Quaderni di
STATISTICA

VOLUME 12 - 2010

LIGUORI EDITORE

Volume 12, anno 2010

ISSN 1594-3739 (printed edition)

Registration n. 5264, 6 December 2001, Court of Joustice at Naples
ISBN-13 **978 - 88 - 207 - 5364-1**

Director: Gennaro Piccolo

First italian edition, December 2010
Printed in Italy in December 2010 by OGL-Naples (IT)

This publication is financially supported by:

- Ministero delle Politiche Agricole, Alimentari e Forestali (gestione ex Centro per la formazione in economia e politica dello sviluppo rurale);
- Dipartimento di Scienze Economiche e Statistiche, Università di Salerno;
- Dipartimento di Scienze Statistiche, Università di Napoli Federico II;
- Dipartimento di Economia, Università di Roma Tre

Our policy is to use permanent paper from mills that operate a sustainable foresty policy and which has been manufactured from pulp that is processed using acid-free and elementary chlorine free practice. Furthermore, we ensure that the materials used have met acceptable environmental accreditations standard.

Index

Quaderni di Statistica
Vol. 12, 2010

Detecting contemporaneous mean co-breaking via ART and PCA

Carmela Cappelli   Francesca Di Iorio
Dipartimento di Scienze Statistiche, Università di Napoli Federico II
E-mail: carcappe@unina.it,   fdiiorio@unina.it

Summary: In this paper we propose a procedure to detect the presence of co-breaking i.e. of a common structural break occurring at an unknown date in a vector of time series. Co-breaking occurs if a linear combination of the time series cancels the break. Our procedure employs a regression tree based approach, called ART, to detect the presence of breaks and Principal Component Analysis to generate the linear combinations of the vector series. On each of these linear combinations ART is performed again to detect the presence of a break. The combination that "hides" the co-breaking time is the one minimizing the employed splitting criterion. The results of a simulation study carried out to evaluate the performance of the proposed approach are presented and discussed.

Keywords: Time series, Co-breaking, Regression Trees.

1. Introduction

In recent years the analysis of common features in economic time series has emerged as a relevant topic of research (for an overview of the common feature literature and recent developments see Urga, 2007). Indeed, economic or financial time series might have several distinctive features such as serial correlation, trends, seasonality, breaks; the knowledge that a group of variables presents the same feature offers valuable information for modelling, policy making and above all forecasting.

The notion of common features was formalized by Engle and Kozicki (1993): *A feature possessed by a group of series is said to be common to* *those series if a linear combination of the series does not have the feature.*

Hendry and Massmann (2007) state that the literature on common features can be classified into two groups: the first concerns the estimation of the relationship that cancels the common feature under the assumption that the number of this relationships is known; the second deals with testing the number of common features and simultaneously estimating the corresponding linear relationship. Engle and Granger (1987) cointegration regression, common feature regression by Engle and Kozicki (1993), "co-trending regression" proposed by Ogaki and Park (1998) and the residual based cointegration test robust under structural breaks (see Gregory and Hansen, 1996) belong to the first group, whereas the second group includes cointegration analysis introduced by Johansen (1988).

Developments of common features include common structural breaks or *co-breaking* whose notion was introduced by Hendry (1996) and Hendry and Mizon (1998). The idea arises from the evidence that deterministic shifts in the conditional mean of economic variables is a frequent feature in empirical economics and when these shifts happen, they affect simultaneously several variables. Indeed, deterministic shifts are often induced by policy makers when they change the level of some variables in order to reach specific goals, or they occur in financial markets when some news affect at the same time several variables.

Co-breaking investigates whether shifts in deterministic terms of individual series cancel under linear combinations of these variables. Detecting co-breaking is useful in forecasting, since its knowledge improves accuracy (see Hendry, 1996; Hendry and Massmann, 2007; Hendry and Clements, 2000 and 2003).

Moreover, there is a strong link between the concept of co-breaking and cointegration as discussed by Hendry and Massmann (2007), as well as a growing interest in literature for co-breaking regressions. An estimation algorithm for unconditional co-breaking (in the underlying process) has been proposed by Massmann (2007), whose Monte Carlo experiments show reasonable power for tests of co-breaking, although the break points were assumed *a priori* known. Note that although the concepts of common features and co-breaking have been developed in the econometric and financial literature, the topic is relevant in several fields of research such as hydrology, environmental sciences, marketing researches, and so on. For example in a lake district such as the Great Lakes in North America, co-breaking analysis might be applied to detect declining mean water levels that are known to cause social, economic and ecosystem disruption and ultimately may signal global warming.

In general, dealing with the presence of a structural break in time series analysis involves two kind of problems: detecting the break date, if it is unknown, and taking the break into account in modelling and forecasting. In the last two decades the literature of structural breaks has flourished; among the others Bai and Perron in various papers (1998, 2003, 2006) have presented a comprehensive discussion of the issue providing estimation methods, testing procedures and confidence intervals for dealing with multiple structural breaks occurring at unknown dates in a regression framework.

In this streamline, Cappelli and Reale (2005) proposed a fast method called ART (Atheoretical Regression Trees) that mimics Bai and Perron (1998, 2003) break estimation procedure, i.e. it finds similar and often the same results by employing Least Squares Regression Trees (Breiman *et al.* 1984) to locate the breaks.

ART is a completely heuristic procedure, thus it is not possible to conduct inference on the break dates. Nevertheless, extensive simulation studies and comparison with Bai and Perron's and other current methods have provided evidence of the usefulness of the approach. The results as well as applications to various real time series can be found in Cappelli *et al.* (2008) and Rea *et al.* (2010) whereas Cappelli and Di Iorio (2007) have fruitfully employed ART in the framework of an empirical strategy to distinguish occasional structural breaks and long memory.

In this paper we present a ART-based procedure to detect contemporaneous mean co-breaking. In particular, ART is employed for the preliminary identification of a break in the mean of each one-dimensional component of a multidimensional time series. The corresponding break dates represent candidate co-breaking dates and thus they are employed to delimit an interval on which Principal Component Analysis (PCA) is recursively performed to generate the linear combinations of the series. On each of these linear combinations ART is performed again to identify the best linear combination that "hides" the co-breaking date. Although the use of PCA is not new in the field of common feature analysis as a tool to evaluate the presence of mean co-breaking occurring in a multidimensional time series process (see Urga 2007 for a review), our approach represents an enhancement as it enables the co-breaking date to be *a priori* unknown i.e. to be estimated along the procedure. According to the current literature on the topic we focus on a single location shift.

The remainder of the paper is organized as follows. In section 2 we describe the ART procedure for detecting level shifts and we outline how it can be employed together with PCA for co-breaking analysis to estimate the presence and timing of a mean common break. In section 3 we report the results of a simulation study carried out to evaluate the ability of the proposed approach to detect co-breaking. Final remarks end the paper.

2. Regression trees for structural break analysis

In this section we discuss the issue of detecting multiple structural breaks and we introduce the ART method proposed by Cappelli and Reale (2005) to locate multiple breaks in the mean occurring at unknown dates; finally we show how the procedure can be employed for co-breaking analysis.

Let y_t be a series characterized by G regimes and $G-1$ breaks, with $t = 1, \ldots, T$. The objective is to estimates a set of a break-dates $(T_1, \ldots, T_g, \ldots, T_{G-1})$ that define a partition of the series

$$P(G) = \{(y_1, ..., y_{T_1}), ..., (y_{T_{(g-1)}+1}, ..., y_{T_g}), ..., (y_{T_{(G-1)}+1}, ..., y_T)\}$$

into subintervals where the series behavior is homogeneous with respect to some characteristic. In case of level shifts, the subseries are such that $\mu_g \neq \mu_{g+1}$, $\forall g = 1, \ldots, G-1$. To identify these types of breaks a common estimation criterion is based on the least square principle that yields to

$$P^{\star}(G) = \arg\min_{P(G)} \sum_{g=1}^{G} \sum_{t=T_{(g-1)}+1}^{T_g} (y_t - \mu_g)^2 \tag{1}$$

with $T_0 = 0$ and $T_G = T$. Cappelli and Reale (2005) showed that Regression Trees can be employed to identify multiple breaks in mean occurring at unknown dates. Their procedure, called Atheoretical Regression Trees (ART), works within the framework of Least Square Regression Trees (LSRT) that are based on the same objective function as (1). In LSRT a node h is split into its left and right descendants h_l and h_r to reduce the deviance of the response variable y_t fitting the mean of corresponding y's values to each node (for complete discussion the reader is referred to Breiman *et al.*, 1984). Thus, the algorithm selects the split that maximizes the *sum of square reduction* i.e. the difference:

$$SS(h) - [SS(h_l) + SS(h_r)] \tag{2}$$

where $SS(h) = \sum_{y_t \in h}(y_t - \mu(\hat{h}))^2$, is the sum of squares for node h, $\hat{\mu}(h)$ is the mean of the y_t values in node h and $SS(h_l)$ and $SS(h_r)$ are the corresponding quantity computed for the left and right descendants, respectively. Note that, since h_l and h_r are an exhaustive partition of node h, $SS(h)$ represents the total sum of squares whereas $[SS(h_l) + SS(h_r)]$ is the within-group sum of squares. Therefore the splitting criterion stated in (2) is equivalent to maximize the between-group sum of squares and it searches for the child nodes that are as far as possible, in terms of distance between the means. Figure 1 displays the graphical output of the splitting of node h in its child nodes h_l and h_r. Further splitting produces a binary tree diagram i.e. an acyclic direct graph.

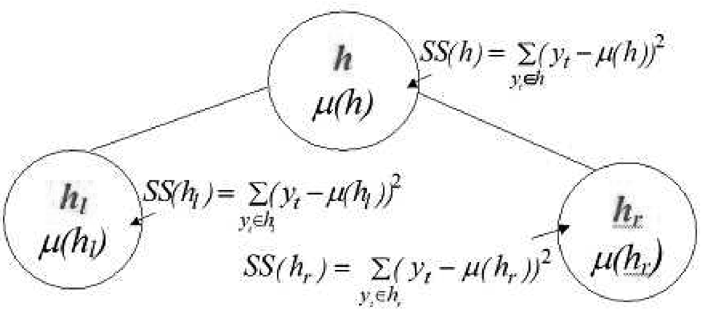


Figure 1. A single split in a binary tree diagram

Once the binary partition of a node is performed, the splitting process is applied to each subgroup separately, and so on recursively until either the subgroups reach a minimum size or no improvement of the criterion can be achieved.

LSRT provides a practical tool for locating structural breaks in the mean of long time series data. At this aim, let w be an arbitrary ascending (or descending) sequence of completely ordered numbers, for sake of simplicity take $w = 1, 2, \ldots, T$. Tree regressing y_t on w yields to a partition of the series into homogeneous subperiods such that $\hat{\mu}_g \neq \hat{\mu}_{g+1}$. The split points in the tree identify the break-dates whereas the different regimes are given by the y_t values falling into the tree terminal nodes.

Note that, according to the common practice, breaks are searched into a proper interval, i.e. $min_{obs} \leq w \leq T - min_{obs}$, where min_{obs} is a minimum number of observations need for modelling, estimating and testing purposes.

As in any tree procedure a criterion to select the relevant of set of splits, i.e. the sets of final breaks and corresponding regimes, is needed. At this aim either retrospective pruning or a stopping rule based on a testing procedure can be employed (for details see Rea *et al.* , 2010).
ART mimics the well known break estimation method of Bai and Perron (1998, 2003) that is based on Fisher (1958) method of exact optimization that produces the optimal partition of a time series in a given number G of subintervals. Bai and Perron's method is computationally expensive because it requires $O(GT^2)$ steps to find the optimal minimum sum of squares residuals partition. Indeed, several financial and econometric time series, such as stock market volatilities, are characterized by low frequencies in data collection and thus they generate a huge amount of observations then computational times and large memory requirements of the Bai and Perron's method make its use impractical on these types of series. On the contrary ART is not a global minimizer as Bai and Perron's method but it's much faster requiring at each node $O(n(h))$ steps to identify the best split, where $n(h)$ is the number of observations in the node h to be split. In other words, ART provides a suboptimal solution for which, being the partitions contiguous, misplacements can occur only on the boundaries. As discussed in Hansen (2001), although structural

breaks are treated as immediate, it's more reasonable to think that they take a period of time to become effect, thus misplacements on the boundaries are not a concern.

Extensive simulation studies and comparison with Bai and Perron's method and other current methods (see Cappelli *et al.* 2008 and Rea *et al.*, 2010) show that ART provides comparable results, in particular it performs well in the simulations for long series for which the computing time for the Bai and Perron's method may be prohibitive. Thus, ART is suitable for applied scientists who routinely analyze large numbers of time series or for those dealing with long series. Moreover ART can be easily implemented in any software.

2.1. Co-breaking analysis via ART and PCA

Following Hendry and Massmanm (2007, section 2.2), the notion of *Common Structural Breaks* can be formalized as follows.

Let $\mathbf{y}_t = (y_{1t}, \ldots, y_{nt})'$ be a n-dimensional series over $t \in \{0, \ldots, T\}$ with unconditional expectation $E[\mathbf{y}_0] = \boldsymbol{\mu}_0$ around an initial parametrization. A location shift is said to occur in $\mathbf{y}_t$ if, for any $t \in \mathcal{T} = \{1, \ldots, T\}$

$$E(\mathbf{y}_t - \boldsymbol{\mu}_0) = \boldsymbol{\mu}_t \in \mathcal{R}^n$$

and $\boldsymbol{\mu}_t \neq \boldsymbol{\mu}_{t-1}$, i.e. the expected value of $\mathbf{y}_t$ around the initial unconditional expectation in one time period differs from that in the previous time period.

Contemporaneous Mean Co-breaking occurs when

$$E[\boldsymbol{\Psi}' \boldsymbol{y}_t - \boldsymbol{\Psi}' \boldsymbol{\mu}_0] = \boldsymbol{\Psi}' \boldsymbol{\mu}_t = \mathbf{0}, \ \ \forall t \in \mathcal{T}$$

where $\boldsymbol{\Psi}$ is a $(n \times r)$ matrix of rank r $(r < n)$. Then the linear transforms $\boldsymbol{\Psi}'\mathbf{y_t}$ are independent of the location shifts and thus they cancel the break.

As pointed out before, so far co-breaking literature has been developed assuming the knowledge of the location shift timing (for an overview see Hendry and Massmann, 2007). On the contrary we propose a procedure to deal with contemporaneous mean co-breaking occurring at an unknown date, i.e. to detect the presence and timing of co-breaking. Thus, our procedure can be regarded as a supporting tool with respect to existing methods for co-breaking analysis to be employed when the co-breaking date cannot be assumed *a priori*.

The procedure that employs Principal Component Analysis (PCA) to generate the linear transforms in the framework of ART can be summarized as follows:

1. Let $\mathbf{Y}_t$ with $t \in \mathcal{T} = \{1, \ldots, T\}$ be the $(T \times n)$ matrix that collects the n observed time series. To each of the one-dimensional component of $\mathbf{Y}_t$ apply ART to detect a single break in the mean and let $\boldsymbol{\tau} = (\tau_1 \ldots \tau_j \ldots \tau_n)\prime$ be the vector of the identified break points i.e. each τ_j is the best split of the j-th series of $\mathbf{Y}_t$ into two subintervals, where $j = 1, ..n$ and $min_{obs} \leq \tau_j \leq T - min_{obs}$.

2. Let τ_m and τ_M be the minimum and maximum among the τ_j's values; thus, they represent the earliest and latest candidate co-breaking date, respectively. Consider the interval $[t_a = (\tau_m - \nu); t_b = (\tau_m + \nu)]$ where $\pm\nu \in N$ is added to expand the searching interval. For each possible partition obtained using $t_a \leq k \leq t_b$ as split point, estimate the PCA linear combination $\mathbf{z_k} = \mathbf{Y_t e}$, where $\mathbf{e}$ is the $(n \times 1)$ eigenvector corresponding to the largest eigenvalue of the correlation matrix of $\mathbf{Y}_t$ (thus r=1);

3. On each linear combination $\mathbf{z_k}$, ART is performed and the best linear combination $\mathbf{z_{k^\star}}$, i.e. the one "hiding" the co-breaking date, is the combination that shows the minimum value of the splitting criterion (2).

4. The estimated co-breaking date is the split point $k^\star$.

It is worth noting that both the preliminary identification of the candidate co-breaking dates and the analysis of the linear combinations benefit from a fast approach such as ART to locate the break due to large memory and computing requirements.

Also note that in this context neither a pruning procedure or a stopping rule is needed because a single contemporaneous mean co-breaking is searched, thus only the first split is performed and evaluated.

3. Simulation study

In order to evaluate the ability of our procedure to detect a contemporaneous mean co-breaking occurring at an unknown date, we have carried out a simulation study considering two alternative Data Generating Process (DGP): a basic $AR(1)$ model and a more general $ARMA(2,1)$ model.

In both cases we generated a 3-dimensional vector of time series $\mathbf{y}_t$ with $t = 1, ..., 200$. Each series has a structural break in ϕ_0 that occurs at time $t = 100$.
The $AR(1)$ model setting is as follows:

$$\boldsymbol{y_t} = \boldsymbol{\phi_{0,t}} + \boldsymbol{\Phi y_{t-1}} + \boldsymbol{\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim \boldsymbol{N(0, I)} \tag{3}$$

$$\phi_{0,t} = \begin{cases} \phi_0 \sim U(1, 1.3), & \text{if } t = 1, \ldots, 100; \\ \phi_0^\star \sim U(3, 3.3), & \text{if } t = 101, \ldots, 200; \end{cases} \tag{4}$$

where $\mathbf{\Phi} = (0.5, 0.7, 0.4)'$ and $U(a, b)$ denotes a continuous Uniform random variable over the support (a, b). Values ϕ_0 and $\phi_0^\star$ are fixed along the Monte Carlo experiment.

The $ARMA(2,1)$ model setting is as follows:

$$\boldsymbol{y_t} = \boldsymbol{\phi_{0,t}} + \boldsymbol{\Phi_1 y_{t-1}} + \boldsymbol{\Phi_2 y_{t-2}} + \boldsymbol{\Theta \epsilon_{t-1}} + \boldsymbol{\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim \boldsymbol{N(0, I)} \tag{5}$$

where $\mathbf{\Phi}_1 = (1.5, 0.7, 0.2)'$, $\mathbf{\Phi}_2 = (-0.7, -0.2, 0.3)'$, $\mathbf{\Theta} = (-0.2, 0.4, 0.7)'$ and, $\phi_{0,t}$ is generated as in (4).

Figure 2 depicts one of the 1000 Monte Carlo replications generated according to above defined DGP.

Also, to give an insight into the problem, in Figures 3 and 4 we report for the two DGP the estimated PCA linear combination associated with the actual co-breaking time $t = 100$ (left panel) and the linear combination associated with $t = 80$ (right panel), that is 20 observations before the true co-breaking point.

As we can see in both cases the linear combination corresponding to the erroneous co-breaking date ($t = 80$) does not cancel the break, on the contrary it shows a false appearance of 2 breaks whereas the linear com-

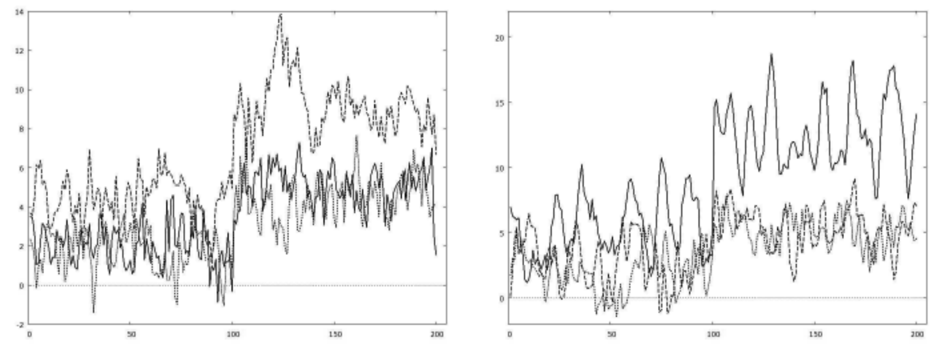


Figure 2. Illustrative example of the simulated 3-dimensional vector of time series. Left panel $AR(1)$ model, right panel $ARMA(2, 1)$ model.

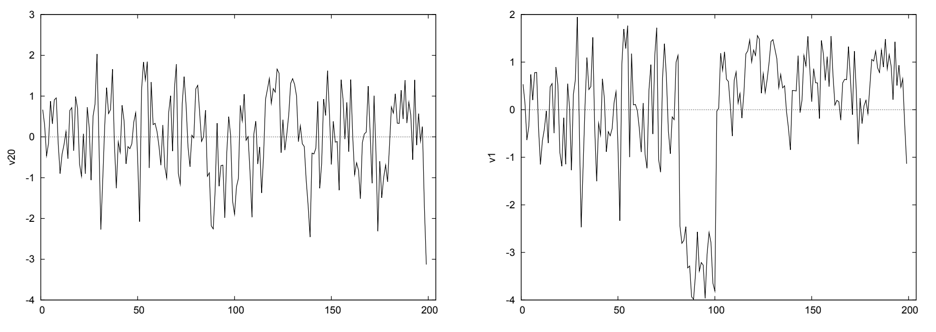
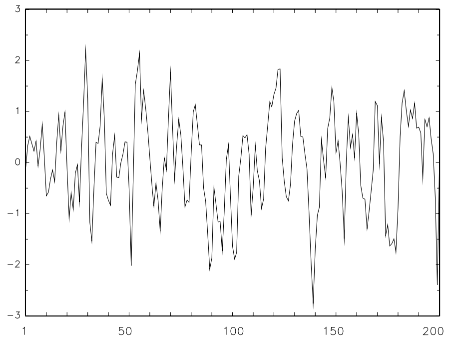


Figure 3. $AR(1)$ model: linear combination corresponding to the actual co-breaking time (t=100) and to (t=80).

bination corresponding to the actual co-breaking date shows no evidence of the breaks.

The procedure described in section 2.1 has been applied by setting $min_{obs} = 80$ and $\nu = \pm 3$ for both DGP. The results, averaged over the 100 replications, are summarized in Table 1 where, for the two DGP we report the Monte Carlo average of the earliest and latest co-breaking date $\bar{\tau}_m$ and $\bar{\tau}_M$, respectively (the associated standard errors are given in brackets) and the percentages of correct identifications $\%ci$ i.e. number of correctly identified contemporaneous mean co-breaking date within a short interval from the actual date.



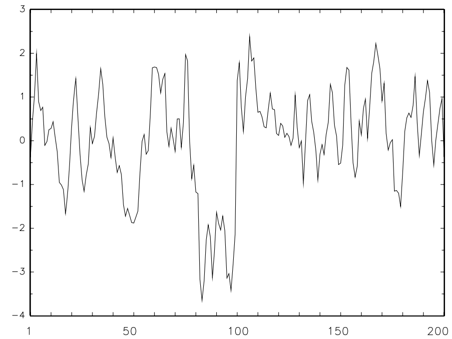


Figure 4. $ARMA(2,1)$ model: linear combination corresponding to the actual co-breaking time (t=100) and to (t=80).

Table 1. Main results averaged over the 1000 MC replications; co-breaking at time 0.5×T ($t = 100$)

	$\bar{\tau}_m$	$\bar{\tau}_M$	$\%ci \pm 6$	$\%ci \pm 4$	$\%ci \pm 2$	$\%ci \pm 1$
$AR(1)$	100.3	101.6	100	94	58	41
	(1.01)	(1.95)				
$ARMA(2,1)$	99.3	103.1	98	80	52	35
	(2.92)	(2.98)				

We see that the values of $\bar{\tau}_m$ and $\bar{\tau}_M$ are very close to each other and to the actual break date $t = 100$, confirming that ART is very effective in identifying the break date on individual series. As to its application to the linear combinations together with the selection criterion described in section 2.1 it provides percentages of correct identifications over the 50% in the ± 2 interval around the actual break date reaching the 100% and 98% in the interval ± 6. As expected the percentages of correct identifications tend to be higher for the $AR(1)$ model but, for both DGP the results show that the proposed approach represents a promising fast tool to detect a relatively short interval to focus on for further analysis and investigations.

In order to illustrate how the selection criterion works, for the first $AR(1)$ Monte Carlo replication we depicted in Figure 5 the sequence of the values of the splitting criterion (2) computed for the 40 PCA linear combinations generated over the whole interval $(min_{obs}, T - min_{obs})$ i.e.

with $80 \leq k \leq 120$ in such a way to better investigate the behavior of the criterion.

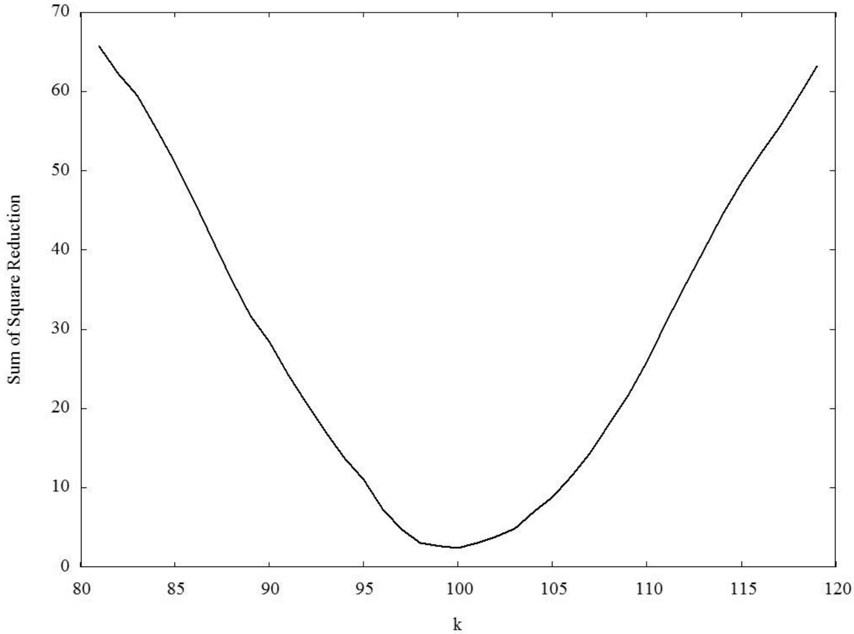


Figure 5. Sequence of the splitting criterion values associated with the linear combinations generated for $80 \leq k \leq 120$.

The plot shows that the linear combinations with very low values (almost 0) of the splitting criterion, i.e. those most likely to contain the co-breaking, are located around the actual common break date confirming that the proposed approach helps identify an interval that "hides" the co-breaking date. In this case the best linear combination is the 20-th, i.e. the minimum is reached at the actual co-breaking time $k^{\star}{=}min_{obs}+20{=}100$.

To evaluate how sensitive the method is to the position of the co-breaking we have carried out a further simulation considering a break at time $t = 150$. The results are reported in Table 2.

As in the previous case the values of $\bar{\tau}_m$ and $\bar{\tau}_M$ are close to each other and, surprisingly, the percentages of correct identifications are (slightly) higher.

Table 2. Results averaged over the 1000 MC replications;co-breaking at time 0.75×T ($t = 150$)

	$\bar{\tau}_m$	$\bar{\tau}_M$	$\%ci \pm 6$	$\%ci \pm 4$	$\%ci \pm 2$	$\%ci \pm 1$
$AR(1)$	150.1	151.7	99.6	96.3	69.2	46.61
	(1.81)	(1.75)				
$ARMA(2,1)$	148.2	153.1	96.7	87.8	55.5	37.5
	(6.68)	(4.00)				

The simulations have been carried out with GAUSS routines and, on a Intel Core Duo 1.8 GHz, the CPU time was 1 min and 16 sec for the break at time $t = 100$ and 1min and 22 sec at time $t = 150$.

Finally note that, in principle, the approach could be applied separately to the subseries identified by the co-breaking date to detect further mean co-breaking although, up to now we have not yet investigated this issue.

4. Conclusions

In this paper we have proposed a procedure to detect a contemporaneous mean co-breaking occurring at an unknown date in multivariate time series. Co-breaking represents a development of common features whose presence implies that a linear combination of the series cancels the feature, and generally assume the form of a level shift. Our approach employs a fast tool, namely ART, for the preliminary identification of breaks in each one-dimensional component of a multidimensional time series. The corresponding break dates represent candidate co-breaking dates and thus they are employed to delimit the interval on which to perform recursively Principal Component Analysis to generate the linear combinations of the series. On each of these linear combinations ART is performed again and the best linear combination that provides the estimated co-breaking date is the one minimizing the splitting criterion. We considered as investigating model various ARMA models in which, according to the literature, a location shift i.e. a change in the constant term occurs at a given time. This kind of model might be regarded as a particular case of an *arranged autoregression* in the framework of the threshold autoregressive (TAR) model (see Tsay, 1989) where a subset of observations belongs to the first regime and the remaining to the second regime; although the TAR model as well as the Markov-Switching model (Hamilton, 1989) are commonly used to allow for stochastic regime shifts whereas, as said before, we focused on a non stochastic regime shift.

The results of the simulation study carried out to evaluate the effectiveness of our proposal have shown that the proposed procedure provides high percentages of correct identification of the contemporaneous mean co-breaking date and thus it helps to identify an interval that is likely to contain the co-breaking date. Further simulations to investigate the performance of the proposed approach when the mean co-breaking is moderate or small, when it occurs in the tails and the case of multiple breaks are the subject of ongoing research, as well as the possibility to extend the procedure to other classes of processes such as the above mentioned TAR and Markov Switching.

Acknowledgements: The authors would like to thank Dipartimento di Scienze Statistiche, Università degli Studi di Napoli Federico II, for financial support.

References

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

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

Bai J., Perron P. (2006), *Multiple structural change models: a simulation analysis*, in: Phillips P.C.B., Corbae D., Durlauf S.N., Hanse B.(eds.) *Econometric theory and practice: frontiers of analysis and applied research*, Cambridge University Press, 212-240.

Breiman L. , Freidman J.H., Olshen R.A., Stone C.J. (1984), *Classification and regression trees*, Wadsworth & Brooks, Monterey.

Cappelli C., Reale M. (2005), Dating multiple structural breaks occurring at unknown dates via Atheoretical Regression Trees, in: Provasi C. (ed.) *S.Co.*

2005: Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione, Cleup, Padova, 479-484.

Cappelli C., Di Iorio F. (2007), Structural breaks *versus* Long Memory: a Simulation Study , *Statistica Applicata*, 17, 285-295.

Cappelli C., Penny R.N., Rea W.S., Reale M. (2008), Detecting multiple mean breaks at unknown points with regression trees, *Mathematics and Computers in Simulation*, 78, 351-356.

Engle, R.F, Granger C.W.J. (1987) Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica*, 55, 251-176.

Engle R.F., Kozicki F. (1993), Testing for common features, *Journal of Business & Economic Statistics*, 11, 369-395.

Fisher W. (1958), On grouping for maximum homogeneity, *Journal of the American Statistical Association*, 53, 789-798.

Gregory, A.W., Hansen B.E. (1996), Residual-based tests for cointegration in models with regime shifts, *Journal of Econometrics*, 70, 99-126.

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

Hansen B.E. (2001), The new econometrics of structural change: dating breaks in the U.S. labor productivity. *Journal of Economic Perspectives*, 15, 117-128.

Hendry D.F. (1996), *A theory of Co-breaking*, mimeo Nuffield College, University of Oxford.

Hendry D.F., Clements M.P. (2000), Economic forecasting in face of structural breaks, in *Econometric modelling: techniques and applications*, Di Sean Holly, Martin Weale (eds), Cambridge University Press, 3-37.

Hendry D.F., Clements M.P. (2003), Economic forecasting: some lessons from recent research, *Economic Modelling*, 20, 301-329.

Hendry D.F., Massmann M. (2007), Co-breaking: recent advances and a synopsis of the literature, *Journal of Business & Economic Statistics*, 25, 33-51.

Hendry D.F., Mizon G.E. (1998), Exogeneity, casuality and co-breaking in economic policy analysis of a small econometric model of money in the UK, *Empirical Economics*, 23, 267-294.

Johansen S. (1988), Statistical Analysis of Cointegrating Vectors, *Journal of Economic Dynamics and Control*, 12, 231-54

Massmann M. (2007), Cobra: A package for co-breaking analysis, *Computational Statistics & Data Analysis*, 52, 663-679.

Ogaki M., Park J.Y. (1997) A cointegration approach to estimating preference parameters, *Journal of Econometrics*, 82, 107-134.

Rea, W.S., Reale M., Cappelli C., Brown J.A. (2010), Identification of changes in mean with regression trees: an application to market research, *Econometric Reviews*, 29(5-6),1-14.

Tsay, R. S. (1989) Testing and Modelling Threshold Autoregressive Processes, Journal of the American Statistical Association, vol. 84, No. 405, 231-240.

Urga G. (2007), Common Features in economics and finance: an overview of recent developments, *Journal of Business & Economic Statistics*, 25, 2-11.