ARCH model for WIG-Banki

Marcin Pawlik

January 7, 2005

1 Introduction

The idea of this project is to conclude tests of ARCH processes existence on the Polish Stock Exchange. The subject of the research is WIG-BANKI index. One can interpret this as an attempt to explain the rate of return of a portfolio that is build in the same way as WIG-BANKI.
The ARCH-class models are of my interest since heard about them and their applications over a year ago during a lecture in advanced econometrics. It is even more temptive to take them into consideration as the inventor of this approach has been awarded a Nobel prize in economics. I think that this paper is a great opportunity to practice some techniques of ARCH modelling. In my model I would like to explain the rate of return of a given commercial paper or changes of index levels. The explicit specification is still the subject of testing procedure.

2 Specification of the model

The test for ARCH effect consists in checking whether the disturbance term ξ has constant conditional variance against the alternative that the conditional variance can be explained by the following equation:

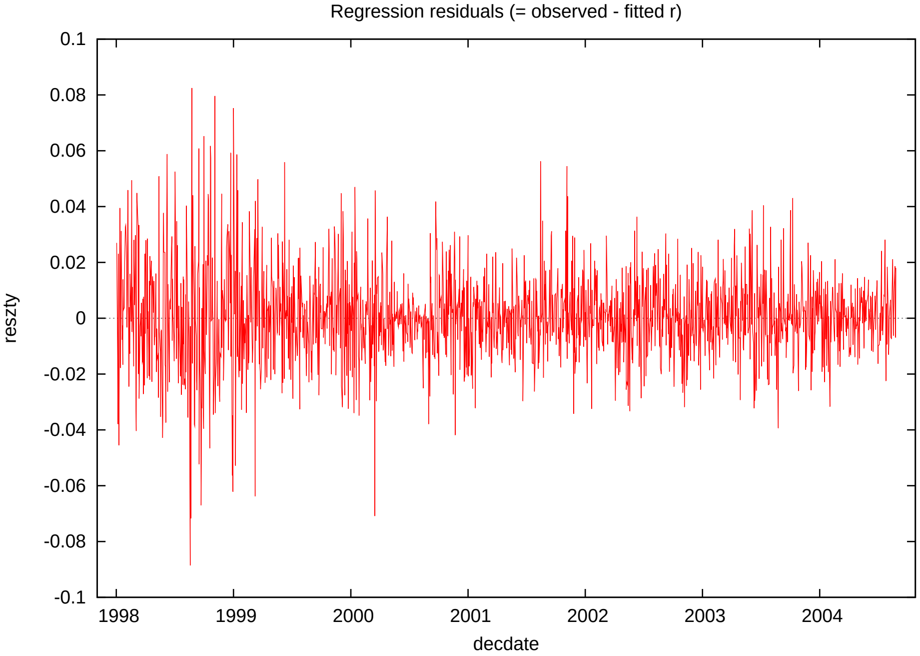
$$h_t = \gamma_0 + \sum_{s=1}^{S} \gamma_s \xi_{t-s}^2 \tag{1}$$

The testing procedure assumes following hypothesis:

$$H_0 : \gamma_1 = \gamma_2 = \ldots = \gamma_S = 0, \qquad H_1 : \gamma_S \neq 0. \tag{2}$$

The test statistic is given with equation:

$$TR^2 : \chi_S^2 \tag{3}$$



with S degrees of freedom, where T is the number of observations, R^2 is the coefficient of determination of the following equation:

$$e_t^2 = \beta_0 + \sum_{s=1}^{S} \beta_s e_{t-s}^2 + \tau_t \tag{4}$$

Where e_t are the residuals obtained by OLS estimation of the first equation. The rejection of H_0 hypothesis means that the ARCH effect is significant. The parameters of ARCH models can be estimated in an OLS procedure because x_t and ξ_t are not correlated and $\upsilon_t : IID(0,1)$. However the maximum likelihood estimator is asymptotically more effective, which begs the question of if its application in the estimation.

The success of a specific investment strategy depends often on the precise forecast of changes in the moves of quotation, not necessarily its valuse. That is why another measure used to reckon the goodness of an ARCH model is the measure of accordance in the changes of direction:

$$Q1 = \frac{N\left(\{r_t \hat{r}_t > 0\}\right)}{N\left(\{r_t \hat{r}_t \neq 0\}\right)} \tag{5}$$

where:

$\hat{r}_t$	theoretical values of the dependent variable,
$N\left(\{r_t \hat{r}_t > 0\}\right)$	number of observations for which $r_t \hat{r}_t > 0$,
$N\left(\{r_t \hat{r}_t \neq 0\}\right)$	number of observations for which $r_t \hat{r}_t \neq 0$.

In this paper I assume initially that the realizations of the rate of change is a process that can be represented as a stochastic differential equation. Actually that means I assume that it is an autoregressive process and so the present values can be explained by those from the past. Eventually the equation to be estimated takes form of:

$$r_t = \alpha_0 + \sum_{k=1}^{K} \alpha_s r_{t-k} + \xi_t \tag{6}$$

$$\xi_t = \upsilon_t \sqrt{\gamma_0 + \sum_{s=1}^{S} \gamma_s \xi_{t-s}^2} \tag{7}$$

for which:

$$E\left(r_t\right) = E\left(r_t|_{t-1} r_{t-1}\right) = \mathbf{x_{(k)t}} \mathbf{a_k} \tag{8}$$

$$D^2\left(r_t|_{t-1}\right) = \gamma_0 + \sum_{s=1}^{S} \gamma_s \xi_{t-s}^2 \tag{9}$$

where $\mathbf{x_{(k)t}}$ means a vector of independent variables and $\mathbf{a_{(k)}}$ a vector of parameters.

3 Data description

In my work I use the data on WIG-BANKI available on the portal www.bossa.pl. The data set ranges from 31st December 1997 to 7th December 2004. The summary statistics of the downloaded data are presented below.

Variable	Mean	Median	Minimum	Maximum
Open	20585	19213	9800,5	33404
High	20659	19218	9800,5	34001
Low	20543	19213	9800,5	33404
Close	20619	19218	9800,5	34001
Volume	1,0648E+05	87311	0,0000	5,0403E+05

The variables used in my model are defined as it follows:

`r`	rate of change (percent), `r = (close - close(-1))/close(-1)` where `close` is the value of WIG-BANKI at the session close and `close(-1)` is lagged value
`r_i`	lagged value where "i"s are the lag lengths
`const`	simply constant value
`yhat1`	fitted values in arch(1)
`uhat1`	residuals in ARCH(1)
`uhat12`	squared lagged residuals in ARCH(1)
`yhat2`	fitted values in garch(1,1)
`uhat2`	residuals in GARCH(1,1)
`uhat22`	squared lagged residuals in GARCH(1,1)
`r2`	squared `r`

And their summary statistics are following:

Variable	Mean	Median	Minimum	Maximum
`r`	0,00076587	0,00035117	-0,097030	0,078977
`r_1`	0,00075557	0,00034174	-0,097030	0,078977
`r_2`	0,00076924	0,00034787	-0,097030	0,078977
`r_3`	0,00076420	0,00034174	-0,097030	0,078977
`r_4`	0,00075392	0,00034174	-0,097030	0,078977
`yhat1`	0,00073143	0,00069608	-0,0076699	0,0074519
`uhat1`	3,4441E-05	-0,00036010	-0,092556	0,078687
`uhat12`	0,00027432	6,9416E-05	0,0000	0,0085666
`yhat2`	0,00069109	0,00064777	-0,0096030	0,0089256
`uhat2`	7,4785E-05	-0,00033579	-0,091343	0,079190
`uhat22`	0,00027385	6,9062E-05	0,0000	0,0083435

4 Estimation results

Assuming that the process genrating `r` values is autoregressive I started with an attempt to determine the lag length. My target was to find as short as possible length bearing in mind however not falling in the trap of autocorrelation. I use "from general to specific" methodology which I find quite reasonable. At first I test lag length of 4.

Model 1: OLS estimates using the 1735 observations 98/01/02–04/08/26
Dependent variable (Y): `r`

Variable	Coefficient	Std. Error	*t*-statistic	p-value
`const`	0,000711952	0,000398679	1,7858	0,0743
`r_1`	0,142054	0,0240241	5,9130	0,0000
`r_2`	-0,0207068	0,0242675	-0,8533	0,3936
`r_3`	-0,00341302	0,0242597	-0,1407	0,8881
`r_4`	-0,0466308	0,0240253	-1,9409	0,0524

Mean of dependent variable	0,000765871
S.D. of dependent variable	0,0167168
Sum of squared residuals	0,473811
Standard error of residuals ($\hat{\sigma}$)	0,0165493
Unadjusted R^2	0,0222022
Adjusted $\bar{R}^2$	0,0199414
$F(4, 1730)$	9,82047
Durbin–Watson statistic	1,99300
First-order autocorrelation coeff.	0,00244773

Then I run F-test to answer whether `r_4`, `r_3` and `r_2` are essential in the specification. The result is as follows:

```
Restriction set
 1: b(r_2) = 0
 2: b(r_3) = 0
 3: b(r_4) = 0

Test statistic: F(3, 1730) = 1,57472, with p-value = 0,193562
```

While the critical values for F(3, 1730) distribution are as it is shown in table:

10%	2,09
5%	2,61
1%	3,79

I can state that these three variables are insignificant with the 5% probability of an first order error. That is why I follow estimation with only one regressor and a constant.

Model 2: OLS estimates using the 1735 observations 98/01/02–04/08/26
Dependent variable (Y): `r`

Variable	Coefficient	Std. Error	t-statistic	p-value
`const`	0,000660184	0,000397914	1,6591	0,0973
`r_1`	0,139796	0,0237923	5,8757	0,0000

Mean of dependent variable	0,000765871
S.D. of dependent variable	0,0167168
Sum of squared residuals	0,475105
Standard error of residuals ($\hat{\sigma}$)	0,0165575
Unadjusted R^2	0,0195321
Adjusted $\bar{R}^2$	0,0189663
Degrees of freedom	1733
Durbin–Watson statistic	1,99190
First-order autocorrelation coeff.	0,00300527

Low values of R^2 are nothing unusual in case of ARCH models. This phenomenon has been noticed by Andersen and Bollerslev for the first time. They showed empirically that this statistics rarely exceeds 0,05 for models based on daily data. Next I run a test for ARCH effect. Actually I run several of them for different orders of ARCH process ranging from 1 to 5. The results are displayed below:

Test for ARCH of order 5
OLS estimates using the 1730 observations 98/01/09-04/08/26
Dependent variable (Y): `ut^2`

Variable	Coefficient	Std. Error	t-statistic	p-value
`const`	0,000129192	1,68581E-05	7,664	< 0,00001 ***
`ut^2(-1)`	0,171040	0,0240757	7,104	< 0,00001 ***
`ut^2(-2)`	0,0920446	0,0243525	3,780	0,000162 ***
`ut^2(-3)`	0,157772	0,0241301	6,538	< 0,00001 ***
`ut^2(-4)`	0,0783104	0,0243205	3,220	0,001306 ***
`ut^2(-5)`	0,0253149	0,0240453	1,053	0,292578

Number of observations=1730, $R^2 = 0{,}115728$
LM test statistic (200,208970) is distributed as Chi-square (5)
Area to the right of LM = 0,000000
ARCH effect is essential at the level of 10% error.

As the lag length proved to be to long (the last variable was insignificant) now I run the test for ARCH effect of order 4

Test for ARCH of order 4
OLS estimates using the 1731 observations 98/01/08-04/08/26
Dependent variable (Y): `ut^2`

Variable	Coefficient	Std. Error	t-statistic	p-value
`const`	0,000132602	1,65445E-05	8,015	< 0,00001 ***
`ut^2(-1)`	0,173117	0,0239874	7,217	< 0,00001 ***
`ut^2(-2)`	0,0965350	0,0240115	4,020	0,000061 ***
`ut^2(-3)`	0,160086	0,0240075	6,668	< 0,00001 ***
`ut^2(-4)`	0,0827806	0,0239568	3,455	0,000563 ***

Number of observations =1731, $R^2 = 0{,}115304$
LM test statistic (199,590417) is distributed as Chi-square (4)
Area to the right of LM = 0,000000
ARCH effect is essential at the level of 10% error.

As one can see there is a strong ARCH effect of order 4. The TR^2 statistic exceeds the critical level for null hypothesis. Finally I reestimate my modle running WLS.

Model 3: WLS (ARCH) estimates using the 1730 observations
98/01/09-04/08/26
Dependent variable (Y): `r`
Variable used as weight: `1/sigma`

Variable	Coefficient	Std. Error	t-statistic	p-value
`const`	0,000677280	0,000351802	1,925	0,054371 *
`r_1`	0,153552	0,0277713	5,529	$< 0,00001$ ***

Statistics based on the weighted data:

Sum of squared residuals	1704,58
Standard error of residuals	0,992914
Unadjusted R^2	0,0173745
Adjusted $\hat{R}^2$	0,0168061
Degrees of freedom	1729
Durbin-Watson statistic	2,01341
First-order autocorrelation coeff.	-0,00774189

Statistics based on the original data:

Mean of dependent variable	0,000770214
Standard deviation of dep. var.	0,0166897
Sum of squared residuals	0,472291
Standard error of residuals	0,0165275

Next step is ARCH estimation using MLE. Here I run estimation for ARCH(2), ARCH(1) and GARCH(1,1). This results from two major issues. The first one is that in practice one usually doesn't estimate ARCH models with larger lag orders. The reason is that those models have not got better forecasting properties. The other is of numerical nature of algorithms used in my software to run this regression. ARCH and GARCH of longer lag orders cannot be calculated because of non-divergence of numerical procedures.
The two first models are ARCH(2) and ARCH(1) respectively.

Model 4: ARCH estimates using the 1735 observations 98/01/02–04/08/26
Dependent varible (Y): `r`
Standard errors based on Hessian

Variable	Coefficient	Std. Error	t-statistic	p-value
`const`	0,000646047	0,000349914	1,8463	0,0650
`r_1`	0,100261	0,0283791	3,5329	0,0004
α_0	0,000161441	8,86749e-06	18,2059	0,0000
α_1	0,246016	0,0369679	6,6549	0,0000
α_2	0,168758	0,0309985	5,4441	0,0000

The results of ARCH(1) estimation are presented below.

Model 5: ARCH estimates using the 1735 observations 98/01/02–04/08/26
Dependent variable (Y): `r`
Standard errors based on Hessian

Variable	Coefficient	Std. Error	t-statistic	p-value
`const`	0,000666477	0,000369165	1,8054	0,0712
`r_1`	0,0859157	0,0337401	2,5464	0,0110
α_0	0,000202857	9,23548e-06	21,9650	0,0000
α_1	0,260930	0,0398457	6,5485	0,0000

Mean of dependent variable	0,000765871
Standard deviation of dep. var.	0,0167168
Log-likelihood	4717,087
AIC	-9424,173

BIC = -9396,879

According to the findings of Andersen and Bollerslev the right accuracy measure in ARCH models is R^2 of the following equation:

$$r_t^2 = \phi_0 + \phi_1 h_t + \xi \qquad \text{where: } h_t = \gamma_0 + \sum_{s=1}^{S} \gamma_s \xi_{t-s}^2. \tag{10}$$

The results of the estimation of that equation are following:

Model 6: OLS estimates using the 1735 observations 98/01/02–04/08/26
Dependent variable (Y): `r2`

Variable	Coefficient	Std. Error	t-statistic	p-value
`const`	0,000206123	1,57492e-05	13,0878	0,0000
`uhat1_1s`	0,269091	0,0238250	11,2945	0,0000

Mean of dependent variable	0,000279877
S.D. of dependent variable	0,000618362
Sum of squared residuals	0,000617574
Standard error of residuals ($\hat{\sigma}$)	0,000596960
Unadjusted R^2	0,0685626
Adjusted $\bar{R}^2$	0,0680251
Degrees of freedom	1733
Durbin–Watson statistic	2,06429

As we can observe R^2 behaves just like in Bollerslev's and Andersen findings. The other measure of ARCH model goodness shows how often model forecasts the change in direction of quotation fluctuation.

$$Q1 = \frac{N\left(\{r_t \hat{r}_t > 0\}\right)}{N\left(\{r_t \hat{r}_t \neq 0\}\right)} \tag{11}$$

And the result for ARCH(1) is: $Q1 = 0{,}53141$.
The last step is estimation of GARCH(1,1). The results are as follows:

Model 7: GARCH estimates using the 1735 observations
98/01/02–04/08/26
Dependent variable (Y): `r`
Standard errors based on Hessian

Variable	Coefficient	Std. Error	t-statistic	p-value
`const`	0,000611500	0,000331533	1,8445	0,0653
`r_1`	0,105271	0,0248113	4,2429	0,0000
α_0	3,06014e-06	1,22877e-06	2,4904	0,0129
α_1	0,0574160	0,0120061	4,7822	0,0000
β_1	0,930372	0,0148734	62,5527	0,0000

Mean of dependent variable	0,000765871
S.D. of dependent variable	0,0167168
Log-likelihood	4825,78
AIC	-9639,5
BIC	-9606,8

Once again I run estimation of proper equation in order to achieve R^2 for GARCH model.

Model 8: OLS estimates using the 1735 observations 98/01/02–04/08/26
Dependent variable (Y): `r2`

Variable	Coefficient	Std. Error	t-statistic	p-value
`const`	0,000206560	1,57610e-05	13,1057	0,0000
`uhat2_1s`	0,267970	0,0239107	11,2071	0,0000

Mean of dependent variable	0,000279877
S.D. of dependent variable	0,000618362
Sum of squared residuals	0,000618227
Standard error of residuals ($\hat{\sigma}$)	0,000597276
Unadjusted R^2	0,0675773
Adjusted $\bar{R}^2$	0,0670393
Degrees of freedom	1733
Durbin–Watson statistic	2,05669
First-order autocorrelation coeff.	-0,0285684

The other measure of GARCH model goodness is Q1 and this time it amounts to: $Q1 = 0{,}53429$

5 Conclusion

The subject of the modelling was the growth rate for WIG-BANKI index. I assumed one substantial specification - autoregressive. I used two types of ARCH models: ARCH(1) and GARCH(1,1).
Statistical test have shown existence of a strong ARCH effect. The research have been done for daily data.
The results of my estimation confirm the fact that R^2 for ARCH models amounts to about 0.05 for daily data. In my research this value was 0,068 and 0,067 for ARCH(1) and GARCH(1,1) respectively. The values of Q1 statistic exceed the level of 50% which gives them some forecasting properties. ARCH(1) model succeded to foresee over 53% of direction in which the index moves. Similar value was achieved by this statistic for GARCH(1,1) model.