Financial Econometrics

Slides-10: Modeling Return Volatility: Testing/Estimating/Forecasting ARCH and Introduction to GARCH

Dr. Rachida Ouysse
School of Economics[1]

Lecture Plan

- ARCH LM-test
- Forecasting with ARCH

- Generalised ARCH: why and how

  Formulation of GARCH: parameter restrictions
- Properties of GARCH(1,1)
  - Mean, variance, ARMA(1,1) representation
  - ML estimation of GARCH

- Forecasting with GARCH

LM test for ARCH effect

Testing for ARCH-effects: **The ARCH-LM test**

- Obtain the residuals $\hat{\mu}_t$ from a regression, e.g.

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \mu_t$$

- Obtain R^2 of the auxiliary regression

$$\hat{\mu}_t^2 = \gamma_0 + \gamma_1 \hat{\mu}_{t-1}^2 + \gamma_2 \hat{\mu}_{t-1}^2 + \dots + \gamma_q \hat{\mu}_{t-q}^2 + \nu_t$$

- Calculate test statistic $T'R^2$ with T' the number of observations in the auxiliary regression.
- Under the null hypothesis of no ARCH, $T'R^2 \sim \chi^2(q)$.

LM test for ARCH effect: Example

eg. NYSE composite return:

1. Estimate the model for mean (eg. AR(1)) and save the residual series $\hat{\mu}_t$.
2. OLS auxilary regression: $\hat{\mu}_t^2 = \gamma_0 + \gamma_1 \hat{\mu}_{t-1}^2 + \cdots + \gamma_q \hat{\mu}_{t-q}^2 + error_t$
   Save the R^2. (q depends on T and data frequency)
3. $T' = T - q$, with $q = 5$ reject when $T'R^2$ exceeds $\chi^2_{(5)}$

eg. NYSE composite return: LM test with q = 5

E

ARCH Test:

F-statistic	37.43273	Probability	0.000000
Obs*R-squared	171.0570	Probability	0.000000

V

ARCH Test:

F-statistic	1.318328	Probability	0.253338
Obs*R-squared	6.589613	Probability	0.252993

Performed on "V" to check the adequacy of variance equation

Forecasting with ARCH Models

- Using repeated substitutions, we can make multi-step forecasts for the return and its volatility
- Example. AR(1)-ARCH(2)

$$
\begin{aligned}
y_t &= c + \phi_1 y_{t-1} + \mu_t, \quad \mu_t|\Omega_{t-1} \sim N(0, \sigma_t^2) \\
\sigma_t^2 &= \alpha_0 + \alpha_1 \mu_{t-1}^2 + \alpha_2 \mu_{t-2}^2 \\
y_{t+1|t} &= c + \phi_1 y_t, \\
y_{t+2|t} &= c + \phi_1 y_{t+1|t}, \cdots \\
\sigma_{t+1|t}^2 &= \alpha_0 + \alpha_1 \mu_t^2 + \alpha_2 \mu_{t-1}^2, \\
\sigma_{t+2|t}^2 &= \alpha_0 + \alpha_1 \sigma_{t+1|t}^2 + \alpha_2 \mu_t^2, \\
\sigma_{t+3|t}^2 &= \alpha_0 + \alpha_1 \sigma_{t+2|t}^2 + \alpha_2 \sigma_{t+1|t}^2, \cdots
\end{aligned}
$$

Forecasting with ARCH models: Example

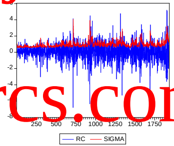
– Forecasting with ARCH models

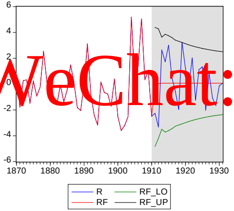
eg. NYSE composite return:

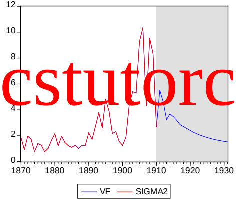
AR(1)-ARCH(5) forecasts

revert to unconditionals

(mean reverting)







Remember the limitations of ARCH!

Advantages of ARCH

- It is able to capture 'clustering' in return series or the autocorrelation in squared returns.
- It facilitates volatility forecasting.
- It explains, partially, non-normality in return series.

Limitations of ARCH

- In ARCH(q), the q may be selected by AIC, SIC or LR test. The correct value of q might be very large. The model might not be parsimonious. (eg. ARCH(1) would not work for the composite return)
- The conditional variance σ_t^2 cannot be negative: Requires non-negativity constraints on the coefficients. Sufficient (but not necessary) condition is: $\alpha_i \geq 0$ for all $i = 0, 1, 2, \cdots q$. Especially for large values of q this might be violated

GARCH Models: Introduction

Generalised ARCH (GARCH) models allow the conditional variance to depend upon previous own lags.

- Let μ_t be the error term or shock in a model.

  ARCH(q): Var$(\mu_t|\Omega_{t-1}) = \sigma_t^2$,

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \alpha_2 \mu_{t-2}^2 + \cdots + \alpha_q \mu_{t-q}^2,$$

  is not parsimonious as a large q is often required.

- If σ_{t-1}^2 is a summary of volatility info in Ω_{t-2}, then $\Omega_{t-1} = \{\mu_{t-1}, \mu_{t-2}, \mu_{t-3}, \cdots\} = \{\mu_{t-1}, \Omega_{t-2}\} \approx \{\mu_{t-1}, \sigma_{t-1}^2\}$ (volatility wise!)

- This leads to the GARCH(1,1) model:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

GARCH: Introduction

- More generally, GARCH(p, q) model

  Var$(\mu_t|\Omega_{t-1}) = \sigma_t^2$,

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \cdots + \alpha_q \mu_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2,$$

  where the parameters should satisfy:

(1) Positivity constraint: $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0$ for all $i = 1, \cdots, q$ and $j = 1, \cdots, p$

(2) Finite Variance $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$

- In practice, the models for asset returns rarely go beyond GARCH(1,1).

Properties of GARCH

The generalisation implied by GARCH can be seen from backward iterating the GARCH(1,1) model:

$$\sigma_t^2 = \frac{\alpha_0}{1-\beta_1} + \alpha_1 \sum_{j=1}^{\infty} \beta_1^{j-1} \mu_{t-j}^2.$$

This shows that the GARCH model is an ARCH(∞) with geometrically declining coefficients (for $|\beta_1| < 1$).

Properties of GARCH(1,1)

Alternatively, if we define the surprise in the squared innovations as $\omega_t = \mu_t^2 - \sigma_t^2$, the GARCH(1,1) model can be rewritten as

$$\begin{aligned} \mu_t^2 - \omega_t &= \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta_1 \left(\mu_{t-1}^2 - \omega_{t-1} \right) \\ \mu_t^2 &= \alpha_0 + (\alpha_1 + \beta_1) \mu_{t-1}^2 + \omega_t - \beta_1 \omega_{t-1} \end{aligned}$$

which shows that the **squared errors** follow an ARMA(1,1) model. As the root of the autoregressive part is $\alpha_1 + \beta_1$, the squared residuals are stationary provided $|\alpha_1 + \beta_1| < 1$.

Under stationarity, $E\left(\mu_t^2\right) = E\left(\mu_{t-1}^2\right) = E\left(\sigma_{t-1}^2\right) = \sigma^2$, the unconditional variance of μ_t is given by

$$\begin{aligned} \sigma^2 &= \alpha_0 + \alpha_1 \sigma^2 + \beta_1 \sigma^2 \\ &= \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \end{aligned}$$

Properties of GARCH(1,1)

Two general cases can be distinguished

- $\alpha_1 + \beta_1 < 1 \;\rightarrow$ the unconditional variance is defined, i.e. finite
- $\alpha_1 + \beta_1 \geq 1 \;\rightarrow$ the unconditional variance is not defined, i.e. infinite

The latter case is denoted **non-stationarity in variance**

- Variance does not converge to an unconditional mean
- The special case where $\alpha_1 + \beta_1 = 1$ is known as a unit root in variance or **integrated GARCH** (IGARCH)

Properties of GARCH(1,1)

- GARCH(1,1): $\mu_t|\Omega_{t-1} \sim N(0, \sigma_t^2)$,

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1$$

- Its conditional variance is time varying:

$$E(\mu_t|\Omega_{t-1}) = 0, \ \text{Var}(\mu_t|\Omega_{t-1}) = \sigma_t^2,$$

$$\text{CI (95\%)} = E(y_{t+1}|\Omega_{t-1}) \pm 2\sigma_t$$

- μ_t is a White Noise: $E(\mu_t) = 0$, $\text{Var}(\mu_t) = \frac{\alpha_0}{1-(\alpha_1+\beta_1)}$, $\text{Cov}(\mu_t, \mu_{t-j}) = 0$
- But it is NOT an independent WN or iid WN. It is NOT unconditionally Normally distributed: $\text{kurt}(\mu_t) > 3$

Properties of GARCH(1,1)

- GARCH(1,1) can be expressed in terms of **standardised shocks** ν_t: $\mu_t = \sigma_t \nu_t$ and $\nu_t \sim$ iid $N(0,1)$

- When model is correct, ν_t^2 should have no autocorrelation.

Advantages of the GARCH model (compared to ARCH)

- Avoids overfitting, i.e. a higher order ARCH model may have a more parsimonious GARCH representation

- Due to less estimated parameters, violations of the non-negativity constraint are less likely

GARCH(1,1) Estimation

Estimating GARCH models

For instance, estimate the following AR(1)-GARCH(1,1) model

$$
\begin{aligned}
y_t &= \mu + \phi y_{t-1} + u_t \\
u_t &= \nu_t \sigma_t \quad \nu_t \sim N(0,1) \\
\sigma_t^2 &= \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta_1 \sigma_{t-1}^2
\end{aligned}
$$

OLS is inappropriate

- OLS minimises the RSS, $\sum \hat{u}_t^2 = \sum \left(y_t - \hat{\mu} - \hat{\phi} y_{t-1} \right)^2$, which is a function of the parameters in the conditional mean equation only and not in the conditional variance equation
- In fact, OLS assumes that the residuals are homoscedastic, i.e. all slope coefficients in the conditional variance equation are set to zero

GARCH(1,1) Estimation

Maximum Likelihood

- Make assumptions about **conditional distribution** of μ_t, e.g.

$$\nu_t \sim N(0,1) \quad \text{such that} \quad \mu_t \sim N\left(0, \sigma_t^2\right)$$

  This means that conditional on information available at $t-1$ μ_t is normally distributed with mean zero and variance σ_t^2 with the latter being known at time $t-1$. Note that this does not imply that the **unconditional distribution** of μ_t is normal, as σ_t becomes a random variable if we do not condition on all information available on $t-1$.

- The conditional distribution of y_t is then also normal, given by

$$f\left(y_t \,|y_{t-1}, \ldots, \mu_{t-1}, \ldots\right) = \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\left(-\frac{1}{2}\frac{\mu_t^2}{\sigma_t^2}\right)$$

  with $\mu_t = y_t - \mu - \phi y_{t-1}$ and $\sigma_t^2 = \alpha_0 + \alpha_1\mu_{t-1}^2 + \beta_1\sigma_{t-1}^2$.

GARCH(1,1) Estimation

- The **loglikelihood function** is given by the sum over all t of the log of the conditional distribution of y_t

$$L = -\frac{T}{2}\log\left(2\pi\right) - \frac{T}{2}\sum_{t=1}^{T}\log\left(\sigma_t^2\right) - \frac{1}{2}\sum_{t=1}^{T}\frac{\mu_t^2}{\sigma_t^2}$$

- The ML estimator is obtained by **maximising** the loglikelihood with respect to the unknown parameters $(\mu, \phi, \alpha_0, \alpha_1, \beta_1)$.
- Analytical solution not possible: use **numerical procedures**
  - These algorithms 'search' over the parameter space, from an **initial guess**, until a maximum for the loglikelihood function is found
  - Potential problem: the loglikelihood function may have several **local maxima** such that alternative initial guesses may yield different results.
  - In practice: use linear regression to get initial estimates of the parameters in the conditional mean equation and choose some (alternative) parameter value for the parameters in the conditional variance equation $\neq 0$.

GARCH(1,1) Estimation

GARCH(1,1) Estimation

- Fortunately, first order conditions are, under some weak assumptions, valid even when ν_t is not normally distributed.
  - The parameter estimates are still consistent
  - Adjustments have to be made to the standard errors, i.e. use Bollerslev-Wooldridge variance-covariance matrix, also known as **Quasi Maximum Likelihood Estimation**, which is robust for non-normality.

Example 1: GARCH(1,1) Estimation

ML Estimation of GARCH(1,1)

eg. NYSE composite return

Dependent Variable: RC
Method: ML - ARCH (Marquardt) - Normal distribution

Sample (adjusted): 3 1931
Included observations: 1929 after adjustments
Convergence achieved after 14 iterations
Bollerslev-Wooldridge robust standard errors & covariance
Variance backcast: OFF
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.072383	0.018494	3.913969	0.0001
AR(1)	0.102468	0.025514	4.016133	0.0001
Variance Equation				
C	0.013305	0.004529	2.937750	0.0033
RESID(-1)^2	0.119181	0.024748	4.815836	0.0000
GARCH(-1)	0.877098	[illegible]	41.71302	0.0000

R-squared	[illegible]	Mean dependent var	[illegible]
Adjusted R-squared	[illegible]	S.D. dependent var	[illegible]
S.E. of regression	1.006280	Akaike info criterion	[illegible]
Sum squared resid	1948.240	Schwarz criterion	2.636091
Log likelihood	-2523.598	F-statistic	1.164694
Durbin-Watson stat	2.055220	Prob(F-statistic)	0.324498
Inverted AR Roots	.10		

ARCH Test: $q = 5$

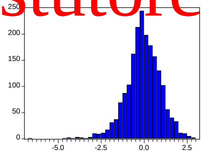
F-statistic	0.764181	Probability	0.575599
Obs*R-squared	3.825239	Probability	0.574842

Test Equation:
Dependent Variable: STD_RESID^2

Correlogram of Standardized Residuals Squared

Sample: 3 1931
Included observations: 1929
Q-statistic probabilities adjusted for 1 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.001	-0.001	0.0005	
		2	[illegible]	0.036	2.4991	0.114
		3	[illegible]	-0.014	2.8675	0.238
		4	[illegible]	-0.012	3.1058	0.376
		5	[illegible]	-0.019	3.8964	0.420
		6	0.009	0.010	4.0531	0.542
		7	0.003	0.004	4.0658	0.668
		8	-0.001	-0.002	4.0670	0.772
		9	-0.011	-0.011	4.3009	0.829
		10	-0.010	-0.010	4.4826	0.877
		11	-0.008	-0.006	4.5929	0.917
		12	-0.016	-0.016	5.1005	0.926
		13	-0.004	-0.004	5.1268	0.954
		14	-0.011	-0.011	5.3589	0.966
		15	-0.019	-0.019	6.0452	0.965
		16	-0.009	-0.009	6.2090	0.976



Series: Standardized Residuals
Sample 3 1931
Observations 1929

Mean	-0.048341
Median	-0.039867
Maximum	2.850528
Minimum	-6.601836
Std. Dev.	0.996820
Skewness	-0.547486
Kurtosis	4.973199
Jarque-Bera	409.3080
Probability	0.000000

Example 1: GARCH(1,1) Estimation

ML Estimation of GARCH(1,1)

eg. NYSE composite return (continued)

- Large β_1 estimate: about 0.9
- Small α_1 estimate: about 0.1
- $\alpha_1 + \beta_1$ estimate: very close to 1

GARCH(1,1) is preferred by AIC and SIC.

	AIC	SIC
AR(1)-ARCH(5)	2.664	2.687
AR(1)-GARCH(1,1)	2.622	2.636

Example 1: GARCH(1,1) Estimation

– ML Estimation of GARCH(1,1)

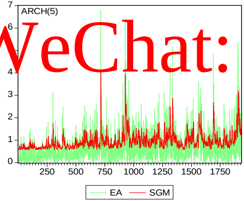
eg. NYSE composite return (continued)

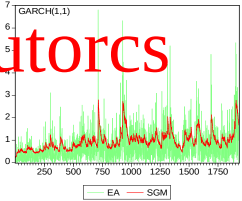
GARCH(1,1) σ_t plot is smoother than ARCH(5).

Large β_1 estimate implies ***persistence***:

σ_t tends to continue at the current level.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$





Summary facts about GARCH models

- GARCH(1,1) is usually preferred to ARCH or higher order GARCH, because of its parsimony.
- Usually, GARCH β_1 estimate is about 0.9 or more and $\alpha_1 + \beta_1$ estimate is very close to 1, for daily returns.

- Standardised residuals are usually non-normal, with negative skewness and excessive kurtosis.
- GARCH(1,1) is able to capture clustering in returns but unable to account for

  Asymmetry: negative returns tend to cause more volatility;

  Non-normality; Structural change
- Coefficient restrictions are hard to impose in MLE