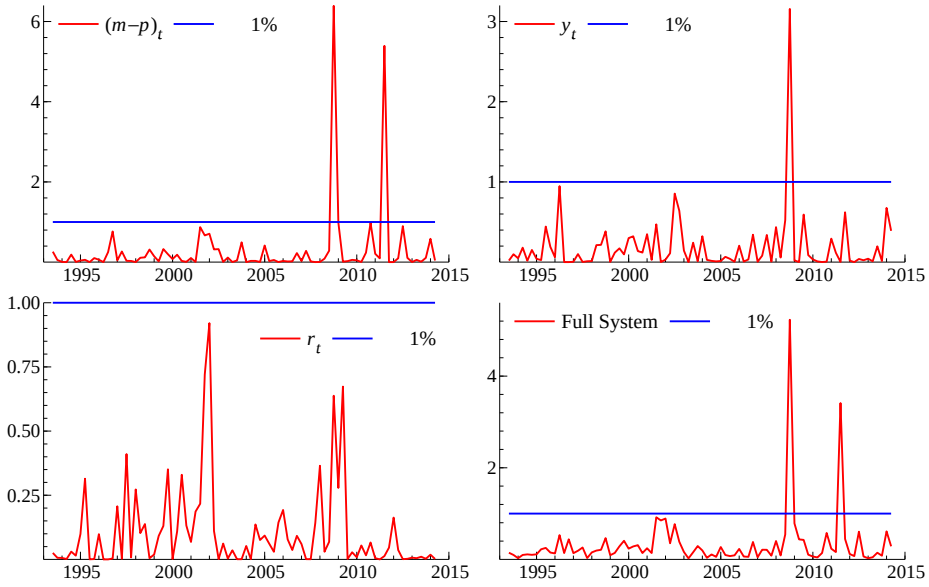


Figure 3.5: 1-step Chow stability test over the period 1981(1)-2014(2).

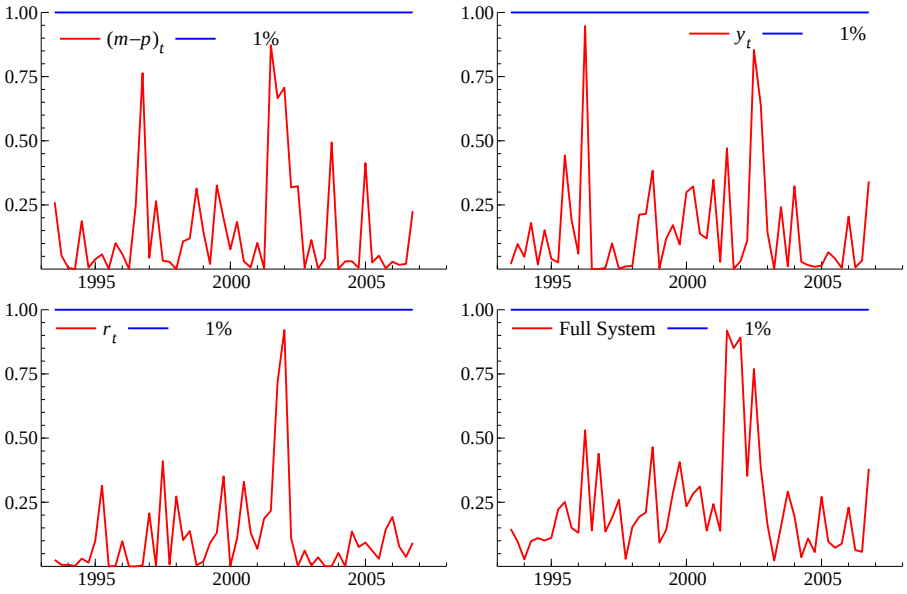
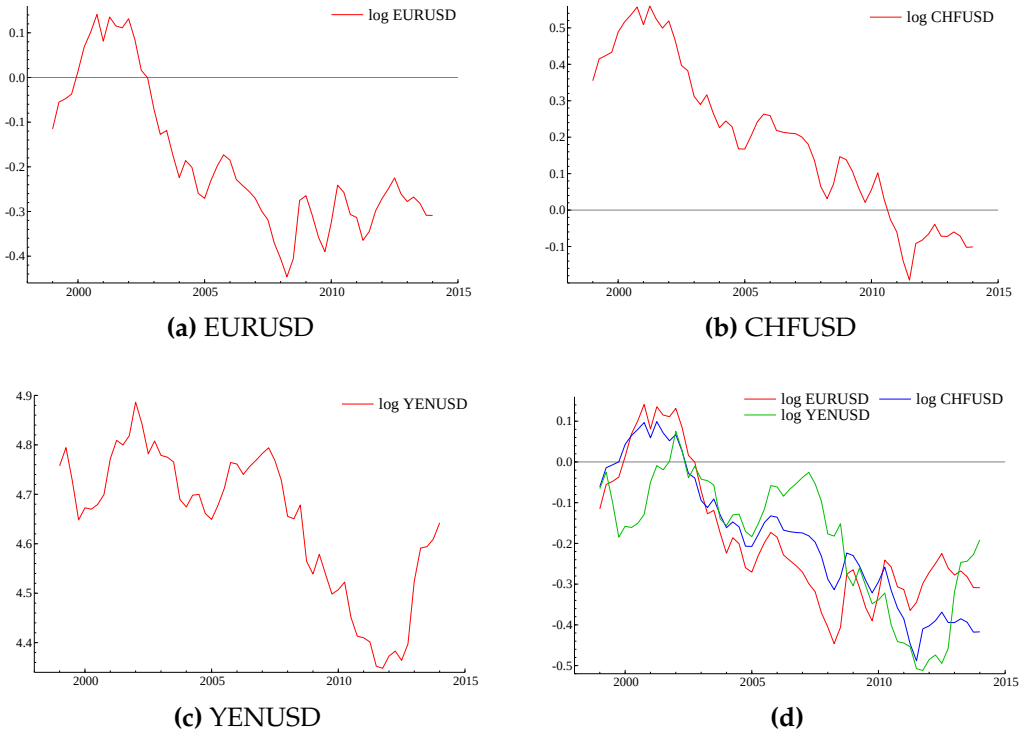


Figure 3.6: 1-step Chow stability test over the period 1981(1)-2006(4).



(a) EURUSD **(b)** CHFUSD

(c) YENUSD **(d)**

Figure 3.7: Log Exchange Rates

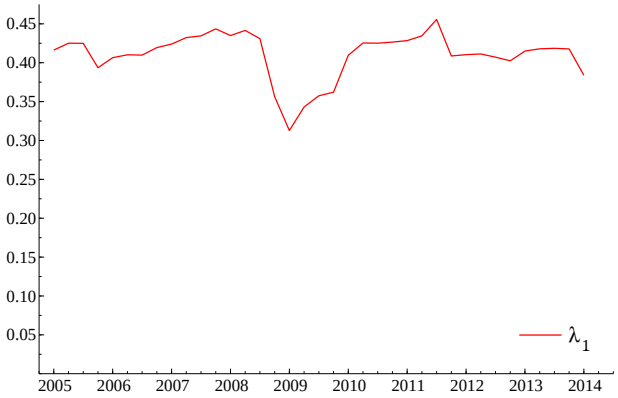


Figure 3.8: Recursive Plot of $\hat{\lambda}_1$ (Eigenvalue associated with the cointegrating vector)

Chapter 4

Robust Estimation of Real Exchange Rate Process Half-life*

4.1 Introduction

The purchasing power parity (PPP) condition holds if the real exchange rate process is stationary and it reverts to its mean in about one to two years, the necessary time to absorb financial and monetary shocks which are thought to affect the exchange rate. Over the last thirty years, the validity of the PPP condition has been subject to several empirical tests by means of alternative econometric techniques. Sarno and Taylor (2002) provide an extensive literature review from the early seventies to recent years (see also James et al., 2012).

*A research paper based on this chapter has been submitted for publication. Preliminary versions of this chapter have been presented at the 15th OxMetrics User Conference (Cass Business School, September 4-5 2014), at the VII International conference of the ERCIM WG (Pisa, 6-8 December 2014), at the XXXIX Spanish Economic Symposium (Palma de Mallorca, 11-13 December 2014), and at the RES post-graduate meeting (University College London, 9-10 January 2015).

The main empirical evidence is that the PPP is valid in the long-run while substantial deviations are usually observed in the short-run. In other terms, the process followed by the real exchange rate is found to be stationary but with high persistence. To measure the degree of persistence of the real exchange rate process, the concept of *half-life* is usually employed (see Mark, 2001; Rossi, 2005; Chortareas and Kapetanios, 2013). This is defined as the period of time necessary for the real exchange rate process to dissipate by half a unitary shock (or its cumulative effect) and it is commonly used as a measure to quantify to which extent the purchasing power parity condition holds. Empirical findings seem to confirm that the half-life of the real exchange rate ranges from about three to five years (see Frankel and Rose, 1996, for an extended study involving over a hundred of countries), giving rise to the so called "PPP puzzle" as defined in Rogoff (1996). Taylor (2001) explores possible sources of bias which might explain the puzzling half-lives measures obtained in the literature and he suggests that the puzzle might be mitigated by taking into account data aggregation issues and by allowing for non-linear dynamics in the real exchange rate process. Notwithstanding, together with the correct specification of the statistical framework used to model the real exchange rate process, correct inference for the model parameters is crucial for obtaining reliable half-life measures.

In this chapter, we argue that the data generating process of the real exchange rate is likely to include outliers that, if not accounted for, distort the estimated half-lives since they alter the autocorrelation structure of the observed time series (see for instance Tsay, 1986). Indeed, from a visual inspection of the USD/GBP real exchange rate time series in Figure 4.1, it is evident that our conjecture is not completely at odds with empirical evid-

ence.

[Figure 4.1 about here.]

This chapter contributes to the literature on the PPP puzzle in three directions.

First, we propose a new framework to model the real exchange rate process. In particular, we allow the real exchange rate to follow an ARMA process contaminated with additive and innovative outliers as well as level shifts. Further, we devise a fast and accurate procedure to estimate the half-life of the ARMA process in this framework.

Second, in order to estimate the outlier-contaminated model, we consider an extension of the Dummy Saturation approach introduced by Hendry (1999) and Hendry et al. (2008) which considers saturating in turn with additive outliers, innovative outliers and level shifts in a maximum likelihood framework. The performance of the procedure, in terms of retention rate of the insignificant outliers, is explored using a Monte Carlo simulation.

Finally, in order to show the severity of the effect of unaccounted outliers on the estimation of the half-life, we carry out empirical application involving US dollar real exchange rates for a group of developed countries. The results we obtain are consistent with our claim. As a matter of fact, we find that half-life estimates can indeed change dramatically when outlying observations are accounted for. In particular, we find that estimated half-lives are considerably shorter for the exchange rates which exhibit the most puzzling behaviour than when computing the half-life disregarding the effect of outliers.

The reminder of the chapter is organised as follows: Section 4.2 briefly reviews the existing literature focusing on the existing approaches to model the real exchange rate process and computing the half-life; Section 4.3 presents our model and describes the estimation approach; Section 4.4 reports the empirical application and Section 4.5 concludes.

4.2 Review of the Literature

The real exchange rate process is usually regarded as a measure of deviation from the PPP relationship. In particular, if we denote with S_t the spot nominal exchange rate while with $P^h_{i,t}$ and $P^f_{i,t}$ the domestic and foreign price of the i^{th} good respectively, the PPP condition can be formulated as a generalisation of the law of one price to a basket of goods, such that

$$S_t = \frac{\sum_{i=1}^{N} w_i P^h_{i,t}}{\sum_{i=1}^{N} w_i P^f_{i,t}} = \frac{\bar{P}^h_t}{\bar{P}^f_t} \qquad t = 1, \dots, T \tag{4.1}$$

where N denotes the number of goods in the basket and w_i, such that $\sum_{i=1}^{N} w_i = 1$, denotes the weight assigned to the i^{th} good. The logarithmic form of the PPP is then given by

$$s_t = \bar{p}^h_t - \bar{p}^f_t\,, \tag{4.2}$$

and from (4.2), the real exchange rate is defined as

$$q_t \equiv s_t - \bar{p}^h_t + \bar{p}^f_t \tag{4.3}$$

which should then be equal to 0 for all $t = 1, \dots, T$ if PPP holds exactly. In general, however, $\{q_t\}$ follows a stochastic process which properties can be

used to study the deviations from PPP. Empirically, as reported in Rogoff (1996), we usually proxy the price of the goods in one country by the corresponding CPI (of traded goods), where $CPI_t^h = \bar{P}_t^h / \bar{P}_0^h$ and $CPI_t^f = \bar{P}_t^f / \bar{P}_0^f$ with the subscript "0" denoting some chosen base year [1]. In practice, what we are actually computing when using the CPI (or every other price index) is then

$$s_t - (\bar{p}_t^h - \bar{p}_0^h) + (\bar{p}_t^f - \bar{p}_0^f) = q_t + \bar{p}_0^h - \bar{p}_0^f = q_t + const \tag{4.4}$$

which implies that the real exchange rate obtained in this way has an expectation different from zero by construction.

In the economic literature, we can find two different ways to model the process followed by $\{q_t\}$. The vast majority of the empirical studies models the real exchange rate process according to linear dynamic models assuming either an AR(1) (Abuaf and Jorion, 1990) or an AR(p) (Rossi, 2005; Chortareas and Kapetanios, 2013) structure. As noted in Chortareas and Kapetanios (2013), assuming a simplistic AR(1) is however suboptimal whenever the dynamics followed by the real exchange rate process can be captured by higher order models although the computation of the half-life complicates. Further, we think that the use of ARMA processes has been largely overlooked in favour of simple AR processes. The reason is not completely clear thus one explanation might lie in the fact that the computation of the half-life is relatively less straightforward. To the best of our knowledge, the only exceptions are Diebold et al. (1991) and Cheung and Lai (2000) where the authors allow $\{q_t\}$ to follow a long-memory process of

[1] We are assuming to be in the ideal situation where the price index is computed from the same basket of goods in the two countries. In practice, this is almost never satisfied and together with the way in which the weights are formed is one of the main source of the index problem.

the ARFIMA class and a stationary ARMA process, respectively. All these studies, with the exception of Chortareas and Kapetanios (2013) who adopt a different half-life measure, find that PPP holds in the long-run whereas substantial deviations occurs in the short-run. Another stream of literature, building on the "bands of inaction" argument raised in Taylor (2001), considers instead non-linear dynamic models like self-exciting threshold autoregression (SETAR) and smooth transition autoregression (STAR). According to Taylor (2001), bands of inaction due to transaction costs, where the real exchange rate behaves like a random walk, would bias upward the autoregressive coefficient giving rise to half-lives bigger than those that would be observed if measuring the half-life outside these bands.

As mentioned in the introduction, testing the PPP condition involves the computation of a measure of persistence of the process followed by $\{q_t\}$. Typically, this is done by computing the *half-life* of the process[2]. According to Mark (2001) and Rossi (2005), the half-life is formally defined as the smallest h such that $\mathbb{E}(q_{t+h} - q_0 | q_{t-s} - q_0, s \leq 0) \leq \frac{1}{2}(q_t - q_0)$, i.e. the time necessary for q_t to revert back at half its initial post shock value. Denoting with $\psi(t)$ for $t \geq 0$ the impulse response function (IRF) of $\{q_t\}$ and considering an initial unitary shock to give $\psi(0) = 1$, the above definition of half-life corresponds to the instant h such that

$$\psi(h) = \frac{1}{2}. \tag{4.5}$$

More in detail, Mark (2001) provides a formula to compute half-lives for stationary AR(p) processes while Rossi (2005) derives an asymptotic approx-

[2]Alternative measures are given in Andrews and Chen (1994).

imation and confidence intervals for the half-life of AR(p) processes when one root of the characteristic polynomial is close to unity. Their results are consistent with the puzzling behaviour of the PPP condition.

Recently, Chortareas and Kapetanios (2013) proposed an alternative definition of half-life based on the decline of the cumulative effect of a shock rather than its point value. In particular, half-life is defined as the instant h such that half of this cumulative effect has dissipated. In terms of the IRF $\psi(t)$, it can be expressed as

$$\int_0^h f[\psi(i)]\mathrm{d}i = \int_h^\infty f[\psi(i)]\mathrm{d}i \tag{4.6}$$

with $f[\psi(i)] = |\psi(i)|$, to accommodate negative values of the IRF, or $f[\psi(i)] = \psi(i)^2$ to take also into account the possibility of long-memory processes. The authors apply this new definition of half-life using US dollar exchange rates for a set of developed countries for the period 1957:Q1-1998:Q4 and finding a reduction in the half-lives with respect to previous studies. Notwithstanding, in this chapter, we adopt the standard definition of half-life for ease of comparison with the previous results and given that the PPP puzzle has been built around this definition. In general, however, it is clear that whichever of the two definitions we decide to adopt, the computation of the half-life depends on the estimated parameters of the underlying model and thus, outliers robust methods are equally advised.

In what follows, we extend the current literature not restricting $\{q_t\}$ to follow only AR(p) processes but considering more general stationary ARMA(p, q) processes contaminated by outliers and level shifts (see Section 4.3). Further, we describe a fast method to compute exact half-lives for ARMA(p, q)

processes according to (4.5) (see Section 4.3.3).

4.3 Real Exchange Rate Process: an Outliers-based Approach

We assume that the real exchange rate process $\{q_t\}$ is well represented by a stationary ARMA(p, q) process contaminated by outliers and level shifts. In particular, let the process followed by $\{q_t\}$ be described by

$$q_t = q_0 + \sum_{i=1}^{k} \delta_i V_i(L)\mathbf{1}(t = T_i) + v_t \tag{4.7}$$

$$\phi(L)v_t = \theta(L)\varepsilon_t \qquad t = 1, \ldots, T \tag{4.8}$$

where q_0 can be thought as the long-run value of the real exchange rate, k denotes the number of outlying events, δ_i is the outlier or level shift size, $V_i(L)$ (with L denoting the lag operator) defines the outlier type, $\mathbf{1}(t = T_i)$ is an impulse indicator assuming value 1 for $t = T_i$ and 0 otherwise, $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$ are lag polynomials with roots outside the unit circle, and $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. The specification of $V_i(L)$ allows to characterise different kinds of outlying observations (see Tsay, 1988). Three specifications are particularly relevant to our analysis:

$V_i(L) = 1$	Additive Outlier (AO)
$V_i(L) = \phi^{-1}(L)\theta(L)$	Innovative Outlier (IO)
$V_i(L) = (1 - L)^{-1}$	Level Shift (LS) (Since $(1 - L)^{-1}\mathbf{1}(t = T_i) = \mathbf{1}(t \geq T_i)$).

The difference between AO and IO concerns with the way they affect the time series. To better understand their impact, assume to be able to distinguish k^A AOs, k^I IOs and k^L LSs, and rewrite (4.7)-(4.8) in terms of ε_t to give

$$q_t = q_0 + \sum_{i=1}^{k^A} \delta_i^A \mathbf{1}(t = T_i) + \frac{\theta(L)}{\phi(L)} \left(\sum_{l=1}^{k^I} \delta_l^I \mathbf{1}(t = T_l) + \varepsilon_t \right) + \sum_{j=1}^{k^L} \delta_j^L \frac{1}{(1-L)} \mathbf{1}(t = T_k). \tag{4.9}$$

From equation (4.9), we can see that an IO affects directly the innovation process and thus it propagates to future observations through the multiplier $\phi(L)^{-1}\theta(L)$ while an AO just affect the single observation at the time of the shock. Finally, a LS produces a shift in the long-run value of the real exchange rate from the time of the shock onwards. Figure 4.2 gives an idea of how the impact of the different outliers looks like using a simulated process.

[Figure 4.2 about here.]

As largely documented in the statistical and econometric literature (see Chang et al., 1988; Tsay, 1988; Chen and Liu, 1993; Sánchez and Peña, 2003; Cavaliere and Georgiev, 2009, and references therein), outlying observations distort the autocorrelation structure of the time series under exam. This has an impact on the identification of the appropriate ARMA order and most importantly it causes biases in the estimated ARMA coefficients. The implications for forecasting follow from the just mentioned problems.

Two major consequences for the PPP relationship can be identified. First, strongly upward-biased autoregressive coefficients might lead to the conclusions that $\{q_t\}$ is nonstationary and thus rejecting the existence of PPP

in a first place[3]. Second, distorted ARMA coefficients such that $\{q_t\} \sim I(0)$ can still lead to distorted half-life measures either upwards or downwards, being the half-life computed from these coefficients (see Section 4.3.3). As noted in Tsay (1986), the exact effect (bias direction) of multiple outliers is difficult to quantify and it depends also on the interaction between the outliers themselves (e.g. two outliers with magnitudes of opposite sign can cancel out) and their location in the sample. Sparse results are only available in the literature for a maximum of two outliers and low order ARMA processes which are of scarce interest for our purposes.

The conclusion is that to which extent the bias induced by unaccounted outliers exacerbates or solves the PPP puzzle, it is worth to be investigated.

4.3.1 Estimation Strategy and Outliers Detection

In this section, we consider the problem of estimating equation (4.9). Using a compact matrix notation, the real exchange rate process contaminated with outliers is given by

$$q_t = q_0 + \mathbf{x}_t^\top \boldsymbol{\delta}^A + \mathbf{y}_t^\top \boldsymbol{\delta}^L + \phi^{-1}(L)\theta(L)(\varepsilon_t + \mathbf{z}_t^\top \boldsymbol{\delta}^I) \qquad t = 1, \ldots, T \tag{4.10}$$

$$= q_0 + \mathbf{x}_t^\top \boldsymbol{\delta}^A + \mathbf{y}_t^\top \boldsymbol{\delta}^L + \tilde{\mathbf{z}}_t^\top \boldsymbol{\delta}^I + v_t \qquad v_t = \phi^{-1}(L)\theta(L)\varepsilon_t, \tag{4.11}$$

where $\mathbf{x}_t$, $\mathbf{y}_t$ and $\mathbf{z}_t$ are vectors containing impulse or step dummies of dimension k^A, k^I and k^L, respectively and $\tilde{\mathbf{z}}_t$ denote the filtered IOs vector.

[3]Consider for instance the low power of unit root tests when applied to a stationary processes with level shifts.

Finally, we can write

$$\mathbf{q} = \mathbf{W}\boldsymbol{\delta} + \mathbf{v} \qquad \mathbf{v} = \phi^{-1}(L)\theta(L)\boldsymbol{\varepsilon}, \tag{4.12}$$

where $\mathbf{W}$ is a matrix of size $(1 + k^A + k^L + k^I) \times T$ mostly made up of 0-1 entries with the exception of the last k^I columns (those corresponding to the innovational outliers). Further, under the assumption that $\{\varepsilon_t\}$ is a sequence of IID normally distributed random variables, it follows that

$$\mathbf{v} \sim \mathcal{N}_T(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{\Omega}) \tag{4.13}$$

where $\boldsymbol{\Omega}$ is a symmetric $T \times T$ Toeplitz matrix containing the autocovariances of the ARMA process followed by $\{v_t/\sigma_\varepsilon\}$. The autocovariances are functions of the ARMA parameters ($\boldsymbol{\phi} = [\phi_1, \ldots, \phi_p]^\top$ and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_q]^\top$) and need to be estimated along with the regression coefficients.

Given the form of (4.12), we are in the framework of a regression with ARMA errors where the exogenous regressors are the outlying observations. Typical estimation methods for such regressions are GLS (in particular see Galbraith and Zinde-Walsh, 1992) or ML based on (4.13). If the time-series parameters ($\boldsymbol{\phi}$, $\boldsymbol{\theta}$, σ_ε^2) are known and under normality, GLS and ML yield the same estimator of the regression coefficients, which take on the following form

$$\hat{\boldsymbol{\delta}}_{GLS} = \hat{\boldsymbol{\delta}}_{ML} = (\mathbf{W}^\top \boldsymbol{\Omega}^{-1} \mathbf{W})^{-1} \mathbf{W}^\top \boldsymbol{\Omega}^{-1} \mathbf{q}. \tag{4.14}$$

Assuming a single outlier at $T_1 < T$, the estimators of the impact of the different kind of outliers can be derived from (4.14) and are given by

$$\text{AO} : \hat{\delta}^A(T_1) = (\Omega^{-1})_{T_1,T_1}(\mathbf{\Omega}^{-1})_{T_1,\cdot}\mathbf{q} \tag{4.15}$$

$$\text{LS} : \hat{\delta}^L(T_1) = \left(\sum_{i=T_1}^{T}\sum_{j=T_1}^{T}(\Omega^{-1})_{i,j}\right)^{-1}\left(\sum_{i=T_1}^{T}(\mathbf{\Omega}^{-1})_{i,\cdot}\right)\mathbf{q} \tag{4.16}$$

$$\text{IO} : \hat{\delta}^I(T_1) = \phi(L)\theta^{-1}(L)q_{T_1} \tag{4.17}$$

where $\mathbf{\Omega}_{i,\cdot}$ denotes the i^{th} row of $\mathbf{\Omega}$. Note that the impact of an IO equals the residual at that time (Chang et al., 1988).

In general, however, $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ and σ_ε^2 are unknown and hence, $\mathbf{\Omega}$ needs to be replaced by a consistent estimator $\hat{\mathbf{\Omega}}$. In this situation, (*iterated feasible*) GLS and ML provide different estimators of $\mathbf{\Omega}$, thus yielding different estimates of $\boldsymbol{\delta}$. Hereafter, we will focus exclusively on ML estimation which has the advantage to allow simultaneous estimation of both the ARMA parameters and the regression coefficients. In particular, the log-likelihood function of model (4.12) under the normality assumption (4.13) is given by

$$\ell(\boldsymbol{\delta},\boldsymbol{\phi},\boldsymbol{\theta},\sigma_\varepsilon^2) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma_\varepsilon^2 - \frac{1}{2}\log|\mathbf{\Omega}| - \frac{1}{2\sigma_\varepsilon^2}(\mathbf{q}-\mathbf{W}\boldsymbol{\delta})^\top\mathbf{\Omega}^{-1}(\mathbf{q}-\mathbf{W}\boldsymbol{\delta}).$$

In our setting, the main challenge is that the regressors matrix ($\mathbf{W}$) is not known as well. Determining $\mathbf{W}$ amounts to the problem of selecting the outlying observations. In the next section, we describe in detail the procedure adopted in the empirical study.

4.3.2 Dummy Saturation in the Presence of AOs, IOs and LSs

In order to locate outlying observations, we employ the Dummy Saturation principle proposed by Hendry (1999) and theoretically explored by Hendry et al. (2008) and Johansen and Nielsen (2009). The original contributions involve saturation with AOs (referred as IIS, *impulse indicator saturation*) while in a recent chapter Doornik et al. (2013) consider saturation with LSs (referred as SIS, *step indicator saturation*).

In this chapter, we consider a ML based procedure which looks for outliers in the real exchange rate process by saturating in turn with AOs, IOs and LSs. The steps of the procedure can be summarised as follows:

Step 1. Start by selecting the ARMA process that best fits the real exchange rate under the assumption of no outliers. To this extent, we estimate via ML processes of the form

$$\phi(L)(q_t - q_0) = \theta(L)\varepsilon_t, \tag{4.18}$$

increasing consecutively the order of the lag polynomials. Finally, we select the model according to the Akaike information criteiron (AIC) and denote the corresponding order with $(\tilde{p}, \tilde{q})$. The ARMA order is kept fixed until Step 3.

Step 2. This step involves the first search for outliers. We look sequentially for AOs, IOs and LSs and we store the selected outliers after each saturation. The significance level adopted in selecting each outlying observations is denoted with α. We can identify three sub-steps:

1. Saturate (4.18) with **AO**s[4]. Following Hendry et al. (2008)[5]:

   (a) Add the first half of AOs, say $x_{j,t}$, $j = 1, \ldots, \lfloor T/2 \rfloor$, and estimate by ML the following regression

   $$q_t = q_0 + \sum_{j=1}^{\lfloor T/2 \rfloor} \delta_j^A x_{j,t} + \phi^{-1}(L)\theta(L)\varepsilon_t. \tag{4.19}$$

   (b) Store all $\mathbf{x}_j$ such that $|\mathrm{t}_{\hat{\delta}_j^A}| > c_{\alpha/2}$. Denote the matrix of retained AOs with $\ddot{\mathbf{X}}_{(1)}$.

   (c) Repeat by saturating with the second half of AOs, i.e. estimating (4.19) with $x_{j,t}$, $j = \lfloor T/2 \rfloor + 1, \ldots, T$, and again define $\ddot{\mathbf{X}}_{(2)}$ the matrix of the outliers for which $|\mathrm{t}_{\hat{\delta}_j^A}| > c_{\alpha/2}$.

   (d) Estimate (4.19) including only the AOs selected at the two previous stages and denote $\ddot{\mathbf{X}}$ the matrix with the statistically significant outliers.

2. Saturate (4.18) with **LS**s. Repeat all the procedure described in steps (a)-(d) in order to get $\ddot{\mathbf{Y}}$, the matrix containing the retained LSs.

3. Saturate (4.18) with **IO**s. Repeat all the procedure described in steps (a)-(d) in order to get $\ddot{\mathbf{Z}}$, the matrix containing the retained IOs.

[4]The choice of starting the saturation with AOs is purely casual as there is no difference in starting with either IOs or LSs in spite of AOs.

[5]Castle et al. (2012) and Bergamelli and Urga (2013) study by simulations the performance of the Dummy Saturation when applied through the model selection algorithm *Autometrics* (see Doornik, 2009a). However, here we resort on a more traditional split algorithm (see for instance Hendry et al., 2008) as estimation of ARMA processes is not currently available through *Autometrics*.

Step 3. In the last step, we select the final model whose time series parameters are then used to obtain a robust half-life estimate. In practice, we start by estimating via ML the following regression with ARMA errors[6]

$$q_t = q_0 + \ddot{\mathbf{x}}_t^\top \boldsymbol{\delta}^A + \ddot{\mathbf{y}}_t^\top \boldsymbol{\delta}^L + \phi^{-1}(L)\theta(L)(\varepsilon_t + \ddot{\mathbf{z}}_t^\top \boldsymbol{\delta}^I) \tag{4.20}$$

and drop the insignificant outliers. Estimation of (4.20) is then iterated until the remaining outliers are all statistically significant. Once the selection of the outliers is terminated, we check whether the coefficients of the lag polynomials are all statistically significant as well and we modify the ARMA order accordingly.

The procedure outlined above extends the existent applications of the Dummy Saturation principle in two directions. First, it considers outliers detection in ARMA models which puts the Dummy Saturation in the context of maximum likelihood estimation. Second, it generalises the search of outlying observations considering also outliers of the innovative form. As noted above, innovative outliers have a slowly decaying effect in contrast with the instantaneous effect of additional outliers. This leads to the intuition that saturating with innovative outliers might help to capture more parsimoniously what otherwise would be captured by a series of additive outliers with decreasing coefficients.

As long as the steps of the procedure are structured, they match rather closely those of other procedures for outliers detection in time series model, in particular the widely used iterative approach as presented in Chen and

[6]The implicit assumption is that the number of outliers at this point is such that there are enough degrees of freedom. If this is not the case an easy solution is to strengthen the significance level used in the saturation.

Liu (1993). The main difference lies in the way the outliers are identified, i.e. saturation with dummies instead of the inclusion of the outliers one-by-one. In particular, we can identified three separate saturations: a first one involving AOs, a second one involving IOs and a third involving LSs. The retained outliers from each single saturation are then combined for the selection of the final model to take place. Additionally, note that the first and the last step are those where the ARMA order is determined.

Finally, when saturating a regression with dummies, an important aspect to keep under control is the retention rate of the outliers. We know from Hendry et al. (2008), Castle et al. (2012) and Doornik et al. (2013) that, given a significance level α and under normality, αT outliers are retained on average under the null of no outliers. Since in our case we are working with three separate saturations, we must set in each single saturation a significance level equal to approximately one third the desired retention rate before the final selection in Step 3. We also stress the fact that though we presented the procedure with two splits of the set of dummies, splitting the set of dummies in more than two parts does not alter the finite-sample properties of the procedure (see Table 4.1).

To verify that the outlined procedure delivers controlled retention rates, we carry out a small simulation study involving an ARMA$(1, 1)$ process, under the null of no outliers. In particular, we simulate $1,000$ paths from the following process

$$y_t = 0.6y_{t-1} + \eta_t + 0.3\eta_{t-1} \qquad t = 1, \ldots, T$$

for $T = \{100, 200, 300\}$ and $\eta_t \sim \mathcal{N}(0, 1)$ for all t. We compute the retention rates[7] of the three sets of dummies setting $\alpha = \{0.01, 0.05\}$. Table 4.1 reports the results using in each experiment a different number of splits of the dummies, i.e. $n = \{2, 5, 10, 20\}$.

[Table 4.1 about here.]

The results allow us to conclude that the retention rates under the null of no outliers are very close to the nominal level α with different block splits for AOs and LSs. For IOs, we observe instead deviations of the retention rates from α when the number of splits is small. In particular, the procedure retains too many IOs with the retention rates converging to the nominal level as the number of splits increases. In particular, the simulations show that the minimum number of splits that allows to reach the nominal level is around 20 which implies that the ideal number of dummies to be considered in each split is about $T/20$. In the empirical application, we will take into account this fact by setting the number of dummies splits to 20. Further, as commonly observed, the results show that the retention rates get closer to the nominal levels as the sample size T increases. As far as the time-series parameters are concerned, Figure 4.3 shows that the distributions of the ARMA coefficients estimates are correctly centred around the coefficients true values.

[Figure 4.3 about here.]

[7] All the computations are carried out using OxMetrics 6.30. The ARFIMA (Doornik and Ooms, 2012) package has been used for estimation purposes.

4.3.3 Half-life Computation for ARMA(p, q) Models

After controlling for outlying observations as described in the previous section, we use the robust ARMA parameters estimates in order to compute the half-life. Starting from (4.10), the outlier-free series is given by

$$\tilde{q}_t \equiv q_t - q_0 - \mathbf{x}_t^\top \boldsymbol{\delta}^A - \mathbf{y}_t^\top \boldsymbol{\delta}^L - \phi(L)^{-1}\theta(L)\mathbf{z}_t^\top \boldsymbol{\delta}^I = \phi^{-1}(L)\theta(L)\varepsilon_t = v_t. \quad (4.21)$$

Further, we can define $\psi(L) = \phi^{-1}(L)\theta(L)$ to give $\tilde{q}_t = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j}$, with $\sum_{j=0}^{+\infty} \psi_j^2 < \infty$ (under the assumption that the roots of $\phi(L)$ all lie outside the unit circle), such that $\lim_{j\to\infty} \psi_j = 0$. As already mentioned, the value assumed by the coefficients ψ_j as a function of time is regarded as the impulse response function (IRF) and denoted $\psi(j) = \psi_j$ with $\psi(0) = 1$. Following the definition of half-life given in (4.5), we are interested in finding the first instant h such that $\psi(h) = \frac{1}{2}$.

To compute the IRF of a general univariate ARMA(p, q) process, it is convenient to arrange its components in a VAR(1) form. Assuming $\{\tilde{q}_t\}_{t=1}^T \sim$

$ARMA(p,q)$,

$$\underbrace{\begin{bmatrix} \tilde{q}_t \\ \tilde{q}_{t-1} \\ \vdots \\ \tilde{q}_{t-p+1} \\ \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-q+1} \end{bmatrix}}_{\boldsymbol{\xi}_t} = \underbrace{\begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_p & \theta_1 & \dots & \theta_q \\ 1 & 0 & \dots & & & & \\ 0 & \ddots & 0 & \dots & & & \\ 0 & \dots & 1 & 0 & \dots & & \\ 0 & \dots & & & & & \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots \\ 0 & 0 & \dots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} \tilde{q}_{t-1} \\ \tilde{q}_{t-2} \\ \vdots \\ \tilde{q}_{t-p} \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \\ \varepsilon_{t-q} \end{bmatrix}}_{\boldsymbol{\xi}_{t-1}} + \underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{G}} \varepsilon_t$$

The IRF can be obtained as

$$\psi(j) = \mathbf{e}(\mathbf{F}^j \mathbf{G}) \tag{4.22}$$

where $\mathbf{e} = [1\ 0\ \dots\ 0]^\top$ is a selection vector to pick up the IRF for $\tilde{q}_t$. Alternatively, it is possible to compute $\psi(j)$ recursively. The half-life is thus given by $\psi(h) = \mathbf{e}(\mathbf{F}^h \mathbf{G}) = 0.5$. For AR(1) processes, the solution is trivially given by $h = \log(0.5)/\log(\rho)$ while for higher order auto-regression we can use the eigenvalues based formula given in Hamilton (1994, p. 12). However, if we want to compute the exact half-life of a general ARMA(p,q) process, we need to use some numerical procedure since no closed form solution exists. Also, the fact that h might no be an integer rules out standard Newton-Raphson methods.

In order to compute the half-life, we propose the following numerical procedure that exploits spline function interpolation and it runs as follows:

(a) Compute ψ_j for $j = 1, \ldots, J$ (typically $50 \leq J \leq 100$), using (4.22).

(b) Interpolate the above values using a spline function to get

$$\boldsymbol{\psi} = [\psi_0, \psi_{0+1/\Delta}, \psi_{0+(2/\Delta)}, \ldots, \psi_1, \psi_{1+(1/\Delta)}, \psi_{1+(2/\Delta)}, \ldots, \psi_J]$$

where $\Delta \in \mathbb{N}$ denotes the number of values in between two impulse responses. The size of $\boldsymbol{\psi}$ is then $\Delta(J-1)+J$. Define the associated index vector $\boldsymbol{\lambda}$ of size $\Delta(J-1)+J$ such that $\lambda_j = j$.

(c) Compute a new vector, $\boldsymbol{\iota}$, of the same size as $\boldsymbol{\psi}$ such that its j^{th} element is defined as

$$\iota_j = \begin{cases} j & \psi_j \geq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

and denote $\tilde{\boldsymbol{\iota}}$ of size $N < \Delta(J-1)+J$ a vector containing the same elements as $\boldsymbol{\iota}$ but with all null entries removed. Note that $\tilde{\boldsymbol{\iota}}$ is the vector of the time periods (x-axis coordinates) for which the IRF is above or equal 0.5.

(d) Define another index vector $\tilde{\boldsymbol{\lambda}}$ such that $\tilde{\lambda}_i = i$ for $i = 1, \ldots, N$ and compute $\boldsymbol{\delta}$ where $\delta_i = I(\tilde{\lambda}_i < \tilde{\iota}_i)$, $I(\cdot)$ denoting the indicator function. The estimated half-life h is given by

$$h = \begin{cases} \lambda_{\max\{\tilde{\iota}\}} & \max\{\boldsymbol{\delta}\} < 1 \\ \lambda_{\min\{\tilde{\iota}\}} | \delta_{\min\{\tilde{\iota}\}} = 1 & \text{otherwise} \end{cases}$$

where the condition is needed to ensure that if the IRF is equal to 0.5 for different time periods, we want the half-life to be the first of these periods.

The procedure described above is computationally very fast and it allows to obtain precise half-life estimates.

4.4 Empirical Application

In order to illustrate the empirical relevance of the impact of unaccounted outliers on the half-life estimation, we analyse US dollar bilateral exchange rates. The data used are taken from the Federal Reserve Bank of Saint Louis database (FRED). Following the literature, the nominal exchange rate is expressed as national currency units in terms of one US dollar (daily averages) while the price indexes are consumer price indexes (CPI) not seasonally adjusted. We consider quarterly data for the following countries: United Kingdom, Germany, France, Italy, Switzerland, Japan, South Africa, Mexico and the Euro Area. The data span the period 1971:1 to 2013:3 though the number of observations varies from a maximum of 171 to a minimum of 59 for the Euro Area. The log real exchange rate for the i^{th} country is computed as $q_{i,t} = s_{i,t} - p_{i,t} + p_{US,t}$ where $s_{i,t}$ is the logarithm of the nominal exchange rate, $p_{i,t}$ the logarithm of the CPI for that country and $p_{US,t}$ the logarithm of the CPI of the United States.

We start by computing the half-life for the above countries disregarding the possibility of outlying observations. Hence, we fit an ARMA model to each real exchange rate series and use the estimated coefficients to compute the IRF response function and the corresponding half-life measure according to the procedure described in Section 4.3.3. To estimate the ARMA models, we use maximum likelihood methods based on (4.13) which also avoids problems related with the finite sample bias of the least squares estimator.

Table 4.2 reports the annualised estimated half-life together with 95% bootstrap confidence intervals, obtained by non-parametric bootstrapping with 299 replications, as well as the ARMA order, the AIC and the Jarque-Bera normality test p-value.

[Table 4.2 about here.]

From a first inspection of the results, we can conclude that, over the timespan considered, the PPP puzzle is more evident for Switzerland, South Africa, Japan and Euro Area with half-lives between three and seven years whereas is decisively less pronounced for UK, Germany, France, Italy and Mexico. In Figure 4.4, we report the estimated IRFs with bootstrapped confidence bands.

[Figure 4.4 about here.]

Next, we re-fit an ARMA model to each real exchange rate time-series following the outlier detection procedure described in Section 4.3.2. In order to select the significant dummies, we set a significance level of $\alpha = 0.01$ in each individual saturation and, based on the simulations results reported in Table 4.1, we use 20 blocks of dummies throughout. This allows to control the retention rate of AOs, IOs and LSs at $0.01 \times T$ individually. Hence, under the null of no outliers we should expect $0.03 \times T$ not significant outliers on average before the combined selection to take place in Step 3 where a significance level of $\alpha = 0.05$ is used. The estimated ARMA coefficients after dummy saturation are then used to obtain robust half-lives estimates. Table 4.3 reports the results while Figure 4.5 the IRFs.

[Table 4.3 about here.]

[Figure 4.5 about here.]

The results after outliers detection by dummy saturation are very interesting. The main evidence is that for the countries where the real exchange rate shows the most puzzling behaviour, modelling outliers seems to drastically reduce the extent of this puzzle. In particular, for Switzerland, South Africa, Japan and the Euro Area, we observe that their half-lives are reduced by a factor of three or two, the most striking result being South Africa which half-life passes from a point value of 5.27 to 1.95. For the countries where instead the PPP puzzle is less evident or even absent, including outliers seem not to change the half-lives estimates. However, the benefit in accounting for outlying observations appears in the tighter confidence intervals and in the restored normality. The only country whose half-life increases after including outliers is Mexico with a point value of 2.05 after saturation compared to 0.99 before saturation. As far as the outliers are concerned, we see from Table 4.3 that the average number of outliers retained for each country is around four while their location is pretty widespread along all the time period considered. The most recurring outlier is the innovative outlier in the fourth quarter of 2008 ("IO:2008(4)") that is in the final model for UK, South Africa, Japan and Mexico. It is straightforward to notice that the mentioned outlier is capturing the effect on the exchange rates of the recent financial crisis which apparently died out – at least for these series – following a decaying effect.

A further investigation of the dynamics followed by the exchange rates is reported in the next section, where we also link our analysis with the "bands of inaction" literature (Taylor, 2001).

4.4.1 Testing for Non-linear Effects

In this section, we test whether the real exchange rate time series considered in the empirical application exhibit non-linear effects or not. This is of interest for at least two reasons. First, as mentioned above, part of the literature uses non-linear models to describe the real exchange rate process based on the "bands of inaction" argument (Taylor, 2001). Testing for non-linearities can thus be seen as an empirical test of this phenomenon. Second, by testing for non-linear effects with and without outliers removal, we can also explore the implications of removing outlying observations on non-linearities. This is of particular interest as we want to explore whether non-linearities might be due to unaccounted outliers or, conversely, outliers inclusion is masking non-linearities.

To this purpose, we consider the Brock-Dechert-Scheinkman test statistic (hereafter, BDS) introduced by Brock et al. (1987, 1996). The statistic is based on the concept of *correlation integral* which aims at measuring the frequency with which temporal patterns are repeated in the data. Briefly, consider the time series $\{x_t\}_{t=1,\dots,T}$ which is embedded in the m-space by forming m-histories $x_t^m = (x_t, x_{t-1}, \dots, x_{t-m+1})$. Brock et al. (1996) shows that the U-statistic

$$C(\eta, m, T) = \frac{2}{(T-m+1)(T-m)} \sum_{m \leq s <} \sum_{t \leq T} \mathbf{1}(|x_{t-i} - x_{s-i}| < \eta : i = 1, \dots, m-1)$$

is a consistent estimator of the correlation integral under fairly general assumptions on $\{x_t\}$. If the process is i.i.d. then it is possible to show that $C(\eta, m, T) = C(\eta, 1, T)^m$. Using this fact with the properties of the U-statistics,

Brock et al. (1996) show that

$$BDS(\eta, m, T) = \sqrt{T}\frac{C(\eta, m, T) - C(\eta, 1, T)^m}{\sigma(\eta, m, T)} \overset{A}{\sim} \mathcal{N}(0, 1)$$

where $\sigma(\eta, m, T)$ is the standard deviation of $\sqrt{T}C(\eta, m, T) - C(\eta, 1, T)^m$ and $\overset{A}{\sim}$ indicates *"asymptotically distributed as"*.

In particular, we apply the BDS test to the residuals of the fitted ARMA models before and after outliers removal[8]. Rejection of the null indicates that unaccounted dynamics is present in the residuals which can be interpreted as unaccounted non-linearities since Brock et al. (1996) show by simulations that the test has high power against non-linear alternatives. Table 4.4 reports the p-values of the BDS test statistic at different embedding dimensions.

[Table 4.4 about here.]

For each country, the first row indicates the p-value of the BDS statistic before outliers removal while the second row the p-value after robust estimation. Empirical evidence of the presence of non-linear effects is not uniform across our sample even before outliers detection. In fact, for only half of the countries considered we fail to accept the null and the results are also dependent on the embedding dimension adopted. Accounting for outliers allows to reduce the number of rejections and in particular, for UK and EMU there is no evidence of non-linear effects using dummy saturation while the null hypothesis is rejected at all embedding dimensions without modelling outliers. However, for countries like South Africa and Italy, where the evid-

[8]The value of the test statistic is obtained using the C code made available by Le Baron at `http://people.brandeis.edu/~blebaron/software/bds/bdsccode/`.

ence of misspecified linear models is quite strong, the inclusion of outliers has no effect on the test outcome.

To summarise, empirical evidence suggests that non-linear effects caused by "bands of inaction" is not uniform across the countries considered, and even before removing outliers this is observed only in half of the sample. Moreover, the number of rejections of the null hypothesis drops further when modelling outlying observations suggesting that model misspecification – at least in some cases – can be reasonably thought to be due to unaccounted outliers. The intuition behind this claim is further confirmed by Figure 4.6 where we compute by simulations the empirical rejection frequency of the BDS test when applied to the residuals of an ARMA model contaminated by outliers and level shifts. Clearly, the test over-rejects the null hypothesis if outlying observations are not taken into account during the estimation process.

[Figure 4.6 about here.]

Nevertheless, in our empirical application there are cases where we fail to reject the null hypothesis disregarding the fact that we take into account the presence of outliers or not, thus suggesting that non-linearities might indeed play a role in such situations.

4.5 Conclusions

In this chapter, we studied to which extent the half-life estimates and the related PPP puzzle found by the existing literature are affected by unaccounted outlying observations. In particular, we modelled the observed

real exchange rate process as a linear ARMA process contaminated by additional and innovative outliers as well as level shifts. To estimates such models, we proposed a sequential Dummy Saturation approach combined with ML estimation of the relevant parameters and we showed that the proposed estimation procedure delivers the correct retention rates of the dummy variables included to capture the outliers.

We illustrate the impact of removing outliers on the half-lives estimates through an empirical application involving a panel of US dollar denominated real exchange rates over a period spanning the last four decades. Our findings can be summarised as follows. First, for countries where the real exchange rate shows the most puzzling behaviour, including outliers seems to drastically reduce the extent of the PPP puzzle. Second, for countries where half-life estimates are contained even without including outliers, the benefit of a robust estimation approach is reflected in tighter confidence intervals for the half-life as well as in the restoration of normality. Therefore, in light of the evidence emerging from the empirical study, the main conclusion is that robust estimation methods allow to reduce the impact of the PPP puzzle and are anyhow beneficial for statistical inference even when the half-life estimates are not affected when accounting for outlying observations. Additionally, motivated by the "bands of inaction" argument of Taylor (2001), we test for the presence of non-linear effects in the real exchange rate process with and without the inclusion of outliers. We find mixed evidence of the presence of non-linear effects across our sample even before removing outlying observations. The evidence stabilises against the presence of non-linearities when accounting for outliers.

Finally, the results in this chapter suggest a number of stimulating avenues for future research. First, from a methodological point of view, it would be of interest to extend the results in Johansen and Nielsen (2009), which relate the Dummy Saturation to robust M-estimation, to our ML-based setting. Second, in light of the empirical application, it will be of some interest to investigate the role of outlying observations in the presence of time series which exhibit genuine non-linear dynamics (see Franses et al., 1996). Third, given the generality of the proposed procedure for outliers detection in a time-series setting, it will be of empirical interest to extend the application to other areas of financial economics like for instance stock return predictability. This is part of an ongoing research agenda.

Table 4.1: Retention Rates of Not Significant Outliers

$T = 100$			$T = 200$			$T = 300$		
AO	IO	LS	AO	IO	LS	AO	IO	LS
			$\alpha = 0.01$					
			$n = 2$					
1.242	6.472	0.901	1.215	6.488	0.983	1.092	6.651	0.989
			$n = 5$					
1.189	2.125	0.989	1.079	2.088	0.973	1.062	2.105	1.000
			$n = 10$					
1.116	1.530	0.993	1.096	1.493	0.972	1.045	1.479	1.004
			$n = 20$					
1.116	1.462	1.041	1.054	1.273	1.005	1.040	1.233	1.002
			$\alpha = 0.05$					
			$n = 2$					
5.260	15.880	4.655	5.223	16.085	4.818	5.103	16.265	4.899
			$n = 5$					
5.245	7.756	4.567	5.096	7.822	4.951	5.092	7.875	4.981
			$n = 10$					
5.206	6.346	4.894	5.188	6.235	4.965	5.038	6.271	5.010
			$n = 20$					
5.332	6.140	4.995	5.096	5.680	4.994	5.132	5.669	5.022

Notes: The reported values are the retention frequencies of the dummies computed across $M = 1,000$ simulations as $\frac{1}{T \times M} \sum_{m=1}^{M} \sum_{j=1}^{T} \mathbf{1}(|\mathsf{t}_{\hat{\delta}_j}| > c_{\alpha/2})$ where $\hat{\delta}_j$, $j = 1, \ldots, T$ denotes the estimated dummy coefficient associated with the j^{th} dummy and $T = \{100, 200, 300\}$ the sample size.

Table 4.2: Half-life ML Estimates Without Outliers Detection

	$\hat{h}$	$\hat{c}_{low}$	$\hat{c}_{upp}$	(p, q)	AIC	J-B
UK	1.78	0.85	2.26	4,3	-584.61	[0.0050]**
Germany	1.86	1.11	4.82	2,2	-400.49	[0.3885]
France	1.82	1.14	4.27	2,2	-419.27	[0.0456]*
Italy	1.87	1.13	4.74	2,2	-414.70	[0.0306]*
Switzerland	7.05	1.89	12.12	1,1	-541.48	[0.9039]
South Africa	5.27	1.40	7.62	2,1	-471.59	[0.0001]**
Japan	6.60	2.31	11.02	5,1	-545.81	[0.0503]
Mexico	0.99	0.31	1.13	3,3	-211.69	[0.0000]**
Euro Area	3.30	0.67	4.38	1,1	-208.00	[0.2370]

Notes: $\hat{h}$ denotes the annualised half-life estimate, $\hat{c}_{low}$ and $\hat{c}_{upp}$ are the lower and upper endpoint of the bootstrapped confidence interval, (p, q) denotes the ARMA order, AIC the Akaike Information Criterion and J-B the p-value of the Jarque-Bera test with '**' and '*' denoting rejection of the null of Normality at 1% and 5% significance level respectively.

Table 4.3: Half-life ML Estimates With Outliers Detection

	$\hat{h}$	$\hat{c}_{low}$	$\hat{c}_{upp}$	(p, q)	AOs	IOs	LSs	AIC	J-B
UK	1.25	0.86	1.93	4,3	1981(3) 1985(1) 1988(2)	2008(4)	1990(3) 1992(4)	-658.67	[0.4640]
Germany	1.89	1.27	3.72	2,2			1974(3) 1975(3) 1984(3) 1988(3)	-412.26	[0.7323]
France	1.85	1.24	3.88	2,2	1985(1)	1991(2)		-425.33	[0.1589]
Italy	1.66	1.03	4.45	2,2	2000(4)	1976(1) 1984(3) 1992(3)		-425.99	[0.0117]*
Switzerland	2.85	1.15	4.82	1,1	1985(1)	1971(1) 1978(1)		-558.30	[0.7162]
South Africa	1.95	0.96	2.42	3,2		2001(4) 2008(4)	1975(4) 1998(3)	-516.03	[0.0000]**
Japan	3.59	1.85	5.07	5,1	1979(4) 1995(2)	1971(1) 1998(4) 2008(4)	1978(3) 2013(1)	-593.94	[0.8416]
Mexico	2.05	0.38	3.26	2,3		1995(1) 2008(4)	1995(2)	-306.87	[0.3196]
Euro Area	1.31	0.51	2.19	1,1	2000(4)		2003(4) 2004(1)	-218.31	[0.2711]

Notes: See below Table 4.2. Further, under the columns tagged "AOs", "IOs" and "LSs", we report the dates of the additive outliers, innovative outliers and level shifts, respectively.

Table 4.4: BDS Test p-values with embedding dimensions $m = \{2, 3, 4, 5\}$.

$\eta = 0.5$	2	3	4	5
UK	0.003**	0.005**	0.002**	0.000**
	0.484	0.313	0.313	0.004**
Germany	0.116	0.025*	0.960	0.097
	0.002**	0.285	0.689	0.136
France	0.447	0.992	0.603	0.711
	0.116	0.313	0.741	0.857
Italy	0.001**	0.000**	0.000**	0.000**
	0.037*	0.006**	0.007**	0.000**
Switzerland	0.749	0.535	0.126	0.022*
	0.772	0.294	0.562	0.352
South Africa	0.003**	0.001**	0.000**	0.000**
	0.001**	0.000**	0.000**	0.000**
Japan	0.689	0.298	0.230	0.478
	0.757	0.407	0.711	0.332
Mexico	0.000**	0.002**	0.052	0.099
	0.000**	0.004**	0.067	0.072
EMU	0.215	0.002**	0.000**	0.000**
	0.555	0.522	0.119	0.555

Notes: ** and * denote presence of non-linear effects at 1% and 5% significance level. For each country, the first row reports p-values without outliers detection while the second row p-values with robust estimation.

Figure 4.1: USD/GBP Real Exchange Rate (log).

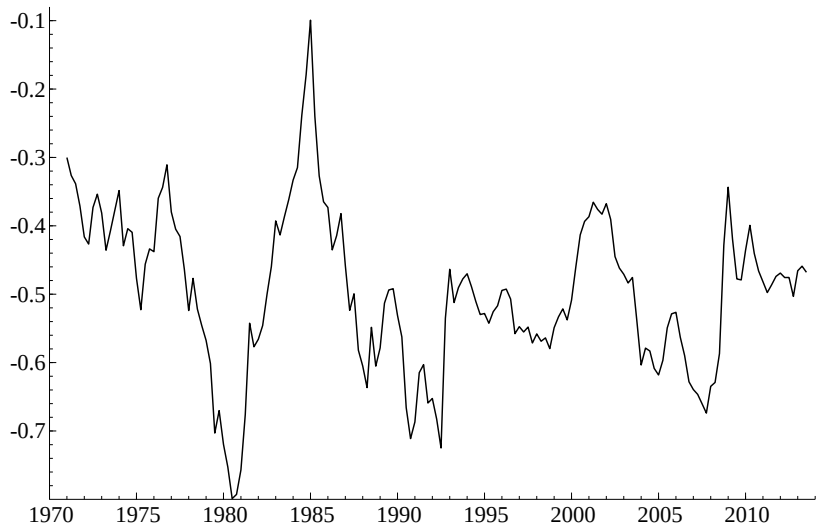
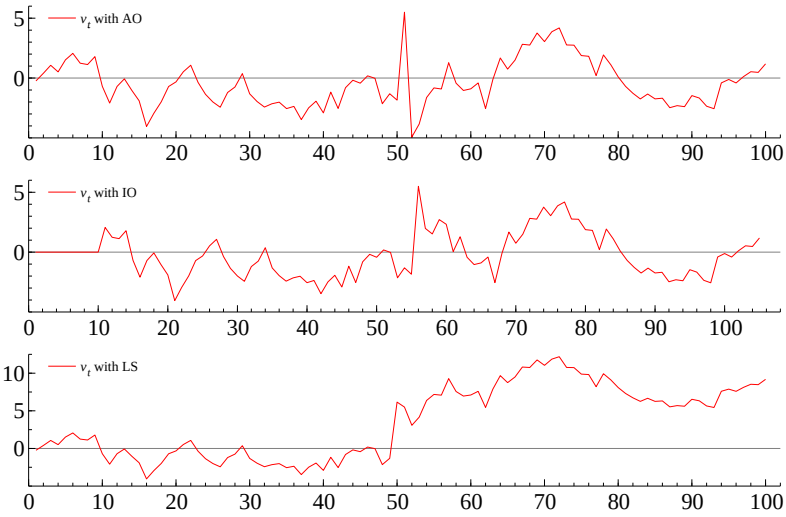
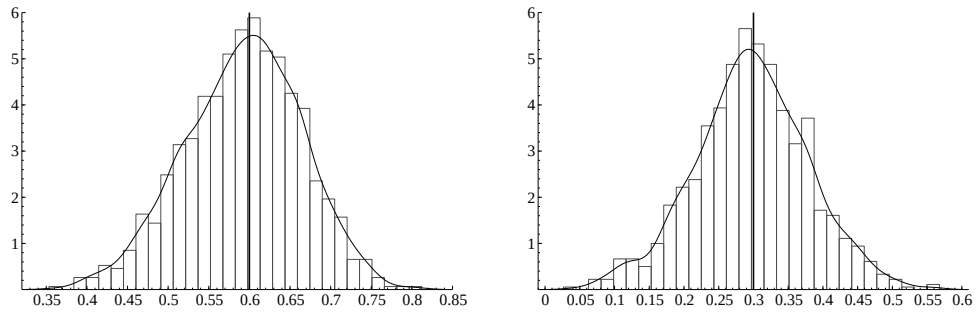


Figure 4.2: Plot of AO, IO and LS at $T_1 = 50$ for $T = 100$ and $v_t = 0.8v_{t-1} + \varepsilon_t$.

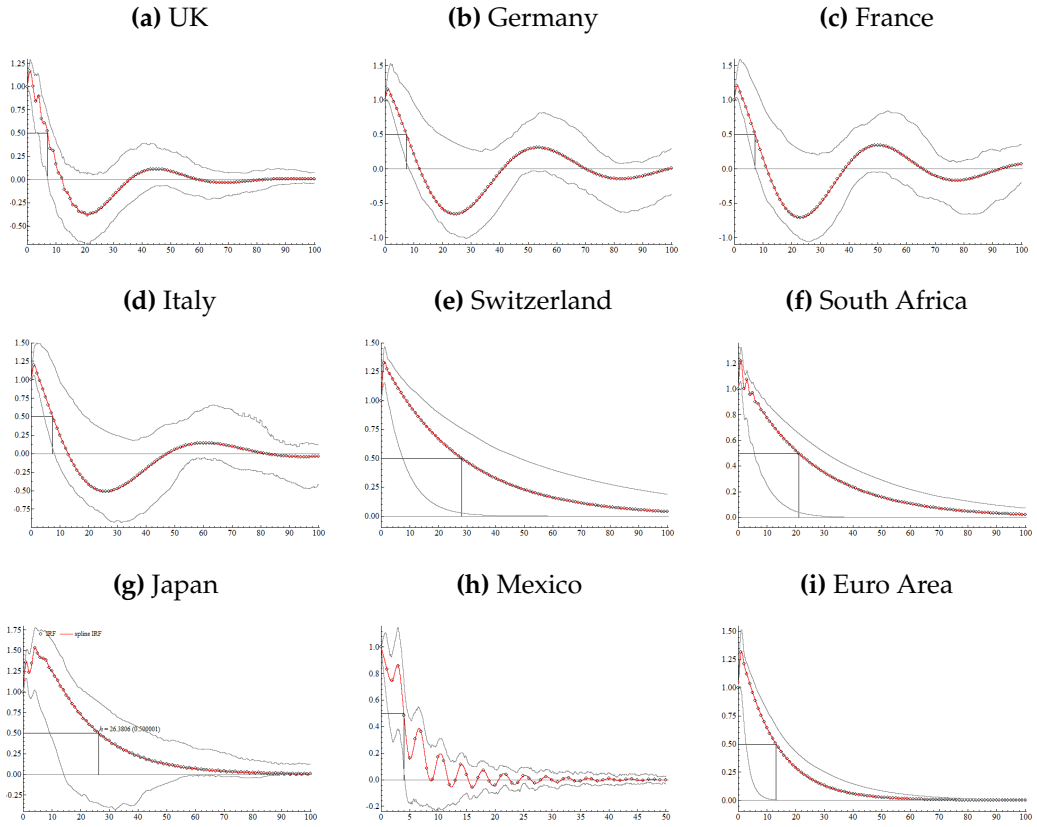




(a) $\phi = 0.6$ **(b)** $\theta = 0.3$

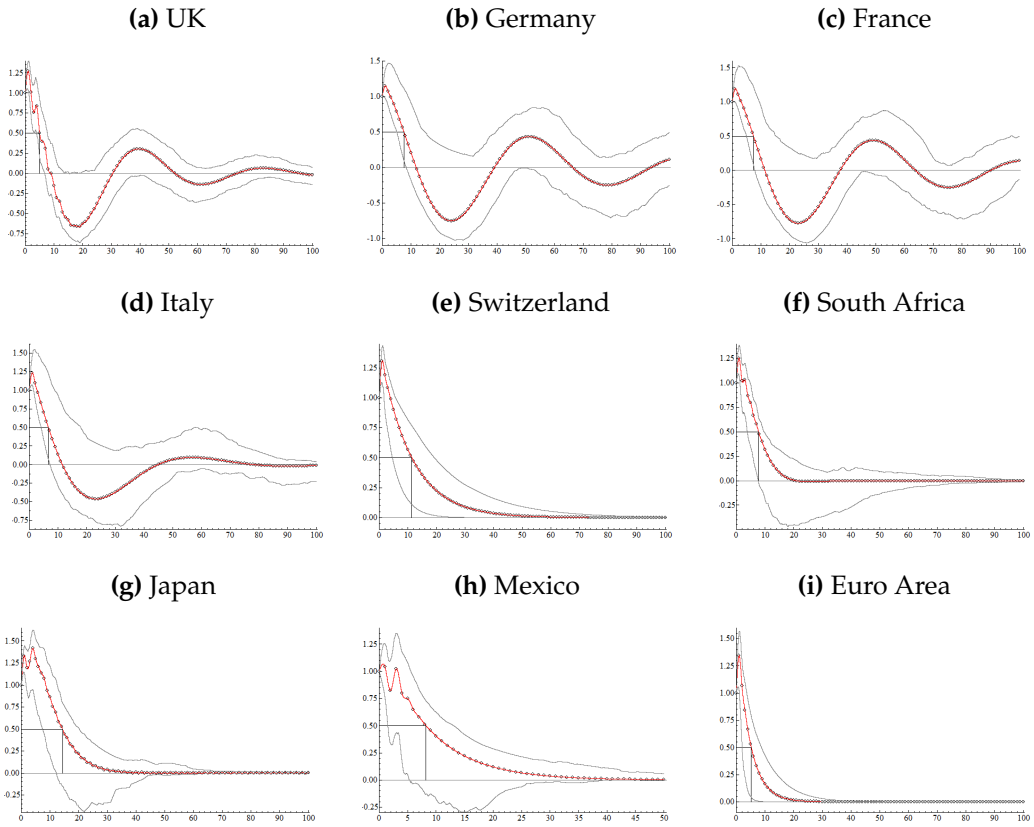
Figure 4.3: Distribution of the ARMA Coefficients Estimates after Dummy Saturation

Figure 4.4: Estimated IRFs Without Outliers Detection

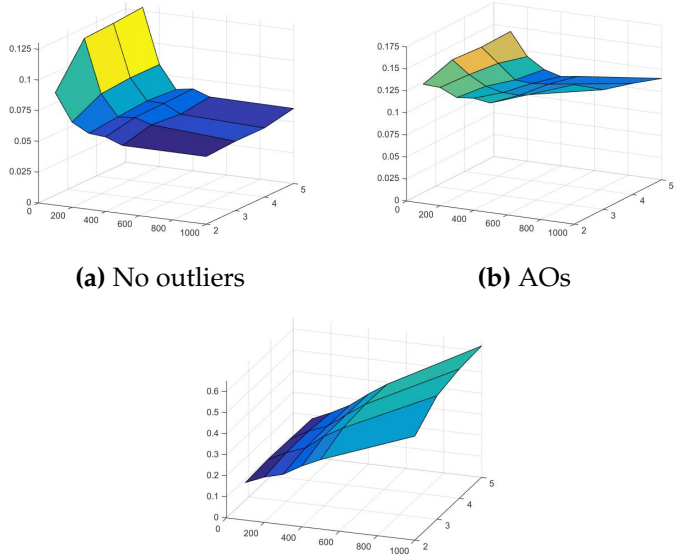


Notes: The black circles denotes point estimates of the IRFs, the red line is an interpolating spline and the grey lines are bootstrapped confidence bands (lower band corresponds to the 2.5$^{\text{th}}$ percentile while the upper band to the 97.5$^{\text{th}}$ percentile). Half-lives are the values (quarterly frequency) on the x-axis in correspondence to the black vertical lines.

Figure 4.5: Estimated IRFs With Outliers Detection



Notes: See below Figure 4.4.



(a) No outliers **(b)** AOs

(c) LSs

Figure 4.6: Empirical rejection frequencies of the BDS test. The plots show how often the BDS test rejects the null hypothesis of no non-linear effects (Z axis) for different sample sizes (X axis) and different embedding dimensions (Y axis). The DGP is an ARMA(1,1) process with no outliers **(a)**, an ARMA(1,1) contaminated by AOs **(b)**, and an ARMA(1,1) contaminated by LSs **(c)**.

Chapter 5

Conclusions and Roadmap for Future Research

In this thesis, we considered a series of unexplored issues faced by econometricians when dealing with structural breaks and outlying observations. The problem of dealing with breaks and outliers is one of the classic topics in the econometric and statistical analysis of time-series, given the striking inferential problems arising when neglecting their presence. Nonetheless, this is still a very active and prolific avenue for research with substantial room for the development of new econometric methods and the refinement of the existing ones.

In Chapter 2, we considered two largely unexplored procedures, the Dummy Saturation (DS) and the Sequential Bootstrapping (SB) of the sup-F statistic to estimate and date multiple structural breaks in the deterministic components of a linear system. Through an extensive Monte Carlo simulation exercise, we evaluated the performance of the two procedures considering several data generating processes ranging from the simple location-scale

model to the case of cointegrating regressions, considering both conditional and marginal processes. Additionally, we study for the first time the finite-sample properties of a new version of the DS procedure which includes also trend dummies. The Monte Carlo exercise allowed us to reach important conclusions. First, we found that the two procedures have good finite-sample properties although the SB approach behaved better in our simulations. Second, we provided useful guidelines for the applied researcher on how to best apply the two procedures based on the simulation evidence. We also illustrated the two approaches by studying empirically to which extent the Fisher effect holds in the United States economy. The application of the DS and the SB procedures led us to two important findings. First, the Fisherian hypothesis is valid only when structural breaks are properly detected and modelled, i.e. the coefficient of the expected inflation is statistically equal to one only when breaks are taken into account. Second, the two procedures detected almost the same break dates affecting the deterministic components of the Fisher equation.

In Chapter 3, we moved from the univariate approach followed in Chapter 2 to consider the problem of multiple breaks testing in the vector error correction model (VECM) framework. In particular, we investigate how to specify the subset of parameters when testing for breaks in that framework. First, we showed that the choice of which parameters of the VECM can be tested for breaks is indeed constrained given that breaks in the long run matrix β imply breaks in the short run impact matrix α, unless weak exogeneity can be imposed. In addition, breaks in β imply also breaks in the covariance matrix of the error term. Monte Carlo simulations showed that ignoring these links amongst the parameters can lead to severely over reject the null

hypothesis (where the number of hypothesized breaks is smaller than under the alternative), leading to the detection of spurious breaks. Second, we developed a new test for multiple structural breaks, by extending the likelihood ratio test proposed in Hansen (2003) to the case of unknown break dates through the specification of several scenarios regarding the number and the location of the breaks. We defined a minimum p-value statistic and a bootstrap procedure to approximate its critical values robust to the presence of breaks in the covariance matrix of the error term. A Monte Carlo study showed that the proposed statistic has the correct size and good power against fixed alternatives where several scenarios concerning the break dates are specified. Furthermore, we illustrated the relevance of taking into account the role of weak exogeneity as well as how to apply the minimum p-value statistic by means of three empirical applications.

Finally, in Chapter 4, we proposed a new procedure to detect outliers and level shifts in ARMA models. The procedure is then applied to obtain robust estimates of the real exchange rate process half-life. Specifically, we studied to which extent the half-life estimates and the related PPP puzzle are affected by unaccounted outlying observations. In particular, we modelled the observed real exchange rate process as a linear ARMA process contaminated by additional and innovative outliers as well as level shifts. In order to estimate such models, we proposed a sequential Dummy Saturation approach combined with ML estimation of the relevant parameters. In a Monte Carlo simulation exercise, we showed that the proposed estimation procedure delivers the correct retention rates of the dummy variables included to capture the outliers. An empirical application involving US dollar real exchange rates showed that the estimated half-lives are consistently

shorter when outlying observations are correctly modelled, thus shedding some light on the PPP puzzle.

Several directions for future research can be identified from each chapter. The analysis of the two novel procedures studied in Chapter 2 still leaves room for further investigation. In particular, for the Dummy Saturation approach the existing theory studies the properties of the resulting robust estimator when under the null hypothesis there are no breaks and outliers. Hence, it would be of great interest to study the properties of the DS estimator when the DGP is affected by breaks and/or outliers and to compare the statistical properties of the resulting break/outlier date estimator with more conventional estimators like the least squares estimator (see Bai, 1994). Both the theoretical and empirical analysis of the DS could be extended further to consider breaks in the slope coefficients. Additionally, it will be useful to compare the performance of the DS with breaks in mean and trend estimated using Autometrics with the robust dummy saturation estimator proposed in Johansen and Nielsen (2009), defined as an M-estimator with a bias corrected variance term. As far as the SB procedure is concerned, a theoretical analysis of the rule to stop the sequential search is an interesting area to be investigated.

Building on the results in Chapter 3, an interesting research direction would be to generalise the minimum p-value statistic to allow for different cointegrating ranks amongst the different regimes. This step will increase the generality of our framework though it will complicate the inference since the LR tests to assess the presence of breaks will not follow a χ^2, even when the break dates are known. Nonetheless, the bootstrapping approach proposed for the minimum p-value statistic is still able in theory to provide

rejection frequencies close to the nominal level (see Dufour et al., 2014). In this more general case, however, the individual LR tests p-values have to be computed either via bootstrap techniques or by using the non-standard distributions which arise as a consequence of the changing cointegrating rank.

Finally, the findings in Chapter 4 suggest several interesting developments. First, from a theoretical point of view, it would be interesting to extend the results in Johansen and Nielsen (2009), which relate the Dummy Saturation to robust M-estimation, to the maximum likelihood set-up. Second, in light of the empirical application, it will be of great interest to investigate the role of outlying observations in the presence of time series which exhibit genuine non-linear dynamics (see Franses et al., 1996). This would allow us to understand whether outlier detection masks non-linearities or vice versa, non-linearities are detected when outliers observations are neglected. Third, given the generality of the proposed procedure for outliers detection in a time-series setting, it will be of empirical interest to extend the application to other areas of financial economics like for instance stock return predictability.

In addition to the research directions suggested above, the econometrics of structural breaks and outliers detection still offers several sparks for future research. For instance, an interesting line of research builds on a generalisation of the concept of structural breaks which allows for gradual changes rather than abrupt ones. A statistical test to assess the presence of gradual changes in linear models could be interesting to develop. Furthermore, the robust statistical literature offers interesting tools that can be used to improve the forecasting performances of factor models. Specifically, the

use of principal components methods robust to outliers and missing data to extract factors from high dimensional dataset is another interesting avenue of research to explore.

Bibliography

Abuaf, N. and Jorion, P. (1990). Purchasing Power Parity in the Long-Run. *Journal of Finance 45*, 157–174.

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

Andrews, D. K. W. and Chen, H. Y. (1994). Approximately Median Unbiased Estimation of Autoregressive Models. *Journal of Business & Economic Statistics 12*, 187–204.

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

Andrews, D. W. K. and Ploberger, W. (1994). Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative. *Econometrica 62*, 1383–1414.

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

Baba, Y. and Hendry, D. F. and Starr, R. S. (1992). The Demand for M1 in the USA. *The Review of Economic Studies 59*, 25–61.

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

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.

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

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

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

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

Bergamelli, M. and Urga, G. (2013). 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 Yélou, C. (2007). Finite sample Multivariate Structural Change Tests with Application to Energy Demand Models. *Journal of Econometrics 141*, 1219–1244.

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

Boswijk, P.H. (1995). Identifiability of Cointegrated Systems. *Working paper, Tinbergen Institute*.

Brock, W.A. and Dechert, W.D. and Scheinkman, J.A. (1987). A Test of Independence Based on the Correlation Dimension. *Working Paper. Department of Economics, University of Winsconsin at Madison, University of Houston, and University of Chicago*.

Brock, W.A. and Dechert, W.D. and Scheinkman, J.A. and Le Baron, B. (1996). A Test of Independence Based on the Correlation Dimension. *Econometrics Reviews 15*, 197–235.

Brown, L.R. and Durbin, J. and Evans, J.M. (1975). Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistic Society, Series B 37*, 147–172.

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.

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

Cavaliere, G. and Georgiev, I. (2009). Robust Inference in Autoregressions with Multiple Outliers. *Econometric Theory 25*, 1625–1661.

Chang, I. and Tiao, C. G. and Chen, C. (1988). Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics 30*, 193–204.

Chang, Y. and Park, J.Y. and Song, K. (2006). Bootstrapping Cointegrating Regressions. *Journal of Econometrics 133*, 703–739.

Chen, C. and Liu, L.-M. (1993). Joint Estimation of Model Parameters ans Outlier Effects in Time Series. *Journal of the American Statistical Association 88*, 284–297.

Cheung, Y. W. and Lai, K. S. (2000). On the Purchasing Power Parity Puzzle. *Journal of International Economics 52*, 321–330.

Chortareas, G. and Kapetanios, G. (2013). How Puzzling is the PPP Puzzle? An Alternative Half-Life Measure of Convergence to PPP. *Journal of Applied Econometrics 28*, 435–457.

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica 28*, 591–605.

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

Davidson, R. and Flachaire, E. (2008). Wild Bootstrap, Tamed at Last. *Journal of Econometrics 146*, 162–169.

Davidson, R. and MacKinnon, J. G. (2005). Bootstrap Methods in Econometrics. In K. Patterson and T. C. Mills (Eds.), *Palgrave Handbook of Econometrics: Volume 1 Theoretical Econometrics*, Chapter 23. Basingstoke: Palgrave Macmillan.

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

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

Diebold, F. X. and Husted, S. and Rush, M. (1991). Real Exchange Rates under the Gold Standard. *Journal of Political Economy 99*, 1252–1271.

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

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

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

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

Doornik, J. A. and Ooms, M. (2012). A Package for Estimating, Forecasting and Simulating Arfima Models: ARFIMA package 1.06 for Ox. *User Manual*.

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

Dufour, J.M. and Khalaf, L. and Voia, M. (2014). Finite-sample Resampling-based Combined Hypothesis Tests, with Applications to Serial Correlation and Predictability. *Communications in Statistics - Simulation and Computation*, forthcoming.

Engle, R. F. and Granger, C. W. J. (1987). Cointegration and Error Correction Representation, Estimation and Testing. *Econometrica 55*, 251–276.

Engle, R. F. and Hendry, D. F. and Richard, J. (1983). Exogeneity. *Econometrica 51*, 277–304.

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

Frankel, J. A. and Rose, A. K. (1996). A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries. *Journal of International Economics 40*, 209–224.

Franses, P.H and van Dijk, D. J.C. and Lucas, A. (1996). Testing for Smooth Transition Nonlinearity in the Presence of Outliers. *Journal of Business & Economic Statistics 17*, 217–235.

Galbraith, J. W. and Zinde-Walsh, V. (1992). The GLS Transformation Matrix and a Semi-Recursive Estimator for the Linear Regression Model with Arma Errors. *Econometric Theory 8*, 95–111.

Gardner, L.A. (1969). On Detecting Changes in the Mean of Normal Variates. *Annals of Mathematical Statistic 10*, 116–126.

Gonzalo, J. (1994). Five Alterntive Methods of Estimating Long-Run Equilibrium Relationships. *Journal of Econometrics 60*, 203–233.

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

Hamilton, J.D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle . *Econometrica 57*, 357–384.

_____ (1994). *Time Series Analysis*. Princeton: Princeton University Press.

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

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

Hansen, 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.

Harvey, A.C. (1989). *Forecasting. Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

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

Hendry, D. F. (1999). An Econometric Analysis of US Food Expenditure, 1931-1989. In J. R. Magnus and M. S. Morgan (Eds.), *Methodology and Tacit Knowledge: Two Experiments in Econometrics*, Chapter 17, pp. 341–361. Chicester: John Wiley & Sons.

Hendry, D. F. and Krozlig, H.-M. (1999). Improving on "Data Mining reconsidered" by K. D. Hoover and S. J. Perez. *Econometrics Journal 2*, 202–219.

_____ (2005). The Properties of Automatic Gets Modelling. *Economic Journal 115*, C32–C61.

Hoffman, D.L. and Rasche, H. R. and Tieslau, M. A. (1995). The Stability of Long-run Money Demand in Five Industrial Countries. *Journal of Monetary Economics 35*, 317–339.

Hoover, K. P. and Perez, S.-J. (1999). Data Mining reconsidered: Encompassing and the General-to-Specific Approach to Specification Search. *Econometrics Journal 2*, 167–191.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics 35*, 73–101.

James, J., I. Marsh, and L. Sarno (Eds.) (2012). *Handbook of Exchange Rates*. Hoboken, New Jersey: John Wiley & Sons.

Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics & Controls 12*, 231–254.

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

_____ (1995). Identifying Restictions of Linear Equations with Applications to Simulatneous Equations and Cointegration. *Journal of Econometrics 69*, 111–132.

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

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

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

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

Mark, N. C. (2001). *International Macroeconomics and Finance: Theory and Empirical Methods*. Blackwell.

Nielsen, B. and Whitby, A. (2008). A Joint Chow Test for Structural Instability. *Working Paper.*.

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

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

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

Pesaran, M.H. and Pettenuzzo, D. and Timmerman, A. (2006). Forecasting Time Series Subject to Multiple Structural Breaks . *Review of Economic Studies 73*, 1057–1084.

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.

Qu, Z. and Perron, P. (2007). Estimating and Testing Structural Changes in Multivariate Regressions. *Econometrica 75*, 459–502.

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

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

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

Rogoff, K. (1996). The Purchasing Power Parity Puzzle. *Journal of Economic Literature 34*, 647–668.

Rosseauw, P. J. and Leroy, A. M. (1986). *Robust Regression and Outlier Detection*. Wiley.

Rossi, B. (2005). Confidence Intervals for Half-Life Deviations from Purchasing Power Parity. *Journal of Business & Economic Statistics 23*, 432–442.

Sánchez, M. J. and Peña. (2003). The Identification of Multiple Outliers in ARIMA Models. *Communications in Statistics 32*, 1265–1287.

Sarno, L. and Taylor, M. (2002). *The Economics of Exchange Rates*. Cambridge, UK: Cambridge University Press.

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

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

Stock, J. H. and Watson, M. W. (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business & Economic Statistics 14*, 11–30.

Taylor, A. M. (2001). Potential Pitfalls for Purchasing Power Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price. *Econometrica 69*, 473–498.

Tippett, L. H. C. (1931). *The Methods of Statistics.* London: Williams & Norgate Ltd.

Tsay, R. S. (1986). Time Series Model Specification in the Presence of Outliers. *Journal of the American Statistical Association 81*, 132–141.

——— (1988). Outliers, Level Shifts, and Variance Changes in Time Series. *Journal of Forecasting 7*, 1–20.