Adv Data Anal Classif (2016) 10:351–373
DOI 10.1007/s11634-015-0216-8

REGULAR ARTICLE

The forward search interactive outlier detection in cointegrated VAR analysis

Tiziano Bellini[1]

Received: 28 June 2014 / Revised: 2 August 2015 / Accepted: 4 September 2015 /
Published online: 18 September 2015

Abstract Cointegration analysis is particularly sensitive to outlying observations. Traditional robust approaches rely on parameter estimates based on weighting schemes to penalize aberrant units. This, in particular, is the idea underlying pseudo maximum likelihood robust estimators. Atypical observations, however, can reveal useful information about the investigated phenomenon. Aiming to detect these observations, we extend the forward search procedure to the cointegrated vector autoregressive model. The analysis is carried out by building up subsets of increasing dimension and monitoring suitable statistics at each subset size. Simulation experiments and real data analysis highlight that our forward search is more effective than the pseudo maximum likelihood in detecting atypical units and data structures.

Keywords Confidence threshold · Data structure detection · Forward search · Outliers · Pseudo maximum likelihood weights · Robust statistics · Vector equilibrium correction model

Mathematics Subject Classification 62F35 · 62-07 · 62P20 · 91B84

1 Introduction

In empirical time series, irregular data patterns often characterize the analysis. These statistical units, however, can be useful to conduct the study and the researcher needs to be conscious about them. Aiming to help this analysis, the goal of our research is to extend the forward search (FS) to the cointegrated vector autoregressive (VAR) model

✉ Tiziano Bellini
tiziano.bellini@unipr.it

[1] Università degli Studi di Parma, Parma, Italy

and compare our procedure to other methods for locating outlying cases. We show through simulation experiments and the analysis of real data the effectiveness of our framework in detecting multiple (masked) outliers. The traditional cointegrated VAR model is applied with the aim to improve the study through a series of informative plots. The researcher is allowed to interactively investigate time series without imposing weighting schemes or black-box mechanics.

Cointegration outlier detection has been investigated from different perspectives. Franses and Haldrup (1994) discussed the properties of the univariate Dickey–Fuller test and the Johansen test (1988, 1991) for the cointegrating rank in the case of additive outlying observations. They showed how dummy variables can be used to remove the influence of such extreme observations.

This analysis was further developed, both from a theoretical as well as from an empirical point of view, by Lucas (1995) who carried out unit root tests based on M-estimators. These tests turned out to be more powerful than OLS-based tests if the innovations are fat tailed. In addition, Lucas (1997) developed pseudo maximum likelihood (PML) estimators to determine the cointegration rank for VAR models for multivariate time series. It was shown that non-Gaussian pseudo likelihood ratio tests generally have a higher power than the test of Johansen, if the innovations demonstrate leptokurtic behaviour.

Franses et al. (1998) considered weekly observed scanning data on a fast moving consumer good, with a specific focus on the relationship between market share, distribution, advertising, price, and promotion. Since weekly scanning data can contain aberrant observations, they applied an outlier robust cointegration method to outline different results across robust and non-robust methods for the long-run marketing effects.

Following the same idea, Franses and Lucas (1998) used an outlier robust technique to reduce the impact of atypical units on the cointegration analysis. As a byproduct of computing the robust estimator, they obtained weights for all observations in the sample. In addition to this approach, Bosco et al. (2010) developed an EM algorithm to estimate the cointegration parameters relying on Student-T innovations. They also introduced a battery of robust inference procedures for unit root and cointegration tests to be applied in the field of multivariate time series analysis.

The effects of innovation outliers and additive outliers were further examined by Nielsen (2004). Particular attention was paid on how outliers can be modelled with dummy variables.

The FS was originally developed by Atkinson and Riani (2000) in linear and non-linear regression. Atkinson et al. (2004) extended the FS to multivariate analysis and further developments have been carried out during the last few years (Johansen and Nielsen 2010, 2015).

What makes the FS particularly appealing over other robust methods is the fact that it is not required to make additional assumptions on parameter estimates, impose weighting schemes or pre-defined patterns. We can simply use the traditional maximum likelihood estimators and monitor how the estimates evolve as the subset size increases (Bellini and Riani 2012).

In Johansen et al. (2000) the focus is on modeling cointegration with piecewise linear trend and known break points. The aim of our research, on the other hand, is to highlight when outliers and level shifts occur. In order to show how the method captures atypical observations, alternative contamination schemes are considered. Additive, level shift and innovation outliers are explored.

We conduct a parallel analysis of the PML and the FS to highlight their similarities and differences.

Thus, we carry out an extensive Monte Carlo simulation to show the effectiveness of the proposed FS procedure in comparison to PML.

Finally, we extend our research to real macro-economic time series.

The paper is organized as follows. Section 2 aims to summarize the motivation of our research. In Sect. 3 we introduce the cointegration framework and we describe robust PML estimators. Section 4 is devoted to the description of FS mechanics, while Sect. 5 aims to compare PML and FS on simulated time series. We devote Sect. 6 to Monte Carlo experiments. Section 7 shows how the FS can be used in practice when applied to real time series. The last Sect. 8 contains concluding remarks and directions for future research.

2 Motivation

In time series analysis, and in particular in econometrics, we can summarize the sources of scepticism towards robust methods as follows.

- There exists a plethora of semi-parametric and non-parametric techniques for outlier analysis that can be used for dealing with contaminated data.
- Outlier robust procedures are not standardized enough, leaving the user too much freedom in choosing parameters to be exploited in the analysis (see, for example, M-estimators).
- The computational effort required to compute robust estimators is higher than for traditional methods.
- Time series analysts tend to consider outlier robust statistics as a black-box procedure where observations are manipulated or weighted to fit the data.

The key advantage of the FS compared to other robust techniques is to show diagnostic tests and model evaluation procedures as a function of the subset size. Deep changes in the monitored statistics occur when atypical units enter in the subset (which consist of the closest units according to a given objective function). Thus, additional assumptions on the model are not required and weighting schemes are not necessary to carry out the FS. In what follows, we describe how the FS addresses the above described questions.

With regards to the first issue, one needs to bear in mind that semi-parametric and non-parametric techniques generally require much more observations to work reasonably well. In addition, they are not free from outlier and data structure problems. In other words, even when applying these techniques, the researcher has to be cautious about atypical observations.

The second argument points out that some robust approaches allow the user to choose among different functional forms or set of parameters to run the estimation. The FS, on the contrary, is not sensitive to this issue due to the fact that it relies on the standard maximum likelihood estimation without requiring any further assumption on the model to be fitted.

When dealing with real data, the need to speed up computations is particularly significant. The FS as well as some other robust methods are subject to intensive computational efforts. Therefore in an attempt to overcome the third problem, we speed up the FS analysis by proposing a fast procedure to obtain confidence thresholds (Riani et al. 2009) without needing Monte Carlo simulations.

Subsequently, to address the fourth issue, we highlight the fact that the FS is not a black-box procedure and manipulation or weighting schemes are not applied. On the contrary, we monitor suitable statistics as the subset size increases. A battery of plots is available for the researcher to make inference on outliers or data structures allowing to decide what to do with these units (Nielsen 2004). In the case of break points, the technique proposed by Johansen et al. (2000) can be used to manage piecewise linear trends.

In the next section we describe the key elements of the unrestricted cointegration framework and we introduce robust PML estimators.

3 Cointegrated VAR analysis and PML estimators

Aiming at defining cointegration, Campbell and Perron (1991) highlight that a $p \times 1$ vector of variables $x_t = (x_{1,t}, \dots x_{p,t})'$, $t = 1, \dots, T$, is said to be cointegrated if at least one non-zero $p \times 1$ vector β_i exists such that $\beta_i' x_t$ is trend stationary and β_i is called cointegrating vector. If r such linearly independent vectors β_i $(i = 1, \dots, r)$ exist, we say that x_t is cointegrated with cointegrating rank r. We than define the $(p \times r)$ matrix of cointegrating vectors $\beta = (\beta_1, \dots, \beta_r)$. The r elements of the vector $\beta' x_t$ are trend-stationary and β is called the cointegrating matrix.

According to Johansen (1996), the cointegrated VAR can be represented in the form of vector equilibrium correction model (VECM) as follows

$$\Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \cdots + \Gamma_{k-1} \Delta x_{t-k+1} + \Phi D_t + \epsilon_t, \tag{1}$$

$$\underbrace{\begin{bmatrix} \Delta x_{1,t} \\ \vdots \\ \Delta x_{p,t} \end{bmatrix}}_{\Delta x_t} = \underbrace{\begin{bmatrix} \pi_{1,1} & \cdots & \pi_{1,p} \\ \vdots & \ddots & \vdots \\ \pi_{p,1} & \cdots & \pi_{p,p} \end{bmatrix}}_{\Pi} \underbrace{\begin{bmatrix} x_{1,t-1} \\ \vdots \\ x_{p,t-1} \end{bmatrix}}_{x_{t-1}} + \underbrace{\begin{bmatrix} \gamma_{1_{1,1}} & \cdots & \gamma_{1_{1,p}} \\ \vdots & \ddots & \vdots \\ \gamma_{1_{p,1}} & \cdots & \gamma_{1_{p,p}} \end{bmatrix}}_{\Gamma_1} \underbrace{\begin{bmatrix} \Delta x_{1,t-1} \\ \vdots \\ \Delta x_{p,t-1} \end{bmatrix}}_{\Delta x_{t-1}} + \cdots$$

$$+ \underbrace{\begin{bmatrix} \gamma_{k-1_{1,1}} & \cdots & \gamma_{k-1_{1,p}} \\ \vdots & \ddots & \vdots \\ \gamma_{k-1_{p,1}} & \cdots & \gamma_{k-1_{p,p}} \end{bmatrix}}_{\Gamma_{k-1}} \underbrace{\begin{bmatrix} \Delta x_{1,t-k+1} \\ \vdots \\ \Delta x_{p,t-k+1} \end{bmatrix}}_{\Delta x_{t-k+1}} + \underbrace{\begin{bmatrix} \phi_{1,1} & \cdots & \phi_{1,g+1} \\ \vdots & \ddots & \vdots \\ \phi_{p,1} & \cdots & \phi_{p,g+1} \end{bmatrix}}_{\Phi} \underbrace{\begin{bmatrix} D_{1,t} \\ \vdots \\ D_{g,t} \\ \mu_{0,t} \end{bmatrix}}_{D_t} + \underbrace{\begin{bmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{p,t} \end{bmatrix}}_{\epsilon_t},$$

where $x_t = (x_{1,t} \dots x_{p,t})'$, $t = 1, \dots, T$, is the $p \times 1$ vector of variables. Δ is the first difference operator, $\Delta x_t = x_t - x_{t-1}$. In the case where Π has a reduced rank, it can be rewritten as $\Pi = \alpha\beta'$, where α and β are $p \times r$ matrices, $r \leq p$. β is the above described cointegrating matrix. $\Gamma_1, \dots, \Gamma_{k-1}$ are $p \times p$ parameter matrices. The vector $D_t = (D_{1,t}, \dots, D_{g,t}, \mu_0)'$ contains g (binary) dummies and a constant.

In addition, it is assumed that errors are Normally distributed $\epsilon_t \sim N(0, \Sigma)$, where Σ represents the $p \times p$ covariance matrix. The overall VECM set of parameter to be estimated, Θ, is represented as follows $\Theta = \{\Pi, \Gamma_1, \ldots, \Gamma_{k-1}, \Phi, \mu_0, \Sigma\}$.

VAR model parameters can be estimated via ML in two steps (Johansen 1988, 1991). Residuals constitute the key element of robust analysis. Thus, once model parameters are estimated we can compute the fitting errors as follows

$$e_t = \Delta x_t - \widehat{\Delta x_t}. \tag{2}$$

In order to reduce the effect of outliers and instead of using the above mentioned procedure, Lucas (1997) proposed a testing procedure based on the pseudo likelihood estimation. Parameters of Eq. (1) are estimated relying on the following Student-T pseudo likelihood with ν degrees of freedom

$$\mathcal{L}(\Theta) = \prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})|\pi\nu\Sigma|^{\frac{1}{2}}} \left(1 + \frac{\epsilon_t^{'}\Sigma^{-1}\epsilon_t}{\nu}\right)^{-\frac{\nu+p}{2}}, \tag{3}$$

where Θ is the vector of unknown parameters, p denotes the number of variables and Σ is the covariance matrix. The degrees of freedom ν are chosen according to Franses and Lucas (1998) in the range $\nu = 3, \ldots, 7$. Lower values of ν provide more protection against atypical observations, but imply higher loss in power if there are no outliers.

A useful byproduct of this approach is that weights are obtained for each observation. According to Franses and Lucas (1998), the Student-T estimator for Eq. (1) can be regarded as the Gaussian PML estimator for a weighted version of the data with weights given by

$$w_t = \left(\frac{\nu + p}{\nu + \epsilon_t^{'}\Sigma^{-1}\epsilon_t}\right)^{\frac{1}{2}}. \tag{4}$$

This representation was exploited by Bosco et al. (2010) to build an estimation process based on the expectation maximization (EM) algorithm made up by the following steps:

1. Initialization. Estimate the parameters of the VECM using standard reduced rank regression.
2. E-Step. Estimate the weights of Eq. (4) using the parameters obtained in the previous step.
3. M-Step. Estimate Eq. (1) parameters by relying on Johansen (1991) procedure, after having multiplied by w_t the variables.
4. Repeat the previous steps until the increment of the log-likelihood is negligible.

With weights corresponding to final estimates, Franses and Lucas (1998) suggest how to calculate their critical values. Under the assumption that ϵ_t are i.i.d. standard Normally distributed, $\epsilon_t^{'}\Sigma^{-1}\epsilon_t$ has a χ^2 distribution with p degrees of freedom. Let $\chi_p(0.005)$ denote the one half per cent value for the χ^2 distribution, weights are extraordinary small if $w_t \leq \left(\frac{\nu+p}{\nu+\chi_p(0.005)}\right)^{\frac{1}{2}}$.

In the next section we describe the FS mechanics. An empirical comparison of PML and FS will then be carried out in Sect. 5. In that section, according to Franses and Lucas (1998), the PML analysis is conducted relying on $\nu = 5$ and comparing w_t to the above described critical value (for $\nu = 5$).

4 FS mechanics

Starting from the FS framework described by Atkinson et al. (2004) as extended to multivariate time series (Bellini and Riani 2012), the focus of our analysis shifts to the criteria to use in the cointegrated VAR analysis.

The basic idea behind the FS is to start from an outlier-free initial subset on which model parameters and residuals are estimated. The FS then progresses by relying on subsets of increasing size until all units are included in the search.

Let $S_*^{(m)}$ be the optimal subset of size m. At each step of the search the vector of parameters $\hat{\Theta}_*^{(m)}$ is estimated relying on $S_*^{(m-1)}$. Residuals, $e_t(m^*)$, and standardized residuals, $r_t(m^*)$, are computed for all units $t = 1, \ldots, T$.

In terms of notation, $e_t(m^*)$ stands for unit t residual at step m (obtained by considering the parameter vector $\hat{\Theta}_*^{(m)}$).

From a computational point of view, standardized residuals are obtained as follows

$$r_t(m^*) = \sqrt{e_t(m^*)'[s^2(m^*)]^{-1}e_t(m^*)}, \tag{5}$$

where $s^2(m^*)$ is estimated as follows

$$s^2(m^*) = \frac{\sum_{t \in S_*^{(m)}} e_t(m^*)e_t(m^*)'}{m-1}. \tag{6}$$

The subset $S_*^{(m)}$ is made up by the m units (not necessarily adjacent) with the lowest standardized residuals.

Relying on $S_*^{(m)}$ we estimate the parameter vector $\hat{\Theta}_*^{(m+1)}$ that is used to compute standardized residuals for the $m+1$ step. We continue until all units are included into the subset.

Let (7) denote the observation with the minimum residual among those not in $S_*^{(m)}$

$$x_{t_{min}} = \arg\min[r_t^*(m^*)] \quad t \notin S_*^{(m)}. \tag{7}$$

To test whether observation $x_{t_{min}}$ is an outlier, we monitor

$$r_{tmin}^*(m^*) = \sqrt{e_{tmin}(m^*)'[s^2(m^*)]^{-1}e_{tmin}(m^*)}. \tag{8}$$

Atkinson and Riani (2006) studied the distribution of minimum deletion residuals in regression, comparing their empirical distribution to approximations based on truncated samples and order statistics. Riani et al. (2009) extended this analysis to the multivariate setting identifying an analytical approximation for standardized residuals. In what follows we apply this latter approach to cointegration.

5 Robust cointegration analysis in practice

We begin our empirical analysis by focusing on simulated uncontaminated time series. Next, we will introduce three different contaminations: additive, level shift and innovation outlier cases. The entire analysis is carried out comparing the FS to Franses and Lucas (1998) weighting scheme implemented according to Bosco et al. (2010). We have developed our software routines in Matlab (R2011a).

The simulated analysis is based on the following bivariate (i.e., $p = 2$) model

$$\begin{pmatrix} \Delta x_{1,t} \\ \Delta x_{2,t} \end{pmatrix} = \alpha\beta' \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \Gamma_1 \begin{pmatrix} \Delta x_{1,t-1} \\ \Delta x_{2,t-1} \end{pmatrix} + \Gamma_2 \begin{pmatrix} \Delta x_{1,t-2} \\ \Delta x_{2,t-2} \end{pmatrix} + \alpha\mu_0 + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \quad (9)$$

where $t = 1, \ldots, 75$. We simulate one path for each of the two time series $x_t = (x_{1,t}, x_{2,t})'$ as highlighted in panel (a) of Fig. 1. The three initial observations of the simulated time series are $(x_{1,t=1}, x_{2,t=1})' = (1.0420, 1.0720)'$, $(x_{1,t=2}, x_{2,t=2})' = (1.0670, 1.1150)'$, $(x_{1,t=3}, x_{2,t=3})' = (1.1290, 1.1570)'$. The cointegration parameters are as follows:

$$\alpha = (0.0113, 0.0177)', \quad \beta = (16.6046, -36.6281)',$$
$$\Gamma_1 = \begin{pmatrix} -0.2313 & 0.4028 \\ -0.0029 & 0.1534 \end{pmatrix} \text{ and } \Gamma_2 = \begin{pmatrix} 0.1310 & 0.0790 \\ 0.0513 & 0.2103 \end{pmatrix}, \quad \mu_0 = 21.9292.$$

The error terms, $\epsilon_{1,t}$ and $\epsilon_{2,t}$, are normally distributed with mean zero and covariance matrix $\Sigma = \begin{pmatrix} 0.0011 & 0.0010 \\ 0.0010 & 0.0013 \end{pmatrix}$. These parameters are aligned with those computed for two Italian stock prices along the period from January to August 2012.

In Fig. 1, panel (a) shows the simulated time series $x_{1,t}$ and $x_{2,t}$ (75 observations). In order to check for the presence of outliers in panel (b) PML weights are highlighted. We concentrate on $\nu = 5$ by showing the related critical value. This panel points out that the critical value is not crossed, thus no outlier is detected. Moving to the FS analysis, the bottom panels of Fig. 1 draw out the evolution of residuals along the search. Panel (c) shows the evolution of standardized residuals $r_t(m^*)$ for each observation. Each line of this plot represents $r_t(m^*)$, for unit t ($t = 1, \ldots, T$), as the subset size m increases.

We remind that $r_t(m^*)$ stands for standardized residual for unit t at step m, obtained relying on the parameter vector $\hat{\Theta}_*^{(m)}$.[1]

[1] It is worth mentioning that the symbol * relates to the parameter vector associated to the optimal subset (of size m) and not to the size of the subset.

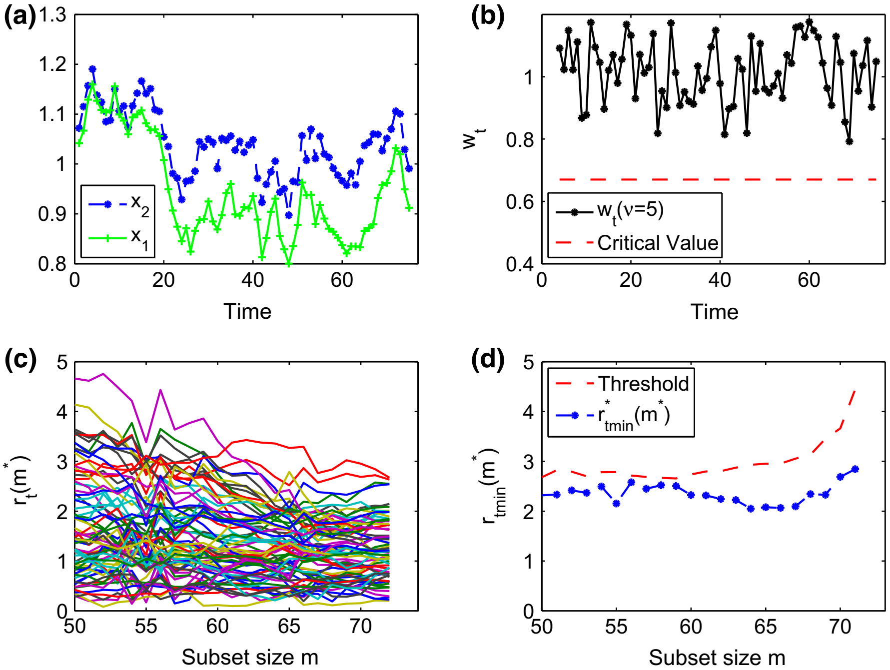

Fig. 1 Simulated time series: uncontaminated analysis. In the *top left panel* **a** the simulated time series $x_{1,t}$ and $x_{2,t}$ are shown. In *panel* **b** PML weights and the critical value (*line*) are shown for $\nu = 5$. No crossing of this line means that no outlier is detected. In *panel* **c** the standardized residuals $r_t(m^*)$ are shown for each unit as the subset size m increases. *Panel* **d** shows $r^*_{tmin}(m^*)$ compared to the 99 % confidence threshold obtained through Monte Carlo simulations. It is evident that $r^*_{tmin}(m^*)$ does not cross the 99 % threshold meaning that no outlier is detected

In other words, on the horizontal axis the subset size is represented whereas the vertical axis describes the standardized residuals $r_t(m^*)$ for each unit. Moving from left to right, at each subset size m, the evolution of residuals is shown through a line connecting $r_t(m^*)$ points.

In Fig. 1, panel (c) points out that all lines are close to each other meaning that there are no separate clusters of observations or atypical units. As we will see shortly, when outliers affect the dataset, $r_t(m^*)$ trajectories are very different from the ones shown in panel (c). In order to make inference on outliers, in panel (d), we compare $r^*_{tmin}(m^*)$ to the 99 % confidence threshold obtained through Monte Carlo simulations. Simulations rely on the parameter vector $\hat{\Theta}$ estimated on the entire set of observations x_t, $t = 1, \ldots, T$. We remark that $r^*_{tmin}(m^*)$ is always below the 99 % threshold meaning that no atypical observation is detected. Thus, both the PML analysis and the FS are correct in not revealing the presence of any outlier.

It is evident from the above analysis that making inference on outliers is one of the key issues of the FS. At this stage, it is useful to check whether the confidence thresholds obtained through Monte Carlo simulations are affected by the choice of the simulation parameters. For this reason, in order to obtain the 99 % confidence

Table 1 Simulation parameters for Θ_s $(s = 1, 2, 3)$

	α_1	α_2	β_1	β_2	μ_0	$\Sigma_{1,1}$	$\Sigma_{1,2}$	$\Sigma_{2,2}$
Θ_1	0.010	0.017	18.392	−38.133	21.702	0.015	0.011	0.013
Θ_2	0.005	0.019	−8.390	−18.136	22.405	0.001	0.001	0.010
Θ_3	0.110	0.010	5.870	−7.130	20.904	0.25	0.110	0.130

The matrices Γ_1 and Γ_2 are the same as described for the simulation analysis

threshold, we consider the parameter vectors Θ_s $(s = 1, 2, 3)$ shown in Table 1. We consider parameter sets which are associated to high roots in order to emphasize that even in these cases confidence thresholds are not affected by the choice of parameters. In Table 1 we omit Γ_1 and Γ_2 matrices because we exploit the same matrices described for the simulated analysis.

In order to obtain fast envelopes, Riani et al. (2009) study the distribution of scaled and unscaled (Mahalanobis) distances in multivariate analysis. Let $Y_{[m+1]}$ be the $(m+1)th$ order statistic from a sample of size T from a distribution with CDF $G(y)$. Then the cumulative distribution function (CDF) of $Y_{[m+1]}$ is given by

$$P(Y_{[m+1]} \leq y) = \sum_{j=m+1}^{T} \binom{T}{j} G(y)^j \{1 - G(y)\}^{T-j}. \tag{10}$$

The required quantile of order γ of the distribution of $Y_{[m+1]}$, say $y_{m+1,T;\gamma}$, can be obtained as follows (Riani et al. 2009)

$$y_{m+1,T;\gamma} = G^{-1}\left\{\frac{m+1}{m+1+(T-m)\chi_{2(T-m),2(m+1);1-\gamma}}\right\}, \tag{11}$$

where $\chi_{2(T-m),2(m+1);1-\gamma}$ is the $1-\gamma$ quantile of the F-distribution with $2(T-m)$ and $2(m+1)$ degrees of freedom.

The envelopes for the scaled Mahalanobis distances are a function of the above described $y_{m+1,T;\gamma}$.

Figure 2 shows that simulated confidence thresholds do not depend on the parameter vector Θ_i on which simulations are carried out. This is due to the fact that we rely on standardized statistics.

At the same time, Fig. 2 provides a comparison of Monte Carlo simulated thresholds and the approximated one obtained relying on Riani et al. (2009) approach. They share the same shape. Carrying out the same analysis considering time series with different lengths and parameters, we obtain the same closeness.

5.1 Contaminated data analysis

In what follows we consider three contaminations. Additive, level shift and innovation outliers are included in the simulated time series studied in the previous section.

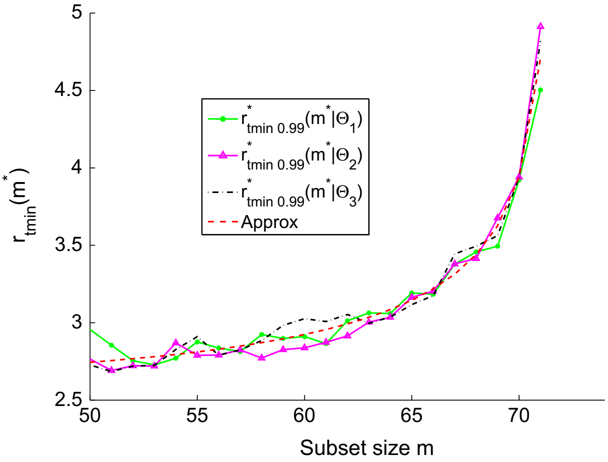

Fig. 2 Simulated vs. approximated 99 % confidence thresholds. Three confidence thresholds are obtained through Monte Carlo simulations with different simulation parameters Θ_S. They are compared to to the approximated threshold

Starting from additive outliers, we contaminate the $x_{1,t}$ time series by adding 0.1 (which represents more than 10 % of its original value) to the observation at time 32. The contaminated time series are shown in Fig. 3. Panel (a) highlights unit $x_{1,32}$ (indicated with $\xi_{1,32}$). This contamination induces two anomalous movements in the first time series, $x_{1,t}$.

From the PML weight distribution of panel (b) it is evident that both unit 32 and 33 are underweighted. However, only observation 32 crosses the critical value and can be considered an outlying unit. Passing to the FS, the effect of the contamination is evident in panel (c) where the evolution of $r_t(m^*)$ is shown.

Two observations, units 32 and 33 move separately from all other units. In panel (d) inference on outliers is made by comparing $r^*_{tmin}(m^*)$ to the 99 % confidence threshold. The plot highlights that units 32 and 33 cross the bound. As suggested by Riani et al. (2009), we super-impose the confidence threshold of size T − 1. We can state that units 32 and 33 are outliers.

In the Monte Carlo Sect. 6, we will use two additive contamination schemes. On the one hand we will contaminate all time series (scheme i.), on the other we will contaminate only one time series (scheme ii.) as in the above described example.

We continue our analysis by introducing a level shift on $x_{2,t}$ time series. In particular, as emphasized in panel (a) of Fig. 4, units $x_{2,\{65,\ldots,75\}}$ are subject to a shift of 0.15. Panel (b) shows that (concentrating on $\nu = 5$) unit 65 crosses the critical value pointing out the level shift. The weights assigned to units from 66 to 75 do not cross the critical value.

Passing to the FS analysis, in panel (c) we notice two separate clusters of trajectories from the beginning until the last few steps of the search. Contaminated units have higher $r_t(m^*)$ along the search, but in the last few steps there is a masking phenomenon. As emphasized in panel (d), the only observation which is identified as an outlier through the $r^*_{tmin}(m^*)$ analysis is unit 65. However, the joint analysis of panels (c) and (d) allows to correctly identify all contaminated units.

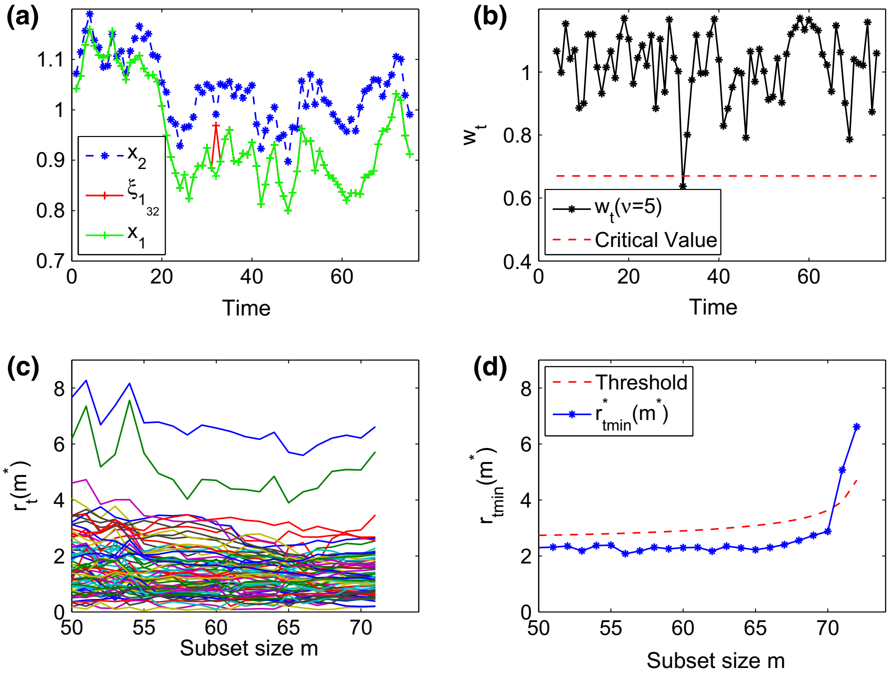

Fig. 3 Simulated time series: additive outliers. *Panel* **a** shows the contaminated time series. Unit $x_{1,32}$ (indicated with $\xi_{1,32}$) is contaminated. *Panel* **b** shows the PML weight distribution and the critical line for $\nu = 5$. Both units 32 and 33 are underweight, but only unit 32 crosses the critical value. *Panel* **c** shows the evolution of $r_t(m^*)$: units 32 and 33 move separately from all other observations. In *panel* **d** the 99 % confidence bound is crossed when units 32 and 33 enter into the subset

The last contamination scheme is associated to innovation outliers. Units $x_{\{1,2\},\{65,\ldots,75\}}$ are contaminated by applying a random shock from a Normal variable with mean 0 and variance 0.12.

In Sect. 6 an extensive Monte Carlo simulation is carried out considering both the contamination all time series (scheme i.) as well as only one of the p time series (scheme ii.).

In Fig. 5, panel (a) highlights innovation outliers $x_{\{1,2\},\{65,\ldots,75\}}$ indicated with $\xi_{\{1,2\},\{65,\ldots,75\}}$. Panel (b) shows that (concentrating on $\nu = 5$) only one of the last time series units crosses the critical value. The weights assigned to other contaminated observations do not cross the critical value and show that no further observations are detected as outliers.

Panel (c) of Fig. 5 highlights that innovation outliers are slightly masked in the last steps of the search. In panel (d), the comparison of $r^*_{tmin}(m^*)$ with the 99 % confidence threshold, emphasizes that all contaminated units (except for the one which is very close to the original pattern) are correctly detected.

Aiming to figure out how the FS helps the researcher to carry out the study, in Figure 6 the top panels show units not belonging to the subset at each step of the search in the innovation outliers analysis. In particular, in the top left panel (a) at

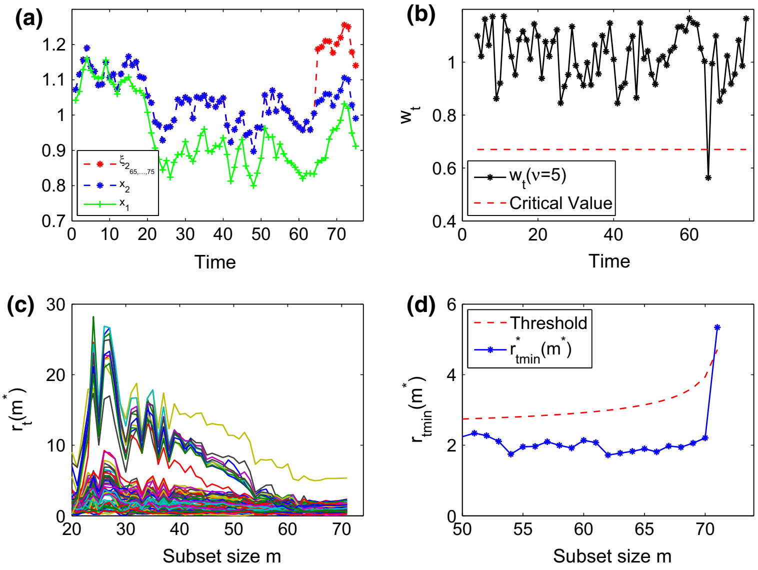

Fig. 4 Simulated time series: level shift outliers. In *panel* **a** the level shift of units $x_{2,\{65,\ldots,75\}}$ are indicated with $\xi_{2,\{65,\ldots,75\}}$. In *panel* **b** only the weight of unit 65 is dramatically low (level shift). *Panel* **c** shows that there are two separate clusters of trajectories: uncontaminated (lowest $r_t(m^*)$) and contaminated (highest $r_t(m^*)$). There is an evident masking effect in the last steps of the search. *Panel* **d** shows that, at the end of the search, only one outlier is detected. The joint analysis of *panels* **c**, **d** highlights all contaminated units

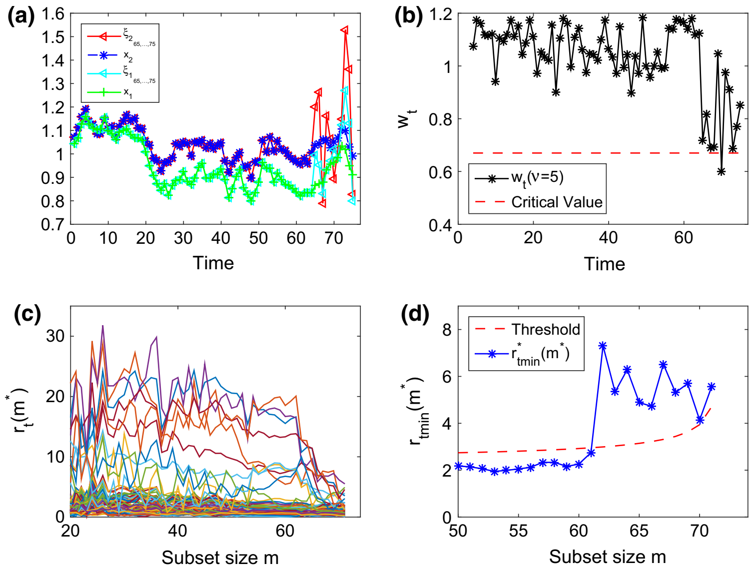

Fig. 5 Simulated time series: innovation outliers. In *panel* **a** innovation outliers $x_{\{1,2\},\{65,\ldots,75\}}$ are indicated with $\xi_{\{1,2\},\{65,\ldots,75\}}$. *Panel* **b** shows that only one unit is detected as atypical. *Panel* **c** shows that there are two separate clusters of trajectories corresponding to the uncontaminated (lowest $r_t(m^*)$) and the contaminated units (highest $r_t(m^*)$). In *panel* **d**, contaminated units are correctly detected

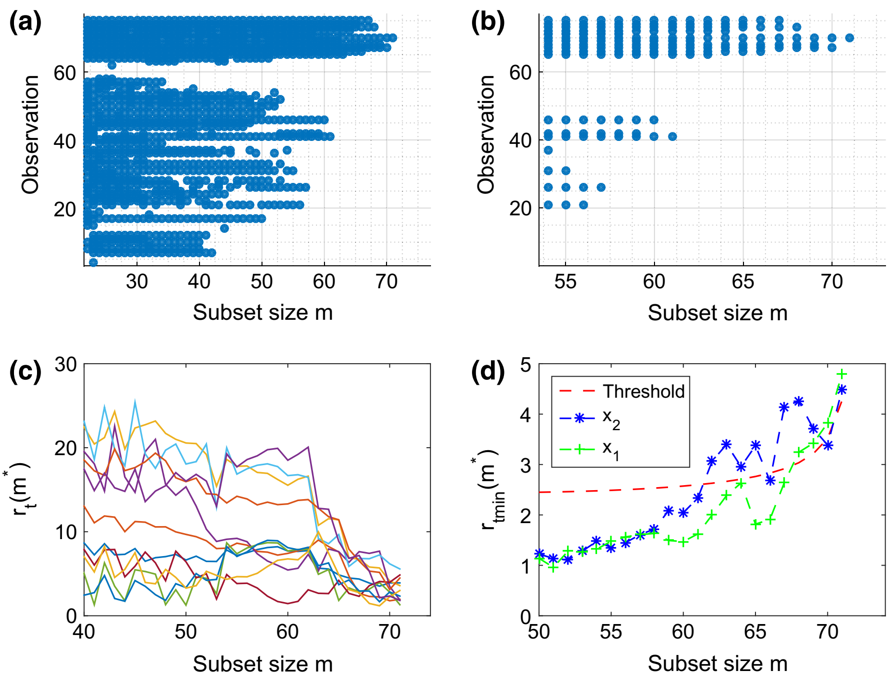

Fig. 6 Simulated time series with innovation outliers: analysis of units not belonging to the subset. In *panel* **a** units outside the subset are indicated with a filled circle for each step of the search. All contaminated units constitute a separate cluster. In *panel* **b** the last few steps are highlighted. *Panel* **c** shows the evolution of $r_t(m^*)$ for the contaminated units. *Panel* **d** highlights the major contribution of $x_{2,t}$ series to the overall outlying pattern

each step of the search we indicate with a filled circle units not belonging to the subset. In panel (b) we focus on the last few steps of the search by emphasizing that all contaminated units enter in the subset in the last steps. In the bottom panels we concentrate on the contaminated units. Panel (c) shows the evolution of $r_t(m^*)$ for the contaminated units. Panel (d) highlights the major contribution of $x_{2,t}$ series to the overall outlying pattern.

Returning to our motivation in Sect. 2, we can additionally underline that our framework can be effectively exploited by researchers who are skeptical about robust methods. As we have so far described throughout the analysis, the assumptions required for running the FS are only those on which the cointegration model relies.

In order to further verify the effectiveness of our procedure in detecting atypical units, the next section focuses on an extensive Monte Carlo analysis.

6 Monte Carlo simulation experiments

As we anticipated in the previous section, some rules are needed to correctly detect atypical units. According to Riani et al. (2009), we will rely on a two-stage process. In the first stage we perceive a signal of the potential outlier (or cluster of outliers), while in the second we decide whether the signal is correct and outliers are detected. In what follows we summarize these rules.

1. Signal perception. There are four conditions leading to the perception of a signal at step $m^\dagger$.
   - In the central part of the search we require three consecutive values of $r^*_{tmin}(m^{\dagger,*})$ above the 99.99 % envelope or one above 99.999 %. $r^*_{tmin}(m^{\dagger,*})$ stands for r^*_{tmin} computed at step $m^\dagger$ relying on the subset $m^{\dagger,*}$.
   - In the final part of the search we need two consecutive values of $r^*_{tmin}(m^{\dagger,*})$ above 99.9 % and one above 99 %.
   - $r^*_{tmin}(T-2)$ above the 99.9 % envelope.
   - $r^*_{tmin}(T-1)$ above the 99 % envelope. In this case a single outlier is detected and the procedure terminates.
2. Outlier detection. With a true signal corresponding to subset size $m^\dagger$ we start superimposing the 99 % envelope computed on $m^\dagger$ observations (instead of T observations). In the case where $r^*_{tmin}(m^{\dagger,*})$ is greater than the 99 % threshold, $(T-m^\dagger)$ outliers are detected.

By considering p-dimensional, $p = 2, 3, 4$, vectors and cointegration rank = 1,2,3, series of 1000 scenarios are generated for each simulation scheme. The analysis relies on standardized time series, $x_{j,t}$, obtained starting from the non-standardized ones (ns). $x_{j,t} = (x^{ns}_{j,t} - \mu^{ns}_j)/\sigma^{ns}_j$, where $j = 1, \ldots, p$, μ^{ns}_j is the average value of x^{ns}_j and σ^{ns}_j is the j-th time series standard deviation.

In Table 2 we describe the criteria exploited to contaminate time series by distinguishing between additive and innovation outliers. We introduce isolated additive contaminations in order to effectively count the outliers avoiding to incur in likewise temporary level shifts. On the other hand, given that the innovation scheme is intrinsically linked to contiguous units, to study innovation outliers we contaminate adjacent units. The level shift can be considered as a special case of the additive contamination scheme.

We can classify units as follows.

- True outliers, TO. Correctly detected outlier.
- True clean observations, TC. Correctly classified clean unit.
- False outliers, FO. Clean unit classified as outlier.
- False clean observations, FC. Outlier classified as clean unit.

In Table 3 we examine the effectiveness of the FS and PML procedures by detecting additive outliers. Starting the analysis from the uncontaminated environment, we emphasize that both FS and PML correctly classify all units as clean units.

Moving to a 1 % contamination, the introduction of one additive outlier implies two abnormal variations (as detailed in Sect. 5.1).

Thus, when we consider a 1 % contamination, 2 % outliers need to be detected. In the case of 5 % contamination, 10 % units need to be detected.

According to contamination scheme i. (all time series are contaminated at time τ_ω), in the 1 % contamination analysis (rank 1) both the FS as well as the PML procedures are very effective in detecting all outliers and correctly classify clean observations.

Table 2 Contamination details for the additive and innovation outlier schemes

Additive	Innovation
1. We contaminate all p time series randomly choosing the position τ_ω, $\omega = 1, 2, \ldots$, of the shock by simulating a uniform random variable in the interval $[1, T]$. The value of the shock is obtained as $\xi \cdot \mathbb{1}_{\{-1,1\}}$ where:	1. We contaminate all p time series. The time τ_ω of the first outlier is obtained from a uniform random variable in the interval $[1, T]$. Thus we contaminate the following $\tau + 1, \tau + 2, \ldots$ units by applying a shock $\varphi + \mathbb{1}_{\{-1,1\}}$ derived as follows:
(a) ξ is exogenously specified	(a) φ is drawn from a normal random variable $N(0, 3)$
(b) $\mathbb{1}$ assumes value $\{-1\}$ or $\{1\}$ according to the original sign of the variation $x_{1,\tau_\omega} - x_{1,\tau_\omega - 1}$. It allows avoiding shocks to be compensated by the original variation. This issue is particularly relevant where the magnitude of the shock ξ is relatively small	(b) $\mathbb{1}$ assumes value $\{-1\}$ or $\{1\}$ according to the original sign of the variation $x_{1,\tau} - x_{1,\tau-1}$
2. We contaminate only the time series j (instead of all time series) randomly choosing the position τ_ω of the shock simulating a uniform random variable in the interval $[1, T]$. The shock is computed as $\xi \cdot \mathbb{1}_{\{-1,1\}}$, where:	2. We contaminate only the time series j. The time τ of the first outlier is obtained from a uniform random variable in the interval $[1, T]$. Thus we contaminate the following $\tau + 1, \tau + 2, \ldots$ units by applying a shock $\varphi + \mathbb{1}_{\{-1,1\}}$ derived as follows:
(a) The time series to contaminate, x_j, is randomly specified for each simulation	(a) The time series to contaminate, x_j, is randomly specified for each simulation
(b) ξ is exogenously specified	(b) φ is drawn from a Normal random variable $N(0, 3)$
(c) $\mathbb{1}$ assumes value $\{-1\}$ or $\{1\}$ according to the original sign of the variation $x_{j,\tau_\omega} - x_{j,\tau_\omega - 1}$	(c) $\mathbb{1}$ assumes value $\{-1\}$ or $\{1\}$ according to the original sign of the variation $x_{j,\tau} - x_{j,\tau-1}$
We consider two alternative values of $\xi = 1, 2$ and two alternative percentage of contaminations 5 and 10 %. In addition, we consider alternative cointegration schemes considering sets of time series with (uncontaminated) rank = 1, 2, 3	We consider two alternative percentage of contaminations 5 and 10 % and we consider alternative cointegration schemes based on (uncontaminated) rank = 1, 2, 3

This holds for all values of $\xi = 1, 2$. Relying on the contamination scheme ii., the detection becomes more difficult for both FS and PML.

When we increase the number of contaminated units passing from 1 to 5 % the FS is always more effective than the PML in detecting outlying units.

The effectiveness of the FS is further confirmed when passing from rank 1 to ranks 2, 3. Emphasizing that Table 3 outputs are aligned with Franses and Lucas (1998) results concerning the PML approach, this table shows that in the most complex environments, the FS is more effective than the PML.

When passing to the innovation outlier contamination, a bunch of consecutive atypical units are generated. The percentage of outlying observations corresponds to the percentage of contaminated units plus one percent.

The FS is capable to catch almost all contaminated units in the contamination scheme i., while, as we expected, it is slightly less effective in the contamination

Table 3 Additive outlier analysis FS and PML comparison

Rank	Contam.%	Scheme	ξ	Approach	TO%	TC%	FO%	FC%
1	0	n.a.	0	FS	0.0	100.0	0.0	0.0
				PML	0.0	100.0	0.0	0.0
	1	i.	1	FS	**2.0**	98.0	0.0	0.0
				PML	2.0	98.0	0.0	0.0
			2	FS	**2.0**	98.0	0.0	0.0
				PML	2.0	98.0	0.0	0.0
		ii.	1	FS	**1.1**	98.0	0.0	0.9
				PML	1.1	98.0	0.0	0.9
			2	FS	**1.3**	98.0	0.0	0.7
				PML	1.2	98.0	0.0	0.8
	5	i.	1	FS	**6.5**	89.3	0.7	3.5
				PML	4.8	90.0	0.0	5.2
			2	FS	**6.6**	87.8	2.2	3.4
				PML	4.8	90.0	0.0	5.2
		ii.	1	FS	**3.8**	89.0	1.0	6.2
				PML	2.8	90.0	0.0	7.2
			2	FS	**4.4**	87.7	2.3	5.6
				PML	3.1	90.0	0.0	6.9
2	0	n.a.	0	FS	0.0	100.0	0.0	0.0
				PML	0.0	100.0	0.0	0.0
	1	i.	1	FS	**0.9**	97.9	0.1	1.1
				PML	0.7	98.0	0.0	1.3
			2	FS	**1.9**	97.3	0.7	0.2
				PML	1.6	98.0	0.0	0.4
		ii.	1	FS	**0.5**	98.0	0.0	1.5
				PML	0.4	98.0	0.0	1.6
			2	FS	**1.5**	97.9	0.1	0.5
				PML	1.2	98.0	0.0	0.8
	5	i.	1	FS	**2.4**	89.9	0.1	7.6
				PML	1.1	90.0	0.0	8.9
			2	FS	**5.3**	88.4	1.6	4.8
				PML	3.1	90.0	0.0	6.9
		ii.	1	FS	**1.2**	90.0	0.0	8.8
				PML	0.6	90.0	0.0	9.5
			2	FS	**4.5**	89.8	0.3	5.5
				PML	1.9	90.0	0.0	8.1
3	0	n.a.	0	FS	0.0	100.0	0.0	0.0
				PML	0.0	100.0	0.0	0.0
	1	i.	1	FS	**1.4**	97.8	0.2	0.6
				PML	1.0	98.0	0.0	1.0

Table 3 continued

Rank	Contam.%	Scheme	ξ	Approach	TO%	TC%	FO%	FC%
			2	FS	**1.9**	97.3	0.8	0.1
				PML	1.7	98.0	0.0	0.3
		ii.	1	FS	**1.2**	97.9	0.1	0.8
				PML	0.8	98.0	0.0	1.2
			2	FS	**1.2**	97.9	0.1	0.8
				PML	0.8	98.0	0.0	1.2
	5	i.	1	FS	**4.0**	89.6	0.4	6.1
				PML	1.5	90.0	0.0	8.5
			2	FS	**5.5**	87.9	2.2	4.5
				PML	3.7	90.0	0.0	6.3
		ii.	1	FS	**3.4**	89.5	0.5	6.6
				PML	1.0	90.0	0.0	9.0
			2	FS	**5.5**	88.1	1.9	4.5
				PML	2.9	90.0	0.0	7.1

Bold values indicate FS results

It is useful to underline that the number of outlier, in the additive scheme, is double of the contaminated units. Column *TO%* shows that the FS is always more effective than PML in detecting outlying units. As we expected the outlier detection is more effective when considering the scheme i. (all time series are contaminated) and for higher values of ξ

scheme ii.. The capability of detecting outliers is confirmed in both the 5 and 10 % contaminations and considering the three cointegration ranks on which the analysis is conducted. It is useful to remark that Franses and Lucas (1998) do not show any Monte Carlo experiment concerning innovation outliers. Table 4 shows that the FS is unequivocally more effective than the PML in detecting clusters of outlying units.

The cointegration analysis is particularly sensitive to the rank specification, in particular for identification issues.

The framework that we propose allows the researcher to conduct the analysis by applying the usual cointegration techniques to estimate model parameters when outlying units are detected.

On the basis of Nielsen (2004), we applied an automatic procedure based on innovation and structural dummy variables to manage the atypical units detected through the FS.

This process allowed us to estimate model parameters and obtain quite often the same rank as in the uncontaminated settings. However, the automatic introduction of dummy variables is far from being part of the FS framework due to the fact that from the FS perspective the researcher needs to be conscious of the presence of atypical units. The issue of how to deal with such observations is the crucial theme of the specific research. Therefore we do not provide any automatic procedure to manage outlying units.

In the next section we show how the FS can be exploited in real time series analysis.

Table 4 Innovation outlier analysis FS and PML comparison

Rank	Contam.%	Scheme	Approach	TO%	TC%	FO%	FC%
1	5	i.	FS	**5.1**	92.6	1.4	0.9
			PML	3.4	94.0	0.0	2.6
		ii.	FS	**4.4**	92.9	1.1	1.6
			PML	3.2	94.0	0.0	2.8
	10	i.	FS	**9.9**	87.8	1.2	1.1
			PML	5.6	89.0	0.0	5.4
		ii.	FS	**7.8**	88.1	1.0	3.2
			PML	4.2	89.0	0.0	6.8
2	5	i.	FS	**5.0**	93.1	0.9	1.0
			PML	3.1	94.0	0.0	2.9
		ii.	FS	**3.9**	93.7	0.3	2.1
			PML	2.4	94.0	0.0	3.6
	10	i.	FS	**10.0**	87.9	1.1	1.0
			PML	5.3	89.0	0.0	5.7
		ii.	FS	**6.3**	88.8	0.2	4.7
			PML	3.3	89.0	0.0	7.7
3	5	i.	FS	**4.5**	92.3	1.7	1.5
			PML	3.2	94.0	0.0	2.8
		ii.	FS	**4.2**	92.9	1.1	1.8
			PML	2.6	94.0	0.0	3.4
	10	i.	FS	**9.8**	87.3	1.7	1.2
			PML	5.7	89.0	0.0	5.4
		ii.	FS	**6.6**	88.1	0.9	4.4
			PML	3.6	89.0	0.0	7.4

It is useful to underline that the number of outlier, in the innovation scheme, correspond to the number of the contaminated units plus one. Column *TO%* shows that the FS is always more effective than PML in detecting outlying units

7 Macro-economic cointegrated analysis

Our empirical analysis relies on Bank of Italy, ISTAT and other financial quarterly datasets from 1993 to 2010.

We consider the following time series: default rates (DR), gross domestic product (GDP), unemployment rate (UR), real estate index (RE), Euro-Dollar exchange rate (ER), Euribor 3 months interest rate (EUR3) and Italian 10 years treasury bond's interest rate (BTP10).

In panel (a) of Fig. 7 we show the standardized time series. In panel (b) we highlight PML weights. This panel points out that observations: 18 (1997, 2nd quarter), 19 (1997, 3rd quarter), 64 (2008, 4th quarter) and 65 (2009, 1st quarter) cross or are very close to the critical value. They can be considered outlying units. Moving on to the FS analysis, the bottom panels of Fig. 7 show the evolution of residuals along the search. Panel (c) highlights the evolution of standardized residuals $r_t(m^*)$ for each observation. This

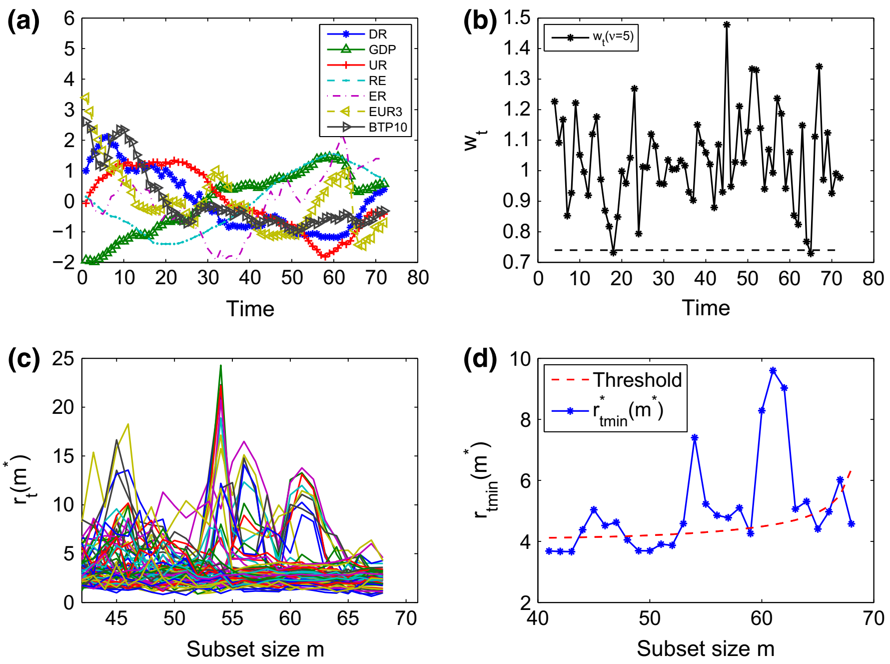

Fig. 7 CVAR time series analysis: PML vs. FS approach. In the *top left panel* **a** the standardized time series X_t are shown. In *panel* **b** PML weights and the critical value (*line*) are shown for $\nu = 5$. Observations: 18, 19, 64 and 65 cross or are very close to the critical value. They can be considered outlying units. In *panel* **c** the standardized residuals $r_t(m^*)$ are shown for each unit as the subset size m increases. *Panel* **d** shows $r^*_{tmin}(m^*)$ compared to the 99 % confidence threshold. A cluster of outlier is detected

plot shows a masking phenomenon. In the final steps of the search the standardized residuals are very small, while a few steps before few units show very high $r_t(m^*)$. Panel (d) confirms what is shown in panel (c) highlighting that a set of atypical units cross (at step 62) the 99 % confidence threshold. According to Riani et al. (2009), we need to analyse $r^*_{tmin}(m^*)$ compared to 99 % threshold from the right to the left hand side. In our case, for the whole set of T observations no outliers are detected. When the subset is composed by $T-1$ units, the 99 % threshold is crossed. Thus, we need to rely on a 99 % threshold built on $T-1$ units (this process is named super-imposition). We do not show the picture, but the super-imposition does not highlight any outlier at step $T-1$. We continue this process until when, at step 62 a cluster of outlier is detected.

Units from 10 (1995, 2nd quarter) to 16 (1996, 4nd quarter) constitute a separate cluster.

The FS suggests to the researcher to introduce dummy variables. In particular, a structural break from unit 10 (1995, 2nd quarter) to 16 (1996, 4nd quarter) is considered. In Table 5 the misspecification tests for the *original model* without dummies and the *robust FS* are compared. The same type of analysis was conducted introducing dummy variables for the units detected as outliers according to the PML approach. However, misspecification and other tests did not change significantly.

Table 5 Multivariate misspecification tests: *original model* vs. *robust FS*

	Original model		Robust FS model	
Autocorr. $\chi^2(49)$, [*p-value*]	72.034	[0.018]	62.929	[0.087]
Norm. $\chi^2(14)$, [*p-value*]	23.503	[0.053]	13.994	[0.450]
ARCH $\chi^2(1568)$, [*p-value*]	1655.197	[0.062]	1627.057	[0.146]
Trace corr.	0.537		0.633	

Table 6 Univariate tests: *original model* vs. *robust FS*

	Mean	Std. dev	Skewness	Kurtosis	ARCH	Norm.	R-Sq.
Original model							
DR	0.000	0.002	0.156	3.201	0.930 [0.628]	1.203 [0.548]	0.548
GDP	0.000	1,292	−0.153	3.510	2.341 [0.310]	2.617 [0.270]	0.594
UR	0.000	0.001	0.168	2.507	1.192 [0.551]	0.773 [0.679]	0.637
RE	0.000	0.401	−0.273	3.966	1.924 [0.382]	5.314 [0.070]	0.935
ER	0.000	0.036	−0.354	3.855	0.021 [0.990]	4.477 [0.107]	0.393
EUR3M	0.000	0.003	−0.073	3.240	1.791 [0.408]	1.275 [0.529]	0.578
BTP10	0.000	0.003	−0.018	3.700	0.598 [0.741]	3.784 [0.151]	0.546
Robust FS model							
DR	0.000	0.002	0.189	3.290	3.852 [0.146]	1.595 [0.450]	0.587
GDP	0.000	1,276	−0.318	3.584	2.014 [0.365]	3.058 [0.217]	0.605
UR	0.000	0.001	−0.038	2.887	1.376 [0.503]	0.179 [0.914]	0.692
RE	0.000	0.340	0.039	2.600	6.186 [0.045]	0.086 [0.958]	0.954
ER	0.000	0.031	0.176	3.222	0.412 [0.814]	1.314 [0.519]	0.549
EUR3M	0.000	0.002	0.113	2.525	1.959 [0.376]	0.418 [0.811]	0.624
BTP10	0.000	0.003	0.499	3.370	0.967 [0.617]	3.217 [0.200]	0.755

From Table 5 it is evident that the introduction of the structural break improves the overall model specification. Moving on from the *original model* to the *robust FS*, autocorrelation test improves while ARCH and Normality cannot be rejected in this latter case.

Concentrating on the univariate analysis, Table 6 shows that the Normality for RE and ER cannot be rejected and there is a reduction of the kurtosis for RE, ER and BPT10. In addition, the fitting is improved essentially for all variables and in particular for RE.

One of the key elements of the cointegration analysis is to figure out the cointegration rank by allowing to divide the data into r relations towards which the process is adjusting and $p-r$ relations which are pushing the process. The former can be considered as equilibrium errors, while the latter are common trends driving the system to its equilibrium.

According to Table 7, the *robust FS* trace test suggests clearly $r = 4$ (p-$value$ = 0.223). The *original model* indicated r = 4 with p-$value$ = 0.119 (p-$value$ = 0.108, r = 5).

Table 7 Trace test for the cointegragion rank: *original model* vs. *robust FS*

p − r	r	Original model			Robust FS model		
		Eig. value	Trace	p-value	Eig. value	Trace	*p-value*
7	0	0.635	217.170	[0.000]	0.79	310.604	[0.000]
6	1	0.493	146.688	[0.000]	0.607	201.243	[0.000]
5	2	0.424	99.200	[0.006]	0.495	135.826	[0.005]
4	3	0.267	60.620	[0.089]	0.39	88.049	[0.068]
3	4	0.203	38.894	[0.119]	0.313	53.452	[0.223]
2	5	0.184	23.050	[0.108]	0.211	27.211	[0.440]
1	6	0.119	8.833	[0.196]	0.141	10.666	[0.474]

Table 8 Hypothesis tests on β and α for the rank $r = 4$

	DR	GDP	UR	RE	ER	EUR3M	BTP10
Tests on β							
Stationarity, $\chi^2(p-r)$	5.583	0.585	1.012	1.619	2.198	1.214	1.879
[*p-value*]	[0.134]	[0.900]	[0.798]	[0.655]	[0.532]	[0.750]	[0.598]
Tests on α							
Weak exogeneity $\chi^2(r)$	20.744	7.738	10.913	26.788	5.644	3.698	50.747
[$p\ text{-}value$]	[0.000]	[0.102]	[0.028]	[0.000]	[0.227]	[0.448]	[0.000]
Unit vector, $\chi^2(p-r)$	10.127	1.749	12.018	1.881	13.848	14.951	7.978
[*p-value*]	[0.018]	[0.626]	[0.007]	[0.597]	[0.003]	[0.002]	[0.046]

As previously highlighted, the *robust FS* approach improves model fitting. It is now interesting to figure out its features. The different hypothesis about α and β [in Eq. (1) $\Pi = \alpha\beta'$] have been discussed at length by Johansen (1991, 1996). It is shown that the asymptotic distribution of the maximum likelihood estimates for β is mixed Gaussian. This implies that the likelihood ratio tests for linear restrictions on β are asymptotically distributed as χ^2. The system of variables should always be chosen because of their economic relevance, not because of their time series properties. Thus the system of variables will often constrain both $I(1)$ or $I(0)$ variables (the inflation rate is for instance sometimes found to be $I(0)$). To test if the variables are stationary we rely on *Stationarity* test in Table 8. In the same table, weak exogeneity is a hypothesis about the rows of α when the researcher is interested in the long run parameters α, β.

According to Table 8, concentrating on β estimates, it is evident that all the variables that we are analysing are stationary. When focusing on the vector α, on the one hand we notice that GDP, ER and EUR3M are weakly exogenous, i.e. they can be regarded as non-equilibrium correcting. On the other hand, the unit vector test emphasizes that GDP and RE are purely adjusting, they do not contribute to common trends.

8 Concluding Remarks

We extended the FS to the cointegration analysis by comparing our approach to the PML framework. We highlighted that the FS can be applied without making any additional assumption on the model investigated. The researcher can be very comfortable in using this procedure given that the FS relies on the traditional maximum likelihood estimators and monitors standardized residuals as a function of the increasing subset size.

In order to speed up the analysis and avoid the researcher to spend much effort on time consuming Monte Carlo simulations, we applied the Riani et al. (2009) procedure to compute confidence thresholds. They effectively approximated those obtained through Monte Carlo simulations.

Considering alternative contamination schemes, the extended Monte Carlo simulation analysis pointed out the effectiveness of the FS in detecting outliers. The FS was shown to be generally more effective than the PML. In particular the FS seemed to be more effective than the PML in detecting clusters of outliers. The real time series analysis further emphasized the effectiveness of the FS in supporting the overall cointegration analysis.

This is the first step in cointegration time series FS analysis. Further research needs to be devoted to study how to deal with questions such as seasonality and other cointegration specific issues.

Acknowledgments I am very grateful to the Editor and two anonymous referees for very constructive suggestions. I also thank Marco Riani, Søren Johansen, Bent Nielsen, Matteo Pelagatti and Daniel Ruediger for valuable comments on earlier drafts.

References

Atkinson AC, Riani M (2000) Robust diagnostic regression analysis. Springer, New York

Atkinson AC, Riani M (2006) Distribution theory and simulations for tests of outliers in regression. J Comput Gr Stat 15:460–476

Atkinson AC, Riani M, Cerioli A (2004) Exploring multivariate data with the forward search. Springer, New York

Bellini T, Riani M (2012) Robust analysis of default intensity. Comput Stat Data Ana 56(11):3276–3285

Bosco B, Parisio L, Pelagatti M, Baldi F (2010) Long-run relations in European electricity prices. J Appl Econom 25:805–832

Campbell JY, Perron P (1991) Pitfalls and opportunities: What macroeconomists should know about unit roots. In: NBER macroeconomic annual. National Bureau of Economic Research, Cambridge, pp 141–218

Franses PH, Haldrup N (1994) The effects of additive outliers on tests for unit roots and cointegration. J Bus Econ Stat 12:471–478

Franses PH, Lucas A (1998) Outlier detection in cointegration analysis. J Bus Econ Stat 16:459–468

Franses PH, Kloek T, Lucas A (1998) Outlier robust analysis of long-run marketing effects for weekly scanning data. J Econom 89:293–315

Johansen S (1988) Statistical analysis of cointegration vectors. J Econ Dyn Control 12:231–254

Johansen S (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59:1551–1580

Johansen S (1996) Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press, Oxford

Johansen S, Nielsen B (2010) Discussion: the forward search: theory and data analysis. J Korean Stat Soc 39:137–145

Johansen S, Nielsen B (2015) Analysis of the forward search using some new results for martingales and empirical processes. Bernoulli J (forthcoming)

Johansen S, Mosconi R, Nielsen B (2000) Cointegration analysis in the presence of structural breaks in the deterministic trend. Econom J 3:216–249

Lucas A (1995) An outlier robust unit root test with an application to the extended Nelson–Plosser data. J Econom 66:153–174

Lucas A (1997) Cointegration testing using pseudolikelihood ratio tests. Econom Theory 13:149–169

Nielsen HB (2004) Cointegration analysis in the presence of outliers. Econom J 7:249–271

Riani M, Atkinson AC, Cerioli A (2009) Finding an unknown number of multivariate outliers. J R Stat Soc Ser B 71:201–221