Volatility Models

Leonid Kogan

MIT, Sloan

15.450, Fall 2010

Outline

Outline

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.

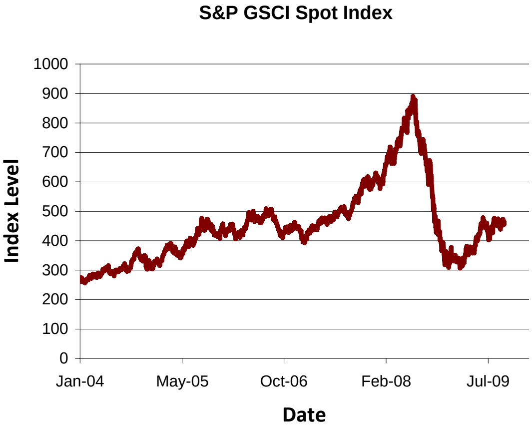
  "The S&P GSCI® provides investors with a reliable and publicly available benchmark for investment performance in the commodity markets. The index is designed to be tradable, readily accessible to market participants, and cost efficient to implement. The S&P GSCI is widely recognized as the leading measure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



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: $x_t = z_t - \widehat{\mathrm{E}}[z_t]$. Set $\overline{p} = 13$.
- 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, $\widehat{\theta} = (\widehat{a}_0, \widehat{a}_1, \ldots, \widehat{a}_p)$.
  - Estimate the var-cov matrix of the coefficients $\widehat{\Omega}$. Don't worry about autocorrelation, since under the null it is not there.
  - Form the test statistic

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

  - Rejection region: $F \geqslant \overline{F}$. 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

GARCH(p,q)

- 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 σ_t^2 follows

$$\sigma_t^2 = a_0 + \sum_{i=1}^{p} a_i x_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2, \quad a_i,\ b_j \geqslant 0 \qquad \text{(GARCH(p,q))}$$

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

GARCH(1,1) Dynamics

Let $\mathrm{E}_t(\cdot)$ denote the conditional expectation given time-t information.

$$\mathrm{E}_t\left[\sigma_{t+1}^2\right] = \mathrm{E}_t\left[a_0 + a_1 x_t^2 + b_1\sigma_t^2\right] = a_0 + (a_1 + b_1)\sigma_t^2$$

$$\begin{aligned}\mathrm{E}_t\left[\sigma_{t+2}^2\right] &= \mathrm{E}_t\left[a_0 + (a_1 + b_1)\sigma_{t+1}^2\right] \\ &= a_0[1 + (a_1 + b_1)] + (a_1 + b_1)^2\sigma_t^2\end{aligned}$$

$$\begin{aligned}\mathrm{E}_t\left[\sigma_{t+3}^2\right] &= \mathrm{E}_t\left[a_0 + (a_1 + b_1)\sigma_{t+2}^2\right] \\ &= a_0[1 + (a_1 + b_1) + (a_1 + b_1)^2] + (a_1 + b_1)^3\sigma_t^2 \\ &\vdots \\ \mathrm{E}_t\left[\sigma_{t+n}^2\right] &= a_0\frac{1 - (a_1 + b_1)^n}{1 - a_1 - b_1} + (a_1 + b_1)^n\sigma_t^2\end{aligned}$$

GARCH(1,1) Dynamics

- Stable dynamics requires

$$a_1 + b_1 < 1$$

- Convergence of forecasts:

$$\lim_{n\to\infty} \mathrm{E}_t\left[\sigma^2_{t+n}\right] = \frac{a_0}{1 - a_1 - b_1}$$

- Average conditional variance:

$$\mathrm{E}\left[x^2_{t+1}\right] = a_0 + a_1 \mathrm{E}\left[x^2_t\right] + b_1 \mathrm{E}\left[\sigma^2_t\right] \;\Rightarrow\; \mathrm{E}\left[x^2_t\right] = \frac{a_0}{1 - a_1 - b_1}$$

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

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

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

$$\sigma_t^2 = a_0 + a_1 x_{t-1}^2 + b_1 \sigma_{t-1}^2$$
$$x_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1), \text{ IID}$$

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

GARCH(1,1) Monte Carlo

MATLAB® Code

```
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:

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

k	1	2	3	4
GARCH(1,1)	0.1540	0.0239	0.0025	0.0002
Gaussian	0.1587	0.0228	0.0013	0.0000
S&P GSCI	0.1351	0.0188	0.0077	0

Outline

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)

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

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

- Need σ_1^2 to complete the definition of $\mathcal{L}(\theta)$.
  - The exact value of σ_1^2 does not matter in large samples, since σ_t^2 converges to its stationary distribution for large t.
  - A reasonable guess for σ_1^2 improves accuracy in finite samples.
  - Use unconditional sample variance: $\sigma_1^2 = \widehat{\mathrm{E}}[x_t^2]$.
- Impose constraints on the parameters to guarantee stationarity.
- MLE-based estimates:

$$\widehat{\theta} = \arg \max_{(a_0, a_1, b_1)} \mathcal{L}(\theta)$$

$$\text{subject to } a_1 \geqslant 0,\ b_1 \geqslant 0,\ a_1 + b_1 < 1$$

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:

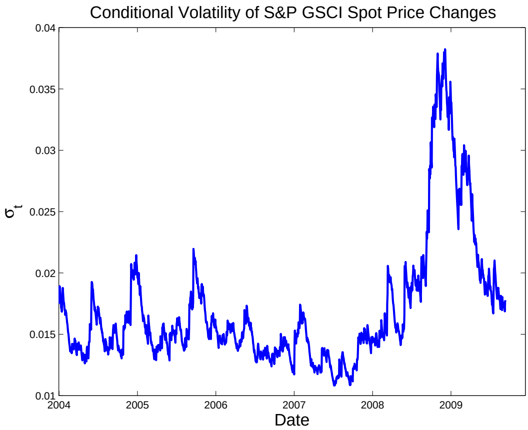
$$a_1 = 0.0453, \quad b_1 = 0.9457$$

- Shocks to conditional variance are persistent, giving rise to *volatility clustering*.

Example: S&P GSCI

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

$$\widehat{\sigma}_t^2 = a_0 + a_1 x_{t-1}^2 + b_1 \widehat{\sigma}_{t-1}^2$$



Example: S&P GSCI

- Extract a series of fitted errors

$$\widehat{\varepsilon}_t = \frac{x_t}{\widehat{\sigma}_t}$$

Tail Probabilities (Prob[$\widehat{\varepsilon}_t > k$])

k	1	2	3	4
Gaussian	0.1587	0.0228	0.0013	0.0000
$\widehat{\varepsilon}_t$	0.1595	0.0209	0.0014	0

- 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

$$\widehat{\mathrm{E}}\left[\frac{\partial \ln p(x_t|\sigma_t;\theta)}{\partial\theta}\right] = 0$$

- Use general formulas for standard errors:

$$\widehat{d} = \widehat{\mathrm{E}}\left[\frac{\partial^2 \ln p(x,\widehat{\theta})}{\partial\theta\partial\theta'}\right], \quad \widehat{S} = \widehat{\mathrm{E}}\left[\frac{\partial \ln p(x,\widehat{\theta})}{\partial\theta}\frac{\partial \ln p(x,\widehat{\theta})}{\partial\theta'}\right]$$

$$T\mathrm{Var}[\widehat{\theta}] = \left(\widehat{d}'\widehat{S}^{-1}\widehat{d}\right)^{-1}$$

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

Outline

GARCH: Non-Gaussian Errors

- Standard GARCH formulation assumes that errors ε_t are Gaussian.
- Assume that x_t follow a different distribution, but still

$$x_t = \sigma_t \varepsilon_t, \quad \mathrm{E}_t[\varepsilon_t] = 0, \quad \mathrm{E}_t[\varepsilon_t^2] = 1$$

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

GARCH: QMLE

- Keep using the objective function

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

- Because the function $x \longmapsto -\ln x - a/x$ is maximized at $x = a$, conditional expectation

$$\mathrm{E}_t\left[-\frac{x_t^2}{2\sigma_t^2(\theta)} - \frac{1}{2}\ln\left(\sigma_t^2(\theta)\right)\right]$$

is maximized at the true value of θ. This means that θ_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

$$p(\varepsilon_t;\nu) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}}\left(1+\frac{\varepsilon_t^2}{\nu-2}\right)^{-(\nu+1)/2}, \quad \nu > 2$$

  $\sqrt{\nu/(\nu-2)}\,\varepsilon_t$ have the Student's *t* distribution with ν degrees of freedom. Γ 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 $\nu \approx 6$.

QMLE vs. MLE: Monte Carlo Experiments

- 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:

$$\sigma_t^2 = a_0 + a_1 x_{t-1}^2 + b_1 \sigma_{t-1}^2$$
$$a_1 = 0.05, \quad b_1 = 0.9$$

  ε_t are IID, Student's t 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 $\widehat{a}_1$, $\widehat{b}_1$, and their standard errors.

Summary Statistics

Compute the following statistics:

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

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

- Average value of each parameter estimate

$$\frac{1}{N}\sum_{n=1}^{N}\widehat{\theta}_n$$

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

$$\frac{1}{N}\sum_{n=1}^{N}1_{[|\widehat{\theta}_n - \theta_0| \leqslant 1.96\ \text{s.e.}(\widehat{\theta})]}$$

Monte Carlo Results

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

- 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

Other GARCH-Type Models: EGARCH

- 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:

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

$$g(z) = |z| - c\,z$$

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.
- **Mi**xed **Da**ta **S**ampling approach forecasts monthly variance directly using daily data.

Mixed Data Sampling

Formulation

- We are interested in forecasting an H-period volatility measure, $V^H_{t+H,t}$ 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)

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

- 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

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

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.

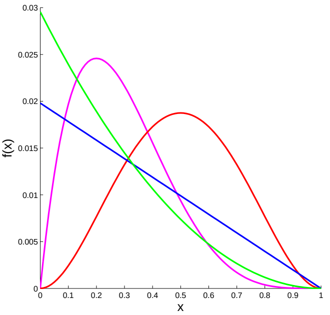
Mixed Data Sampling

Example

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

$$b_H(k,\theta) = \frac{f\left(\frac{k}{K}, \alpha, \beta\right)}{\sum_{j=0}^{K} f\left(\frac{j}{K}, \alpha, \beta\right)}, \quad f(x,\alpha,\beta) = x^{\alpha}(1-x)^{\beta}$$

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



Outline

Multivariate Volatility Models

Overview

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

$$x_t = \Omega_t^{1/2} \varepsilon_t, \quad \varepsilon_t \overset{\text{IID}}{\sim} \mathcal{N}(0, I)$$

- 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

$$\Omega_t = D_t \Gamma D_t$$

  Γ 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 $\widehat{u}_t$, $\widehat{u}_t = \widehat{D}_t^{-1} x_t$

$$\widehat{\Gamma} = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_t \widehat{u}_t'$$

Dynamic Conditional Correlations (DCC)

- Model

$$\Omega_t = D_t \Gamma_t D_t$$

  Γ_t is the *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 Γ_t as

$$(\widehat{\Gamma}_t)_{ij} = \frac{(Q_t)_{ij}}{\sqrt{(Q_t)_{ii}(Q_t)_{jj}}}, \quad Q_t = (1-a-b)\overline{\Gamma} + a(\widehat{u}_{t-1}\widehat{u}'_{t-1}) + bQ_{t-1}$$

    Estimate the parameters $\overline{\Gamma}$, a, b by QMLE on the $\widehat{u}_t$ series. As before, $\widehat{u}_t = \widehat{D}_t^{-1} x_t$.

Summary

- Volatility models are important for risk management, asset allocation, derivative pricing.
- GARCH models are convenient for extracting time-varying volatility and for frecasting.
- 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

- Campbell, Lo, MacKinlay, 1997, Sections 12.2 (Introduction), 12.2.1.

  Note: there are typos in eq. (12.2.19).
- Tsay, 2005, Sections 3.3-3.5, 3.8.
- T. Andersen, T. Bollerslev, P. Christoffersen, F. Diebold, 2006, "Volatility and Correlation Forecasting," in G. Elliott, C. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*. Amsterdam: North-Holland, 778-878.
- E. Ghysels, P. Santa-Clara, R. Valkanov, 2006, "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics* 131, 59-95.

15.450 Analytics of Finance

Fall 2010