Boston University Questrom School of Business

MF793 – Fall 2020
Eric Jacquier

Introduction to GARCH Models

AR C H: AutoRegressive Conditional Heteroskedasticity
G ARCH: Generalized ARCH

1. Limitation of the log-normal distribution
2. Simple extensions do not work
3. Basic ARCH
4. GARCH
5. Testing, Estimating, Diagnostic,
6. Forecasting
7. *Crucial Extensions*
8. *Contrast with SV (stochastic volatility) models*

1 Limitation of the Lognormal distribution: Four stylized facts on asset returns

1. Fat tails in the unconditional distribution of asset returns

Even it the conditional daily distribution of returns is normal with a different variance every day, the unconditional distribution is not normal:

A mixture distribution $p(R_t) = \int N(R_t \,|\mu\,, \sigma_t)\; p(\sigma_t)\, d\sigma_t$

One reason the **un**conditional distribution of financial returns is not normal is that the distribution of returns is a **mixture** of (possibly normal) conditional distributions.

2. Volatility Clustering:

"Periods of high variance tend to follow periods of high variance. (low … low)"

Need to look at a time series plot of a measure of volatility … Realized volatility, VIX
Careful not to misinterpret the spurious autocorrelation due to overlapping data.

3. *Leverage effect* aka Volatility feedback effect

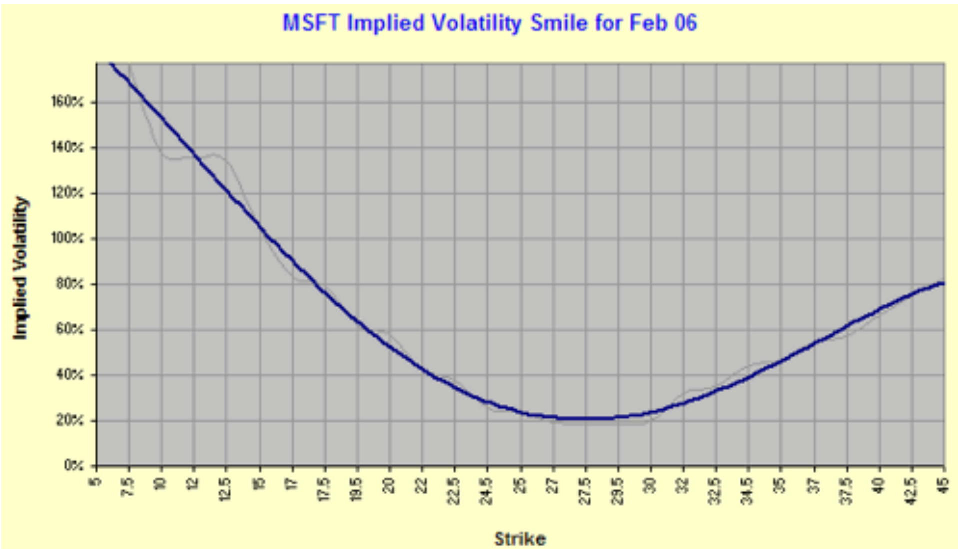
Need to plot a measure of volatility vs. returns
Do we see more volatility after negative returns ?

4. Non uniform information arrival – open / close – week-ends – holidays
French & Roll (JFE 1986)

Daily close to close:	9:30-16:00	7.5 hrs open	16.5 hrs closed
Friday close to Monday close:		7.5 hrs open	64.6 hrs closed

5. and .. a fifth not so stylized fact:
Smile in BSM implied volatility vs. Ln(S / PV(X))

Theoretically consistent with time-varying volatility as in Hull&White, Heston&Nandi.
also consistent with other departures from BSM, e.g., illiquidity



As of Jan 2006:
MSFT = 26$

2. Simple extensions like the CEV Model do not help much

- $dS / S = \mu\, dt + \sigma\, S^{\alpha}\, dz$, $-1 < \alpha < 0$

- $\alpha=0$ homoskedastic returns

  $\alpha<0$ If S ↑ variance ↓, contains Black (1976)'s inverse relationship, meant to capture a "leverage effect"

- Tests: Regression (Stan Beckers, J. Finance June 1980)

  $\text{Log}(\sigma(S_{t+dt}/S_t)) = \alpha_0 + \alpha \ln S_t + u_t$

  ?

  Used a block sampling estimator: intra-day returns to compute daily variance

  For 47 stocks: Average R^2: 3.1% , max R^2: 9.5% , $\alpha<0$ for 38 stocks

  => Specification still not flexible enough.

3 Basic ARCH

- ARCH(1) model $r_t = \mu + u_t = \mu + \sqrt{h_t}\,\varepsilon_t$, $\varepsilon_t \sim$ i.i.d. $N(0, 1)$

$$h_t = \omega + \alpha_1 u^2_{t-1}, \quad \alpha_1, \omega > 0$$

- ARCH is a **filter**, it writes ***unobserved*** volatility as function of a *past* ***observable*** u^2_{t-1}
  Contrast with stochastic volatility: $\log h_t = \alpha_0 + \alpha_1 \log h_{t-1} + \sigma_v v_t, \; v_t \sim N(0,1)$

- ARCH(1) implies the following process on the squared innovation u_t^2.

$$\mathbf{u^2_t} = h_t + u^2_t - h_t = \omega + \alpha_1 u^2_{t-1} + h_t \varepsilon^2_t - h_t$$

$$\mathbf{u^2_t = \omega + \alpha_1 u^2_{t-1} + h_t(\varepsilon^2_t - 1)} \qquad \textbf{[1]}$$

Looks like an AR(1) in u^2_t

- $\nu_t = h_t(\varepsilon^2_t - 1)$ is a noise:

$$\mathbf{E(\nu_t)} = E(h_t)\,E(\varepsilon^2_t - 1) = E(h_t)(1-1) \mathbf{= 0}$$

Can easily show: $\mathbf{Cov(\nu_t, \nu_{t-k}) = 0}$

ARCH(1) implies that Squared return noise u^2_t follows an AR(1)

- $\mathrm{Corr}(u^2_t, u^2_{t-s}) = \alpha_1^s$ by AR(1) property

- AR(1) result in u^2_t: $E(u^2_t) = V(u_t) = \omega / (1-\alpha_1)$ Stationarity: need $|\alpha_1| < 1$

  The unconditional variance of r_t.

- Kurtosis: Can show: $\mathrm{Kurt}(u_t) = 3\,(1-\alpha^2_1) / (1-3\alpha^2_1)$ exists only if $3\alpha^2_1 < 1$

Kurtosis always > 3 => fat tailed unconditional returns

- It was quickly recognized that one needed **more than one lag** of u^2_t: ARCH(q)

- ARCH(q) was not very convenient with q large, too many parameters to estimate

  A generalized model with very few lags did the trick ..

4 The GARCH model

- GARCH(p,q): $r_t = \mu + u_t = \mu + \sqrt{h_t}\,\varepsilon_t$, $\varepsilon_t \sim$ i.i.d. $N(0, 1)$

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

The ARCH(q) part — And now, lags in h_t

- A **GARCH(1,1)** is often sufficient to explain the lag-structure of volatility.

$$\mathbf{h_t = \omega + \alpha u^2_{t-1} + \beta h_{t-1}}$$

- Need $\omega, \alpha, \beta > 0$ for the model to imply positive variance (otherwise makes no sense !)

- Model for u_t^2:

$$\begin{aligned} u^2_t &= h_t && + u^2_t - h_t \\ &= \omega + \alpha u^2_{t-1} + \beta h_{t-1} && + \nu_t \\ &= \omega + \alpha u^2_{t-1} + \beta(h_{t-1} - u^2_{t-1} + u^2_{t-1}) && + \nu_t \\ &= \omega + \alpha u^2_{t-1} - \beta \nu_{t-1} + \beta u^2_{t-1} && + \nu_t \end{aligned}$$

$$\mathbf{u^2_t = \omega + (\alpha + \beta) u^2_{t-1} + \nu_t - \beta \nu_{t-1}} \quad \text{ARMA(1,1) in squared noise} \qquad \textbf{[2]}$$

- Stationarity of u^2_t: u^2_t is stationary if $\sum(\alpha_i + \beta_i) < 1$.

- The **integrated I**GARCH(1,1) : $\mathbf{h_t = \omega + \alpha u^2_{t-1} + (1\text{-}\alpha) h_{t-1}}$ .. is not stationary

5 Testing, Estimating, Diagnostic

- **Testing:** The ACF of the squared returns !

- **Estimation** done by numerical maximum likelihood, no analytical results available

  => $\widehat{\alpha}$, $\widehat{h}_t$, $t = 2, \dots, T$ Why t=2 ?

- **Diagnostic:** To be successful, ARCH should explain the non-normality of unconditional returns

  If $R_t \mid h_t \sim N(\mu, h_t)$ => $\frac{R_t - \widehat{\mu}}{\sqrt{\widehat{h}_t}} \sim i.i.d\ N(0, 1)$ **[3]**

  Diagnostic: Check the normality of the standardized residuals as in [3]

  Note the asymptotics: h_t supposedly estimated in large sample, Student-t assumed approximately normal.

- Empirical Results: GARCH models
  1) **explain a lot of, but not all** the fat-tailness of financial returns.
  2) have been quite successful at forecasting variance in the short to medium term.

6 Forecasting: why it matters for risk management

- VaR before time-varying volatility … used the unconditional variance:

$$\text{Prob}[R_{t+1} < \mu - 1.65\,\sigma] = 0.05 \qquad \text{Naive VaR} \qquad [4]$$

Couls be ... very ... very **wrong:**

Period of high volatility: VaR is optimistic

Period of low volatility: VaR is pessimistic

- Volatility varies with time:

  One must use the relevant conditional distribution to make conditional forecasts

  Conditioning: information I at time t about t+1

  $R_{t+1} \mid I_t \sim N(\mu, \mathbf{h_{t+1}})$

- VaR with time-varying volatility ARCH(1)

  $$h_{t+1} = \omega + \alpha\, u_t^2 + \beta\, h_t$$

  $$\text{Prob}[R_{t+1} < \boldsymbol{\mu - 1.65\sqrt{h_{t+1}}}] = 0.05 \qquad \text{GARCH based VaR} \qquad [4']$$

- GARCH and other filters for time-varying volatility allow the use of the relevant **conditional distribution** of stock returns for **conditional volatility forecasts**.

- **Multiperiod** (multi step ahead) **forecast** borrows from stationary AR method

ARCH(1) $h_{t+1} = \omega + \alpha\, u^2_t$ t+1 known

$h_{t+2} = \omega + \alpha\, u^2_{t+1}$

$E_t(h_{t+2}) = \omega + \alpha\, E_t(u^2_{t+1}) = \omega + \alpha\, h_{t+1}$

$= \omega + \alpha\, (\omega + \alpha\, u^2_t) = \omega\, (1+\alpha) + \alpha^2\, u^2_t$

Or :

$E_t(h_{t+k}) = E_t(u^2_{t+k})$ recall $E_t(u^2_{t+k}) = E_t(\varepsilon^2_t)\, E_t(h^2_{t+k})$

$\mathbf{E_t(h_{t+k}) = \omega\ (1+ \alpha + \ldots \alpha^{k-1}) + \alpha^k\, u^2_t}$ per stationary AR(1) in u^2_t in [1]

GARCH(1,1) $E_t(h_{t+k}) = E_t(u^2_{t+k}) = \mathbf{\omega\ [1+ (\alpha+\beta) + \ldots (\alpha+\beta)^{k-1}] + (\alpha+\beta)^k\, u^2_t}$

- Compare with Risk Metrics:

Using RM to forecast variance for tomorrow $s^2_{t+1} = \lambda\, s^2_t + u^2_{t+1}$? We don't know u_{t+1}

$s^2_{t+1} = \lambda\, s^2_t + (1-\lambda)\, E_t(u^2_{t+1}) = \lambda\, s^2_t + (1-\lambda)\, s^2_t$

$s^2_{t+1} = \lambda\, s^2_t + (1-\lambda)\, s^2_t$

=> $\mathbf{s^2_{t+1} = s^2_t}$ $\mathbf{s^2_{t+k} = s^2_t}$

Risk–Metrics forecasts are similar to a random walk with no drift
RM can not recognize a term-structure of volatility

RM is similar to an IGARCH with no intercept

7 Crucial extensions

- Model : $R_t = \mu + u_t$, $u_t \sim N(0, \sqrt{h_t})$

  $h_t = \omega + \alpha_1 u^2_{t-1} + \beta_1 h_{t-1}$ a GARCH(1,1)

- Estimate => $\hat{\mu},\ \widehat{\omega},\ \widehat{\alpha}\ , \widehat{\beta},\ \widehat{h_t},\ t = 2, \ldots, T$ => $\widehat{u_t}$ $t = 2, \ldots T$

- If model is correct: $\hat{u}_t / \sqrt{\hat{h}_t} \sim$ i.i.d. N(0,1) the GARCH ***standardized residuals***

- We test that there is no autocorrelation left in squares or absolute values of residuals
  One rarely needs more than 1 lag of each GARCH(1,1)

- We test normality of the "standardized residuals" of the GARCH model
  Residuals are more normal that u_t's but most often not quite normal yet.

7.1 Crucial Extension 1: Fat-tailed conditional returns

- For financial returns: $\hat{u}_t / \sqrt{\hat{h}_t}$ still has fat tails after GARCH modeling:

- That is, in $R_t = \mu_t + \sqrt{h}_t \varepsilon_t$, ε_t is **not** normally distributed, it has fat tails.

- ε_t is fat-tail => shortfall probabilities, VAR are incorrect:

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

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

  $\text{Prob}[R_{t+1} < \mu_{t+1} - 1.81\sqrt{h}_{t+1}] = 0.05$ for $\nu = 10$

- ν must be estimated for best fit.

- Results: Evidence of non normality of the conditional distributions, ν found to be below 10 or 20.

- Normal-based GARCH VaRs and confidence intervals are optimistic since ε_t has fat tails.

7.2 Crucial Extension 2: Asymmetric volatility

- EGARCH : $R_t = \mu + u_t$, $u_t = \sqrt{h}_t\, \varepsilon_t$, $\varepsilon_t \sim N(0,1)$ (Dan Nelson)

$$\log h_t = \omega + \boldsymbol{\varphi\, \varepsilon_{t-1}} + \alpha\ [\ \varepsilon^2_{t-1}\ -\ E(\varepsilon^2_{t-1})\ ]\ )\ +\ \beta \log h_{t-1}$$

Parameters to estimate: ω, α,β, φ,

Why log formulation ?

No restriction needed on parameters to enforce $h_t > 0$

Volatility cycles are possible with GARCH(1,2) and < 0 parameters

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

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

- **Asymmetric GARCH** (Glosten, Jagannathan, Runkle GJR):

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

$$h_t = \omega + \boldsymbol{\alpha_1 D_{t-1} u^2_{t-1}} + \alpha_2 u^2_{t-1} + \beta_1 h_{t-1} \quad \text{where } D_t = 0 \text{ if } \varepsilon_t < 0$$

8 Stochastic Volatility Model (SV), contrast with GARCH

SV(1): $R_t = \sqrt{h_t}\varepsilon_t,$

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

- Why logarithm? Discrete time model with unbounded noise, need to force positivity of h_t
- SV: close to continuous time models used in option and asset pricing (Hull & White, Heston)
- SV: More flexible than GARCH: v_t is an **un**observable noise, it adds variability to h_t
  Can induce more kurtosis than GARCH.
  GARCH kurtosis and persistence are linked via α, β. They can't be modeled separately
  SV: kurtosis is modeled via σ_v, persistence via δ.
  Less need for conditional fat-tail return ε_t
- SV: Variance remains an unobservable state variable
  More difficult to estimate than GARCH, requires Bayesian methods
- **Why does GARCH work relatively well ?**

Nelson & Foster show that GARCH filters, even if mis-specified, have good filtering properties for the true continuous time model: asymptotically consistent estimation of σ_t

Among other things, Nelson shows that GARCH filters are asymptotically (in continuous time) more efficient than the Kalman filter