ROMA TRE
UNIVERSITÀ DEGLI STUDI

COLLANA DEL
DIPARTIMENTO DI ECONOMIA

BLUE CHIP ITALIAN BANK STOCKS: CHAIN GRAPH MODELS FOR VAR AND MARCH PARAMETERS SHRINKING

Andrea Pierini

ISSN 2279-6916 Working papers

(Dipartimento di Economia Università degli studi Roma Tre) (online)

Working Paper n° 205, 2016

I Working Papers del Dipartimento di Economia svolgono la funzione di divulgare tempestivamente, in forma definitiva o provvisoria, i risultati di ricerche scientifiche originali. La loro pubblicazione è soggetta all'approvazione del Comitato Scientifico.

Per ciascuna pubblicazione vengono soddisfatti gli obblighi previsti dall'art. l del D.L.L. 31.8.1945, n. 660 e successive modifiche.

Copie della presente pubblicazione possono essere richieste alla Redazione.

esemplare fuori commercio
ai sensi della legge 14 aprile 2004 n.106

REDAZIONE:
Dipartimento di Economia
Università degli Studi Roma Tre
Via Silvio D'Amico, 77 - 00145 Roma
Tel. 0039-06-57335655 fax 0039-06-57335771
E-mail: dip_eco@uniroma3.it
http://dipeco.uniroma3.it

DIPARTIMENTO DI ECONOMIA

BLUE CHIP ITALIAN BANK STOCKS: CHAIN GRAPH MODELS FOR VAR AND MARCH PARAMETERS SHRINKING

Andrea Pierini

Blue Chip Italian Bank Stocks: Chain Graph models for VAR and MARCH parameters shrinking

Andrea Pierini

Abstract In this paper Chain Graph models are applied to a multivariate time series of blue chip Italian bank stock returns, in order to construct graphs with minimum BIC among decomposable graphs.
Firstly a chain graph is built for present and past values of the time series in order to reduce the parameters of a VAR(1) model estimation, setting to zero the parameters related to non-edges in the graphs.
Then another chain graph is built for present and past values of the squared residuals of the previous estimated model.
An MARCH(1) model is so constructed by restricting to zero the parameters which are not indicated by these graphs. In this way a great reduction of parameters is put in place, using the opportune multidimensionality only if it is necessary.
The parameter shrinking doesnt affect the return and standard deviation forecasts while improving the efficiency of the estimations. .

Key words: Chain Graph Model, VAR, MARCH.

JEL codes : C320, C510,C580 .

I thank the full professor Marilena Barbieri, Department of Economics, University of Roma Tre, for the peer review of the paper.

Andrea Pierini
University of Roma Tre, Department of Economics
Via Silvio D'Amico, 00145 Roma, e-mail: andrea.pierini@uniroma3.it

1 Introduction

For an investor holding multiple stocks, the dynamic relationships between returns of his/her portfolio play an important role in its decision making (selling or holding stocks).
To this end treating his/her stocks jointly can be an advantage with respect to treat them separately, and the multivariate time series is the natural framework to do it.
Even if is not universally agreed (Ding and Engle, 2001), the multiple financial time series are treated in literature with heteroskedastic models ($MARCH$) , (Bollerslev, Engle and Nelson, 1994), and vector autoregressive models (VAR) which rests on the predictability of the average return on stocks, (Campbell, Chan and Viceira, 2003).
There are various interesting approach to model VAR trough the use of directed acyclic graph, (Awokuse and Bessler, 2003), (Babula ,Bessler, Reeder and Somwaru, 2004),(Moneta, 2008), (Oxley, Reale and Wilson, 2004), (Oxley, Reale and Wilson, 2009), (Reale and Wilson, 2001) or undirected graph, (Songsiri, 2010), or to model in a bayesian framework a dynamic heteroskedastic linear model trough the use of undirected, (Carvalho and West, 2007).
However the use of chain graph is not exploited to tackle the problem of the heteroskedastic dynamic description of the data so that the time causality is not specified or it requires a specific graph redefinition.
In other approaches a specific graphical model is defined to solve the problem of time series, (Eichler, 2000), (Gao and Tian, 2010). In this paper the use of chain graph is adopted in order to relay on the long theoretical history of it.

The $VAR-MARCH$ type models suffer of computational burden as the number of parameters to be estimated are large with respect to univariate models. This paper suggests the use of Chain Graphs to identify only the significant parameters to be estimated for the $VAR-MARCH$ models whenever a subset of the variables are considered connected. The Chain Graphs are able to exploit the time ordering of the variables, that is some variables are prior in time to others. So the variables are divided into blocks and the modelling respects this ordering between this blocks without assuming any ordering within the blocks (time contemporaneity). Firstly a chain graph is built based on the returns and its immediate past.In case of connected variables, *Restricted* $VAR(1)$ model is estimated taking into account only the connections identified before.
Then a full $VAR(1)$ model is also estimated and the number of parameters are matched with the restricted case. Moreover the log-Likelihood, AIC, MSE of the return forecasts are also matched.
As the residuals exhibit heteroskedasticity, $MARCH(1)$ models are then applied to them in order to take this particular aspect into due consideration. A second chain graph is built based on the residuals and its immediate past.In case of connected residuals, *Restricted* $MARCH(1)$ models are estimated using only the connections previously identified. For the subset of variables which are not connected at all, univariate $AR(1)-ARCH(1)$ are estimated.

Again a full $MARCH(1)$ model is also estimated and the number of parameters are matched with the restricted case. Moreover the log-Likelihood, AIC, MSE of the standard deviation forecasts are also matched.
As the residuals exhibit heteroskedasticity, $MARCH(1)$ models are then applied to them in order to take this particular aspect into due consideration.

In this way a great reduction of calculation is obtained and a more efficient estimation is reached with respect to multidimensional full models.The orders m,M of $VAR(m),ARCH(M)$ can be augmented leaving the approach unchanged. In that case $m+1$ blocks, containing the present and lag $i, i=1,...,m$ past values ,should be built, with the minimum BIC chain graphs edging the nodes.
Moreover the application of multidimensional models for the residuals in the case of detected connection, can be prohibitive due to the dimension of the problem if it is done without simplification. The time series considered regard the Blue Chip Bank sector Stock weekly figures from 3 January 2000 to 9 March 2015 taken by Yahoo finance.
The resulting $VAR-MARCH$ models were stable. As proved in (Ding and Engle, 2001), there is failure to converge on a solution if there are more than five series . So the dimension is chosen to be 5. However, the Chain Graph can be applied to a bigger dimension set in order to find possible subsets of stocks that can be treated altogether with VAR-MARCH multivariate methods.

2 Chain Graph Description

This paper exploits the fact that the set of variables under study partitions into two groups, (X,Y), where the variables in Y can be considered response and the variables in X explanatory.
Indeed the variables in Y are time successive to the variables in X and are referred to the same stocks but in two different times, say t and $t-1$. Here the data are monthly figure so the time t is a month of observation.
This asymmetry has to be incorporated into the model, see (Whittaker, 1990).
The additional structure is part of the initial graphical model specification . Generally speaking the graphical models are based on a partition of the variables into a chain of ordered block, $(X_{b_1},...,X_{b_{\tilde{m}}})$, see (Whittaker, 1990). In the special case $\tilde{m}=2$, past and present blocks.
A graphical model with response variables Y and explanatory variables X is a specification of a density function $f_{Y|X}$ that incorporates a specified subset of intra-response pairwise conditional independence statements:

$$Y_i \perp Y_j | (Y_{-(i,j)}, X)$$

and response-explanatory pairwise conditional independence statements:

$$Y_i \perp X_j | (Y_{-i}, X_{-j})$$

The conditional independence graph for $f_{Y|X}$ is identical to the graph for $f_{(Y,X)}$ the in which the subgraph corresponding to X is complete.
Thus the chain graph is built by combining separate graphs models for each block of the response variables given the explanatory. Specifically, in the case of two blocks, the first step is to find an undirected graph for the explanatory variables. Then, given all the edges among the explanatory variables (saturated model), a graph is found for the response-explanatory variables. Lastly the chain graph is found by combining in a single graph, the two precedent graphs with directed edges from the explanatory block to the response one.
So the chain graph model and the joint graph model, see (Pierini, 2013) for an analogue application, are distinct with neither model a subset of the other .
If the normality of the conditional density is assumed, like the present case, then the hypothesis that the *i*-th regression coefficient is zero is equivalent, in the case of a single response Y, to the hypothesis the Y and X_i are conditionally independent given the remaining variables.
Thus if we consider the multiple time series model equation by equation there seems to be an equivalence with the chain graph,, see (Whittaker, 1990), (Lutkepohl, 2006).

Let's now deal with the specific case at hand. A multivariate time series of the log return for the stock index i at time t is $r_{i,t}$ $i = 1,...,k$ and $t = 1,...,T$, that's to say $r_{i,t} = log(p_{i,t}/p_{i,t-1})$, $p_{i,t}$ is the index value at time t. We can then define the time series of the past $r^{lag1}_{i,t} = r_{i,t-1}$, $i = 1,...,k$ and $t = 1,...,T$ where $r_{i,0} = NA$. A *graph* is defined as a couple $G = (V,E)$ where V is a set of vertices or nodes and E is a set of edges, (Hojsgaard, Edwards and Lauritzen, 2012). Two vertices α and β of G are said *adjacent*, $\alpha \sim \beta$, if there is an edge between them. A *cycle* is a sequence of nodes $v_1, v_2.., v_{n-1}, v_1$ of length n where $v_i \sim v_{i+1}$. If a cycle has adjacent elements $v_i \sim v_j$ and $j \notin \{i-1, i+1\}$ then it is said to have a *chord*. If it has no cords it is said *chordless*. A graph with no cordless cycles of length ≥ 4 is called *triangulated*. A generalization of DAGs and undirected graphs are the *chain graphs*. Chain Graphs are graphs that have undirected edges or directed edges associated with their pair of nodes, subject to the constraint that no cycle containing directed edge(s) is permissible (Whittaker, 1990). It can be seen that such graphs have a block representation which separates vertices into blocks such that inter-block edges are arrows, intra-block edges are undirected and the blocks are completely ordered. Let A,B,C be subset of V and f a joint density function of $(X_v)_{v \in V}$ then $A \perp B | C$ if $f(x_A, x_B | x_C) = f(x_A | x_C) f(x_B | x_C)$ for each possible value x_A of X_A,x_B of X_B,x_C of X_C. In the hypothesis of multivariate gaussian variables , $f = N(\mu, \Sigma), K = \Sigma^{-1}$, if V represents the set of this variables, it is possible to define a dependence graph $G_K = (V, E_K)$ with $k_{u,v} = 0$ whenever there is no edge between a pair of vertices $u, v \in V$ .So if two vertices $u, v \in V$ are not adjacent, that's to say there's no edge, it holds that $u \perp v | V \setminus \{u, v\}$, u and v are *conditionally independent* given the others . Firstly V contains the set of variables Y given by $r_i, i = 1,...,k$ followed by $r^{lag1}_i, i = 1,...,k$ with $r_i = (r_{1,1},...,r_{1,T})$, $r^{lag1}_i = (r^{lag1}_{1,1},...,r^{lag1}_{1,T})$, $f = N(\mu, \Sigma)$. The

method used to select a chain graph, that can deal with high dimension, is to search among the decomposable graph. *Decomposable* models are graphs that are triangulated. For this type of graphs a closed-form for the maximum likelihood estimates are available allowing computational simplification. The starting graph is the minimal AIC or BIC forest. A *forest* is an acyclic undirected graph, that is, an undirected graph with no cycle. Then search the edge giving the maximal AIC or BIC reduction is added until further reduction are not possible. Here the criterion is $-2logL+kdim(G)$, where $dim(G)$ is the number of edges of G, L is the likelihood, $k=2$ (AIC), $k=log(n)$ (BIC). Then it is possible to split the set of variables $\Omega_t=\{r_i, i=1,...,k\}$ in two parts. A subset of $m<k$ variables $C=\{r_i, i=i_1,...,i_m\}$ which are connected to some element of $\Omega_{t-1}=\{r_i^{lag1}, i=1,...,k\}$, with the exception of their own past. A subset of $k-m$ variables $\overline{C}=\{r_j, j=j_1,...,j_{k-m}\}$ which are not connected to any element of $\Omega_{t-1}=\{r_i^{lag1}, i=1,...,k\}$, with the exception of their own past. For the variables in $\overline{C}$ we impose the link with their own past, to treat them all together with the variables in C in a multiple restricted regression and compare with the full counterpart. That is called an *adjustment*.
As the concurrent correlations cannot be used in the forecast (Tsay, 2010), only the past to present information is adjusted and used.
However information among the present (contemporaneous) values are given in the chain graphs and could be used to restrict a *SVAR* model. In this case it doesn't seem possible to describe the time dependency of the variance-covariance matrix as it would be already estimated in a constant-time way in the *SVAR* model.

3 R-VAR and R-MARCH Description

Having said that, the equation of each variable in the model *RVAR* is a multiple normal regression, so the following result, which links regression and graph theory, can be used: the i-th regression coefficient is 0 if only if the dependent variable is conditional independent from the i-th regression variable given the remaining variables,(Lutkepohl, 2006). Therefore the following model is estimated:

$$\begin{bmatrix} r_{1,t} \\ \vdots \\ r_{m,t} \end{bmatrix} = \begin{bmatrix} \phi_{0,1} \\ \vdots \\ \phi_{0,1} \end{bmatrix} + \begin{bmatrix} \phi_{1,1} & \cdots & \phi_{1,m} \\ \vdots & \ddots & \vdots \\ \phi_{m,1} & \cdots & \phi_{m,m} \end{bmatrix} \cdot \begin{bmatrix} r_{1,t-1} \\ \vdots \\ r_{m,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ \vdots \\ a_{m,t} \end{bmatrix} \tag{1}$$

where $a_{j,t}$ are gaussian errors, $\phi_{i,j}$ are parameters. The elements $\phi_{i,j}$ are restricted to 0 whenever no connection was founded between pair of variables and the past with the chain graph approach. Briefly it can be written as

$$\mathbf{r}_t = \phi_0 + \Phi_1 \mathbf{r}_{t-1} + \mathbf{a}_t \tag{2}$$

where $\mathbf{r}_t = (r_{1,t}, ..., r_{m,t})'$, $\mathbf{a}_t = (a_{1,t}, ..., a_{m,t})'$ is a gaussian error, ϕ_0 is a m-dimensional vector, Φ_1 is a $m \times m$ matrix. This estimate has smaller asymptotic variance than the unrestricted estimator. This model is then used to find the return forecast at time t for the future $t+1$:

$$\hat{\mathbf{r}}_{t+1} = \hat{\phi}_0 + \hat{\Phi}_1 \mathbf{r}_t \tag{3}$$

As ARCH tests on the residuals indicate that data exhibit volatility cluster, that is, variance may be high for certain time periods and low for other periods, and variance jump are rare, for $\hat{a}_i$, a $MARCH(1)$ model is needed, see (Tsay, 2010):

$$\begin{bmatrix} a_{1,t}^2 \\ \vdots \\ a_{m,t}^2 \end{bmatrix} = \begin{bmatrix} \psi_{0,1} \\ \vdots \\ \psi_{0,1} \end{bmatrix} + \begin{bmatrix} \psi_{1,1} & \cdots & \psi_{1,m} \\ \vdots & \ddots & \vdots \\ \psi_{m,1} & \cdots & \psi_{m,m} \end{bmatrix} \cdot \begin{bmatrix} a_{1,t-1}^2 \\ \vdots \\ a_{m,t-1}^2 \end{bmatrix} + \begin{bmatrix} b_{1,t} \\ \vdots \\ b_{m,t} \end{bmatrix} \tag{4}$$

where $\psi_{i,j}$ are parameters, $b_{j,t}$ is consider an error such that $E(b_{i,t}) = 0$, $cov(b_{i,t}, b_{i,t-j}) = 0, j \geq 1$ (martingale difference series), but it is generally not an i.i.d sequence.
The elements $\psi_{i,j}$ are restricted to 0 whenever no connection was founded between pair of squared residuals and their past using a second chain graph approach as before. Even now it can be briefly written as

$$\mathbf{a}_t^2 = \psi_0 + \Psi_1 \mathbf{a}_{t-1}^2 + \mathbf{b}_t \tag{5}$$

where $\mathbf{a}_t^2 = (a_{1,t}^2, ..., a_{m,t}^2)'$, ψ_0 is a m-dimensional vector, Ψ_1 is a $m \times m$ matrix, $\mathbf{b}_t = (b_{1,t}, ..., b_{m,t})$ is consider an error.
This estimate has smaller asymptotic variance than the unrestricted estimator. As the squared residuals $a_{i,t}^2$ are unbiased estimate of the variance $\sigma_{i,t}^2$, $i = 1, ...m$, the preceding model is used to find the standard deviation forecasts at time t for the future $t+1$:

$$\hat{\mathbf{a}}_{t+1}^2 = \hat{\psi}_0 + \hat{\Psi}_1 \hat{\mathbf{a}}_t^2 \tag{6}$$

The estimates $\hat{\phi}_0, \hat{\Phi}_1$ of the $RVAR(1)$ model and $\hat{\psi}_0, \hat{\Psi}_1$ of the $RMARCH(1)$ model are obtained in two steps.
Firstly the EGLS (Estimated Generalized Least Squares) estimates $\hat{\phi}_0, \hat{\Phi}_1$ are found and the residuals of the $RVAR(1)$ model are calculated. Then the EGLS estimates $\hat{\psi}_0, \hat{\Psi}_1$ are found by considering the residuals calculated before.
This two steps estimation procedure is asymptotically equivalent to the procedure where all the parameters $\hat{\phi}_0, \hat{\Phi}_1, \hat{\psi}_0, \hat{\Psi}_1$ are estimated at once.
This is due to the fact that the asymptotic information matrix of the VAR and $MARCH$ is block diagonal so that the estimators of the former are asymptotically independent of the latter estimators, (Lutkepohl, 2006).
To restrict the VAR model (1), it is necessary to rewrite and re-parametrize it. It can be rewritten as follow:

$$\tilde{Y} = BZ + U \tag{7}$$

where $\tilde{Y} = (r_1, ..., r_T)$, $r_t = (r_{1,t}, ..., r_{m,t})'$, $Z = (Z_0, ..., Z_{T-1})$, $Z_t = (1, y_t)'$, $B = (\phi_0, \phi_1)$, $U = (u_1, ..., u_T)$, $u_t = (a_{1,t}, ..., a_{m,t})'$.
The linear constraint for ϕ_1 can be given in the form

$$\beta = vec(B) = R\gamma + \tilde{r} \tag{8}$$

for an opportune R fixed matrix of dimension $m(m+1) \times \tilde{M}$, with $\tilde{M}$ the number of restrictions, $\tilde{r} = 0$ a vector of dimension $m(m+1)$, γ a $\tilde{M}$ vector of unknown unrestricted parameters.
The *vec* operator transforms a matrix B into a vector β by stacking its columns.
Sometimes an equivalent way to represent constraint for β is $C\beta = c$ for opportune matrix C and vector c.
Through (8) it is possible to impose the constraints for β, by re parametrizing (7) as follow

$$\begin{aligned} \tilde{y} = vec(\tilde{Y}) &= (Z' \otimes I)vec(B) + vec(U) = \\ &= (Z' \otimes I)(R\gamma + \tilde{r}) + u = \\ &= (Z' \otimes I)R\gamma + u \end{aligned} \tag{9}$$

Thus it possible to minimize the generalized sum of squared errors $S(\gamma) = u(I \otimes \Sigma^{-1})u'$, obtaining a generalized estimator $\hat{\hat{\gamma}}$ and so $\hat{\hat{\beta}}$ which has smaller asymptotic variance than the unrestricted estimator $\hat{\beta}$.
The EGLS estimator $\hat{\hat{\gamma}}$ of γ is

$$\hat{\hat{\gamma}} = [R'(ZZ' \otimes \hat{\Sigma}^{-1})R]^{-1}R'(Z \otimes \hat{\Sigma}^{-1})z$$

where $\hat{\Sigma} = \hat{U}\hat{U}'/T$.
So the EGLS estimator $\hat{\hat{\beta}}$ of β is

$$\hat{\hat{\beta}} = R\hat{\hat{\gamma}} + \tilde{r}$$

Now the asymptotic covariance matrix of the unrestricted estimator $\hat{\beta}$ is $\Sigma_{\hat{\beta}} = (\Gamma^{-1} \otimes \Sigma)$ with $\Gamma = lim^p_{T \to +\infty} ZZ'/T$ whereas the asymptotic covariance matrix of the restricted estimator $\hat{\hat{\beta}}$ can be written as

$$\Sigma_{\hat{\hat{\beta}}} = \Sigma_{\hat{\beta}} - (\Gamma^{-1} \otimes \Sigma)C'[C(\Gamma^{-1} \otimes \Sigma)C']^{-1}C(\Gamma^{-1} \otimes \Sigma) \tag{10}$$

that is a positive semi-definite matrix is subtracted from $\Sigma_{\hat{\beta}}$ to obtain the asymptotic covariance matrix of the restricted estimator. Thus the latter can be smaller or equal to the former. Imposing restrictions is an advantageous in terms of asymptotic efficiency.

To restrict the MARCH model (4), the same procedure as above is applied to the squared residuals a_j^2 instead of the returns r_j.
To verify the null hypothesis of no ARCH effects in $\mathbf{a}_t$, the equivalent formulation

of the null hypothesis of no serial correlation in $\mathbf{v}_t = vech(\mathbf{a}_t\mathbf{a}_t')$ is used, where *vech* is the operator that transforms a matrix square A into a vector a by stacking its elements on and below the main diagonal.
So the multivariate Ljung-Box test is applied to $\mathbf{v}_t$:

$$H_0 : \rho_1 = ... = \rho_h = \mathbf{0}$$

$$H_1 : \exists j : \rho_j \neq \mathbf{0}$$

$$Q_m(h) = T^2 \sum_{j=1}^{h} \frac{1}{T-j} tr(\hat{\Gamma}_j' \hat{\Gamma}_0^{-1} \hat{\Gamma}_j \hat{\Gamma}_0^{-1}) \tag{11}$$

where $tr(A)$ is the trace of A, that is the sum of its diagonal elements, $\hat{\Gamma}_j$ is estimate of the cross-covariance matrix $\Gamma_j = E[(\mathbf{v}_t - \mu_v)(\mathbf{v}_{t-j} - \mu_v)']$ of $\mathbf{v}_t$

$$\hat{\Gamma}_j = \frac{1}{T} \sum_{t=j+1}^{T} (\hat{\mathbf{v}}_t - \bar{\mathbf{v}})(\hat{\mathbf{v}}_{t-j} - \bar{\mathbf{v}})'$$

whereas ρ_j is the lag-j cross-correlation matrix of $\mathbf{v}_t$ defined as

$$\rho_j = D^{-1} \Gamma_j D^{-1}$$

with $D = diag(\sqrt{\Gamma_0(1,1)}, ..., \sqrt{\Gamma_0(m,m)})$. The asymptotic distribution $Q_m(h)$ is $\chi^2(m^2(h-1))$ so that is possible to calculate the theoretical quantile of the statistic. Moreover the inverse of the roots of the reverse characteristic polynomial, that is the eigenvalues λ_Φ of Φ_1 andλ_Ψ of Ψ_1, are found in order to ascertain the stability of the $(R)VAR$ and $(R)MARCH$ models:

$$det(I_m - \Phi_1 \frac{1}{\lambda_\Phi}) = 0$$

$$det(I_m - \Psi_1 \frac{1}{\lambda_\Psi}) = 0 \tag{12}$$

Whenever $|\frac{1}{\lambda_\Phi}| > 1$ or $|\lambda_\Phi| < 1$ the $(R)VAR$ is stable and whenever $|\frac{1}{\lambda_\Psi}| > 1$ or $|\lambda_\Psi| < 1$ the $(R)MARCH$ is stable.
An important need for a precise estimation of the variance-covariance matrix Σ_a of $\mathbf{a}_t$ is that the MSE (Mean Squared Error) $\Sigma_r(h)$ of the forecast $\hat{r}_{t+h}$ at the time t for the future time $t+h$ is a function of Σ_a:

$$\Sigma_r(h) = \sum_{j=0}^{h} \Phi_1^j \Sigma_a (\Phi_1^j)'$$

Thus the confidence intervals for r_{t+h} will be affect by ,at least, its diagonal estimation $\hat{\sigma}_{k,t}(h) = (\hat{\Sigma}_r(h))_{k,k}$:

$$[\hat{r}_{k,t+h} - z_{\alpha/2} \hat{\sigma}_{k,t}(h), \hat{r}_{k,t+h} + z_{\alpha/2} \hat{\sigma}_{k,t}(h)]_{1-\alpha}$$

The *RMARCH* model takes care of the variance-covariance matrix estimation $\hat{\Sigma}_a$ in a time dependent way, more efficiently than its full version *MARCH*, taking into consideration the ARCH effects which are typical of the financial dataset case. However it may be that the shrinking of the parameters could worsen the ARCH effects modelling. Thus it is checked by matching the Ljung-Box test results in both cases. The model equation (5) generalizes many univariate volatility models to the multivariate case, being one of the multivariate GARCH models, (Tsay, 2010). In this case, in order to keep the number of volatility equations low, as in (Bollerslev, 1990) it is considered the special case in which the correlation coefficient is time-invariant.

4 Results

The time series considered regard the Blue Chip (that is the best capitalized) Bank Stock returns weekly figures from 3 January 2000 to 9 March 2015 taken by Yahoo finance.
Specifically those stocks are the 5 bank stock as follows:

UNICREDIT, BANCAPOPOLAREDIMILANO, CREDITOEMILIANO, INTESASANPAOLO, MEDIOBANCA.

In **fig. 1** the time series are depicted. It suggests that some periods have a greater volatility (high peaks) than others (low peaks). So the MARCH approach seems necessary to tackle the time-varying volatility.

Firstly a chain graph is built for the returns $\mathbf{r}_t = (r_{1,t}, ..., r_{m,t})'$, by minimizing the *BIC* criteria, in two steps.
The Schwarz's Bayesian criterion $BIC = -2logLikelihood + log(T)p$, where p is the number of parameters. This criteria gives a bonus for a parsimonious description of the data with respect to other criteria such the AIC or FPE.
The first step for a chain graph to be used in the *RVAR* model, is to build a graph for the past values of the returns $\mathbf{r}_{t-1} = (r_{1,t-1}, ..., r_{m,t-1})'$, which play the role of the explanatory block.
The **fig. 2** represents thus the explanatory variables minimum *BIC* graph.

The second step is to build a minimum *BIC* graph for the past to present values of the returns $\mathbf{r}_t = (r_{1,t}, ..., r_{m,t})'$ which play the role of the response variables , having a saturated model (that is all the edges) for the explanatory variables $\mathbf{r}_{t-1} = (r_{1,t-1}, ..., r_{m,t-1})'$.
The **fig. 3** represents thus the explanatory-response variables minimum *BIC* graph.

Finally, the chain graph for the past to present values of the returns $\mathbf{r}_t = (r_{1,t}, ..., r_{m,t})'$ with the explanatory variables $\mathbf{r}_{t-1} = (r_{1,t-1}, ..., r_{m,t-1})'$ is built by

combining the first preceding graphs with the second preceding graph without its saturated part.
The **fig. 4** represents thus the chain graph for the returns $\mathbf{r}_t$.

The directed edges in figure 3 are used to constraint the $VAR(1)$ model for the returns $\mathbf{r}_t$.

In the same way as above, the **fig. 5,6,7** represent respectively the explanatory variables minimum *BIC* graph, explanatory-response variables minimum *BIC* graph and the chain graph for the $\mathbf{a}_t^2$
.

Please note that for the figures lag1 is attached at the end of the variables' names.

The parameters estimates, standard errors, t values, p-values for the full $VAR(1)$ model, **tab.1**, and restricted $RVAR(1)$ model, **tab.2**, are represented equation by equation :

Names	Estimates	Standard Errors	t values	p-values
UNICREDIT				
UNICREDIT.l1	-0.229	0.055	-4.165	0
BANCAPOPOLAREDIMILANO.l1	-0.035	0.059	-0.594	0.553
CREDITOEMILIANO.l1	0.298	0.072	4.146	0
INTESASANPAOLO.l1	-0.155	0.079	-1.963	0.05
MEDIOBANCA.l1	0.156	0.084	1.853	0.064
BANCAPOPOLAREDIMILANO				
UNICREDIT.l1	-0.031	0.048	-0.642	0.521
BANCAPOPOLAREDIMILANO.l1	-0.121	0.052	-2.347	0.019
CREDITOEMILIANO.l1	0.235	0.063	3.757	0
INTESASANPAOLO.l1	-0.097	0.069	-1.41	0.159
MEDIOBANCA.l1	0.123	0.073	1.681	0.093
CREDITOEMILIANO				
UNICREDIT.l1	-0.022	0.041	-0.542	0.588
BANCAPOPOLAREDIMILANO.l1	-0.064	0.044	-1.452	0.147
CREDITOEMILIANO.l1	0.003	0.054	0.057	0.954
INTESASANPAOLO.l1	-0.086	0.059	-1.454	0.147
MEDIOBANCA.l1	0.187	0.063	2.972	0.003
INTESASANPAOLO				
UNICREDIT.l1	-0.048	0.045	-1.061	0.289
BANCAPOPOLAREDIMILANO.l1	-0.029	0.048	-0.609	0.543
CREDITOEMILIANO.l1	0.094	0.059	1.6	0.11
INTESASANPAOLO.l1	-0.192	0.064	-2.992	0.003
MEDIOBANCA.l1	0.161	0.069	2.346	0.019
MEDIOBANCA				
UNICREDIT.l1	-0.072	0.037	-1.962	0.05
BANCAPOPOLAREDIMILANO.l1	-0.011	0.039	-0.286	0.775
CREDITOEMILIANO.l1	0.129	0.048	2.702	0.007
INTESASANPAOLO.l1	-0.089	0.053	-1.688	0.092
MEDIOBANCA.l1	0.073	0.056	1.309	0.191

Tab. 1: Estimates, standard errors, t values, p-values for $VAR(1)$

Names	Estimates	Standard Errors	t values	p-values
UNICREDIT				
UNICREDIT.l1	-0.17	0.044	-3.856	0
BANCAPOPOLAREDIMILANO.l1	0.055	0.051	1.071	0.285
BANCAPOPOLAREDIMILANO				
BANCAPOPOLAREDIMILANO.l1	-0.025	0.037	-0.667	0.505
CREDITOEMILIANO				
CREDITOEMILIANO.l1	-0.025	0.037	-0.657	0.512
INTESASANPAOLO				
INTESASANPAOLO.l1	-0.107	0.037	-2.886	0.004
MEDIOBANCA				
UNICREDIT.l1	-0.078	0.032	-2.432	0.015
MEDIOBANCA.l1	0.082	0.049	1.675	0.094

Tab. 2: Estimates, standard errors, t values, p-values for $RVAR(1)$

From the preceding tables 1,2 it can be seen that the chain model are not equivalent to the t test shrinking usual approach in that it can find different significant relationships.

The same type of tables can be represented for the $MARCH(1)$ and $RMARCH(1)$ models.

Thus it can be seen from the figures the parameters saving: $p_{var} = 25$ parameters are needed with respect to a full $VAR(1)$, whereas $p_{rvar} = 7$ parameters are needed for a $RVAR(1)$ model. The parameters saving $s_r = (p_{var} - p_{rvar})/p_{var} = 72\%$.
Also $p_{march} = 25$ parameters are needed with respect to a full $MARCH(1)$, whereas $p_{rmarch} = 6$ parameters are needed for a $MARCH(1)$ model. The parameters saving $s_{a^2} = (p_{march} - p_{rmarch})/p_{march} = 76\%$.

All the models are stable. Indeed the following **tab. 3** contains the eigenvalues $\lambda_\Phi, \lambda_\Psi$ and they result less than 1 in module (see equations (12) above):

model	$\lambda_\Phi(1)$	$\lambda_\Phi(2)$	$\lambda_\Phi(3)$	$\lambda_\Phi(4)$	$\lambda_\Phi(5)$
VAR	0.23	0.12	0.11	0.11	0.03
$RVAR$	0.17	0.11	0.08	0.02	0.02
model	$\lambda_\Psi(1)$	$\lambda_\Psi(2)$	$\lambda_\Psi(3)$	$\lambda_\Psi(4)$	$\lambda_\Psi(5)$
$ARCH$	0.51	0.16	0.16	0.08	0.08
$RARCH$	0.39	0.35	0.35	0.30	0.26

Tab. 3: Eigenvalues $\lambda_\Phi, \lambda_\Psi$ for $VAR(1)$,$RVAR(1)$,$MARCH(1)$,$RMARCH(1)$

The AIC for the full $VAR(1)$ is -12356.26 where for the $RVAR(1)$ is -12332.62. So there is a slightly improvement by using the unrestricted model, even if they are very much near one another.
The AIC for the full $MARCH(1)$ is -25205.81 where for the Restricted March is

-25038.35. So there is a slightly worsening by using the restricted model, even if they are very much near one another.

It can be seen that the *VAR* back forecasts are near the *RVAR* back forecasts and with some exceptions near to the actual values, see **fig. 8** for the UNICREDIT return case. The back forecasts, intended as in-sample fitted values, are the forecasts obtained by the models with the parameters estimated using the entire dataset. Then for each time index t, $t = 1, ..., T$ the back-forecast $\hat{\mathbf{r}}_t$ is obtained using the actual values until time $t-1$.

As far as the standard deviation is concerned, the actual values is not observable, (Tsay, 2010), so, a rolling mean with a windows of size 69 is considered as if it were the actual values, (Tsay, 2010), instead of other possible proxy for it. The **fig. 9** depicts the UNICREDIT standard deviation forecast case.

The predictive ability of the models is seen in terms of mean squared error, [4], where MSE^{back} uses the back forecasts and and MSE^{out} uses the out-of-sample forecasts which are obtained with a reduced dataset of $h = 20$ observations

$$MSE_\delta = \frac{\sum_{i\in\{1,...,m\}} \sum_{t\in\{1,...,T\}} (\hat{\delta}_{i,t} - \delta_{i,t})^2}{m \cdot T} \tag{13}$$

As the focus is on weekly data, the investor has in mind a short horizon . So 20 weeks (5 months) of out-of- sample data are chosen.
It results $MSE^{back}_{full,r} = 0.0033$, $MSE^{out}_{full,r} = 0.0082$, $MSE^{back}_{rest,r} = 0.0033$, $MSE^{out}_{rest,r} = 0.0081$ with $\delta_{i,t} = r_{i,t}$ in the equation (13). So there is an equivalence of the restricted and the restricted model.
To statistically prove this MSE_r equivalence a Giacomini-White for testing the null hypothesis $H_0^{(1)}$ equal predictive ability test is used, (Giacomini and White, 2006). The GW p-value is 0.2785 , $\rightarrow H_0^{(1)}$ is not rejected.

The equation (13) with $\delta_{i,t} = \sigma_{i,t}$ gives $MSE^{back}_{full,\sigma} = 0.001$, $MSE^{out}_{full,\sigma} = 0.0082$, $MSE^{back}_{rest,\sigma} = 0.001$, $MSE^{out}_{rest,\sigma} = 0.0082$. The rolling σ_t are considered as the actual values. So there is an equivalence of the restricted and the restricted model.
To statistically prove this MSE_σ equivalence a Giacomini-White for testing the null hypothesis $H_0^{(1)}$ equal predictive ability test is used, (Giacomini and White, 2006). The GW p-value is 0.07944 , $\rightarrow H_0^{(1)}$ is not rejected.

The multivariate Ljung-Box test (11) is applied to the standardized residuals $\hat{\mathbf{a}}_t/\hat{\sigma}_t$ to check for adequacy of the *VAR* and the *RVAR* with respect to the ARCH effect up to lag 1 gives the following statistics:
for the *VAR* ,$Q_m(h) = Q_5(1) = 2.9430$, p-value of 1 , $\rightarrow H_0^{(1)}$ is not rejected
for the *RVAR* ,$Q_m(h) = Q_5(1) = 59.7890$, p-value of 0.00011 , $\rightarrow H_0^{(1)}$ is rejected .
Thus the *VAR* model seems to have a better behaviour with respect to lag correlation

effects in this case.
However if monthly returns are considered instead of weekly returns, the models have the same behaviour and no correlation effects in the standardize residuals:
for the *VAR* ,$Q_m(h) = Q_5(1) = 3.0473$, p-value of 1 , $\rightarrow H_0^{(1)}$ is not rejected
for the *RVAR* ,$Q_m(h) = Q_5(1) = 33.0783$, p-value of 0.1291 , $\rightarrow H_0^{(1)}$ is not rejected.
The multivariate Ljung-Box test (11) is applied to the standardized residuals $\hat{\mathbf{a}}_t^2/\hat{\sigma}_t^2$ to check for adequacy of the *MARCH* and the *RMARCH* with respect to the ARCH effect up to lag 1 gives the following statistics:
for the *VAR* ,$Q_m(h) = Q_5(1) = 161.9274$, p-value of 0 , $\rightarrow H_0^{(1)}$ is rejected
for the *RVAR* ,$Q_m(h) = Q_5(1) = 185.6905$, p-value of 0 , $\rightarrow H_0^{(1)}$ is rejected .
Thus the models seem to have the same behaviour with respect to the arch effects in this case.
However if monthly returns are considered instead of weekly returns, the models have the same behaviour and no arch effects in the standardized residuals:
for the *VAR* ,$Q_m(h) = Q_5(1) = 24.0866$, p-value of 0.514 , $\rightarrow H_0^{(1)}$ is not rejected
for the *RVAR* ,$Q_m(h) = Q_5(1) = 32.8796$, p-value of 0.1341 , $\rightarrow H_0^{(1)}$ is not rejected.
Thus the Chain Graph-VAR-MARCH model has the same behavior of the existing VAR-MARCH model in terms of ARCH effects (with less parameters) and solve the problem with monthly data.

5 Concluding remarks

The great reduction in parameters seems not to affect the predictive power of the *VAR* and *MARCH* models neither the goodness of fit as the $MSEs_\delta$ are similar and *AICs* are similar too.
The Giacomini-White test confirms the equivalence of the two models in terms of forecasts. To the other hand the great shrinking of the parameters' number implies that the so restricted models are more efficient so that they are to be preferred.
Even if conditional normal distributions are chosen for a_t, their unconditional distributions result non normal in both cases thanks to the *MARCH*. Moreover, the models seem to have the same behaviour with respect to ARCH effects in this case.
So it seems that for large problem and forecast use the restricted model has to be chosen to improve the efficiency of the estimation.
Lastly the high frequency of the weekly data can require a larger lag order to describe the the process, even if the unrestricted and the restricted approaches have the same adaptation. However, the use of lower frequency data such as monthly data is sufficient to get the two approaches appropriate to describe the process.

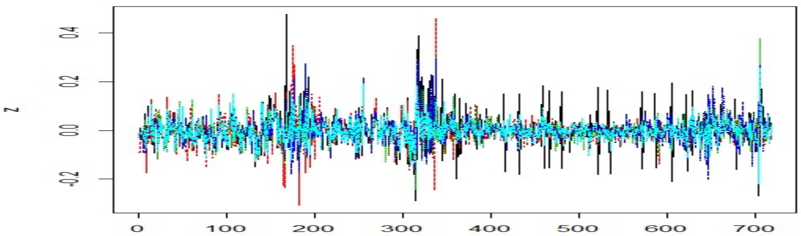

Fig. 1 Weekly 5 bank stock Time Series

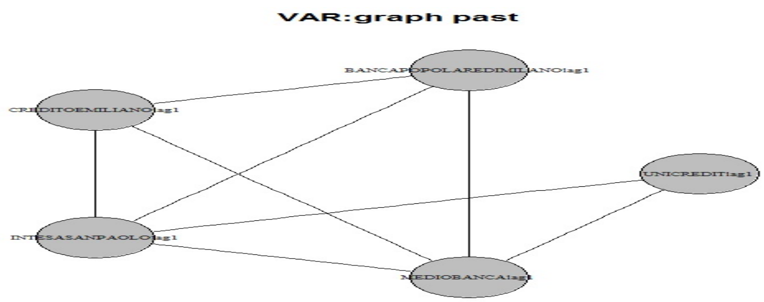

Fig. 2 Minimum *BIC* graph for the explanatory variables, $RVAR(1)$

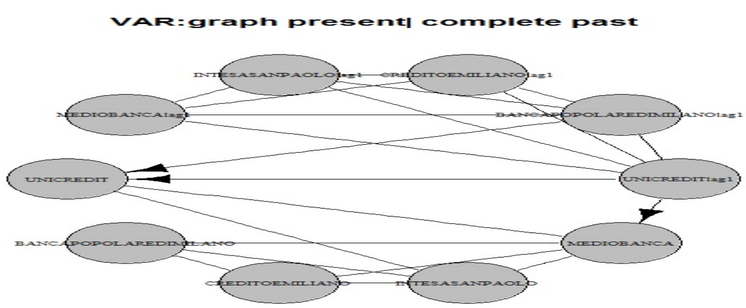

Fig. 3 Minimum *BIC* graph for the explanatory-response variables, $RVAR(1)$

Fig. 4 Chain Graph for the returns $\mathbf{r}_t$, $RVAR(1)$

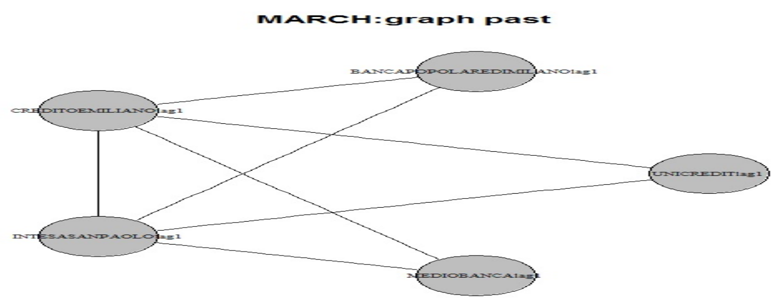

Fig. 5 Minimum BIC graph for the explanatory variables, $RMARCH(1)$

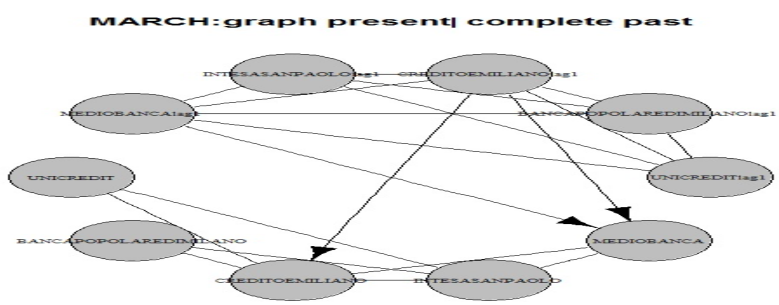

Fig. 6 Minimum BIC graph for the explanatory-response variables, $RMARCH(1)$

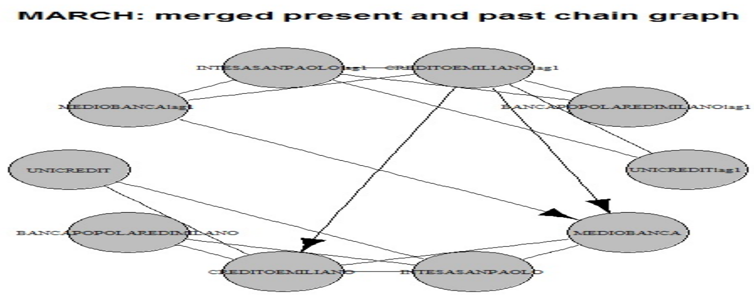

Fig. 7 Chain Graph for the returns $\mathbf{a}_t^2$, $RMARCH(1)$

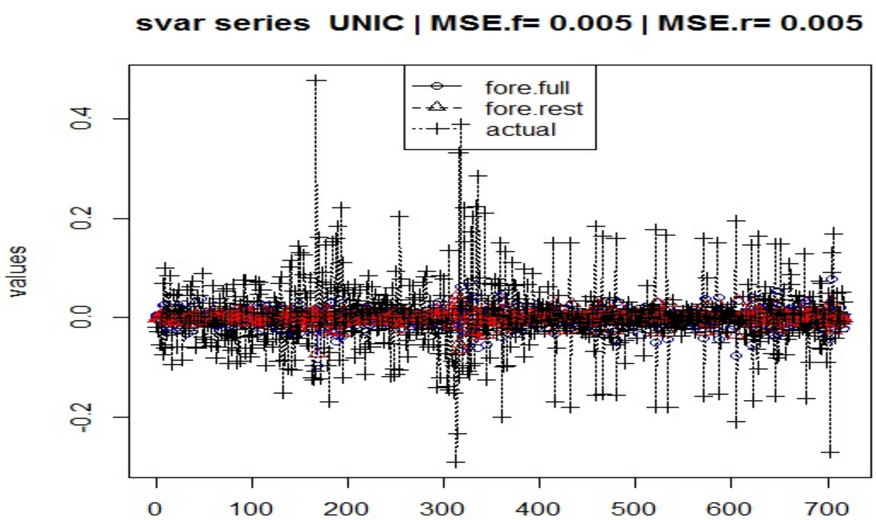

Fig. 8 Returns one period head back-forecasts of $\mathbf{r}_t$ with VAR and $CG-RVAR$, UNICREDIT

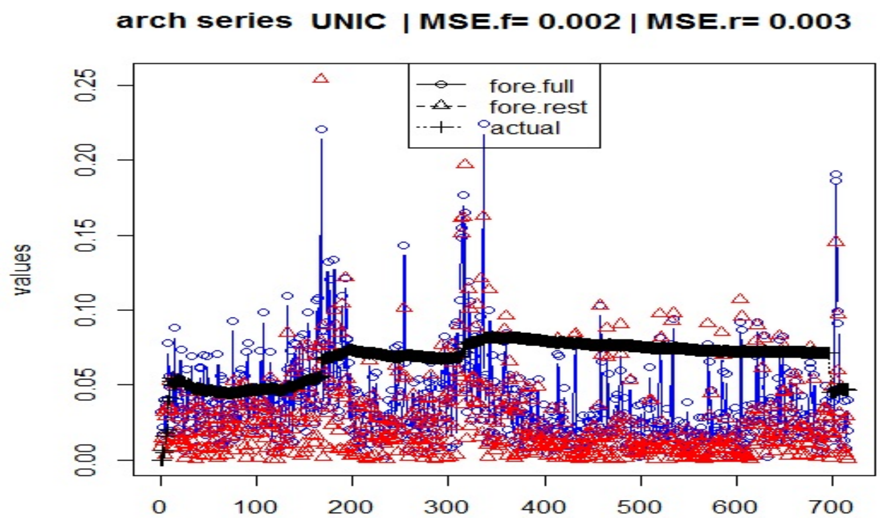

Fig. 9 Standard Deviations one period head back-forecasts of σ_t with $MARCH$ and $CG-RMARCH$, UNICREDIT

References

1. Awokuse T.,Bessler D., Vector Autoregressions, policy analysis, and directed acyclic graphs: an application to U.S. economy, Journal of Applied Economics, vol.6, n.1, 1-24, (2003).
2. Babula R.,Bessler D., Reeder J., Somwaru A., Modeling U.S. soy-based markets with directed acyclic graphs and time series econometrics: evaluating the U.S. market impatcs of high soy meal prices, WP, Office of Industry, U.S. International Trade Commission, (2004).
3. Bollerslev T., Modeling the coherence in short-term nominal exchange rates: A multivariate, generalized ARCH approach. Review of Economics and Statistics, vol.72, n.3, 498-505, (1990).
4. Bollerslev T., Engle R.F., Nelson D.B., ARCH Model, Handbook of Econometrics, Elsevier Science, Amsterdam, vol.4, 2959-3038, (1994).
5. Campbell, J.Y., Chan Y.L., Viceira L.M., A Multivariate Model of Strategic Asset Allocation, Journal of Financial Economics, vol.67, n.1, 41-80, (2003).
6. Carvalho C., West M., Dynamic Matrix-Variate Graphical Models, Bayesian Analysis, vol.2. n.1, 69-98, (2007).
7. Ding Z., Engle R., Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing, NYU Working Paper No. S-DRP-01-07, (2001).
8. Eichler, M., Granger-causality graphs for multivariate time series. Technical report, University of Heidelberg, Germany, (2000).
9. Fama E. , The Behaviour of Stock-Market Prices, Journal of business, vol.38, n.1, 34-105, (1965).
10. Hojsgaard S., Edwards D.,Lauritzen S., Graphical Models with R, Springer, (2012).
11. Gao W., Tian Z., Latent ancestral graph of structure vector autoregressive models,Journal of Systems Engineering and Electronics, vol.21, n.2, 233-238, (2010).
12. Giacomini R., White H., Test of Conditional Predictive Ability, Econometrica, vol.74, n.6, (2006).
13. Lutkepohl H., New Introduction to Multiple Time Series Analysis, Springer, (2006).
14. Moneta A., Graphical causal models and VARs: an empirical assessment of the real business cycles hypothesis, Empirical Economics, vol.35, n.2, 275-300, (2008).
15. Oxley L., Reale M., Wilson T., Finding directed acyclic graphs for vector autoregressions, Symposium, (2004).
16. Oxley L., Reale M., Wilson T., Graphical models for structural VARMA representations, 18th World IMACS MODSIM Congress, Cairns, Australia, (2009).
17. Pierini A., Undirected Gaussian Graphical Models for VAR parameters reduction: EU Indexes case, Sco13 Conference Proceedings, Politecnico di Milano, (2013).
18. Reale M., Wilson T., Identification of vector AR models with recursive structural errors using conditional independence graphs, Statistical Methods and Applications, vol.10, n.1, 49-55, (2001).
19. Songsiri J., Topology Selection in Graphical Models of Autoregressive Processes,Journal of Machine Learning Research, vol.11, 2671-2705,(2010).
20. Tsay R., Analysis of Financial Time Series, Wiley and Sons, (2010).
21. Wang H., Carvalho C., Simulation of hyper-inverse Wishart distributions for non-decomposable graphs,Electronic Journal of Statistics, vol.4, 1470-1475, (2010).
22. Whittaker J., Graphical Models in Applied Multivariate Statistics, Wiley and Sons, (1990).