Financial Econometrics

Sol Yates

March 4, 2024

Contents

Lecture 5: Model Estimation and Forecasting

Mon 26 Feb 08:58

[FE-L5] [Reading]

0.1 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

$$\mathcal{L}|_{\mu,\sigma^2} = \prod_{t=1}^{T} f(x_t\mu, \sigma^2)$$
$$= f(x_1|\mu, \sigma^2) \cdot f(x_2|\mu, \sigma^2) \ldots$$

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

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

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

0.2 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.

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

$E[c_t|\mathcal{F}_{t-1}]$ 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 $t = -m + 1, -m + 2, \ldots, 0$

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

0.3 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 1: 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 $\lambda = 0.94$

0.4 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?)

Tutorial 1. 5 Last week simulated GARCH, this week estimating GARCH and forecasting based on the estimates. In PS4 we have simulated $y_t = c + \psi y_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t$ Chose some parameters, then simulated based upon those parameters, and once we had innovations, we simulated for values of ARMA parameters, simulated the ARMA recursions This week have the daily log returns of SNP500, estimate ARMA and GARCH parameters which are coming from data, why cant we just plug these parameters in and use them in the simulation? Taking our $\tilde{\sigma^2}$ and simulate returns, why cant we do this and why instead do we forecast where $\hat{\sigma^2_t} = \hat{w} + \gamma r^2_{t-1} + \hat{\beta}\sigma^2_{t-1}$ Simulated series which resemble data properties is defined as $E[\sigma^2_{t+1}|\mathcal{F}_t]$

Lecture 6: Kalman Filter

Mon 04 Mar 09:04

$$E[\sigma^2_{t+1}|\mathcal{F}_t] = w + \alpha\varepsilon^2_t + \beta\sigma^2_t$$

If we think about the classical ARMA-GARCH framework, we have

1. returns with some conditional mean $+\varepsilon_t$ where

$$\begin{aligned} r_t &= E[r_t|\mathcal{F}_{t-1}] + \varepsilon_t \\ \varepsilon_t &= \mathcal{L}_t \cdot \sigma_t(\mathcal{F}_{t-1}) \\ \mathcal{L}_t &\sim \mathcal{N}(0,1) \end{aligned}$$

If we would like to assume that returns is driven by

$$\begin{aligned} r_t &= \varepsilon_t = \mathcal{L}_t\sigma_t \\ \sigma^2_t &= f(\mathcal{F}_{t-1}) + q_t \\ \varepsilon_t|\mathcal{F}_{t-1} &\sim \mathcal{N}(0,1) \\ f(\varepsilon_t) &= \frac{1}{\sqrt{2\pi\sigma^2_t}}\exp\left(-\frac{\varepsilon^2_t}{2\sigma^2_t}\right) \end{aligned}$$

Multivariate normal distribution

$$\begin{bmatrix} r \\ y \end{bmatrix} \approx N\left(\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}\begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{pmatrix}\right)$$

We need the $x|y$ distribution to derive the Kalman filter, the transformation can get us certain properties

Expected value (these are population parameters, fixed values, numbers measuring fixed variance)

$$\begin{aligned}
E[z] &= E[x] - \Sigma_{xy}\Sigma_{yy}^{-1}(E[y] - \mu_y) \\
&= E[Z'Z] \quad \text{if scalar, then} \quad E[z^2] \\
&= E\left[\left(x - \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)\right)\left(x - \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)\right)\right] \\
&\quad \text{algebra} \ldots \\
\operatorname{cov}(x,y) &= E[xy] - E[x]E[y] \\
&\to u_x u_y - E[xy'] = -\operatorname{cov}(x,y) = -\Sigma_{xy} \\
&\quad \text{and so the whole term} \\
&= E[X'X] - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}' \\
\operatorname{cov}(z,z) &= E[ZZ'] - E[Z]E[Z]' \\
&= E[xx']
\end{aligned}$$

We have 2 RV with joint distribution, we want to understand the conditional distribution

$$X|y \approx N\left(\ldots\right)$$

Then if we take the transformation z

$$z = x - \Sigma_{xy}$$

1. $E[z] = E[x]$
2. $V[z] = \Sigma_{xx} - \Sigma_{yy}^{-1}\Sigma_{xy}$
3. $\operatorname{Cov}(y,z) = 0$

$x = z + \Sigma_{xy} + \Sigma_{yy}^{-1}\left(y - \mu y\right)$
$E[x|y] = E[z|y] + \Sigma_{xy}\Sigma_{yy}^{-1}(E[y|y] - \mu_y)$
First term $= E[z]$, second term is same third is y so

$$E[x|y] = \mu_x = \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$$

$V[x|y]$

$$x|y \approx N\left(\mu_x\right)$$

Local Trend Model

In order to understand the SV model, consider a simple local trend model first

$$y_t = \mu_t = e_t \quad e_t \sim \mathcal{N}(0, \sigma_e^2)\mu_{t+1} = \mu_t + n_t$$

Idea is that if you have state equation with large variance, you wont be able to recover much.

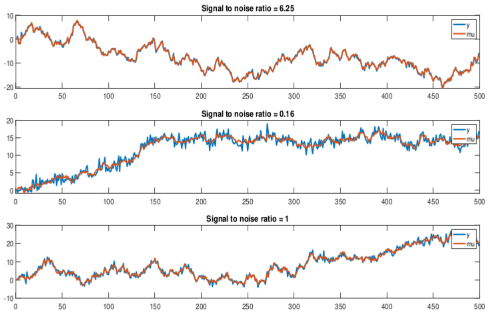

Figure 2: State Space Model

If signal to noise ratio = 0.16, observe blue try to recover red. It is not very informative, if ratio is 6 then signal is very informative

Kalman Filter

Three estimates of interest

1. Filtering : recover μ_t (remove measurement error)
2. Smoothing : estimate μ_t given all available information up to time T
3. Prediction : forecast μ_{t+k}

Properties of Multivariate Normal Distribution Considering a multivariate normal distribution

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{pmatrix}\right)$$

Kalman filter is tool which characterises the conditional distribution of μ_t given the data. Given the date we observe what is the distribution of μ_t

then the conditional distribution of x given y is

$$x|y \sim\sim \mathcal{N}(\mu_x + \Sigma_x y \Sigma_y y^{-1}(y - \mu_y), \Sigma-)$$

$$\begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} |_{\mathcal{F}_{t-1}}$$

Goal is the conditional distribution $\mu_t|F_t$ based on new data y_t and the conditional distribution $\mu_t|F_{t-1}$

$$\begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{t|t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{t|t-1} & \Sigma_{t|t-1} \\ \Sigma_{t|t-1} & V_t \end{pmatrix}\right)$$

Prediction

Initialise idea on conditional mean and variance of unobserved latent state variable (signals), take data minus initial value/expectation.

Look at forecast error variance

How does it work?

First remove measurement error then estimate μ_t given all available information, then forecast. Distribution of μ_t given information set $\mathcal{F}_t$ today. In order to today recover the value of μ_t need to update conditional expectation so that take into account signal to noise ratio. How much new noise contributes to the conditional variance expectation.

Filter latent process based on information $t-1$ then we update forecast once new information has arrived. Kalman game measures how much information does the new shock at time t add to uncertainty (?). Dent take information as given y_t has noise itself e so we only update conditional expectations proportionally to the signal to noise ratio.

Recover, then smoothing re estimating μ_t (trying to mitigate effect of starting values), then based on this forecast latent process. All based on one property of MVR norm.

After we know this we can write it down given this formula