Volatility Models

Leonid Kogan

MIT, Sloan

15.450, Fall 2010
Outline

1. Heteroscedasticity
2. GARCH
3. GARCH Estimation: MLE
4. GARCH: QMLE
5. Alternative Models
6. Multivariate Models
Outline

1. Heteroscedasticity
2. GARCH
3. GARCH Estimation: MLE
4. GARCH: QMLE
5. Alternative Models
6. Multivariate Models
Example: S&P GSCI Index

- Model daily changes in S&P GSCI index.
- The S&P GSCI index is a composite commodity index, maintained by S&P.

“The S&P GSCI® provides investors with a reliable and publicly available benchmark for investment performance in the commodity markets. The index is designed to be tradable, readily accessible to market participants, and cost efficient to implement. The S&P GSCI is widely recognized as the leading measure of general commodity price movements and inflation in the world economy.”

Source: Standard & Poor’s.

- Changes in daily spot index levels:

\[ z_t = \ln \frac{P_t}{P_{t-1}} \]
Example: S&P GSCI Index

S&P GSCI Spot Index

Date

Index Level

© Leonid Kogan (MIT, Sloan)
Volatility Models
15.450, Fall 2010
Example: S&P GSCI Index

- Daily changes from 02-Jan-2004 to 23-Sep-2009.
- First, fit an $AR(p)$ model to the series $z_t$ to extract shocks.
- De-mean the series: $x_t = z_t - \hat{E}[z_t]$. Set $p = 13$.
- BIC criterion shows that $z_t$ has no AR structure. AIC criterion is virtually flat.
- AR coefficients are very small.
- Treat $x_t$ as a serially uncorrelated shock series.
Example: S&P GSCI Index

- While $x_t$’s may be uncorrelated, they may not be IID.
- Look for evidence of heteroscedasticity: time-varying conditional variance.
- Perform the Engle test, e.g., Tsay, 2005 (Section 3.3.1).
Engle Test for Conditional Heteroscedasticity

- The idea of the test is simple: fit the AR(p) model to squared shocks and test the hypothesis that all coefficients are jointly zero.

\[ x_t^2 = a_0 + a_1 x_{t-1}^2 + \ldots + a_p x_{t-p}^2 + u_t \]

- One way to derive the test statistic:
  - Estimate the coefficients of the AR(p) model, \( \hat{\theta} = (\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p) \).
  - Estimate the var-cov matrix of the coefficients \( \hat{\Omega} \). Don’t worry about autocorrelation, since under the null it is not there.
  - Form the test statistic

\[
F = (\hat{a}_1, \ldots, \hat{a}_p) \left[ \hat{\Omega}_{\hat{a}_1,\ldots,\hat{a}_p;\hat{a}_1,\ldots,\hat{a}_p} \right]^{-1} \begin{pmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_p \end{pmatrix}
\]

- Rejection region: \( F \geq F^* \). Size of the test based on the asymptotic distribution: \( F \sim \chi^2(p) \).
Engle Test

MATLAB® code

Lags = [1:1:5];
[H, pValue, ARCHstat, CriticalValue] = archtest(x, Lags, []);

MATLAB® output

pValue =

1.0e-006 *

0.1031  0.0000  0  0  0
Outline

1. Heteroscedasticity
2. GARCH
3. GARCH Estimation: MLE
4. GARCH: QMLE
5. Alternative Models
6. Multivariate Models
Consider a widely used model of time-varying variance: GARCH(p,q) (generalized autoregressive conditional heteroskedasticity).

Consider a series of observations

\[ x_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1), \text{ IID} \]

Assume that the series of conditional variances \( \sigma_t^2 \) follows

\[ \sigma_t^2 = a_0 + \sum_{i=1}^{p} a_i x_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2, \quad a_i, b_j \geq 0 \]

Focus on a popular special case GARCH(1,1).
GARCH(1,1) Dynamics

Let $E_t(\cdot)$ denote the conditional expectation given time-$t$ information.

\[
E_t \left[ \sigma^2_{t+1} \right] = E_t \left[ a_0 + a_1 x_t^2 + b_1 \sigma_t^2 \right] = a_0 + (a_1 + b_1) \sigma_t^2
\]

\[
E_t \left[ \sigma^2_{t+2} \right] = E_t \left[ a_0 + (a_1 + b_1) \sigma^2_{t+1} \right]
= a_0[1 + (a_1 + b_1)] + (a_1 + b_1)^2 \sigma_t^2
\]

\[
E_t \left[ \sigma^2_{t+3} \right] = E_t \left[ a_0 + (a_1 + b_1) \sigma^2_{t+2} \right]
= a_0[1 + (a_1 + b_1) + (a_1 + b_1)^2] + (a_1 + b_1)^3 \sigma_t^2
\]

\[
\vdots
\]

\[
E_t \left[ \sigma^2_{t+n} \right] = a_0 \frac{1 - (a_1 + b_1)^n}{1 - a_1 - b_1} + (a_1 + b_1)^n \sigma_t^2
\]
GARCH(1,1) Dynamics

- Stable dynamics requires
  \[ a_1 + b_1 < 1 \]

- Convergence of forecasts:
  \[
  \lim_{n \to \infty} E_t \left[ \sigma_{t+n}^2 \right] = \frac{a_0}{1 - a_1 - b_1}
  \]

- Average conditional variance:
  \[
  E \left[ x_{t+1}^2 \right] = a_0 + a_1 E \left[ x_t^2 \right] + b_1 E \left[ \sigma_t^2 \right] \Rightarrow E \left[ x_t^2 \right] = \frac{a_0}{1 - a_1 - b_1}
  \]
GARCH(1,1) Monte Carlo

- Unconditional distribution of $x_t$ has heavier tails than the conditional (Gaussian) distribution.
- Monte Carlo experiment: simulate GARCH (1,1) process with parameters

$$a_0 = 1, \quad a_1 = 0.1, \quad b_1 = 0.8$$

- Initiate $\sigma_1 = \sqrt{\frac{a_0}{1-a_1-b_1}}$.
- Generate a sample of 100,000 observations using dynamics

$$\sigma_t^2 = a_0 + a_1 x_{t-1}^2 + b_1 \sigma_{t-1}^2$$

$$x_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1), \text{ IID}$$

- Drop the first 10% of the simulated sample (burn-in) and analyze the distribution of the remaining sample.
MATLAB® Code

```matlab
sigma(1) = InitValue;  % Initialize
for t = 1:1:T
    x(t) = sigma(t)*randn(1,1);
    sigma(t+1) = sqrt(a0 + b1*sigma(t)^2 + a1*x(t)^2);
end

x(1:floor(T/10)) = [];
% Drop burn-in sample
x = x./std(x);
% Normalize x
for k=1:1:4
    A(1,k) = mean(x>k);
    % Estimate tails of x
    A(2,k) = 1-normcdf(k);
    % Compare to Gaussian distribution
end
```
GARCH(1,1) Monte Carlo

- Compare the tails of the simulated sample to the Gaussian distribution:

\[
\text{Prob} \left[ \frac{x_t}{\sqrt{\mathbb{E} (x_t^2)}} > k \right]
\]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH(1,1)</td>
<td>0.1540</td>
<td>0.0239</td>
<td>0.0025</td>
<td>0.0002</td>
</tr>
<tr>
<td>Gaussian</td>
<td>0.1587</td>
<td>0.0228</td>
<td>0.0013</td>
<td>0.0000</td>
</tr>
<tr>
<td>S&amp;P GSCI</td>
<td>0.1351</td>
<td>0.0188</td>
<td>0.0077</td>
<td>0</td>
</tr>
</tbody>
</table>
Outline

1. Heteroscedasticity
2. GARCH
3. GARCH Estimation: MLE
4. GARCH: QMLE
5. Alternative Models
6. Multivariate Models
MLE for GARCH(1,1)

- Focus on GARCH(1,1) as a representative example.
- Estimate parameters by maximizing conditional log-likelihood.
- Form the log-likelihood function:

\[
\mathcal{L}(\theta) = \sum_{t=1}^{T} \ln p(x_t|\sigma_t; \theta)
\]

- \(p(x_t|\sigma_t; \theta)\) is the normal density

\[
p(x_t|\sigma_t; \theta) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{x_t^2}{2\sigma_t^2}}
\]
MLE for GARCH(1,1)

Likelihood function for GARCH(1,1)

\[ \mathcal{L}(\theta) = \sum_{t=1}^{T} -\ln \sqrt{2\pi} - \frac{x_t^2}{2\sigma_t^2} - \frac{1}{2} \ln (\sigma_t^2) \]

\[ \sigma_t^2 = a_0 + a_1 x_{t-1}^2 + b_1 \sigma_{t-1}^2 \]

- Need \( \sigma_1^2 \) to complete the definition of \( \mathcal{L}(\theta) \).
  - The exact value of \( \sigma_1^2 \) does not matter in large samples, since \( \sigma_t^2 \) converges to its stationary distribution for large \( t \).
  - A reasonable guess for \( \sigma_1^2 \) improves accuracy in finite samples.
  - Use unconditional sample variance: \( \sigma_1^2 = \hat{\text{E}}[x_t^2] \).

- Impose constraints on the parameters to guarantee stationarity.
- MLE-based estimates:

\[ \hat{\theta} = \arg \max_{(a_0,a_1,b_1)} \mathcal{L}(\theta) \]

subject to \( a_1 \geq 0, b_1 \geq 0, a_1 + b_1 < 1 \)
Example: S&P GSCI

- Fit the GARCH(1,1) model to the series of S&P GSCI spot price changes.
- Use MATLAB® function `garchfit`. `garchfit` constructs the likelihood function and optimizes it numerically.
- Parameter estimates:
  \[ a_1 = 0.0453, \quad b_1 = 0.9457 \]
- Shocks to conditional variance are persistent, giving rise to volatility clustering.
Example: S&P GSCI

- Fitted time series of *conditional volatility* $\hat{\sigma}_t$ computed using

$$\hat{\sigma}_t^2 = a_0 + a_1 x_{t-1}^2 + b_1 \hat{\sigma}_{t-1}^2$$
Example: S&P GSCI

- Extract a series of fitted errors

\[ \hat{\varepsilon}_t = \frac{x_t}{\hat{\sigma}_t} \]

<table>
<thead>
<tr>
<th>k</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>0.1587</td>
<td>0.0228</td>
<td>0.0013</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \hat{\varepsilon}_t )</td>
<td>0.1595</td>
<td>0.0209</td>
<td>0.0014</td>
<td>0</td>
</tr>
</tbody>
</table>

- Fitted errors conform much better to the Gaussian distribution than the unconditional distribution of \( x_t \) does.

- In case of S&P GSCI spot price series, can attribute heavy tails in unconditional distribution of daily changes to conditional heteroscedasticity.
Standard Errors

- We treat MLE as a special case of GMM with moment conditions:
  \[
  \hat{E} \left[ \frac{\partial \ln p(x_t | \sigma_t; \theta)}{\partial \theta} \right] = 0
  \]

- Use general formulas for standard errors:
  \[
  \hat{d} = \hat{E} \left[ \frac{\partial^2 \ln p(x, \hat{\theta})}{\partial \theta \partial \theta'} \right], \quad \hat{S} = \hat{E} \left[ \frac{\partial \ln p(x, \hat{\theta})}{\partial \theta} \frac{\partial \ln p(x, \hat{\theta})}{\partial \theta'} \right]
  \]
  \[
  TVar[^{\hat{\theta}}] = \left( \hat{d}' \hat{S}^{-1} \hat{d} \right)^{-1}
  \]

- How to compute derivatives, e.g., \( \frac{\partial \ln p(x, \hat{\theta})}{\partial \theta} \)?
  - Use finite-difference approximations (\textit{garchfit}).
  - Compute derivatives analytically, recursively (discussed in recitations).
Outline

1. Heteroscedasticity
2. GARCH
3. GARCH Estimation: MLE
4. GARCH: QMLE
5. Alternative Models
6. Multivariate Models
Standard GARCH formulation assumes that errors $\varepsilon_t$ are Gaussian.

Assume that $x_t$ follow a different distribution, but still

$$x_t = \sigma_t \varepsilon_t, \quad E_t[\varepsilon_t] = 0, \quad E_t[\varepsilon_t^2] = 1$$

Two approaches:
- QMLE estimation, treating errors as Gaussian.
- MLE with an alternative distribution for $\varepsilon_t$, e.g. Student's $t$. 
GARCH: QMLE

- Keep using the objective function

\[ \mathcal{L}(\theta) = \sum_{t=1}^{T} -\ln \sqrt{2\pi} - \frac{x_t^2}{2\sigma_t^2} - \frac{1}{2} \ln (\sigma_t^2) \]

- Because the function \( x \mapsto -\ln x - a/x \) is maximized at \( x = a \), conditional expectation

\[ E_t \left[ -\frac{x_t^2}{2\sigma_t^2(\theta)} - \frac{1}{2} \ln (\sigma_t^2(\theta)) \right] \]

is maximized at the true value of \( \theta \). This means that \( \theta_0 \) maximizes the unconditional expectation as well, and hence we can estimate it by maximizing \( \mathcal{L}(\theta) \).
GARCH: MLE with Student’s $t$ Shocks

- One prominent example of GARCH with non-Gaussian errors is the GARCH model with Student’s $t$ error distribution.
- Assume that
  \[ p(\varepsilon_t; \nu) = \frac{\Gamma[(\nu + 1)/2]}{\Gamma(\nu/2)\sqrt{\pi(\nu - 2)}} \left(1 + \frac{\varepsilon_t^2}{\nu - 2}\right)^{-(\nu+1)/2}, \quad \nu > 2 \]

  \[ \sqrt{\frac{\nu}{(\nu - 2)}}\varepsilon_t \] have the Student’s $t$ distribution with $\nu$ degrees of freedom. $\Gamma$ is the Gamma function, $\Gamma(x) = \int_0^\infty z^{x-1}e^{-z} \, dz$.

- Likelihood function for GARCH(1,1):
  \[ \mathcal{L}(\theta) = \sum_{t=1}^T \ln \left(\frac{\Gamma[(\nu + 1)/2]}{\Gamma(\nu/2)\sqrt{\pi(\nu - 2)}}\right) - \frac{\nu + 1}{2} \ln \left(1 + \frac{x_t^2}{(\nu - 2)\sigma_t^2}\right) - \ln(\sigma_t^2)/2 \]
GARCH: Non-Gaussian Errors

- Student’s $t$ distribution has heavier tails than the Gaussian distribution.
- The number of degrees of freedom can be estimated together with other parameters, or it can be fixed.
- GARCH models generate heavy tails in the unconditional distribution, Student’s $t$ adds heavy tails to the conditional distribution.
- Daily S&P 500 returns: capture unconditional distribution of shocks as Student’s $t$ with $\nu \approx 3$; GARCH(1,1) captures conditional distribution of shocks as Student’s $t$ with $\nu \approx 6$. 
How effective is the QMLE approach when dealing with non-normal shocks?
We can gain intuition using Monte Carlo experiments.
Beyond this particular context, our Monte Carlo design illustrates a typical simulation experiment.
Monte Carlo Design

- Data Generating Process:
  \[ \sigma_t^2 = a_0 + a_1 x_{t-1}^2 + b_1 \sigma_{t-1}^2 \]
  \[ a_1 = 0.05, \quad b_1 = 0.9 \]

  \( \varepsilon_t \) are IID, Student’s \( \nu \) distribution with \( \nu = 6 \).

- Simulate \( N = 1,000 \) samples of length \( T = 1,000 \) or \( 3,000 \).

- In each case, start with \( \sigma_1 = \sqrt{\frac{a_0}{1-a_1-b_1}} \) and use a burn-in sample of 500 periods.

- Perform MLE and QMLE estimations for each simulated sample and save point estimates \( \hat{a}_1, \hat{b}_1 \), and their standard errors.
Summary Statistics

Compute the following statistics:

- Root-mean-squared-error (RMSE) of each parameter estimate

\[
\text{RMSE}(\hat{\theta}) = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (\hat{\theta}_n - \theta_0)^2}
\]

- Average value of each parameter estimate

\[
\frac{1}{N} \sum_{n=1}^{N} \hat{\theta}_n
\]

- Estimated coverage probability of the confidence interval for each parameter estimate

\[
\frac{1}{N} \sum_{n=1}^{N} 1[|\hat{\theta}_n - \theta_0| \leq 1.96 \text{ s.e.}(\hat{\theta})]
\]
Monte Carlo Results

<table>
<thead>
<tr>
<th>$T$</th>
<th>Method</th>
<th>RMSE($\hat{\theta}$)</th>
<th>Mean($\hat{\theta}$)</th>
<th>C.I.($\hat{\theta}$) Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\hat{a}_1$ $\hat{b}_1$</td>
<td>$\hat{a}_1$ $\hat{b}_1$</td>
<td>$\hat{a}_1$ $\hat{b}_1$</td>
</tr>
<tr>
<td>1,000</td>
<td>QMLE</td>
<td>0.0286 0.1951</td>
<td>0.0551 0.8335</td>
<td>0.9240 0.8930</td>
</tr>
<tr>
<td>1,000</td>
<td>MLE</td>
<td>0.0228 0.1396</td>
<td>0.0538 0.8613</td>
<td>0.9170 0.8920</td>
</tr>
<tr>
<td>3,000</td>
<td>QMLE</td>
<td>0.0149 0.0356</td>
<td>0.0513 0.8920</td>
<td>0.9380 0.9200</td>
</tr>
<tr>
<td>3,000</td>
<td>MLE</td>
<td>0.0115 0.0289</td>
<td>0.0503 0.8940</td>
<td>0.9370 0.9390</td>
</tr>
</tbody>
</table>

- Both QMLE and MLE produce consistent parameter estimates.
- At $T = 1,000$ there is a bias, which disappears at $T = 3,000$.
- MLE estimates are more efficient: smaller RMSE.
- QMLE estimates do not rely on the exact distribution, more robust.
- QMLE confidence intervals are reliable, GMM formulas work.
Outline

1. Heteroscedasticity
2. GARCH
3. GARCH Estimation: MLE
4. GARCH: QMLE
5. Alternative Models
6. Multivariate Models
Empirically, conditional volatility of asset returns often reacts asymmetrically to the past realized return shocks.

**Leverage effect**: conditional stock market volatility increases following a stock market decline.

EGARCH\((p,q)\) model captures the asymmetric volatility response:

\[
\ln \sigma_t = a_0 + \sum_{i=1}^{p} a_i g \left( \frac{x_{t-i}}{\sigma_{t-i}} \right) + \sum_{j=1}^{q} b_j \ln \sigma_{t-j} \quad \text{(EGARCH}(p,q))
\]

\[g(z) = |z| - c \cdot z\]
Mixed Data Sampling (MIDAS)

Motivation

- Suppose we want to predict realized variance over a single holding period of the portfolio, which is a month.
- GARCH approach:
  - Use monthly historical data, ignore the available higher-frequency (daily) data; or
  - Model daily volatility and extend the forecast to a one-month period. Sensitive to specification errors.
- **Mixed Data Sampling** approach forecasts monthly variance directly using daily data.
Mixed Data Sampling

Formulation

- We are interested in forecasting an $H$-period volatility measure, $V_{t+H,t}^H$, e.g., sum of squared daily returns over a month ($H = 22$).
- Model expected monthly volatility measure as a weighted average of lagged daily observations (e.g., use squared daily returns)

$$V_{t+H,t}^H = a_H + \phi_H \sum_{k=0}^{K} b_H(k, \theta) X_{t-k, t-k-1} + \varepsilon_{Ht}$$

- Significant flexibility:
  - $X$ can contain squared return, absolute value of returns, intra-day high-low range, etc.
  - Weights $b_H(k, \theta)$ can be flexibly specified.
Mixed Data Sampling

Estimation

- The model
  \[ V_{t+H,t}^H = a_H + \phi_H \sum_{k=0}^{K} b_H(k, \theta) X_{t-k, t-k-1} + \varepsilon_{Ht} \]

  Estimate using nonlinear least squares (NLS).

- Alternative specification:
  \[ r_{t+H,t} \sim \mathcal{N}\left( \mu, a_H + \phi_H \sum_{k=0}^{K} b_H(k, \theta) X_{t-k, t-k-1} \right) \]

  Estimate the parameters using QMLE.
**Mixed Data Sampling**

**Example**

- Beta-function specification of the weights $b_H(k, \theta)$:

$$b_H(k, \theta) = \frac{f \left( \frac{k}{K}, \alpha, \beta \right)}{\sum_{j=0}^{K} f \left( \frac{j}{K}, \alpha, \beta \right)}$$

$$f(x, \alpha, \beta) = x^\alpha (1 - x)^\beta$$

Weights $b_H(k, \theta)$ have flexible shape.
Outline

1. Heteroscedasticity
2. GARCH
3. GARCH Estimation: MLE
4. GARCH: QMLE
5. Alternative Models
6. Multivariate Models
Multivariate Volatility Models

Overview

Model the dynamics of conditional variance-covariance matrix of the time series

\[ x_t = \Omega_t^{1/2} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, I) \]

Many multivariate generalizations of GARCH framework. Main challenge is parameter proliferation.

Use factor structures to treat high-dimensional cases.

Averaging of realized covariances (exponentially weighted moving average, MIDAS framework).
MGARCH

Example

- Multivariate GARCH analog

\[ \Omega_t = C + a(x_{t-1}x'_{t-1}) + b\Omega_{t-1} \]

- Estimate using QMLE, analogous to GARCH(1,1).

- Limitation: all covariances have the same persistence.
Constant Conditional Correlations (CCC)

- Model
  \[ \Omega_t = D_t \Gamma D_t \]
  \( \Gamma \) is the *constant* matrix of conditional correlations; 
  \( D_t \) is the diagonal matrix of conditional standard deviations.

- Two-step estimation method:
  - Fit a scalar GARCH(1,1) to each component of \( x \) to estimate \( D_t \);
  - Estimate the unconditional correlation matrix of \( \hat{u}_t, \hat{u}_t = \hat{D}_t^{-1} x_t \)

\[ \hat{\Gamma} = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{u}'_t \]
Dynamic Conditional Correlations (DCC)

Model

\[ \Omega_t = D_t \Gamma_t D_t \]

\( \Gamma_t \) is the \textit{time-varying} conditional correlation matrix;
\( D_t \) is the diagonal matrix of conditional standard deviations.

Two-step estimation method:

- Fit a scalar GARCH(1,1) to each component of \( x \) to estimate \( D_t \);
- Model \( \Gamma_t \) as

\[
(\hat{\Gamma}_t)_{ij} = \frac{(Q_t)_{ij}}{\sqrt{(Q_t)_{ii}(Q_t)_{jj}}}, \quad Q_t = (1 - a - b)\bar{\Gamma} + a(\hat{u}_{t-1}\hat{u}_{t-1}') + bQ_{t-1}
\]

Estimate the parameters \( \bar{\Gamma}, a, b \) by QMLE on the \( \hat{u}_t \) series. As before, \( \hat{u}_t = \hat{D}_t^{-1} x_t \).
Summary

- Volatility models are important for risk management, asset allocation, derivative pricing.
- GARCH models are convenient for extracting time-varying volatility and for forecasting.
- GARCH models can be estimated using QMLE or MLE.
- Mixed-frequency data can be used in forecasting. MIDAS. Straightforward using NLS or QMLE.
- Multiple extensions of GARCH, multivariate models.
Readings

  
  Note: there are typos in eq. (12.2.19).

- Tsay, 2005, Sections 3.3-3.5, 3.8.

