Clustering Heteroskedastic Time Series By Model-Based Procedures

Classificazione di serie temporali eteroschedastiche tramite procedure model-based

Edoardo Otranto
Dipartimento di Economia, Impresa e Regolamentazione, Università di Sassari
e-mail: eotranto@uniss.it

Keywords: agglomerative algorithm, AR metrics, cluster analysis, GARCH models, Wald test

1. Introduction

The interest toward the classification of time series has recently received a lot of contributions (see Piccolo, 2007, for a review). Most of these studies are devoted to capture the structure of the mean of the process hypothesized as generator of the data, whereas little attention has been devoted to the variance. When dealing with heteroskedastic time series, in which the (conditional) variance follows a stochastic process (typically a GARCH process), the comparison of the dynamics of the variances is fundamental. This is particularly important dealing with financial time series, when the investor has a very large opportunity set (hundreds of stocks) and he would like to have groups of series with similar characteristics (similar unconditional variance, similar dynamics, etc.). Moreover, the volatility of a return is generally considered as a proxy of the risk of the same return; in other words the classification of returns of several assets is equivalent to classify the assets in clusters with similar risk.

In this paper we propose a clustering procedure based on simple statistical tools. In particular, we consider the squared disturbances of the returns of a financial time series as a measure of the volatility of the series. Then, we use the GARCH representation of the conditional variance to derive the model underlying the squared disturbances. We classify the series with similar unconditional volatility, similar time-varying volatility and/or similar volatility structure, using classical Wald statistics.

2. The clustering algorithm

Let us consider a time series \( y_t \); we suppose that it is the sum of a constant term and a heteroskedastic disturbance \( \varepsilon_t \), with mean zero; its conditional variance \( h_t \) follows a GARCH(p,q) process:

\[
    h_t = \gamma + \alpha_1 \varepsilon_{t-1}^2 + \alpha_p \varepsilon_{t-p}^2 + \beta_1 h_{t-1} + \ldots + \beta_q h_{t-q},
\]

with \( \gamma > 0 \), \( 0 \leq \alpha_i < 1 \), \( 0 \leq \beta_j < 1 \) \( (i = 1, \ldots, p; j = 1, \ldots, q) \), \( \left( \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j \right) < 1. \)

\((\star)\) Work supported by Italian MIUR under Grant 2006137221_001
As well known, we can represent the time series $\varepsilon_t^2$ by an ARMA($p^*, q$) model, with $p^* = \max(p, q)$. From simple algebra, we can obtain the AR($\infty$) representation:

$$\varepsilon_t^2 = \frac{\gamma}{1 - \sum_{j=1}^{q} \beta_j} + \sum_{k=1}^{\infty} \pi_k \varepsilon_{t-k}^2 + (\varepsilon_t^2 - h_t).$$

We classify the heteroskedastic time series using the unconditional volatility of (2) and the time-varying volatility, expressed as (this expression is derived by the AR distance of Piccolo, 1990, and the GARCH distance of Otranto, 2004):

$$\left[ \sum_{k=1}^{\infty} \pi_{rk}^2 \right]^{1/2}.$$

We define three levels of clustering. The steps of the clustering algorithm can be synthesized in the following way (the equalities are verified by Wald tests):

1. First Level (*clusters with equal unconditional volatility*):
   (a) order the series in terms of increasing unconditional volatility and choose the series with minimum unconditional volatility as the initial benchmark; (b) insert in the same cluster the series with unconditional volatility not significantly different from the benchmark volatility; the series with minimum unconditional volatility which does not enter in this cluster is considered the benchmark for the successive cluster; (c) go on until no series remain;

2. Second Level (*clusters with equal unconditional volatility and equal time-varying volatility*): (a) in each first level cluster, order the series in terms of increasing time-varying volatility; the series with minimum time-varying volatility of each cluster is the benchmark of the cluster; (b) form sub-clusters with series having equal time-varying volatility with respect to the benchmark; the series with minimum time-varying volatility which does not enter in a sub-cluster is considered the benchmark for the successive sub-cluster. (c) go on until no series remain;

3. Third Level (*clusters with equal volatility structure*) (a) in each second level sub-cluster verify the equality of the GARCH processes (by the Wald test) for each pair of series; (b) if all the p-values are less than the nominal size of the test, the series considered have different volatility structures; otherwise, select the pair of series with maximum p-value and repeat the test of equal volatility structure adding a series to the pair selected; (c) select the group of series with maximum p-value and go on until the hypothesis of equal volatility structure is rejected; (d) repeat steps (b) and (c) for the remaining series; (e) go on until no series remain.

**References**

