Financial Econometrics

Volatility

Gerald P. Dwyer

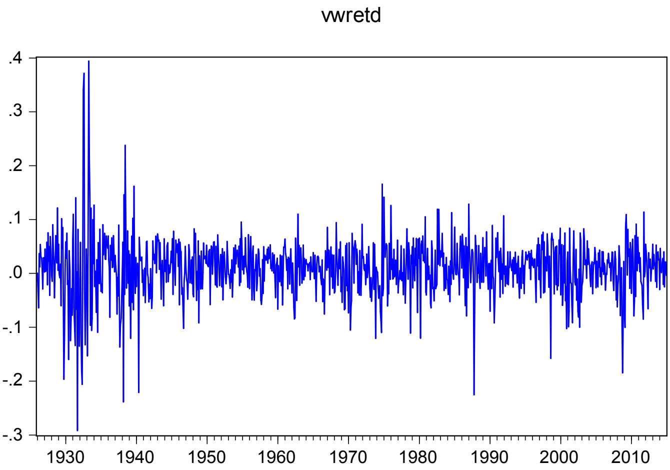
Trinity College, Dublin

January 2016

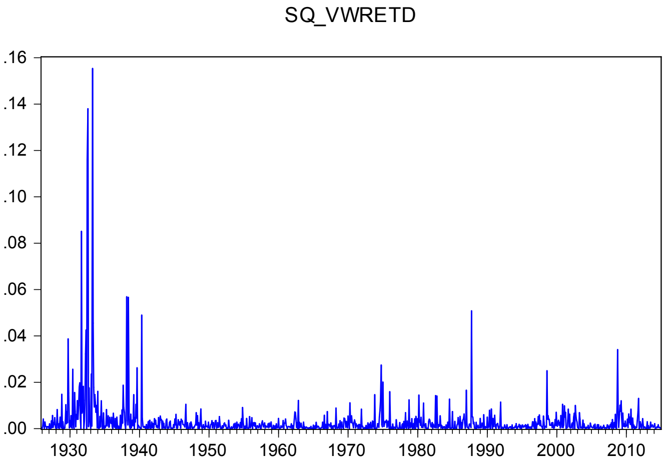
Outline

1. Univariate Volatility
   - Does variance change?
   - Heteroskedasicity
   - Exponentially Weighted Moving Average
   - Use of Shorter-term Variance
   - Use of intraday changes to estimate volatility
   - ARCH
   - More complicated models of ARCH
     - GARCH
     - IGARCH
     - GARCH-M
     - Aysmmetric GARCH
     - TGARCH
     - Tests of Restrictions
   - Stochastic volatility
   - Summary of Univariate Volatility
2. Multivariate Volatility
   - Formal Representation
   - Simplify GARCH models with nonzero correlations of innovations
   - Constant conditional-correlation GARCH

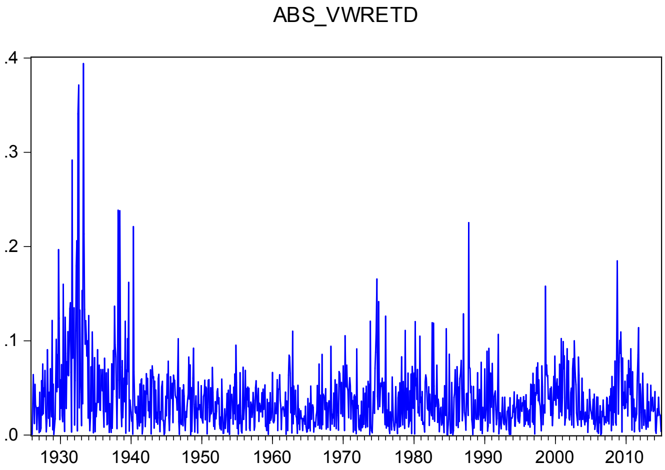
Returns for CRSP monthly



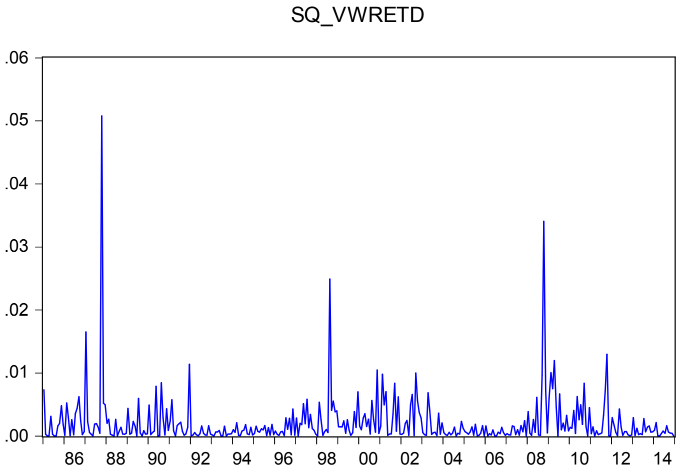
Squared log returns for CRSP monthly



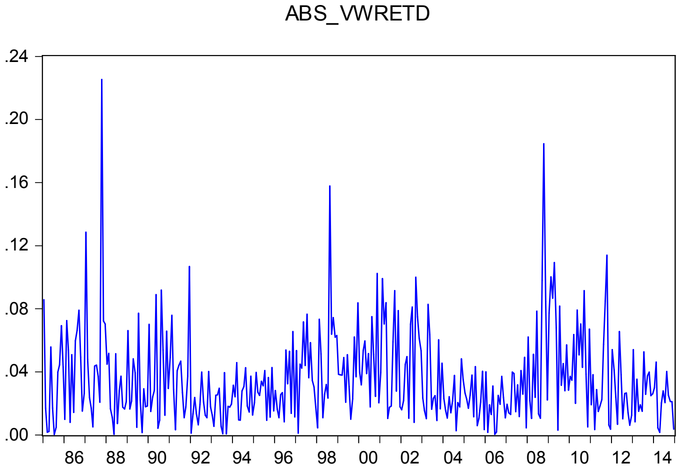
Absolute value of log returns for CRSP monthly



Squared log returns for CRSP monthly from 1985 to 2014



Absolute value of log returns for CRSP monthly from 1985 to 2014



Heteroskedasticity over time

- These graphs suggest heteroskedasticity over time
  - Time-varying volatility of returns
    - Of interest in itself to characterize returns
    - Matters for prices of options and some other financial instruments
  - Volatility clustering
- These graphs are suggestive but don't tell us too much
  - Using individual observations on squared changes and absolute value to estimate variance and standard deviation as it changes
  - Similar to using each individual observation to estimate mean as it changes
  - Can't forecast anything going forward

Exponentially weighted moving average

- Exponentially weighted average assumes today's variance forecast is a weighted average of variance yesterday and variance forecasted for yesterday
$$s_t^2 = (1 - \lambda)(r_t - \bar{r})^2 + \lambda s_{t-1}^2$$
- And yesterday's variance forecast is a weighted average of variance the day before that and variance forecasted for day before that, and so on
$$s_{t-1}^2 = (1 - \lambda)(r_{t-1} - \bar{r})^2 + \lambda s_{t-2}^2$$
- $\lambda = 0.94$ daily frequency has been suggested

Use shorter-term returns to estimate variance over longer periods

- Use shorter-term variance to calculate variance for a longer period
- Called realized volatility
- For example, for a monthly variance of log returns
  - Let r_t^m be the monthly log return in month t
  - Let $r_{t,i}$ be the daily log return on day i in month t
  - Suppose that daily returns are serially uncorrelated and the daily variance is constant
  - Then

$$r_t^m = \sum_{i=1}^{n} r_{t,i}$$

$$\mathrm{Var}\left[r_t^m\right] = n\,\mathrm{Var}\left[r_{t,i}\right]$$

  - and $\widehat{\mathrm{Var}\left[r_{t,i}\right]} = \frac{\sum_{i=1}^{n}(r_{t,i}-\overline{r_t})^2}{n-1}$ where $\overline{r_t}$ is the mean of the daily returns
  - The estimated monthly variance thus is

$$\widehat{\sigma}_t^{m2} = n\frac{\sum_{i=1}^{n}\left(r_{t,i}-\overline{r_t}\right)^2}{n-1}$$

Daily variance to estimate monthly variance

- The estimated monthly variance is simple to calculate

$$\widehat{\sigma}_t^{m2} = n\frac{\sum_{i=1}^{n}\left(r_{t,i} - \overline{r_t}\right)^2}{n-1}$$

- This becomes more complicated if the daily returns are serially correlated, but it's still manageable
- If daily log returns have high excess kurtosis and serial correlations, then this estimator may not be consistent

Garman-Klass estimator of daily variance

- Use high, low, opening, and closing prices to estimate variance
  - Can estimate daily variance just knowing opening, high, low and closing prices
  - Assume that price follows a random walk
  - Let c_t be the logarithm of the closing price so $r_t = c_t - c_{t-1}$
  - Conventional estimator is

$$\sigma_t^2 = \mathrm{E}\left[\left(c_t - c_{t-1} \right)^2 \right]$$

  - Using only closing price
  - High H_t, low L_t, and open O_t also often are available
  - Can estimate daily variance of price (not log price) from

$$\widehat{\sigma}_{GK}^2 = 0.12 \frac{\left(O_t - C_{t-1} \right)^2}{f} + 0.88 \frac{0.5 \left(H_t - L_t \right)^2 + 0.386 \left(C_t - O_t \right)^2}{1 - f}$$

    where f is the fraction of the day that the market is closed
  - Minimum variance unbiased estimator for a random walk with no drift

Yang and Zhang estimator

- Use high, low, opening, and closing prices to estimate variance of log prices over a longer period
  - Define

$$\begin{aligned} o_t &= \ln O_t - \ln O_{t-1} \\ h_t &= \ln H_t - \ln O_{t-1} \\ \ell_t &= \ln L_t - \ln O_{t-1} \\ c_t &= \ln C_t - \ln O_{t-1} \end{aligned}$$

  - Monthly variance based on n days of trading is

$$\widehat{\sigma}^2_{YZ} = \widehat{\sigma}_o + k\widehat{\sigma}_c + (1-k)\,\widehat{\sigma_{rs}}$$

where $\widehat{\sigma}_o$ and $\widehat{\sigma}_c$ are is the estimated variances of o_t and c_t and

$$\begin{aligned} \widehat{\sigma}^2_{rs} &= \frac{1}{n}\sum\left[h_t\left(h_t - c_t\right) + \ell_t\left(\ell_t - c_t\right)\right] \\ k &= \frac{0.34}{1.34 + (n+1)\,/\,(n-1)} \end{aligned}$$

and k was chosen to minimize the variance of the estimator $\widehat{\sigma}^2_{YZ}$

Serial correlation

- Change in logarithm of value-weighted CRSP index
- Serial correlation of squared changes in logarithm of value-weighted CRSP index
- Serial correlation of absolute values of change in logarithm of value-weighted CRSP index

Autoregressive conditional heteroskedasticity (ARCH)

is intended to deal with this

- Simple ARCH model for returns

$$r_t = \mu_t + u_t \quad \mathrm{E}\, u_t = 0, \mathrm{E}\, u_t^2 = \sigma_t^2$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

  where r_t is a log return, σ_t^2 is the variance of a_t conditional on past values of the squared innovations, a_{t-1}^2
- $r_t = \mu_t + u_t$ is the mean equation for r_t
- $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$ is the variance equation for r_t
- u_t is the innovation in r_t
- A slightly different version of the same equations is

$$r_t = \mu_t + h_t^{1/2}\varepsilon_t \quad \mathrm{E}\, \varepsilon_t = 0, \mathrm{E}\, \varepsilon_t^2 = 1$$
$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2$$

  where $u_t = h_t^{1/2}\varepsilon_t$

Estimating an ARCH model

$$r_t = \mu_t + u_t \quad \mathrm{E}\, u_t = 0, \mathrm{E}\, u_t^2 = \sigma_t^2$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

- Steps in estimating an ARCH model

1. Estimate a model for the mean equation
2. Use the residuals of the mean equation to test for ARCH effects
3. Specify a variance model with ARCH effects if it seems warranted
4. Check the fitted model and refine as suggested by diagnostic statistics

Mean equation

- In general, there is no reason the mean equation can't be as complicated as we like

$$r_t = \mu_t + u_t$$

- r_t can be a complicated ARMA(p,q) or can have other variables included
  - r_t is stationary in mean
  - r_t may be first difference of original series
  - for example $r_t = p_t - p_{t-1}$
- May be mis-specified if ignore conditional heteroskedasticity of u_t

Testing for ARCH

- Simple model

$$
\begin{aligned}
r_t = u_t &= \sigma_t \varepsilon_t \quad \mathrm{E}\,\varepsilon_t = 0, \mathrm{E}\,\varepsilon_t^2 = 1 \\
\sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2
\end{aligned}
$$

Three tests for ARCH

- Box-Ljung test applied to squared residuals, u_t^2, for some pre-specified number of lags k
- Engle test based on a regression for the squared residuals

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_k u_{t-k}^2 + e_t \tag{1}$$

  where e_t is the error term in the regression for squared residuals
  - Test whether $\alpha_1 = \alpha_2 = \ldots = \alpha_k$ using $T \cdot R^2$ where T is the number of observations and R^2 is the R^2 in equation (1) which has a χ_k^2 distribution
  - Common to use one lag
- F-test for regression (1)

Properties of ARCH models

Mean

- Simple model

$$r_t = u_t = \sigma_t \varepsilon_t \quad \mathrm{E}\,\varepsilon_t = 0, \mathrm{E}\,\varepsilon_t^2 = 1$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

where $\alpha_0 > 0$ and $\alpha_1 \geq 0$. Why?

- Let F_{t-1} denote the set of all information available in $t-1$ and earlier, especially $r_{t-1}, r_{t-2}, \ldots, u_{t-1}, u_{t-2,\ldots}$

$$\begin{aligned}
\mathrm{E}\left[u_t\right] &= \mathrm{E}\left[\mathrm{E}\left(u_t | F_{t-1}\right)\right] \text{ (application of law of iterated expectations)} \\
&= \mathrm{E}\left[\mathrm{E}\left(\sigma_t \varepsilon_t | F_{t-1}\right)\right] \\
&= \mathrm{E}\left[\mathrm{E}\left(\sigma_t | F_{t-1}\right) \mathrm{E}\left(\varepsilon_t | F_{t-1}\right)\right] \\
&= \mathrm{E}\left[\mathrm{E}\left(\sigma_t | F_{t-1}\right) \cdot 0\right] \\
&= 0
\end{aligned}$$

Properties of ARCH models

Variance

- Simple model

$$
\begin{aligned}
r_t &= \sigma_t \varepsilon_t \quad \mathrm{E}\,\varepsilon_t = 0, \mathrm{E}\,\varepsilon_t^2 = 1 \\
\sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 \\
\alpha_0 &> 0, \alpha_1 \geq 0 \\
\mathrm{E}\left[u_t\right] &= 0
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}\left[u_t\right] &= \mathrm{E}\left[u_t^2\right] = \mathrm{E}\left[\mathrm{E}\left(u_t^2 | F_{t-1}\right)\right] \\
&= \mathrm{E}\left[\mathrm{E}\left(\alpha_0 + \alpha_1 u_{t-1}^2 | F_{t-1}\right)\right] \\
&= \mathrm{E}\left[\alpha_0 + \alpha_1 u_{t-1}^2\right] = \alpha_0 + \alpha_1 \mathrm{E}\,u_{t-1}^2 \\
&= \alpha_0 + \alpha_1 \mathrm{E}\,u_t^2
\end{aligned}
$$

- Therefore, if $0 \leq \alpha_1 < 1$,

$$
\mathrm{Var}\left[u_t\right] = \frac{\alpha_0}{1 - \alpha_1}
$$

Properties of ARCH models

Kurtosis – fourth moment

- Simple model

$$r_t = \sigma_t \varepsilon_t \quad \mathrm{E}\,\varepsilon_t = 0, \mathrm{E}\,\varepsilon_t^2 = 1$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$
$$\alpha_0 > 0, \alpha_1 \geq 0$$
$$\mathrm{E}\left[u_t\right] = 0$$
$$\mathrm{Var}\left[u_t\right] = \frac{\alpha_0}{1-\alpha_1}$$

- Tail behavior
  - Assume ε_t is normally distributed
  - Do we get fatter tails than from the normal distribution?

Fourth moment, tail behavior

$$\mathrm{E}\, u_t^4 = \frac{3\alpha_0^2 (1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$$

- $\mathrm{E}\, u_t^4 > 0$ obviously must hold and therefore α_1 must satisfy $(1-3\alpha_1^2) > 0$ and therefore $0 \leq \alpha_1^2 \leq 1/3$
  - Already assumed $1 > \alpha_1$
- The unconditional kurtosis of u_t with normally distributed ε_t is

$$\frac{\mathrm{E}\, u_t^4}{\mathrm{Var}\,[u_t]^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2}$$

  - Therefore,

$$\frac{\mathrm{E}\, u_t^4}{\mathrm{Var}\,[u_t]^2} > 3$$

  - This implies fatter tails than for a normal distribution

Properties of ARCH models

Restrictions on estimated variance equation in practice

- Simple model

$$
\begin{aligned}
r_t &= \sigma_t \varepsilon_t \quad \mathrm{E}\,\varepsilon_t = 0, \mathrm{E}\,\varepsilon_t^2 = 1 \\
\sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 \\
\alpha_0 &> 0, \alpha_1 \geq 0 \\
\mathrm{E}\left[u_t\right] &= 0 \\
\mathrm{Var}\left[u_t\right] &= \frac{\alpha_0}{1-\alpha_1} \\
\frac{\mathrm{E}\,u_t^4}{\mathrm{Var}\left[u_t\right]^2} &= 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3, \quad 0 \leq \alpha_1^2 \leq \frac{1}{3}
\end{aligned}
$$

- α_i need not all be positive when more than one lag in the variance equation
  - Sufficient to make sure all the estimated conditional volatilities $\widehat{\sigma_t^2} > 0$
  - If one $\widehat{\sigma_t^2}$ is negative, the estimated results make no sense

Limitations of ARCH models

1. Symmetric effects of shocks. This is too restrictive for stock returns, where negative shocks have a larger effect on future variance than positive shocks
2. Returns tend to have some clusters of high and low variance, whereas ARCH models tend to predict slow decay to mean from any current variance
3. Restrictive parameterization, e.g. $0 \leq \alpha_1^2 \leq \frac{1}{3}$ for kurtosis to be well defined for ARCH(1)
4. Deterministic equation for variance; no error term in $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$
5. Provides no evidence on source of changes in variance

Estimation of ARCH model

1. Use partial autocorrelation function of u_t^2 to determine order of ARCH specified
2. Maximize the likelihood given the distribution of ε_t
   - Distributions
     - Normal distribution
     - t-distribution with degrees of freedom υ
     - generalized error distribution
   - Quasi-maximum likelihood estimation
     - Consistent estimates of parameters
     - Issue of correct standard errors of coefficients
3. u_t/σ_t is a sequence of IID variables if correctly specified

- Use $\widehat{u}_t/\widehat{\sigma}_t$ to examine whether the serial correlation is adequately estimated
  - Autocorrelation functions of $\widehat{u}_t/\widehat{\sigma}_t$ and $(\widehat{u}_t/\widehat{\sigma}_t)^2$ with Engle regression test on $(\widehat{u}_t/\widehat{\sigma}_t)^2$
  - Compare distribution to the one assumed using Kolmogorov-Smirnoff tests
  - These tests are asymptotically correct but have nontrivial estimation error – p-values are only a guide to decisions

GARCH

- ARCH models can require many lags
  - Reduce lags in mean equations by using ARMA models
    - MA terms can substitute for several AR terms
  - Maybe including something like MA terms in ARCH equation can reduce number of lags
- GARCH (Generalized ARCH) model
  - ARCH Model

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_k u_{t-m}^2$$

  - Instead, try GARCH, here a GARCH(m,s) (order of lags often not consistent)

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_k u_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_k \sigma_{t-s}^2$$

    - Lag lengths are m for the part analogous to the moving average and s for the part analogous to an autoregression
    - May be able to reduce number of parameters substantially by having both sets of terms

Properties of GARCH models

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_k u_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_k \sigma_{t-s}^2$$

- Restrictions on parameters

$$\alpha_i > 0, \beta_i > 0, \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$$

- Properties of estimates and relation to parameters

$$\mathrm{E}\left[u_t^2\right] = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i)}$$

- For GARCH(1,1), $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, with $1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2 > 0$

$$\frac{\mathrm{E}\left[u_t^4\right]}{\left(\mathrm{E}\left[u_t^2\right]\right)^2} = \frac{3\left[1 - (\alpha_1 + \beta_1)^2\right]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3$$

IGARCH

- What if the variance is very persistent?

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_k u_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_k \sigma_{t-s}^2$$

  with $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) \approx 1$, suggesting a unit root in the variance process
- Actually pretty common with returns
  - Change in logarithm of value-weighted CRSP index 1/2/1985 to 12/31/2014
  - dlnvwcrsp $= 0.000724 + \widehat{a}_t$
  - $\widehat{\sigma}_t^2 = 1.76 \cdot 10^{-6} + 0.100067 \widehat{a}_{t-1}^2 + 0.884653 \widehat{\sigma}_{t-1}^2$
  - standard errors of coefficients are $1.33 \cdot 10^{-7}$, 0.0045 and 0.0040
  - sum of coefficients is $0.984720 \approx 0.98$
- Suggests IGARCH model, IGARCH(1,1)

IGARCH model

- IGARCH model, IGARCH(1,1)

$$
\begin{aligned}
u_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1)\, u_{t-1}^2 \\
0 &< \beta_1 < 1
\end{aligned}
$$

Properties of IGARCH model

- IGARCH(1,1)

$$u_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1) u_{t-1}^2$$
$$0 < \beta_1 < 1$$

  - Unconditional variance is undefined
  - Stationary
  - Constant term is similar to constant term for a random walk – a trend
    - ★ A nonzero constant term suggests a trend in variance
    - ★ Why? One-step-ahead forecast at h is forecast of σ_{h+1}^2
    - ★ Suppose that estimate of σ_h^2 and u_h^2 are available and $\sigma_h^2(1)$ is the forecast made of σ^2 made at h for one step ahead

$$\sigma_h^2(1) = \alpha_0 + \beta_1 \sigma_h^2 + (1 - \beta_1) u_h^2$$
$$\sigma_h^2(2) = \alpha_0 + \beta_1 \sigma_{h+1}^2 + (1 - \beta_1) u_{h+1}^2$$

    - ★ Best forecast of σ_{h+1}^2 is $\sigma_h^2(1)$ and best forecast of u_{h+1}^2 from $u_{h+1}^2 = \sigma_{h+1}^2 \varepsilon_t^2$ is $\sigma_h^2(1)$, so

$$\sigma_h^2(2) = \alpha_0 + \sigma_h^2(1)$$

GARCH-M

- GARCH-M is GARCH in mean – simple one is

$$r_t = \mu + c\sigma_t^2 + u_t, \quad u_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

  - r_t is a return on an asset and c is called a risk premium parameter
  - Could use σ_t or $\ln \sigma_t$ instead of σ_t^2
  - Question: When will the variance of an asset's return reflect its risk?

Asymmetric GARCH

- Glosten, Jagannathan and Runkle (1993)
- Represent the asymmetry in returns that negative shocks create more future variance

$$r_t = \mu + c\sigma_t^2 + u_t, \quad u_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma u_{t-1}^2 I_{t-1}$$

  - I_{t-1} is an indicator with $I_{t-1} = 1$ if $u_{t-1} < 0$ and $I_{t-1} = 0$ if $u_{t-1} \geq 0$
  - Greater effect of negative shocks if estimate $\gamma > 0$

EGARCH

- Exponential GARCH – allows for asymmetry
- An EGARCH(1,1) with ε_t iid and normally distributed

$$r_t = \mu + u_t, u_t = \sigma_t \varepsilon_t$$

$$(1 - \beta \, \mathrm{L}) \ln \sigma_t^2 = \left\{ \begin{array}{c} \alpha_* + (\gamma + \theta) \frac{u_{t-1}}{\sigma_{t-1}} \text{ if } u_{t-1} \geq 0 \\ \alpha_* + (\gamma - \theta) \frac{|u_{t-1}|}{\sigma_{t-1}} \text{ if } u_{t-1} < 0 \end{array} \right\}$$

$$\alpha_* = (1 - \alpha_1) \, \alpha_0 - \sqrt{2\pi}\gamma$$

where the lag operator L is such that $\mathrm{L}\, x_t = x_{t-1}$ and $\mathrm{L}^i \, x_t = x_{t-i}$

EGARCH explained

- An EGARCH model starts from the function $g(\varepsilon_t)$

$$g(\varepsilon_t) = \theta\varepsilon_t + \gamma(|\varepsilon_t| - \mathsf{E}|\varepsilon_t|)$$
$$\mathsf{E}\, g(\varepsilon_t) = 0$$

  - which can be rewritten as

$$g(\varepsilon_t) = \left\{ \begin{array}{ll} (\theta+\gamma)\,\varepsilon_t - \gamma\,\mathsf{E}|\varepsilon_t| & \text{if } \varepsilon_t \geq 0 \\ (\theta-\gamma)\,\varepsilon_t - \gamma\,\mathsf{E}|\varepsilon_t| & \text{if } \varepsilon_t < 0 \end{array} \right\}$$

  - If θ is negative, then $g(\varepsilon_t)$ is larger for $\varepsilon_t < 0$ than for $\varepsilon_t \geq 0$

EGARCH model

- Define the lag operator L such that $\mathrm{L}\, x_t = x_{t-1}$ and $\mathrm{L}^i\, x_t = x_{t-i}$
  - EGARCH model, EGARCH(m,s)

$$
\begin{aligned}
r_t &= \mu + u_t, \quad u_t = \sigma_t \varepsilon_t \\
g\left(\varepsilon_t\right) &= \theta \varepsilon_t + \gamma\left(\left|\varepsilon_t\right| - \mathrm{E}\left|\varepsilon_t\right|\right) \\
\ln \sigma_t^2 &= \alpha_0 + \frac{1 + \beta_1\, \mathrm{L} + \ldots + \beta_{s-1}\, \mathrm{L}^{s-1}}{1 - \alpha_1\, \mathrm{L} - \ldots - \alpha_m\, \mathrm{L}^m} g\left(\varepsilon_{t-1}\right)
\end{aligned}
$$

Test for asymmetric volatility

- Estimate a mean equation and get residuals $\widehat{u}_t$
- Define $I_{t-1} = 1$ if $\widehat{u}_{t-1} < 0$ and $I_{t-1} = 0$ if $\widehat{u}_{t-1} \geq 0$
- Run regression

$$\widehat{u}_t^2 = \phi_0 + \phi_1 I_{t-1} + v_t$$

- If $\phi_1 > 0$, then there is an asymmetric effect of negative shocks

TGARCH

- Threshold GARCH – TGARCH(m,s)

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{s} (\alpha_i + \gamma_i I_{t-i})\, u_{t-i}^2 + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^2$$

$$I_{t-i} = 1 \text{ if } u_{t-i} < 0$$

$$I_{t-i} = 0 \text{ if } u_{t-i} \geq 0$$

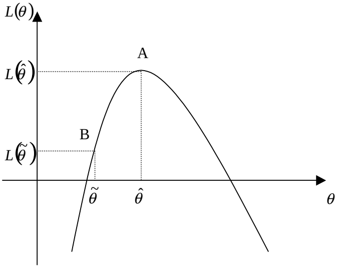
$$\alpha_i, \gamma_i, \beta_j \geq 0$$

- This also allows for bigger effects of negative shocks

Testing restrictions in nonlinear models

- Usual t-ratios are estimated for maximum likelihood
  - Inverse of information matrix provides estimator of variance for these tests
- Likelihood ratio, Wald and Lagrange multiplier tests
- Consider estimating a parameter θ
  - Maximum likelihood estimate is $\widehat{\theta}$
  - Restricted estimate is $\widetilde{\theta}$

Comparison of likelihood ratio, Wald and Lagrange multiplier tests



- Vertical distance from B to A is basis of likelihood ratio tests
- Horizontal distance from B to A is basis of Wald tests (e.g. t-tests)
- Slope of likelihood function at B is basis of Lagrange multiplier test

Stochastic Volatility

- ARCH is restrictive
  - Evolution of variance is deterministic except for influence of innovations in mean equation
- Stochastic volatility
  - The evolution of volatility is not a deterministic function of only past volatility and innovations to the mean equation
  - Innovations to the variance affect variance independent of mean equation

Relatively simple example of stochastic volatility

$$
\begin{aligned}
r_t &= \mu + u_t = \mu + \sigma_t \varepsilon_t \\
\ln\left(\sigma_t\right)^2 &= \alpha + \beta \ln\left(\sigma_{t-1}\right)^2 + \sigma_\eta \eta_t \\
&\alpha > 0,\ \ \varepsilon_t \sim N(0,1),\ \ \eta_t \sim N(0,1)
\end{aligned}
$$

- Two innovations for every observation r_t
  - How can that be?
  - The innovations reflect different aspects of the series
  - If they did not reflect different aspects of series, there would be a problem

$$
\begin{gathered}
r_t = \mu + u_{1,t} + u_{2,t} \\
u_{1,t} \sim N(0, \sigma_1^2),\, u_{2,t} \sim N(0, \sigma_2^2)
\end{gathered}
$$

    - ★ Never would be able to tell how much of variance of r_t is due to $u_{1,t}$ and how much is due to $u_{2,t}$

Summary

- Volatility itself varies over time
- There are quite a few ways of estimating volatility
- Simple ways without a lot of complication
  - Exponential smoothing
  - Realized volatility
- ARCH and GARCH are relatively straightforward to implement
- There are many variants of GARCH models to allow for different ways variance can change over time

Multivariate GARCH

- Multivariate return series

$$\mathbf{r}_t = \boldsymbol{\mu}_t + \mathbf{u}_t$$

where the vectors are simple generalizations of a univariate process

- The vectors are m by one with m asset returns
- $\boldsymbol{\mu}_t = \mathrm{E}\left[\mathbf{r}_t | F_{t-1}\right]$
- $\mathbf{u}_t$ is the innovation in the returns in period t with $\mathrm{E}\,\mathbf{a}_t = \mathbf{0}$ and

$$\boldsymbol{\Sigma}_t = \mathrm{Cov}\left[\mathbf{u}_t | F_{t-1}\right]$$

  - $\boldsymbol{\Sigma}_t$ is m by m with $m(m+1)/2$ distinct elements
- The number of parameters in $\boldsymbol{\Sigma}_t$ increases with the square of m because there are $\left(m^2 + m\right)/2$ distinct elements
  - The number of observations is mT and increases linearly with m

Curse of Dimensionality

Estimation of all parameters unconstrained not feasible for large systems

- The implication of the number of parameters in $\boldsymbol{\Sigma}_t$ increasing with the square of m
  - Five assets implies $m(m+1)/2 = 15$ parameters
  - Twenty assets implies $m(m+1)/2 = 210$ parameters
  - One hundred assets implies $m(m+1)/2 = 5,050$ parameters
- Suppose have 250 observations on each asset
  - Five assets implies $1,250$ observations
  - Twenty assets implies $5,000$ observations
  - One hundred assets implies $25,000$ observations
- Even one hundred assets is not a lot
- If have 1,000 assets
  - Then 500,500 parameters in the covariance matrix
  - Only 250,000 observations
  - Sometimes makes more sense to look at volatility of portfolio, fortunately

Constrained systems

- Want to find a way to allow for correlations but not have too many parameters estimated with poor precision

Strategies to estimate multivariate volatility models

- Exponentially smoothing estimate
  - Includes all the parameters in the variance-covariance matrix
  - Adds only one additional parameter to get a time-varying variance-covariance matrix
- Multivariate GARCH models
  - Diagonal VECH model
    - GARCH(1,1) for each term in variance-covariance matrix
- Reparameterizations
  - Replace covariances by correlations
  - Cholesky decomposition

Exponential smoothing estimate

- Exponential smoothing forecasts are based on simple forecasts that smooth values of the data
  - E.g. the forecast at $t-1$ of the value of a series x_t is ${}_{t-1}f_t = (1-\lambda)\, x_{t-1} + \lambda\ {}_{t-2}f_{t-1}$
  - where ${}_{t-1}f_t$ is the forecast at $t-1$ for period t
- Apply this idea to the variance-covariance matrix
  - Estimate time-varying variance-covariance matrix as
  - $\widehat{\boldsymbol{\Sigma}}_t = (1-\lambda)\, \mathbf{u}_{t-1}\mathbf{u}'_{t-1} + \lambda\widehat{\boldsymbol{\Sigma}}_{t-1}$

VECH GARCH model

- Conditional variances and conditional covariances depend on all innovations and cross-products
- Two variable example with one lag of innovations squared and all lagged variance terms
  - Let *a*s and *b*s represent parameters

$$
\begin{aligned}
h_{11,t} &= c_{11,0} + a_{11,1}u_{1,t-1}^2 + a_{12,1}u_{2,t-1}^2 + a_{13}u_{1,t-1}u_{2,t-1} \\
&\quad + b_{11,1}h_{1,t-1}^2 + b_{12,1}h_{2,t-1}^2 + b_{13}h_{12,t-1} \\
h_{22,t} &= c_{21,0} + a_{21,1}u_{1,t-1}^2 + a_{22,1}u_{2,t-1}^2 + a_{23}u_{1,t-1}u_{2,t-1} \\
&\quad + b_{21,1}h_{1,t-1}^2 + b_{22,1}h_{2,t-1}^2 + b_{23}h_{12,t-1} \\
h_{12,t} &= c_{31,0} + a_{31,1}u_{1,t-1}^2 + a_{32,1}u_{2,t-1}^2 + a_{33}u_{1,t-1}u_{2,t-1} \\
&\quad + b_{31,1}h_{1,t-1}^2 + b_{32,1}h_{2,t-1}^2 + b_{33}h_{12,t-1}
\end{aligned}
$$

- Everything depends on everything
- Problems with this model
  - Curse of dimensionality will overtake this quickly
  - No guarantee of positive definite variance-covariance matrix every period

Curse of Dimensionality with GARCH

- Suppose a general GARCH in which all variances and covariances depend on all lagged variances and cross-products of errors
- Two asset case

$$
\begin{aligned}
h_{11,t} &= c_{11,0} + a_{11,1} u_{1,t-1}^2 + a_{12,1} u_{2,t-1}^2 + a_{13} u_{1,t-1} u_{2,t-1} \\
&\quad + b_{11,1} h_{1,t-1}^2 + b_{12,1} h_{2,t-1}^2 + b_{13} h_{12,t-1} \\
h_{22,t} &= c_{21,0} + a_{21,1} u_{1,t-1}^2 + a_{22,1} u_{2,t-1}^2 + a_{23} u_{1,t-1} u_{2,t-1} \\
&\quad + b_{21,1} h_{1,t-1}^2 + b_{22,1} h_{2,t-1}^2 + b_{23} h_{12,t-1} \\
h_{12,t} &= c_{31,0} + a_{31,1} u_{1,t-1}^2 + a_{32,1} u_{2,t-1}^2 + a_{33} u_{1,t-1} u_{2,t-1} \\
&\quad + b_{31,1} h_{1,t-1}^2 + b_{32,1} h_{2,t-1}^2 + b_{33} h_{12,t-1}
\end{aligned}
$$

  - The number of parameters in each equation is $2m + 1$
  - The number of equations is $m(m+1)/2$
  - The total number of parameters is $m(m+1)(2m+1)/2$
  - Two assets implies $m(m+1)(2m+1)/2 = 15$ parameters
  - Five assets implies $m(m+1)(2m+1)/2 = 165$ parameters
  - Ten assets implies $m(m+1)(2m+1)/2 = 1105$ parameters

Diagonal VECH GARCH model

- The Diagonal VECH GARCH model has each term in the variance-covariance matrix evolve independently according to a GARCH(1,1)
- Two-variable example
  - Let *a*s and *b*s represent parameters

$$\begin{aligned} h_{11,t} &= c_{11,0} + a_{11,1} u^2_{1,t-1} + b_{11,1} h^2_{1,t-1} \\ h_{22,t} &= c_{21,0} + a_{22,1} u^2_{2,t-1} + b_{22,1} h^2_{2,t-1} \\ h_{12,t} &= c_{31,0} + a_{33} u_{1,t-1} u_{2,t-1} + b_{33} h_{12,t-1} \end{aligned}$$

- Still no guarantee that the estimated variance-covariance matrix will be positive definite in every period

BEKK

Use of Cholesky decomposition in computations can guarantee positive definiteness

- We want the estimated variance-covariance matrix to be positive definite every period
  - Models such as the Diagonal VECH GARCH model with each term in the variance-covariance matrix evolving independently according to a GARCH(1,1) will not necessarily satisfy positive definiteness
- Transforming the estimation problem using a Cholesky decomposition ensures positive definiteness of the covariance matrix every period
- BEKK model ensures that the variance-covariance matrix is positive definite

Constant-correlation GARCH model

- Constant-correlation model
  - A constant correlation is not an obvious constraint to impose with time-varying variances
    - Implies that the covariance varies proportionately to the standard deviations of the innovations
    - This implication can be seen from

$$\rho = \frac{\text{Cov}\left[u_{1t}, u_{2t}\right]}{\text{SD}\left[u_{1t}\right]\text{SD}\left[u_{2t}\right]} = \frac{h_{12,t}}{\sqrt{h_{11,t}}\sqrt{h_{22,t}}}$$

    - Does allow for multivariate aspect of process
  - A two-variable GARCH model is

$$\begin{bmatrix} h_{11,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} \alpha_{10} \\ \alpha_{20} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} u_{1,t-1}^2 \\ u_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{22,t-1} \end{bmatrix}$$

  - $h_{11,t}$ and $\text{h}_{22,t}$ are the conditional variances of the variables
  - Do not need a term for covariance because its value each period is implied by the estimated constant correlation

Constant-correlation GARCH model

- Two-variable example
  - The constant-correlation model can be written

$$h_{11,t} = \alpha_{10} + \alpha_{11} u_{1,t-1}^2 + \alpha_{12} u_{2,t-1}^2 + \beta_{11} h_{11,t-1} + \beta_{12} h_{22,t-1}$$
$$h_{22,t} = \alpha_{20} + \alpha_{21} u_{1,t-1}^2 + \alpha_{22} u_{2,t-1}^2 + \beta_{21} h_{11,t-1} + \beta_{22} h_{22,t-1}$$

  - Bollerslev determined the conditions necessary for this model to be stationary

Time-varying correlation GARCH

- Can have a time-varying correlation

Constant conditional-correlation GARCH model

- Constant conditional-correlation GARCH

$$\rho = \frac{\text{Cov}\left[\varepsilon_{1t}, \varepsilon_{2t}\right]}{\text{SD}\left[\varepsilon_{1t}\right]\text{SD}\left[\varepsilon_{2t}\right]} = \frac{\sigma_{12,t}}{\sqrt{\sigma_{11,t}\sigma_{22,t}}}$$

  - A constant correlation is not an obvious constraint to impose with time-varying variances
    - Implies that the covariance varies proportionately to the standard deviations of the innovations
    - Does allow for multivariate aspect of process

- A two-variable Constant conditional-correlation GARCH model in matrix form is

$$\begin{bmatrix} \sigma_{11,t} \\ \sigma_{22,t} \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} \varepsilon^2_{1,t-1} \\ \varepsilon^2_{2,t-1} \end{bmatrix} + \begin{bmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{bmatrix} \begin{bmatrix} \sigma_{11,t-1} \\ \sigma_{22,t-1} \end{bmatrix}$$

Constant conditional-correlation GARCH model

- The Constant conditional-correlation GARCH model

$$
\begin{aligned}
\sigma_{11,t} &= \gamma_{10} + \gamma_{11}\varepsilon_{1,t-1}^2 + \gamma_{12}\varepsilon_{2,t-1}^2 + \delta_{11}\sigma_{11,t-1} + \delta_{12}\sigma_{22,t-1} \\
\sigma_{22,t} &= \gamma_{20} + \gamma_{21}\varepsilon_{1,t-1}^2 + \gamma_{22}\varepsilon_{2,t-1}^2 + \delta_{21}\sigma_{11,t-1} + \delta_{22}\sigma_{22,t-1} \\
\sigma_{12,t} &= \rho\left(\sigma_{11,t}\sigma_{22,t}\right)^{1/2}
\end{aligned}
$$

- Do not need an equation for covariance because its value each period is implied by the estimated constant correlation
- Bollerslev determined the conditions necessary for this model to be stationary
- Quasi-maximum likelihood estimator is consistent under certain conditions and the estimators are asymptotically normally distributed

Dynamic conditional-correlation GARCH model

- Instead of the constant conditional-correlation GARCH model

$$
\begin{aligned}
\sigma_{11,t} &= \gamma_{10} + \gamma_{11}\varepsilon^2_{1,t-1} + \gamma_{12}\varepsilon^2_{2,t-1} + \delta_{11}\sigma_{11,t-1} + \delta_{12}\sigma_{22,t-1} \\
\sigma_{22,t} &= \gamma_{20} + \gamma_{21}\varepsilon^2_{1,t-1} + \gamma_{22}\varepsilon^2_{2,t-1} + \delta_{21}\sigma_{11,t-1} + \delta_{22}\sigma_{22,t-1} \\
\sigma_{12,t} &= \rho\left(\sigma_{11,t}\sigma_{22,t}\right)^{1/2}
\end{aligned}
$$

- A simple dynamic conditional-correlation GARCH model (Engle 2002)

$$
\begin{aligned}
\sigma_{11,t} &= \gamma_{10} + \gamma_{11}\varepsilon^2_{1,t-1} + \delta_{11}\sigma_{11,t-1} \\
\sigma_{22,t} &= \gamma_{20} + \gamma_{22}\varepsilon^2_{2,t-1} + \delta_{22}\sigma_{22,t-1} \\
\sigma_{12,t} &= \rho_{12,t}\left(\sigma_{11,t}\sigma_{22,t}\right)^{1/2}
\end{aligned}
$$

Standardized innovations and the relationship between conditional correlation and conditional covariance I

- As above, examine the problem with two series, $i = 1, 2$

$$
\begin{aligned}
r_{i,t} &= \mu_{i,t} + \varepsilon_{i,t} \\
\varepsilon_{i,t} &= \sigma_{i,t} v_{i,t} \\
v_{i,t} &\sim \text{iid}\,(0,1) \\
\mathrm{E}_{t-1}\, v_{i,t} &= 0 \\
\mathrm{E}_{t-1}\, v_{i,t}^2 &= 1 \\
\mathrm{E}\, v_{1,t} v_{1,t} &= \rho_{12}^{v}
\end{aligned}
$$

- To simplify notation, I will use the convention that $\sigma_{i,t} = \sigma_{ii,t}^{(1/2)}$
- The conditional correlation of the two series is

$$
\rho_{12,t} = \frac{\mathrm{Cov}_{t-1}\left[r_{1,t}, r_{2,t}\right]}{\mathrm{SD}_{t-1}\left[r_{1,t}\right]\mathrm{SD}_{t-1}\left[r_{2,t}\right]}
$$

Standardized innovations and the relationship between conditional correlation and conditional covariance II

- For simplicity, suppose that

$$\mu_{i,t} = 0$$

which implies

$$r_{i,t} = \varepsilon_{i,t}$$

- The conditional correlation of the two series is

$$\rho_{12,t} = \frac{\text{Cov}_{t-1}\left[\varepsilon_{1,t}, \varepsilon_{2,t}\right]}{\text{SD}_{t-1}\left[\varepsilon_{1,t}\right]\text{SD}_{t-1}\left[\varepsilon_{2,t}\right]}$$

which, because all innovations have zero mean, is

$$\rho_{12,t} = \frac{\text{E}_{t-1}\left[\varepsilon_{1,t}\varepsilon_{2,t}\right]}{\left(\text{E}_{t-1}\left[\varepsilon_{1,t}^2\right]\text{E}_{t-1}\left[\varepsilon_{2,t}^2\right]\right)^{1/2}}$$

Standardized innovations and the relationship between conditional correlation and conditional covariance III

- Note that

$$\varepsilon_{i,t} = \sigma_{i,t} v_{i,t}$$

which implies

$$\begin{aligned}
\rho_{12,t} &= \frac{\mathrm{E}_{t-1}\left[\varepsilon_{1,t}\varepsilon_{2,t}\right]}{\left(\mathrm{E}_{t-1}\left[\varepsilon_{1,t}^{2}\right]\mathrm{E}_{t-1}\left[\varepsilon_{2,t}^{2}\right]\right)^{1/2}} \\
&= \frac{\mathrm{E}_{t-1}\left[\sigma_{1,t}v_{1,t}\sigma_{2,t}v_{2,t}\right]}{\left(\mathrm{E}_{t-1}\left[\sigma_{11,t}v_{1,t}^{2}\right]\mathrm{E}_{t-1}\left[\sigma_{22,t}v_{2,t}^{2}\right]\right)^{1/2}} \\
&= \frac{\mathrm{E}_{t-1}\left[v_{1,t}v_{2,t}\right]}{\left(\mathrm{E}_{t-1}\left[v_{1,t}^{2}\right]\mathrm{E}_{t-1}\left[v_{2,t}^{2}\right]\right)^{1/2}} \\
&= \mathrm{E}_{t-1}\left[v_{1,t}v_{2,t}\right]
\end{aligned}$$

Standardized innovations and the relationship between conditional correlation and conditional covariance IV

- This means that the conditional correlation between the unstandardized innovations is the same as the conditional covariance of underlying zero-mean unit-variance innovations
- It proves that $\rho_{12,t} = \sigma^{v}_{12,t}$ where $\sigma^{v}_{12,t}$ is the covariance of the standardized innovations v
- The constant-conditional correlation GARCH model assumes that the underlying innovations have a constant correlation
- This need not be the case though
- One easily can imagine the correlation of the shocks changing over time
- The expression above shows that allowing the conditional correlation of the ε's to change is the same as allowing the conditional correlations between the v's to change over time
- There are various ways of specifying this correlation and covariance

Standardized innovations and the relationship between conditional correlation and conditional covariance V

- Engle (2002) suggests specifying simple GARCH processes for the variances and the correlation
- For example, specify GARCH(1,1) models for the underlying variances and conditional correlation
- A simple GARCH process that could work for many assets is

$$\sigma_{11,t} = \gamma_{1,0} + \gamma_{1,1}\varepsilon^2_{1,t-1} + \delta_{1,1}\sigma_{11,t-1}$$
$$\sigma_{22,t} = \gamma_{2,0} + \gamma_{2,1}\varepsilon^2_{2,t-1} + \delta_{2,1}\sigma_{22,t-1}$$
$$\sigma^v_{12,t} = \gamma_{3,0} + \gamma_{3,1}v_{1,t-1}v_{2,t-1} + \delta_{3,1}\sigma^v_{12,t-1}$$

- The last equation is the same as the correlation because the variances of the standardized innovations are unity

Restriction of unconditional correlation to sample correlation

- Rather than let the implied unconditional (also "'long run") correlation float, the equation can be constrained so that the correlation in the data $\bar{\rho}_{12}$ is implied
- This requires modifying the third equation to

$$\sigma^{v}_{12,t} = \bar{\rho}_{12} + \gamma_{3,1}\left(v_{1,t-1}v_{2,t-1} - \bar{\rho}_{12}\right) + \gamma_{3,2}\left(\sigma^{v}_{12,t-1} - \bar{\rho}_{12}\right)$$

- This estimator has very nice properties
- In particular, a positive definite variance-covariance matrix
  - More complicated models are possible
  - The key is to keep the unconditional correlation equal to the sample correlation

Estimation of Dynamic Conditional Correlation GARCH model

- Engle shows that the model can be estimated consistently by maximum likelihood
- Other estimation methods are possible and relatively simple
- For example, estimate the GARCH equations for the variances first and then estimate the equation for the correlation
  - First estimate GARCH equations for variances by maximum likelihood
  - Then estimate GARCH equation for correlation by maximum likelihood conditional on the estimated GARCH equations for the variances
  - Engle (2002) shows that this estimator is consistent under standard conditions although it is inefficient
  - With this simple structure, the process is not subject to the curse of dimensionality that affects a general process

A simple way of dealing with ARCH

when it is not the center of attention

- Suppose we have a multivariate relationship between variables that we want to estimate
  - Vector Autoregression or Error Correction Mechanism or some other possibly more complicated relationship
- We discover that the variables have a time-varying variance
- We are not especially interested in the variance
- We think it will complicate things and we will have trouble with estimating the mean model and GARCH

A simple way of dealing with ARCH I

when it is not the center of attention

- How to deal with ARCH and not estimate it at the same time as multivariate model
- A simplification: Suppose that each variance has a diagonal GARCH representation
- This representation can be written for each variable i as

$$r_{i,t} = \mu_{i,t} + u_{i,t}$$
$$\sigma^2_{i,t} = \alpha_i + \gamma_i u^2_{i,t-1} + \delta_i \sigma^2_{i,t-1}$$

where $\mu_{i,t}$ is a predictable part of the mean and $u_{i,t}$ has GARCH

- Estimate this model separately for each of the variables
- After creating these estimates, standardize the original variable by the estimated variance

$$r^s_{i,t} = \frac{r_{i,t}}{\sigma^2_{i,t}}$$

A simple way of dealing with ARCH II

when it is not the center of attention

- Now use $r^s_{i,t}$ is the multivariate model
- If GARCH removes all autoregressive conditional heteroskedasticity, the standardized variable has a constant variance and estimation can proceed on that assumption
- This really is a simplification and estimates of the variance function and the multivariate mean function are estimated consistently only if those parts of the process are uncorrelated
- Even so, this is used by various people and works reasonably well when a multivariate GARCH model would be intractable

Summary

- Multivariate GARCH models can reflect very complex interactions
- Multivariate GARCH models can have an incredibly large number of parameters
- Multivariate GARCH models can have problems with the analogue of a positive variance: a positive definite variance-covariance matrix
- As a result of these issues, they are seldom used with more a few assets
- With many assets, what seems to be the most common solution is to estimate IGARCH models and check for positive definiteness of the estimated variance-covariance matrix

The end

Notes on running svol.prg

- In Windows Explorer, go to directory with svol.prg and double-click it
- The rest just happens