Stochastic Volatility
From GARCH to the SV Model

Boston University Questrom School of Business

MF840 – Spring 2021

Eric Jacquier

TABLE OF CONTENTS

1. ARCH Family of Models

1.1 Limitation of the Lognormal distribution

Stylized facts on asset returns returns的问题

- Fat tails 肥尾

- Clustering: "Periods of high variance tend to follow periods of high variance." Mandelbrot. Volatility 聚集

- Leverage effect: Higher volatility when Returns are (more) negative μ negative, $\sigma \uparrow$

- Non uniform information arrival - French&Roll (JFE 1986) 信息不对称

- Not so stylized fact:
  Smile in BS-ISD vs. Ln(S / PV(X)) 波动率微笑
  Time varying volatility as in Hull & White can also cause this.

- Recall: Normal mixture with random σ is fat-tailed (Student-t if $\sigma \sim I_tG$)

1.2 Basic GARCH model

GARCH模型是说，今天的方差不仅依赖于过去的收益率的变化 [公式]，还依赖于过去的方差 [公式]，而ARCH只说这个方差依赖于过去收益率的变化 [公式]。

- GARCH(p,q)

$\mu_t + \sqrt{h_t}\,\varepsilon_t$

$$R_t = \mu_t + u_t = \mu_t + \text{sqrt}(h_t)\,\varepsilon_t, \quad u_t \sim N(0,h_t)$$

$$h_t = \omega + \alpha_1 u^2_{t-1} + \alpha_2 u^2_{t-2} + .. + \alpha_q u^2_{t-q} + \beta_1 h_{t-1} + \beta_2 h_{t-2} + .. + \beta_p h_{t-p}$$

1.3 Estimation, forecasting, diagnostic and forecasting

Estimation by maximum likelihood or Bayesian MCMC methods

- Given the parameters and max(p,q) initial values, volatility is observable
- Constrained optimization needed: $\alpha>0$, $\beta>0$, $\Sigma(\alpha+\beta) < 1$ If not, no sense
- Likelihood can be flat for small samples 小样本不大靠谱

Diagnostics

$R_t = \mu_t + u_t$, $u_t \sim N(0,\sigma_t)$

$h_t = \omega + \alpha_1 u^2_{t-1} + \alpha_2 u^2_{t-2} +..+ \alpha_q u^2_{t-q} + \beta_1 h_{t-1} + \beta_2 h_{t-2} +..+ \beta_p h_{t-p}$

Estimation $\quad => \hat{\alpha}, \hat{\beta}, \hat{h}_t, t = 1,\ldots,T$

$=> \hat{u}_t, t = 1,\ldots,T$

If model is correct

$=> \hat{u}_t / \sqrt{\hat{h}_t} \sim$ i.i.d. $N(0,1)$

⇨ test normality

⇨ test ACF

GARCH(1,1) Rarely need more than 1 lag of each

1.4 Crucial extension: Fat-Tailed conditional returns

- For financial returns: $\hat{u}_t / \sqrt{\hat{h}_t}$ is $\boxed{\text{not normal}}$ after GARCH modeling:
- That is, in $R_t = \mu_t + \sqrt{h_t}\,\varepsilon_t$, ε_t is **not** normally distributed
- ε_t is fat-tail => shortfall probabilities, VAR are incorrect:

  i.e., $\text{Prob}[R_{t+1} < \mu_{t+1} - 1.65\sqrt{h_{t+1}}] = 0.05$

Instead: $\varepsilon_t \sim \text{Student}(\nu)$

(normal)

$\varepsilon_t \sim \text{GED}(\nu)$ Dan Nelson's EGARCH generalized error dist

$$\text{GED: } f(\varepsilon_t) = \frac{\nu \exp\{-0.5|\varepsilon/\lambda|^{\nu}\}}{\lambda 2^{[\nu+1/2]}\Gamma(1/\nu)},$$

关于 ε_t 了

$$with\,\lambda = \sqrt{2^{-2/\nu}\Gamma(1/\nu)/\Gamma(3/\nu)}$$

$\nu = 2;$ GED ~ N(0,1)

$\nu < 2;$ Fat-tails

EGARCH是从GARCH衍生出的模型，是为了解释“杠杆效应”。
先说一下“杆杆效应”，经验性的分析表明，金融资产收益率的涨跌这个定性结果对未来波动性的影响是不同的。注意，这个现象GARCH(ARCH)模型是不能解释的。因为在GARCH(ARCH)中，历史数据是以平方的形式影响未来波动率的，所以涨或跌对未来波动率的影响是一样的。
为了可以解释“杠杆效应”，必须舍弃平方这个对称函数，历史数据应该通过一个“非对称函数”影响未来波动率，EGARCH正是这样处理历史数据的。那么新的问题来了，非对称函数可能会打破GARCH模型的“合理性”，简单地讲，未来波动率可能出现“负值”。解决方法也简单，用波动率的对数代替波动率，合理性就得到了保证。
总之，杠杆效应是理解EGARCH的关键词，蔡老师的《金融时间序列分析》有一节专门讲EGARCH，阁下可以仔细研读一下。

- ν must be estimated for best fit.

Conditional dist is non-normal.

- Results: Strong evidence of non normality of the conditional distributions

- Normal based GARCH standard errors generally understate standard errors if ε_t is fat-tailed.

Fat-tail condition Garch
考虑到了肥尾

1.5 Crucial Extension: Asymmetric volatility

EGARCH : $R_t = \mu_t + u_t$, $u_t \sim N(0,h_t)$

$$\ln h_t = \omega + \alpha_1 \left(\underbrace{\varphi\, \varepsilon_{t-1} + \gamma \left[\, |\varepsilon_{t-1}| - E|\varepsilon_{t-1}| \,\right]}_{g(\varepsilon)} \right) + \beta_1 \ln h_{t-1}$$

⇒ 条件均值为0

No restriction on parameters for $h_t>0$

所有参数均可为+或-，因为是log，e回去就都为正了.

Volatility cycles are possible

If $\alpha_i \varphi<0$, h_t rises after a <0 shock more than after a positive shock.

Recall: The leverage effect is a misnomer, better term: Volatility Feedback

: 市场对负面消息的影响更大

Assymetric GARCH (Glosten, Jagannathan, Runkle GJR): 也考虑到了正负收益的不对称性

$$R_t = \mu_t + u_t, \quad u_t \sim N(0, h_t)$$

$$h_t = \omega + \alpha_I D_{t-1} \varepsilon^2_{t-1} + \alpha_2 \varepsilon^2_{t-1} + \beta_1 \ln h_{t-1} \text{ where } D_t = 0 \text{ if } \varepsilon_t < 0$$

more shock on negative return

1.6 Multivariate Factor GARCH

Best first pass at a multivariate GARCH model

Start with a classic K factor model.

Allow the covariance matrix of factors to vary with time.

$$\underset{n\times 1}{R_t} = \underset{n\times 1}{\mu} + \underset{n\times k}{B}\, \underset{k\times 1}{f_t} + \underset{n\times 1}{v_t}$$

k factors << n stocks

Factors $f_{1t}, \ldots, f_{Kt}$:

Observable,
Or Unobservables (Principal Components, Gibbs Sampling) 见 EJ. 13

$E(f_{kt}\ f_{lt}) = 0$ $E(F_k' F_l) = 0$ Diagonal Factor Covariance Matrix

$E_t(f^2_{k,t+1}) = h_{k,t+1} = \omega_k + \beta_k h_{k,t} + \alpha_k (f_{k,t})^2$ GARCH on factor volatility

$E_t(v_{t+1}v^T_{t+1}) = D_{t+1}$ If needed GARCH on diagonal idiosyncratic covariance matrix

$E[\varepsilon_{t+1}\varepsilon'] = D_{t+1}$

$\Omega_{t+1} = E_t(u_{t+1}, u^T_{t+1}) = B\ E_t(f_{t+1}f^T_{t+1})\ B^T + D_{t+1}$

2 Stochastic Volatility Model (SV)

2.1 Difference with GARCH

- SVOL(1): $R_t = \text{sqrt}(h_t)\ e_t$,
  $\log h_t = \alpha + \delta \log h_{t-1} + v_t$, $v_t \sim N(0,\sigma_v)$

  Why logarithm? It is a discrete time model

- v_t adds variability to h_t
  induces more kurtosis than the ARCH.

- GARCH: kurtosis is linked to persistence via α, β. Can't model separately 不一样
  SVOL: kurtosis is modeled separately via σ_v

Why does GARCH work relatively well?

Many GARCH models, even if misspecified, have some good filtering properties for the true continuous time model: asymptotically consistent estimation of σ_t 因为渐进

- ARCH filter is asymptotically more efficient than the Kalman filter often used for estimating SV models!

SV model

2.2 Estimation

$R_t \quad = h_t^{0.5}\, e_t$,

$\text{Log } h_t = \alpha + \delta \log h_{t-1} + v_t$, $v_t \sim N(0,\sigma_v)$

Parameters: $\theta = (\alpha, \delta, \sigma_v)$

- Maximum Likelihood:

$$L(\theta \mid R) = p(R \mid \alpha, \delta, \sigma_V) \;\; ??$$

$$= \int \ldots \int p(R \mid h, \alpha, \delta, \sigma) \;\; p(h \mid \alpha, \delta, \sigma)\, dh$$

Very difficult numerically,

- Bayesian: $p(\theta \mid R) \sim p(R \mid \theta)\, p(\theta)$ Not easier !

- Trick: Data Augmentation: $p(\theta, h \mid R)$

Consider the hs as parameters to estimate

- Bayes Theorem again: $p(\theta, \mathbf{h} \mid R) \sim p(R \mid \theta, \mathbf{h}) \quad p(\theta, \mathbf{h})$
  $p(\mathbf{h} \mid \theta)\, p(\theta)$

$$P(R \mid \theta, h)\, \underbrace{P(h \mid \theta)\, P(\theta)}_{P(\theta, h)}$$

- Gibbs Cycle:

[1] $p(\overset{\theta}{\alpha, \delta, \sigma} \mid \overset{h}{\mathbf{h}}, R)$

[2] $p(\underset{h}{\mathbf{h}} \mid \underset{\theta}{\alpha, \delta, \sigma}, R)$

[1] a linear regression Easy

[2] Smoothing: $p(\mathbf{h} \mid R, .)$ A T-dimensional vector ??

Trick 2: break it further: $p(h_t \mid h_{-t}, R, \theta)$ 去掉 h_t

Overall, a T+3 dimensional Gibbs cycle

- Simplification: $p(h_t \mid h_{-t}, R, \theta) = p(h_t \mid h_{t-1}, h_{t+1}, R, \theta)$

- A bit of Bayes theorem to find $p(h_t \mid .)$:

$$p(h_t \mid \theta, r_t, h_{t-1}, h_{t+1})$$

$$\propto p(r_t \mid \theta, h_t, h_{t-1}, h_{t+1})\, p(h_t \mid \theta, h_{t-1}, h_{t+1})$$

$$\propto p(r_t \mid \theta, h_t) \qquad p(h_{t+1} \mid \theta, h_t, h_{t-1})\, p(h_t \mid \theta, h_{t-1})$$

$$\propto p(r_t \mid \theta, h_t) \quad p(h_{t+1} \mid \theta, h_t) \quad p(h_t \mid \theta, h_{t-1})$$

- We have a Normal (r_t) and two Lognormals (h_{t+1} and h_t).

$$p(h_t \mid \theta, R, h_{-t}) \propto p(h_t \mid \theta, r_t, h_{t-1}, h_{t+1})$$

$$\propto \underbrace{p(h_t \mid \theta, h_{t-1})}_{LN} \; \underbrace{p(h_{t+1} \mid \theta, h_t)}_{LN} \; \underbrace{p(r_t \mid h_t)}_{I_s G}$$

$$\propto \frac{1}{h_t} \exp\left\{ -\frac{(\log h_t - \mu_t)^2}{2\sigma^2} \right\} \frac{1}{h_t^{1/2}} \exp\left\{ -\frac{r_t^2}{2h_t} \right\}$$

$$\begin{cases} \sigma^2 \equiv \sigma^2(\theta) \\ \mu_t \equiv \mu_t(h_{t-1}, h_{t+1}, \theta) \end{cases}$$

$$\mu_t = (\alpha(1-\delta) + \delta(\log h_{t+1} + \log h_{t-1}))/(1+\delta^2)$$
$$\sigma^2 = \sigma_v^2/(1+\delta^2).$$

- How to draw from $p(h_t \mid ..)$? Metropolis

  Independence: Choose $q(h_t)$ as Inverse Gamma (fatter tail than Log-normal)

  Random walk: No need to approximate $p^*(h_t)$

选$q(h_t)$为IG，比log肥尾.

2.3 Needed Extensions to the SV model

2.3.1 Asymmetric volatility

$R_t = \mu + h_t^{0.5} \varepsilon_t,$ $\varepsilon_t \sim N(0,1)$
$\log h_t = \alpha + \delta \log h_{t-1} + v_t$, $v_t \sim N(0,\sigma_v)$

$\rho(v_t, e_t) < 0$ as in Hull & White or Heston.

Result: Needed for stock indices and interest rates
Not so crucial for most exchange rates and individual stocks

2.3.2 Conditional Fat tails

$R_t = h_t^{0.5} \lambda_t^{0.5} \varepsilon_t,$ $\varepsilon_t \sim N(0,1),$ $\nu / \lambda_t^{0.5} \sim \chi^2(\nu)$
$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t,$ $v_t \sim N(0,1)$

Algorithm => $p(\lambda_t | R, h_t)$, $p(h_t | R)$, $p(\alpha, \delta, \sigma_v, \nu | R)$

Results:
- ν higher than for GARCH 比garch的v高
- Still crucial around periods of high variance to control outliers
- Omitting this extension leads to biased forecasts..