Financial Econometrics

Sol Yates

February 26, 2024

Contents

Lecture 4: GARCH

Mon 19 Feb 09:00

Recap

ARMA

1. $E[\varepsilon_t] = 0$
2. $V\varepsilon_t = \sigma^2$
3. $\mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0$ that is no serial correlation

Tutorial 2 : S&P 500 Daily log returns → ARMA(p,q) → BIC then use residual diagnostics

$$@_t = y_t - \hat{E}[y_t|F_{t-1}] \to MA(\mathcal{L})$$

Week 3
NP / Rob Engel 2003

$$\begin{aligned} \varepsilon_t &= \sigma_t \mathcal{L}_t \\ \mathcal{L}_t &\sim \mathcal{N}(0,1) \\ \sigma_t^2 &= w + \alpha\varepsilon_{t-1}^2 < - \end{aligned}$$

1. $a \geq 0$ and $\omega > 0$ - to ensure positivity of conditional variance
2. $|\alpha| < 1$ Stationarity of conditional variance

ARCH(1)

$$\begin{cases} \sigma = w + \alpha\varepsilon_{t-1} < - \\ \varepsilon_t + \mathcal{L}_t\sigma_t \\ \quad \text{rewrite } \ \sigma_t^2 = w + \alpha\varepsilon_{t-1}^2 + \varepsilon_t^2 - \varepsilon_t^2 \\ AR(1) \ in \ \varepsilon_t^2 \to \varepsilon_t^2 = w + \alpha\varepsilon_{t-1}^2 + \left(\varepsilon_t^2 - \sigma_t^2\right) \end{cases} \qquad Video \begin{cases} E[V_t] = 0 \\ V[v_t] < \infty v_t = \sigma^2 \\ cov(v_t, v_{t-s}) = 0 \end{cases}$$

Pros

- Volatility clustering (video)
- rise persistence at the cost of ARCH (p)
- Leptokurtic property $\alpha^2 \in (0, \frac{1}{3})$

Cons

- leverage effect : $E[\mathcal{L}_t^3] = 0$
- Long memory (ACF)

What can we do with our Garch models to capture all remaining things in ACF?

0.1 GARCH

A process σ_t^2 is called an GARCH(1, 1) process if

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

with $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$

Properties

- ε_t^2 is stationary if $\alpha + \beta < 1$
- both processes ε_t and ε_t^2 are stationary and $E[\varepsilon_t] = 0$ then the unconditional variance of ε_t $V[\varepsilon_t]$ which is equal to the unconditional mean of ε_t^2
- no leverage effects as in the ARCH
-

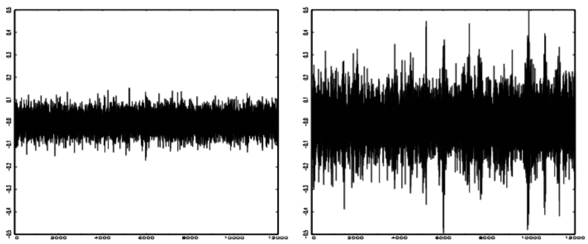

Figure 1: Simulated GARCH Models

left $\alpha = 0.01$ and $\beta = 0.8$. Right $\alpha = 0.08$ and $\beta = 0.9$ If allow close to 1 then can generate longer persistence, usually the memory of the daily log returns is us more persistent. Most have very low memory, thus people came up with GARCH(p, q)

M1 GARCH(1, 1)

- it takes into account / able to model more persistent conditional volatility processes
- mitigating the tradeoff between generating a leptokurtic distribution of ε_t and the persistence iof the ACF if ε_t^2 as compared to ARCH(1)

M2 ARCH(1)

GARCH captures over kurtosis, even if we could like sum of $\alpha + \beta$ to 1, we still have an opportunity to generate a over kurtosis (>3)

We can also show GARCH reveals larger excess kurtosis than the arch model, we can compare which is larger than the other, $\frac{6\alpha^2}{1-2a^2-(\alpha+\beta)^2)}$

can show A(1) is equal to $MA(\infty)$, same applies for GARCH for $ARCH(\infty)$

$\alpha + \beta$ providers the necessary information on the degree of volatility clustering

GARCH(p, q)

Just extension of GARCH(1, 1), key notation is polynomial for lag operator, lags shift an observation 1 period ahead (power 2 = 2 period ahead). But except for notation, nothing fundamental changes.

To lie outside of the root circle, in practice to estimate such a model, ensure positivity constraints, then also have to ensure process modelling is stationary - the constraints on stationary on highly non linear. This very quickly becomes a complicated non linear constraint, thus a numerical issue driven by Stationarity constraint (non linear) imposed by IRMA (p, q), but if allow for more p and q lags, then model is able to generate over kurtosis then the persistence of the series, the properties become better but at the cost of optimising over something with highly non-linear constraint.

Further Types of GARCH models

ARCH providers an exponential decay, have to know GARCHS for risk modelling.

Integrated GARCH(1, 1)

- specific to high frequency time series
- describes a very large persistence in the conditional variance
- is strictly stationary
- propose α and β sum upto 1, GARCH STRUCTURE there to ensure non stationary process
- risk metrics assumes that daily log returns follows process with infinite variance, that is we are not dealing with well defined statistical processes in real life, as seen by lack of first 2 moments

RisksMetrics^{TM} A special case of the IGARCH(1, 1) process

- From estimating the
- Gives forecast
- λ calibrates on loads of different stocks in the 90s
- Fix the β with λ

Exponential GARCH aimed at capturing asymmetric shocks, now modelling h_{t-1} log transformation of σ_t^2, assuming it follows GARCH looking process, and modify the ARCH part

- modelling logs of variance because we want to get rid of parameter constraints, if modelling logs can be positive, negative, get rid of these issues by modelling logs
-

Threshold GARCH TGARCH (1, 1) with indicator function, if shock was negative, bit easier to look at, if γ is positive, then ...

Tgarch, E garch if model left skewed

Tgarch(1,1) GJR-Garch

usual garch(1, 1) : $\sigma_t^2 = w + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$

tgarch(1, 1) : $\sigma_t^2 = w + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$

News impact curve : $NIC(\varepsilon_t|\sigma_{t-1}^2 = \sigma_{t-2}^2, \ldots, = \sigma_t^2)$

GARCH(1, 1) : $w + \beta\sigma_t^2 + \alpha\varepsilon_{t-1}^2$ TGARCH(1, 1) =

$$\begin{cases} w + \beta\sigma_t^2 + \alpha\varepsilon_{t-1}^2 \varepsilon_{t-1} < 0 \\ w + \beta\sigma_t^2 + \alpha + \delta\varepsilon_{t-1}^2, \varepsilon_{t-1} < 0 \end{cases}$$

NIC : Egarch(1,1)

$$h_t = \ln(\sigma_t^2) = w + \alpha\mathcal{L}_{t-1} + \gamma(|z_{t-1}| - \sqrt{\frac{2}{a}}) \exp(h_t) = \sigma_t^2 = \exp^w \cdot \exp^{\alpha z_{t-1}} \cdot \exp^{\gamma(|z_{t-1}| - \sqrt{\frac{2}{a}})} \sigma_t^2 = \exp^w \cdot \sigma^2$$

$$\varepsilon_t > 0$$
$$\varepsilon_t < 0$$

If shock positive then $\exp^{\alpha+\gamma} \cdot \varepsilon_t/\sigma_t$

NIC: once you write down NIC, then it becomes more evident what model parameters give you which response, EGarch $\alpha < 0$, z_t between 0 and 1

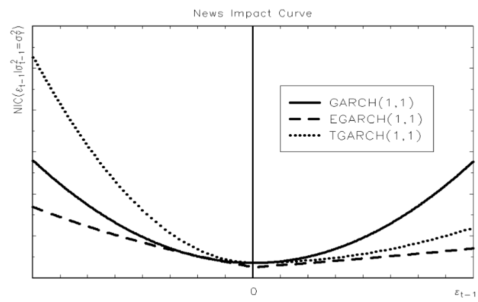

Figure 2: News Impact Curve

Model quality based on one picture, isn't exactly true, in order to plot NIC. plug in σ, γ, β (ω) , plot is based on one set of parameters, can easily be reversed.

So has something to do with data rather than overall quality of model,

Recap ARCH, GARCH, IGARCH, EGARCH, TGARCH. Financial econometrics model conditional second moment, but what about first moment?

- conditional mean? (1st moment), why are we interested in the second moment?
- We are risk averse etc, but

- in week 2 we have talked about how to model, ARMA - expected value of y_t then T2 we estimated conditional mean models, but the returns are on average 0, there is very slight autoregressive coefficients, but overall there is **no time series structure** in the conditional mean :

$$E[r_t|F_{t-1}] = 0$$

- WE have compared the ACF for daily log returns r_t, but in the actual return series, the history of returns is completely uninformative of the future
- in autocorrelation function few squared return we see a lot going on, and it doesn't die out, squared return is a proxy of conditional variance

Why do we model conditional second moment?

There is no time series structure to first moment, but there is in conditional second moment. Then we think how can we model our conditional variance of return process?

Nobel prize given for ARMA framework where ε_t can be white noise process. Then, even GARCH is not enough.

Then RiskMetrics comes and assumes infinite variance of daily returns, albeit a popular way of thinking. How much does turbulence persevere in market, how long after do we have to be conservative in our risk approaches

EGARCH, TGARCH. TGARCH more intuitive, EGARCH model the log variances and so can relax the positivity constraints, we don't care whether shocks are negative. Essentially a philosophical introduction to risk-modelling

$\mathbb{R}$- IS A REAL number?

Lecture 5: Model Estimation and Forecasting

Mon 26 Feb 08:58

0.2 Recap

Week 1 : Leverage effects (skewness + testing whether neg) , volatility clustering (time series) , long memory (ACF of squared returns series) , leptokurtic property (sample skewness testing against 3). Properties (plots/ test)

Week 2 : limitations of ARMA modelling, which assume innovations are white noise - nothing about conditional heteroskedacity). Unconditional - variance of innovations is constant over time, but evidence empirically that conditional 2nd moment seems to be time variant.

Week 3 : Rob engles ARCH ARCH(1) model $\sigma_t^2 = f(\{_{t-1})$. Pro - volatility clustering, con - leverage of $\{_t$, but long memory for very large p, kurtosis $\alpha^2 \in (0, \frac{1}{s})$

Week 4 : GARCH(1,1) - pro - volatility clustering and long memory and overkurtisis, con - leverage

TGARCH, EGARCH $\rightarrow$ leverage. M (IGARCH) .

Maximum Likelihood

Quasi Maximum Likelihood

Maximum likelihood - have data $x_1, \ldots, x_t$ then **assume** this data follows *some* distribution.

Which is function of the parameters, say $x_t \sim N(\mu, \sigma^2)$ and $\Theta = (\mu, \sigma^2)$

Then have PDF of data $f(x_t, \mu, \sigma^2) = -\frac{1}{\sqrt{2\pi\sigma^2}\exp(-\frac{(x-\mu)^2}{2\sigma})}$. If assume normal dist,then each and every value of x_t you know probability this data came from this distributing, then voter the entire sample you can take the likelihood function

$$\mathcal{L}|_{\mu,\sigma^2} = \prod_{t=1}^{T} f(x_t\mu, \sigma^2)$$

$$= f(x_1|\mu, \sigma^2) \cdot f(x_2|\mu, \sigma^2) \dots$$

So take log likelihood that is a function of data for given value of parameters μ, σ^2

$$\log \mathcal{L}(xq, \dots, x_t|\mu, \sigma^2) = \ln(\prod_{t=1}^{T} f(x_t), |\mu, \sigma^2)$$

In any time series we work with quasi likelihood, in classical ML you must be able to evaluate likelihood function at each and every point. At an autoregressive process of order 1 (AA(1)).

Have $\varepsilon_t \sim N(0, \sigma^2)$ so $\varepsilon_t = y_t - c - \phi y_{t-1}$ which us $N(0, 1)$

Then we have

Why quasi-likelihood?

Likelihood for first population : $f(e_1|c, \phi, \sigma^2)$, we assume y_0 is ...

Now likelihood function becomes function of data and parameters, but *also initial values* depending on how many autoregressive lags are there. $\rightarrow$ it is not really a likelihood. The conditioning makes it a quasi-likelihood

0.3 Estimation, Model choice and forecasting

Use knowledge of Max likelihood to ascertain which model fits the data best

Assume

$$r_t = c + \varepsilon_t$$
$$e_t = \mathcal{L}_t \sigma_t \sigma_t^2 = w + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

Estimate with Max likelihood.

$$\varepsilon_t = r_t - ce_t|_{\mathcal{F}_{t-1}} \sim N(0, \sigma_t^2)$$

$E[c_t|\mathcal{F}_{t-1}]$ and $V[\varepsilon_t|\mathcal{F}_{t-1}]$

where $\sigma_t = f(\mathcal{F}_{t-1})$ and $\varepsilon_t|_{F_{t-1}} = \mathcal{L}|F_{t-1}\sigma_t|\mathcal{F}_{\sqcup-\infty}$

where $f(\varepsilon|F_{t-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp(-\frac{e_t^2}{s\sigma_t^2})$

and

$$\theta = (c, w, \alpha, \rho)$$

$$= \frac{1}{\sqrt{2\pi(w + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2)}} \cdot \exp(-\frac{(v_t - c)^2}{2(w + \alpha\varepsilon_{t-1}^2 + \beta\sigma_t^2 - 1)})$$

$$\sigma_0^2 = \frac{w}{1 - \alpha - \beta} = \frac{1}{T} 2(r_t - \hat{\mu})^2$$

Normally distributed innovations. From likelihood theory, the best is the one with the largest likelihood.

Estimation of GARCH Models

Model : $Y_t = X_t'\gamma + \varepsilon_t$

The conditional variance of ε_t follows a GARCH(p, q) model

- m = max(p, q) numbers of initial observations $t = -m + 1, -m + 2, \dots, 0$

Conditional maximum likelihood

normal Z_t

Student t Z_t

Assume $z_t \sim T(v)$ (std student t dist) , then :

$$E[Z_t] = 0$$
$$V[Z_t] = \frac{v}{v-2}$$
$$\text{Density Function} \quad \frac{\Gamma[(\nu+1)/2]}{(\pi\nu)^{1/2}\Gamma[\nu/2]}\left[1+\frac{z_t^2}{\nu}\right]^{-(\nu+1)/2}$$

Often estimated using standardised student t distribution, which is symmetric so expected value is 0, and

In PS1, there was ex on student t distribution with different degrees of freedom - the larger the dof, the closer to normal RV, smaller the hevier the tails (more outliers). 1 dof - Cauchy distribution

0.4 Model Choice and Diagnostics

Verify if there are ARCH effects in

- the origunal series of intrest Y_t
- the residuals froma mean regression $\hat{\varepsilon}_t$ the residuals standardised by the estimated GARCHS $\hat{z}_t = \frac{\hat{\varepsilon_t}}{\sqrt{\hat{\sigma_t^2}}}$

Test for ARCH effects

ARCH-M test

Auxiliary regression on the series of interest $\overline{x}_t$ (original series, residuals, standardised residuals):

$$\overline{x}_t^2 = \psi + \alpha_1\overline{x}_{t-1}^2 + \alpha_2\overline{x}_{t-2}^2 + \ldots + \alpha_m\overline{x}_{t-m}^2 + \varepsilon_t$$

With $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$ and $H_A : H_0$ is not true

Standardised Residual Diagnostics

assuming you already estimate a GARCH model for series

Verify if there are still ARCH effects left in the series (if the estimated GARCH model is correctly specified) by performing standardised residual diagnostic tests on the residuals standardised by the estimated GARCH conditional volatility ($\hat{z}_t = \frac{\varepsilon_t}{\sqrt{\hat{\sigma^2}_t}}$)

				parameters					
Model	c	ω	α	β	γ	d	AIC	ARCH-LM test	JB test
ARCH(1)	7.94E-05	0.0002	0.2592				−5.388	71.985	29533.63
	(0.5838)	(0.0000)	(0.0000)					(0.0000)	(0.0000)
GARCH(1,1)	0.0004	2.41E-6	0.0603	0.9339			−5.543	3.948	15747.28
	(0.0019)	(0.0000)	(0.0000)	(0.0000)				(0.4130)	(0.0000)
Risk Metrics (RM)	0.0004			0.9615			−5.529	10.395	22530.63
	(0.0109)			(0.0000)				(0.0349)	(0.0000)
EGARCH(1,1)	0.0001	−0.1474	−0.0475	0.9920	0.1099		−5.565	5.998	8276.66
	(0.3297)	(0.0000)	(0.0000)	(0.0000)	(0.0000)			(0.1117)	(0.0000)
TGARCH(1,1)	0.0001	2.57E-6	0.0274	0.9343	0.0653		−5.557	2.6153	8865.977
	(0.2409)	(0.0000)	(0.0000)	(0.0000)	(0.0000)			(0.624)	(0.0000)
FIGARCH(0,d,1)	0.0003	1.87E-6		0.2898		0.370	-5.502	3.248	18833.38
	(0.0081)	(0.000)		(0.0000)		(0.000)		(0.4251)	(0.0000)

Note: The round parentheses give the p-values; The JB test is done on the standardized residuals.

Figure 3: Estimation of different GARCH Models

Arch(1) is capturing overkurtosis, since it is able to generate outliers (α is sig diff from 0). But intercept is not sig different from 0.

Arch-LM test and JB test are tested on ..., both tests are redirected, there is remaining heteroskedacity, α relatively mild.

GARCH - passing arch lm test, decay in ACF is very slow, α, β close to 1, very persistent, but able to measure conditional heteroskedacity

RM - re estimated on data, p val for ARCH lm is 0.04, depends on confidence interval determines rejection. But none are looking like norm RV

E(T) GARCH - neagtive shocks (response to future volaltity) γ positive. Egarch model log variances,

EGARCH - α - if shock negative then log of variance should be multiplied with negative variance (asymmetric response, how much is shock diffent from abs value of expected shock)

α and γ? negative and positive for egarch - at 5% sig level, all garchs seem to model sufficiently long memory using model parameters, out of these (ignoring fact dont past JB test of normality)

When we talked about ARMA we talked about AIC, BIC allowing us to compare different models estimated using ML, but different models have different parameters, so to control for this have different penalty functions (k denotes parameters).

Even asymmetric GARCH are unable to account for negative (β), we see in the data. Thus we require advanced financial econometrics

Garch loved since it is easy to forecast risk with them, central banks require risk forecasting on daily basis - using GARCH(1,1) is very easy for this.

Exercise 1. TGARCH(1,1) Estimated $\hat{\sigma_t^2} = \hat{w} + \hat{\alpha}\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \dots$

$$E[\sigma_{t-1}^2|\mathcal{F}_t] = \hat{w} + \alpha r_t^2 + \hat{\beta}\hat{\sigma_t^2} + \dots$$
$$E[\sigma_{t-1}^2|\mathcal{F}_t] = w + aE[\varepsilon_t^2|\mathcal{F}_t] + \beta E[\sigma_{t+1}^2|\mathcal{F}_t] + \dots$$

Expected value

$$w + \alpha E[\sigma_{t+1}^2|\mathcal{F}_t] + \beta E[\sigma_t^2|\mathcal{F}_t] + \gamma E[\pi(z_t)]$$

Forecasting with Risk Metrics

Let σ_t^2 follow a riskmetrics model:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)Y_{t-1}^2$$

where $\lambda = 0.94$

0.5 Variance Forecast Evaluation

σ^2 is not observed, it may be replaced by proxies such as

- $\sigma^2_{t+h} = r^2_{t+h}$ (squared daily returns)
- $\sigma^2_{t+h} = RV_{t+h}$ daily realised variance

Or alternatively, we evaluate the variance forecasts within economic applications :

- value at risk, expected shortcuts,
- asset pricing etc

good forecasting performance does not translate to good in sample fit (tradeoff?)