

Robust methods for the identification of ARCH components

Metodi robusti per l'identificazione di componenti ARCH

Luigi Grossi

Dipartimento di Economia
Università di Parma
luigi.grossi@unipr.it

Fabrizio Laurini

Dipartimento di Scienze Statistiche
Università di Padova
flaurini@stat.unipd.it

Riassunto: Nel presente lavoro si propone una versione robusta del test LM per la individuazione di componenti ARCH. Il contributo principale è l'introduzione della *forward search* nella modellistica ARCH. Lo scopo è di monitorare l'influenza di ogni osservazione sul test LM, attraverso l'utilizzo di procedure robuste. Attraverso l'analisi di dati simulati e di serie finanziarie, si può notare che il numero di componenti ARCH individuate dalle procedure tradizionali dipende largamente da un sottoinsieme ristretto di osservazioni influenti. La presenza di *outliers* non intacca il p -value del test; la stima dei coefficienti e le t -statistiche dipendono principalmente da poche osservazioni estreme. Tutto ciò, può portare a considerare un errato numero di componenti ARCH.

Keywords: Lagrange Multiplier (LM) test, Robustness, Masking, Forward Search.

1. Introduction

Volatility clustering and the presence of several extreme observations are relevant characteristics of financial returns. These features have led to the introduction of the ARCH model family (Engle, 1982). Before fitting an ARCH model, it makes sense to test for the presence of ARCH components in the series. Engle developed a Lagrange Multiplier (LM) test based on a OLS regression of the squared residuals of a model for the conditional mean. We propose a method to improve the robustness of the LM test using the forward search for OLS regression (Atkinson and Riani, 2000). The main purpose of the work is to show how influential observations may produce misleading conclusions on the order of the ARCH model to fit to the data. In the next section the LM test will be introduced and the forward search is extended in the ARCH context. Finally, the robust analysis of LM test is applied to simulated and financial data.

2. A robust LM test for ARCH through the forward search

Let y_t be a time series of returns with $t = 1, \dots, T$. For a polynomial in the backward operator L , $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ having all roots outside the unit circle and for parameters $\alpha_i \geq 0$ ($i = 0, \dots, q$), we define the AR(p)-ARCH(q) model as follows

$$\begin{aligned}\phi(L)(y_t - \mu) &= \varepsilon_t \\ \varepsilon_t &= h_t z_t, \quad z_t \stackrel{i.i.d.}{\sim} N(0, 1)\end{aligned}\tag{1}$$

$$h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2.\tag{2}$$

For stationarity $\sum_{i=1}^q \alpha_i \leq 1$. The test proposed by Engle for the presence of ARCH components is based on the following consideration: the conditional variance h_t in the ARCH(q) model in (2) is constant if the parameters corresponding to the lagged squared shocks ε_{t-i}^2 , $i = 1, \dots, q$ are equal to zero. Therefore, the null hypothesis of conditional homoscedasticity can be formulated as $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_q = 0$. Given the following regression of the squared residuals of model (1) under the null hypothesis on a constant and q of its lags

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \dots + \alpha_q \hat{\varepsilon}_{t-q}^2 + v_t, \quad (3)$$

the LM test for ARCH components can be computed as $\lambda = TR^2$, where R^2 is the multiple correlation coefficient of model (3). Under the null hypothesis, the LM test statistic has an asymptotic $\chi^2(q)$ distribution. The diagnostic test after fitting model (1) can lead to consider several ARCH components; presumably too many.

Engle's LM test is affected by the presence of outliers, additive outliers in particular, as shown by Franses and Van Dijk (2000). They propose a backward method to identify influential observations. In general backward methods are affected by masking, so we suggest to use a forward search method. The forward search approach is based on a robust and efficient estimator and has three steps: the first concerns the choice of an initial subset free from outliers (a Clean Data Set, *CDS*), the second refers to the way in which we progress in the forward search and the third relates to the monitoring of the statistics during the search's progress. The initial subset is defined using the least median of squared (LMS) regression (Rousseeuw and Leroy, 1987). The forward search test statistic for ARCH components λ_{FS} is defined as a sequence of test statistics derived from least squares estimators in each step of the forward search, that is

$$\lambda_{FS} = (\lambda_{(q+1)}, \dots, \lambda_{(T)}) \quad (4)$$

where $\lambda_{(m)}$ is the test statistic computed on m observations. To move from step m to $m+1$ we put all observations in a non decreasing order with respect to the corresponding OLS residuals. The procedure is repeated until $m = T$, that is until all observations are included in the *CDS*. In the first steps, forward search avoids the inclusion of outliers and immediately highlights units far away from the bulk of data. During the forward search we graphically monitor $\lambda_{(m)}$, coefficients $\hat{\alpha}_{i(m)}$ and t -statistics of coefficients. Sharp changes of the curves point out the introduction of influential observations. Monitoring t -statistics of each component provides more information than in the standard LM test. The robustness of this method derives from a progressive inclusion of units in a subset which, in the first steps, is outliers-free.

3. Simulated data

We test the robustness of the forward method using simulated data. In particular we simulated 1,000 observations from an AR(1)–ARCH(1) process with the following set of parameters

$$y_t = 0.5y_{t-1} + z_t \sqrt{0.1 + 0.8y_{t-1}^2}, \quad z_t \sim N(0, 1).$$

In addition we put 10 additive outliers (AO) at random positions. Franses and Van Dijk suggest to add a noise magnitude of 7 (in absolute value).

Step Pos.	971 861	...	988 355	989 972	990 210	991 745	...	999 356	1000 706
$\hat{\alpha}_0$	0.259		0.368	0.340	0.325	0.310		0.730	0.795
t	(11.81)		(8.26)	(5.92)	(4.23)	(3.29)		(3.96)	(4.11)
$\hat{\alpha}_1$	0.081		0.081	0.085	0.084	0.082		0.101	0.100
t	(21.55)		(11.14)	(9.07)	(6.66)	(5.30)		(3.36)	(3.16)
$\hat{\alpha}_3$	0.007		0.007	0.008	0.007	0.090		0.084	0.082
t	(1.76)		(0.98)	(0.77)	(0.56)	(5.84)		(2.77)	(2.59)
$\hat{\alpha}_{10}$	0.001		0.002	0.059	0.135	0.134		0.132	0.132
t	(0.23)		(0.31)	(6.08)	(10.75)	(8.76)		(4.38)	(4.15)

Table 1: Simulated data: last steps of the forward search applied to the squares of the residuals after the fit of an AR(1) to the simulated data. Coefficients, t —statistics (round brackets) and influential units are monitored.

Fitting an AR(1) model to the data and applying the LM test, you could not refuse the null of having 10 ARCH components, where $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_3$ and $\hat{\alpha}_{10}$ are 1% significantly; the p —value of the test is 0.000171.

The results achieved, applying the forward search to the square of the residuals after the AR(1) fitting, show that the test is strongly influenced by the presence of big values (as expected). In Table 1 we emphasize the most relevant steps out of the last 30, for all the ARCH coefficients involved in the traditional approach. It is possible to see that the t —statistic of $\hat{\alpha}_{10}$ becomes statistically relevant at the step 989, when we observe the value 6.54 in the original (simulated) series. The presence of the 10 additive outliers introduced in the simulated series keeps the t —statistics significant until the last step.

The same consideration applies for $\hat{\alpha}_3$, starting from the step 991 where a contaminated value is observed. What we argue is that any traditional approach would fail in detecting the sensitivity of the test to a large number of big values.

4. Application to financial data

The data we use to illustrate the suggested robust test is the series of FTSE100, the stock index of the 100 English blue chips, sampled at daily frequency. The sample period runs from 6 January 1995 until 12 December 2001 for a total number of 1756 observations. The data have been obtained from Datastream. For the application of the test we have computed the logarithmic returns, that is $y_t = \ln(P_t/P_{t-1})$, where P_t is the stock index at time t . The order of the autoregressive model (1) selected by the AIC is $p = 2$. We have monitored the LM test λ for $q = 1, 5, 10$ starting from residuals of an AR(2) model. Major changes during the last steps of the forward search have been observed for the coefficients of model (3) and for the corresponding t —statistics, while p —values of the LM test remain almost stable at the end of the procedure and lead to the rejection of the null hypothesis of homoscedasticity.

Figure 1 reports one of the more interesting plots of the performed analysis with $q = 10$ (other plots can not be shown for lack of space). Each panel of the figure refers to the t —statistic of coefficients $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ and $\hat{\alpha}_5$ during the last 257 steps of the forward search, that is from step 1500 to step 1756. Dashed lines indicate the limit of 2.58 for significance of parameters at 1% level. It is worth noting that residuals are ordered during the procedure according to their degree of outlyingness and that, in the last step of the forward search, all observations are included in the estimation procedure. Thus, the last value in each plot is the t —statistic obtained considering also outlying observations. As

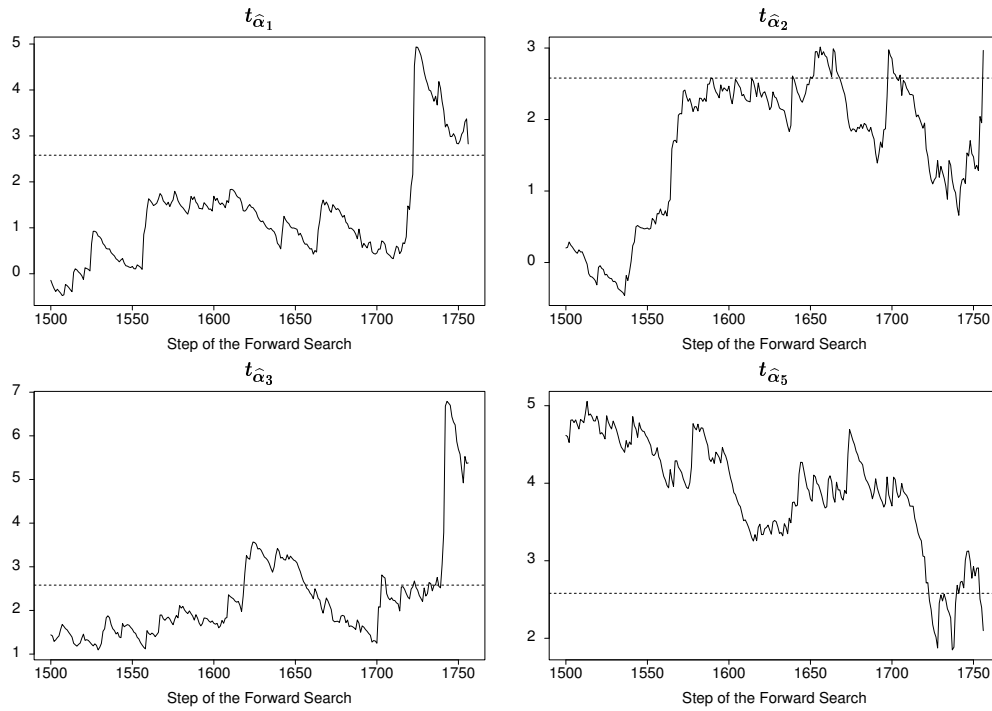


Figure 1: Plot of the t -statistics for the ARCH parameters $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ and $\hat{\alpha}_5$. Notice the sharp movements at the entrance of influential observations and the change of significance's level in the parameters.

can be observed, t -statistics are strongly influenced by the presence of single or small patches of outliers. Plot of $t_{\hat{\alpha}_1}$ indicates a non significant coefficient for most steps of the forward search, while the entrance of one single observation at step 1722 increases t -value from 2.16 to 4.52. A similar behaviour can be stressed for $t_{\hat{\alpha}_3}$: in three steps (from 1739 to 1742) it moves from 2.51 to 6.67. The estimates of the second ARCH component remain under the significance level until step 1755 and only the inclusion of the most outlying observation increases its value from 1.95 to 2.97. On the contrary, it may be noticed that $t_{\hat{\alpha}_5}$ is significantly different from zero up to the last steps of the search, where it crosses several times the 1% level. As a conclusion we can say that the classical LM test, without consideration of outliers, leads to the identification of an ARCH(3) disturbance term which is based only on a few most outlying observations.

References

- Atkinson A.C. and Riani M. (2000) *Robust Diagnostic Regression Analysis*, Springer Verlag, New York.
- Engle R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation, *Econometrica*, 50, 987–1007.
- Franses P.H. and Van Dijk D. (2000) *Non-Linear Time Series Models in Empirical Finance*, Cambridge University Press, Cambridge.
- Rousseeuw P.J. and Leroy A.M. (1987) *Robust Regression and Outlier Detection*, Wiley, New York.