Modeling Volatility and Correlation

Yanping YI

November 22, 2012

Volatility

- What is asset volatility?
- Answer: conditional standard deviation of the asset returns
- A key characteristic: Not directly observable!!
- The volatility is **latent** rather than observed, and so is modeled indirectly.

Volatility

Why is volatility important?
Has many important applications

- Many derivative securities depend explicitly on volatility, e.g., Black-Scholes formula
- Risk management measures such as value at risk (VaR) and expected shortfall (ES) depend explicitly on volatility
- Asset allocation, e.g., minimum-variance portfolio
- Modeling the volatility of a time series can improve the efficiency in parameter estimation (e.g. feasible GLS) and the accuracy in interval forecasting (i.e, provide correct standard error bands for forecasts)

Stylized Statistical Properties of Asset Returns

- Absence of autocorrelations: (linear) autocorrelations of asset returns are often insignificant, except for very small intraday time scales for which microstructure effects come into play.
- Heavy tails: the (unconditional) distribution of returns displays heavy tails
- Gain/loss asymmetry: one observes large drawdowns in stock prices and stock index values but not equally large upward movements.

Stylized Statistical Properties of Asset Returns

- Aggregational Gaussianity: as one increases the time scale over which returns are calculated, their distribution looks more and more like a normal distribution. In particular, the shape of the distribution is not the same at different time scales.
- Volatility clustering: different measures of volatility display a positive autocorrelation over several days, which quantifies the fact that high-volatility events tend to cluster in time.
- Volume/volatility correlation: trading volume is correlated with all measures of volatility.

Stylized Statistical Properties of Asset Returns

- Conditional heavy tails: even after correcting returns for volatility clustering (e.g. via GARCH-type models), the residual time series still exhibit heavy tails. However, the tails are less heavy than in the unconditional distribution of returns.
- Slow decay of autocorrelation in absolute returns: the autocorrelation function of absolute returns decays slowly as a function of the time lag (a sign of long-range dependence)
- Leverage effect: most measures of volatility of an asset are negatively correlated with the returns of that asset.
- Volatility co-movements: evidence of common factors to explain volatility in multiple series

Motivation for Volatility Modeling

- The linear structural (and time series) models cannot explain a number of important features common to financial data
- How to measure and model volatility?
- There are several definitions of volatility. Our use of conditional standard deviation is just one of them.
- Econometric modeling: use daily or monthly returns. How to incorporate the information in modeling volatility?

Basic Idea of Econometric Modeling

- Shocks of asset returns are NOT serially correlated, but dependent. That is, the serial dependence is nonlinear.
- As shown by the ACF of returns and absolute returns of some assets

Motivating Example

The daily closing index of the S&P500 index from 1950 to 2010.
The log returns follow approximately an MA(2) model

$$r_t = 0.00028 + a_t + 0.039a_{t-1} - 0.051a_{t-2}$$

- How about the volatility?
- Is volatility constant over time?
- NO! See the ACF of squared residuals!
- Testing for ARCH effects.

Basic Structure

$$r_t = \mu_t + a_t, \quad \underbrace{\mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i a_{t-i}}_{the\ \ mean\ \ equation}$$

Volatility models are concerned with time-evolution of

$$\underbrace{\sigma_t^2 = Var(r_t|F_{t-1}) = Var(a_t|F_{t-1})}_{the\ \ volatility\ \ equation}$$

the conditional variance of a return based on a past information set.

Basic Structure

- Volatility of a return $\sigma_t^2 = Var(r_t|F_{t-1}) = Var(a_t|F_{t-1})$ is a one-period ahead estimate for the variance calculated based on any **past information** thought relevant.
- Relevant past information:
  1. Information about volatility during the previous periods $\{a_{t-1}^2, a_{t-2}^2, \cdots\}$
  2. The fitted variance from the model during the previous periods $\{\sigma_{t-1}^2, \sigma_{t-2}^2, \cdots\}$
  3. Any other past information that you think is relevant
- ARCH models use information only from the first category.
- GARCH models use information both from the first category and from the second category.

How to model the evolving volatility?

The manner under which σ_t^2 evolves over time distinguishes one volatility model from another.

Two general categories

- "Fixed function" of the available information.
- Stochastic function of the available information.

Univariate volatility models discussed

- Autoregressive conditional heteroscedastic (ARCH) model of Engle (1982)
- Generalized ARCH (GARCH) model of Bollerslev (1986)
- GARCH-M models
- IGARCH models (used by RiskMetrics)
- Exponential GARCH (EGARCH) model of Nelson (1991)
- Threshold GARCH model of Zakoian (1994) or GJR model of Glosten, Jagannathan, and Runkle (1993)

Univariate volatility models discussed

- Asymmetric parametric ARCH (APARCH) models of Ding, Granger and Engle (1994), [TGARCH and GJR models are special cases of APARCH models.]
- Stochastic volatility (SV) models of Melino and Turnbull (1990), Harvey, Ruiz and Shephard (1994), and Jacquier, Polson and Rossi (1994).

Autoregressive Conditionally Heteroscedastic (ARCH) Models

- Recall the definition of volatility of a return

$$\sigma_t^2 = Var(r_t|F_{t-1}) = Var(a_t|F_{t-1})$$

- What could the current value of the variance of the errors plausibly depend upon? — Previous squared error terms.
- This leads to the autoregressive conditionally heteroscedastic model for the variance of the errors:

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2$$

where $\alpha_1 \geq 0$ (large shocks tend to be followed by another large shock). This is known as an **ARCH(1)** model.

ARCH Model

We can easily extend ARCH(1) to the general case where the error variance depends on m lags of squared errors:

$$\begin{aligned} r_t &= E(r_t|F_{t-1}) + a_t \quad a_t = \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2 \end{aligned}$$

- This is an ARCH(m) model
- where $\{\epsilon_t\}$ is a sequence of iid r.v. with mean 0 and variance 1, $\alpha_0 > 0$, and $\alpha_i \geq 0$ for $i > 0$.
- Distribution of ϵ_t : standard normal, standardized student-t, generalized error distribution (GED), or their skewed counterparts.

Properties of ARCH Errors

- $\{a_t, F_{t-1}\}$ is a Martingale difference series with conditionally heteroskedastic errors

$$\begin{aligned} E(a_t|F_{t-1}) &= E(\sigma_t\epsilon_t|F_{t-1}) = \sigma_t E(\epsilon_t|F_{t-1}) = 0 \\ Var(a_t|F_{t-1}) &= E(a_t^2|F_{t-1}) = \sigma_t^2 E(\epsilon_t^2|F_{t-1}) = \sigma_t^2 \end{aligned}$$

$\{a_t\}$ is an uncorrelated process:

$$E(a_t a_{t-j}) = E[E(a_t a_{t-j}|F_{t-1})] = E[a_{t-j}E(a_t|F_{t-1})] = 0 \quad j \geq 1$$

Properties of ARCH Errors

- Re-parameterization:
  Let $\eta_t = a_t^2 - \sigma_t^2$. $\{\eta_t\}$ is an uncorrelated series with mean 0. The ARCH model becomes

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2 + \eta_t$$

  This is an AR(m) form for the squared series a_t^2. PACF of a_t^2 is a useful tool to determine the ARCH order m.

Properties of ARCH Errors

The error $\{a_t\}$ is stationary with mean zero and constant unconditional variance with constraints to ARCH parameters.

$$\begin{aligned} E(a_t) &= E[E(a_t|F_{t-1})] = 0 \\ Var(a_t) &= E(a_t^2) = E[E(a_t^2|F_{t-1})] = E(\sigma_t^2) \end{aligned}$$

Properties of ARCH Errors

Assuming stationarity

$$\begin{aligned} E(\sigma_t^2) &= \alpha_0 + \alpha_1 E(a_{t-1}^2) + \cdots + \alpha_m E(a_{t-m}^2) \\ &= \alpha_0 + \alpha_1 E(\sigma_t^2) + \cdots + \alpha_m E(\sigma_t^2) \end{aligned}$$

which implies that

$$E(\sigma_t^2) = \bar{\sigma}^2 = \frac{\alpha_0}{1 - \alpha_1 - \cdots - \alpha_m} \quad \textit{provided that } \alpha_1 + \cdots + \alpha_m < 1$$

Properties of ARCH(1) models

Consider an ARCH(1) model

$$a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2$$

where $\alpha_0 > 0$, and $\alpha_1 \geq 0$.

- $E(a_t) = 0$
- $Var(a_t) = \frac{\alpha_0}{1-\alpha_1}$ if $0 \leq \alpha_1 < 1$
- Under normality,

$$m_4 = E(a_t^4) = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$$

provided $0 \leq \alpha_1^2 < 1/3$.

Properties of ARCH(1) models

- The unconditional kurtosis of a_t

$$\frac{E(a_t^4)}{[Var(a_t)]^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3 = kurtosis(normal)$$

which implies heavy tails.

Building an ARCH Model

1. Modeling the mean effect and testing for ARCH effects
   H_0: no ARCH effects versus H_a : ARCH effects
   Use Q-statistics of squared residuals $\{\hat{a}_t^2\}$
2. Order determination
   Use PACF of the squared residuals $\{\hat{a}_t^2\}$
3. Estimation: conditional MLE (joint estimation of the mean and volatility equations)
4. Model checking: Q-stat of standardized residuals and squared standardized residuals. Skewness & Kurtosis of standardized residuals.

Testing for ARCH Effects

Consider testing the hypotheses
H_0 : (No ARCH) $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$
H_a : (ARCH) at least one $\alpha_i \neq 0$

- Step 1: Compute residuals $\{\hat{a}_t\}$ from mean equation regression
- Step 2: Apply the usual Ljung-Box statistics $Q(m)$ to $\{\hat{a}_t^2\}$ series

Testing for ARCH Effects

Consider testing the hypotheses
H_0 : (No ARCH) $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$
H_a : (ARCH) at least one $\alpha_i \neq 0$
Engle derived a simple LM test

- Step 1: Compute residuals $\{\hat{a}_t\}$ from mean equation regression
- Step 2: Estimate auxiliary regression

$$\hat{a}_t^2 = \alpha_0 + \alpha_1 \hat{a}_{t-1}^2 + \cdots + \alpha_m \hat{a}_{t-m}^2 + error_t$$

Obtain $R^2 \equiv R^2_{AUX}$ from this regression.

Testing for ARCH Effects

- Step 3 : Form the LM test statistic

$$LM_{ARCH} = T \cdot R^2_{AUX}$$

where T = sample size from auxiliary regression. Under H_0 : (No ARCH), LM_{ARCH} is asymptotically distributed as $\chi^2(m)$.

Remark: Test has power against GARCH(p, q) alternatives

Model Checking

For a properly specified ARCH model, the standardized residuals

$$\tilde{a}_t = \frac{a_t}{\sigma_t}$$

form a sequence of i.i.d. random variables. Therefore, we can check the adequacy of a fitted ARCH model by examining the series $\{\tilde{a}_t\}$.

- The Ljung-Box statistics of $\{\tilde{a}_t\}$ can be used to check the adequacy of the mean equation.
- The Ljung-Box statistics of $\{\tilde{a}_t{}^2\}$ can be used to check the adequacy of the volatility equation.
- The skewness, kurtosis, and QQ plot of $\{\tilde{a}_t\}$ can be used to check the validity of the distribution assumption.

Forecasting

Forecasts of the ARCH(m) model can be obtained recursively as those of an AR model. At the forecast origin h,

- 1-step-ahead forecast of σ_{h+1}^2 is

$$\sigma_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \cdots + \alpha_m a_{h+1-m}^2$$

- 2-step-ahead forecast is

$$\sigma_h^2(2) = \alpha_0 + \alpha_1 \sigma_h^2(1) + \alpha_2 a_h^2 + \cdots + \alpha_m a_{h+2-m}^2$$

- l-step-ahead forecast is

$$\sigma_h^2(l) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \sigma_h^2(l-i)$$

where $\sigma_h^2(l-i) = a_{h+l-i}^2$, if $l - i \leq 0$

Example 3.1.

- simple returns $\rightarrow$ log returns : $logrtn = log(1 + rtn)$
- This series does not have significant serial correlations, therefore $E_{t-1}(r_t) = \mu$. Regress r_t on a constant by OLS and save the residual series
- Test for ARCH effect:
  1. view $\rightarrow$ residual tests $\rightarrow$ correlogram squared residuals : strong ARCH effects with $Q(12) = 89.85$, the p value of which is close to zero.
  2. view $\rightarrow$ residual tests $\rightarrow$ heteroskedasticity tests $\rightarrow$ arch : LM test shows strong ARCH effects with p value being close to zero.
  3. The sample ACF and PACF of the squared residuals clearly show the existence of ARCH effect.

Example 3.1.

- Order determination : The sample PACF of squared residuals indicates that an ARCH(3) model might be appropriate.
- Quick $\rightarrow$ Estimate Equation $\rightarrow$ ARCH
- Model checking : standardized residuals
    1. check adequacy of the mean equation : view $\rightarrow$ residual tests $\rightarrow$ correlogram- Q-statistics
    2. check adequacy of the volatility equation : view $\rightarrow$ residual tests $\rightarrow$ correlogram squared residuals
- normal innovation v.s. t innovation
- It turns out that a GARCH(1,1) model fits better for the data.

ARCH Models

Advantages

- Simplicity
- ARCH/GARCH-type can model the volatility clustering effect since the conditional variance is autoregressive. Such models can be used to forecast volatility.
- Heavy tails (high kurtosis)

Weaknesses

- Symmetric between positive & negative prior returns
- Restrictive on parameter space
- Provides no explanation
- Not sufficiently adaptive in prediction (respond slowly to large isolated shocks to the return series)

Motivation

- The required value of m for ARCH(m) might be very large.
- Non-negativity constraints might be violated.
- A natural extension of an ARCH(m) model which gets around some of these problems is a GARCH model.
- Due to Bollerslev (1986), allow the conditional variance to be dependent upon previous own lags.
- GARCH models are more parsimonious - avoids overfitting.

GARCH Model

Idea: ARCH is like an AR model for volatility. GARCH is like an ARMA model for volatility. $GARCH(m, s)$ model:

$$\begin{aligned} r_t &= E(r_t|F_{t-1}) + a_t \quad a_t = \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \end{aligned}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. r.v. with mean 0 and variance 1, $\alpha_0 > 0$, and $\alpha_i \geq 0$, $\beta_j \geq 0$ for $i > 0$, and $\sum_{i=1}^{max(m,s)} (\alpha_i + \beta_i) < 1$.

GARCH Model

$\{a_t, F_{t-1}\}$ is a stationary martingale difference series with finite unconditional variance provided that $\sum_{i=1}^{max(m,s)}(\alpha_i + \beta_i) < 1$

$$\begin{aligned} E(a_t) &= 0 \\ Var(a_t) &= E(a_t^2) = \frac{\alpha_0}{1 - (\sum_{i=1}^{m} \alpha_i) - (\sum_{j=1}^{s} \beta_j)} \end{aligned}$$

GARCH Model

Re-parameterization:
Let $\eta_t = a_t^2 - \sigma_t^2$. $\{\eta_t\}$ un-correlated series. The GARCH model

becomes

$$a_t^2 = \alpha_0 + \sum_{i=1}^{max(m,s)} (\alpha_i + \beta_i) a_{t-i}^2 + \eta_t - \sum_{j=1}^{s} \beta_j \eta_{t-j}$$

This is an ARMA form for the squared series a_t^2.
Use it to understand properties of GARCH models, e.g. moment equations, forecasting, etc.

GARCH Model

- The modeling procedure of ARCH models can also be used to build a GARCH model.
- Specifying the order of a GARCH model is not easy. Only lower order GARCH models are used in most applications, say, GARCH(1,1), GARCH(2,1), and GARCH(1,2) models.
- In general a GARCH(1,1) model will be sufficient to capture the volatility clustering in the data.

GARCH(1,1) model

Focus on a GARCH(1,1) model

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

- Weak stationarity: $0 \leq \alpha_1, \beta_1 < 1$, $(\alpha_1 + \beta_1) < 1$
- Unconditional variance : $\bar{\sigma}^2 = \frac{\alpha_0}{1-\alpha_1-\beta_1}$
- Volatility clusters
- Heavy tails: if $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$, then

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

GARCH(1,1) model

- For 1-step ahead forecast,

$$\sigma_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2$$

- For multi-step ahead forecasts : use $a_t^2 = \sigma_t^2 \epsilon_t^2$ and rewrite the model as

$$\sigma_{t+1}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2 + \alpha_1 \sigma_t^2(\epsilon_t^2 - 1)$$

- 2-step ahead volatility forecast

$$\sigma_h^2(2) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(1)$$

GARCH(1,1) model

- In general, we have

$$\sigma_h^2(l) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(l-1), \qquad l > 1$$

This result is exactly the same as that of an ARMA(1,1) model with AR polynomial $1 - (\alpha_1 + \beta_1)L$

GARCH Model : Interval Forecasting

Confidence bands for the 1-step ahead forecast at the forecast origin h

- The conditional distribution of r_{h+1} given the information available at time h is $D\left[E(r_{h+1}|F_h), Var(r_{h+1}|F_h)\right]$, where D depends on the distribution of ϵ_t.
- $var(r_{h+1}|F_h) = \sigma_h^2(1) = \sigma_{h+1}^2$.
- Under normality, 95% confidence interval of the prediction is $[\ E(r_{h+1}|F_h) - 1.96\sigma_h(1), E(r_{h+1}|F_h) + 1.96\sigma_h(1)\ ]$.

GARCH Model : Interval Forecasting

Consider an AR(1)-GARCH (1,1) model

$$\begin{aligned} r_t &= \phi_0 + \phi_1 r_{t-1} + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim i.i.d.\ N(0,1) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned}$$

At the forecast origin h, 95% confidence interval for 1-step ahead forecast is $\left[\ \phi_0 + \phi_1 r_h \mp 1.96\sqrt{\alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2}\ \right]$

GARCH Model : Interval Forecasting

$$r_{h+2} = \phi_0(1+\phi_1) + \phi_1^2 r_h + \phi_1 a_{h+1} + a_{h+2}$$

2-step ahead forecast :

$$\begin{aligned} r_h(2) &= E(r_{h+2}|F_h) = \phi_0(1+\phi_1) + \phi_1^2 r_h \\ var(r_{h+2}|F_h) &= \phi_1^2\sigma_h^2(1) + \sigma_h^2(2) \\ &= \phi_1^2\sigma_h^2(1) + \alpha_0 + (\alpha_1+\beta_1)\sigma_h^2(1) \end{aligned}$$

95% confidence interval for 2-step ahead forecast is

$$\left[\ \phi_0(1+\phi_1) + \phi_1^2 r_h \mp 1.96\sqrt{\alpha_0 + (\alpha_1+\beta_1+\phi_1^2)\sigma_h^2(1)}\ \right]$$

Example 3.3.

Monthly excess returns of S&P 500 index starting from 1926 for 792 observations.

- The mean equation : AR(3)
- Test for ARCH effects
- A joint estimation of the AR(3)-GARCH(1,1)
- All AR coefficients are statistically insignificant.
- A simplified model: AR(0)-GARCH(1,1)
- Model checking: Q tests for $\{\tilde{a}_t\}$ and $\{\tilde{a}_t{}^2\}$
- Student-t innovation ; estimation of degrees of freedom
- Forecast

Motivations

The high persistence often observed in fitted GARCH(1,1) models suggests that volatility might be nonstationary implying that $\alpha_1 + \beta_1 = 1$, in which case the GARCH(1,1) model becomes the integrated GARCH(1,1) or IGARCH(1,1) model.

IGARCH(1,1) Model

- An IGARCH(1,1) model:

$$a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1) a_{t-1}^2$$

- For the monthly excess returns of the S&P 500 index, an estimated IGARCH(1,1) model is

$$r_t = 0.007 + a_t, \quad \sigma_t^2 = 0.0001 + 0.806 \sigma_{t-1}^2 + 0.194 a_{t-1}^2$$

- The parameter estimates are close to those of the GARCH(1,1) model shown before, but there is a major difference between the two models.

IGARCH(1,1) Model

- For an IGARCH(1,1) model,

$$\sigma_h^2(l) = \sigma_h^2(1) + (l-1)\alpha_0, \quad l \geq 1,$$

  where h is the forecast origin.
- Effect of $\sigma_h^2(1)$ on future volatilities is persistent, and the volatility forecasts form a straight line with slope α_0.
- Special case: $\alpha_0 = 0$ is of particular interest in studying the IGARCH(1,1) model.
  used in RiskMetrics to VaR calculation.

IGARCH Model

- In the IGARCH(1,1) model the unconditional variance is not finite and so the model does not exhibit volatility mean reversion. However, it can be shown that the model is strictly stationary provided $E\left[\log(\alpha_1\epsilon_t^2 + \beta_1)\right] < 0$.
- Diebold and Lopez (1996) argued against the IGARCH specification for modeling highly persistent volatility processes for two reasons. They argue that,
  1. the observed convergence toward normality of aggregated returns is inconsistent with the IGARCH model.
  2. observed IGARCH behavior may result from misspecification of the conditional variance function. For example, ignored structural breaks in the unconditional variance can result in IGARCH behavior.

Conditional Mean Specification

- $E(r_t|F_{t-1})$ is typically specified as a constant or possibly a low order ARMA process to capture autocorrelation caused by market microstructure effects.
- If extreme or unusual market events have happened during sample period, then dummy variables associated with these events are often added to the conditional mean specification to remove these effects.

Conditional Mean Specification

The typical conditional mean specification is of the form:

$$r_t = ARMA(p, q) + \beta' X_t$$

where X_t are explanatory variables. The choice of X_t is flexible, for example,

- a dummy variable can be used for the Mondays to study the effect of the weekend on daily stock returns
- the market return
- volatility

Explanatory Variables in the Conditional Variance Equation

- Exogenous explanatory variables may also be added to the conditional variance formula

$$\sigma_t^2 = GARCH(p,q) + \delta' Z_t$$

  where Z_t are explanatory variables.
- Variables that have been shown to help predict volatility are trading volume, interest rates, macroeconomic news announcements, implied volatility from option prices and realized volatility, overnight returns, and after hours realized volatility.

The GARCH-M Model

- Idea: Modern finance theory suggests that volatility may be related to risk premia on assets
- The GARCH-M model allows time-varying volatility to be related to expected returns

$$\begin{aligned}
r_t &= \mu + c g(\sigma_t) + a_t \quad a_t = \sigma_t \epsilon_t \\
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
g(\sigma_t) &= \begin{cases} \sigma_t \\ \sigma_t^2 \\ \log(\sigma_t^2) \end{cases}
\end{aligned}$$

where c (expected to be positive) is referred to as **risk premium**.

Example

- Monthly excess returns of the S&P 500 index from January 1926 to December 1991
- GARCH(1,1)-M model with Gaussian innovations
- The estimated risk premium for the index return is positive but is not statistically significant at the 5% level.

Leverage Effect

- A stylized fact of financial volatility is that bad news (negative shocks) tends to have a larger impact on volatility than good news (positive shocks).
- Black (1976) attributed this effect to the fact that bad news tends to drive down the stock price, thus increasing the leverage (i.e., the debt-equity ratio) of the stock and causing the stock to be more volatile. Based on this conjecture, the asymmetric news impact on volatility is commonly referred to as the **leverage effect**.

EGARCH Model

- In the basic GARCH model, since only squared residuals a_{t-i}^2 enter the conditional variance equation, the signs of the residuals or shocks have no effect on conditional volatility.
- EGARCH Models allow for asymmetry in responses to past positive and negative returns.

EGARCH Model

- The weighted innovation

$$g(\epsilon_t) = \theta\epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)] \quad \textit{with } E[g(\epsilon_t)] = 0$$

  where θ and γ are real constants, $\{\epsilon_t\}$ are zero-mean i.i.d. sequences.

- The asymmetry of $g(\epsilon_t)$ can easily be seen by rewriting it as

$$g(\epsilon_t) = \begin{cases} (\theta + \gamma)\epsilon_t - \gamma E(|\epsilon_t|) & \textit{if } \epsilon_t \geq 0, \\ (\theta - \gamma)\epsilon_t - \gamma E(|\epsilon_t|) & \textit{if } \epsilon_t < 0. \end{cases}$$

EGARCH Model

An EGARCH(m, s) model:

$$a_t = \sigma_t \epsilon_t, \quad log(\sigma_t^2) = \alpha_0 + \frac{1 + \beta_1 L + \cdots + \beta_{s-1} L^{s-1}}{1 - \alpha_1 L - \cdots - \alpha_m L^m} g(\epsilon_{t-1})$$

$$or \quad log(\sigma_t^2) = \tilde{\alpha} + \sum_{i=0}^{s-1} \beta_i g(\epsilon_{t-i-1}) + \sum_{j=1}^{m} \alpha_j log(\sigma_{t-j}^2)$$

where $\beta_0 = 1$ and $\tilde{\alpha} = \alpha_0(1 - \alpha_1 - \cdots - \alpha_m)$.
Some features of EGARCH models:

- uses log transformation to relax the positiveness constraint
- asymmetric responses
- the unconditional mean of $log(\sigma_t^2)$ is α_0

EGARCH(1,1) Model

Consider an EGARCH(1,1) model

$$a_t = \sigma_t \epsilon_t, \quad (1-\alpha L) log(\sigma_t^2) = (1-\alpha)\alpha_0 + g(\epsilon_{t-1})$$

Under normality, $E(|\epsilon_t|) = \sqrt{2/\pi}$ and the model becomes

$$(1-\alpha L) log(\sigma_t^2) = \begin{cases} \alpha_* + (\gamma+\theta)\epsilon_{t-1} & if \ \epsilon_{t-1} \geq 0, \\ \alpha_* + (\gamma-\theta)(-\epsilon_{t-1}) & if \ \epsilon_{t-1} < 0. \end{cases}$$

where $\alpha_* = (1-\alpha)\alpha_0 - \sqrt{2/\pi}\gamma$. This is a nonlinear function similar to that of the threshold autoregressive (TAR) model of Tong (1978, 1990).

EGARCH(1,1) Model

The conditional variance evolves in a nonlinear manner depending on the sign of a_{t-1}. Specifically, we have

$$\sigma_t^2 = \sigma_{t-1}^{2\alpha} exp(\alpha_*) \begin{cases} exp\left[(\gamma+\theta)\frac{a_{t-1}}{\sigma_{t-1}}\right] & if \;\; a_{t-1} \geq 0, \\ exp\left[(\gamma-\theta)\frac{|a_{t-1}|}{\sigma_{t-1}}\right] & if \;\; a_{t-1} < 0. \end{cases}$$

- The coefficients $(\gamma+\theta)$ and $(\gamma-\theta)$ show the asymmetry in response to positive and negative a_{t-1}.
- The model is, therefore, nonlinear if $\theta \neq 0$. θ is referred to as the **leverage parameter**. Since negative shocks tend to have larger impacts, we expect θ to be negative.
- The leverage parameter θ shows the effect of the sign of a_{t-1} whereas γ denotes the magnitude effect.

EGARCH Model

An alternative form for the EGARCH(m, s) model is

$$a_t = \sigma_t \epsilon_t, \quad log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{s} \alpha_i \frac{|a_{t-i}| + \gamma_i a_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{m} \beta_j log(\sigma_{t-j}^2)$$

- Here a positive a_{t-i} contributes $\alpha_i(1+\gamma_i)|\epsilon_{t-i}|$ to the log volatility, whereas a negative a_{t-i} gives $\alpha_i(1-\gamma_i)|\epsilon_{t-i}|$
- γ_i parameter thus signifies the leverage effect of a_{t-i}.
- We expect γ_i to be negative in real applications.

EGARCH Model: Specification in Eviews

$$log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{s} \alpha_i \frac{|a_{t-i}|}{\sigma_{t-i}} + \sum_{k=1}^{r} \gamma_k \frac{a_{t-k}}{\sigma_{t-k}} + \sum_{j=1}^{m} \beta_j log(\sigma_{t-j}^2)$$

The presence of leverage effects can be tested by the hypothesis that $\gamma_k < 0$. The impact is asymmetric if $\gamma_k \neq 0$.

Forecasting Using an EGARCH Model

If interested, check Tsay's textbook pp. 147-148

Example

- Section 3.8.2
- Section 3.8.3 : Simple returns need to be transformed to log returns

The Threshold GARCH (TGARCH) or GJR Model

Another volatility model commonly used to handle leverage effects is the threshold GARCH (or TGARCH) model.

TGARCH(s,m) or GJR(s,m) model is defined as

$$\begin{aligned} r_t &= \mu_t + a_t, \quad a_t = \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{s} (\alpha_i + \gamma_i N_{t-i}) a_{t-i}^2 + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^2 \end{aligned}$$

The Threshold GARCH (TGARCH) or GJR Model

- N_{t-i} is an indicator variable such that

$$N_{t-i} = \begin{cases} 1 & if \ a_{t-i} < 0, \\ 0 & if \ a_{t-i} \geq 0. \end{cases}$$

- When a_{t-i} is positive, the total effects are $\alpha_i a_{t-i}^2$
- When a_{t-i} is negative, the total effects are $(\alpha_i + \gamma_i) a_{t-i}^2$
- One expects γ_i to be positive so that prior negative returns have higher impact on the volatility.
- The model uses zero as its threshold to separate the impacts of past shocks. Other threshold values can also be used.

Example

- Specification in Eviews :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{s} \alpha_i a_{t-i}^2 + \sum_{k=1}^{r} \gamma_k a_{t-k}^2 N_{t-k} + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^2$$

- The monthly log returns of IBM stock from 1926 to 2003 : TGARCH(1,1) model with conditional GED innovations

The Asymmetric Power ARCH (APARCH) Model

This model was introduced by Ding, Engle and Granger (1993) as a general class of volatility models. The basic form is

$$\begin{aligned} r_t &= \mu_t + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim D(0,1) \\ \sigma_t^\delta &= \alpha_0 + \sum_{i=1}^{s} \alpha_i (|a_{t-i}| - \gamma_i a_{t-i})^\delta + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^\delta \end{aligned}$$

where δ is a non-negative real number.

The Asymmetric Power ARCH (APARCH) Model

- Leverage effect implies that $\gamma_i > 0$.
- $\delta = 2$ gives rise to the TGARCH model.
- If $\delta = 2$ and $\gamma_i = 0$ for all i, the PARCH model is simply a standard GARCH specification.
- $\delta = 1$ gives a model for σ_t and is more robust to outliers than when $\delta = 2$.
- δ can be fixed at a particular value or estimated by MLE.
- Example : the monthly log returns of IBM stock from 1926 to 2003

The Difference between SV Models and GARCH Models

- The conditional variance equation of a GARCH specification is completely **deterministic** given all information available up to that of the previous period.
- In other words, there is no error term in the volatility equation of a GARCH-type model, only in the mean equation.
- Stochastic volatility models contain a second error term, which enters into the conditional variance equation.

Stochastic Volatility (SV) Models

A (simple) SV model is

$$
\begin{aligned}
r_t &= E(r_t|F_{t-1}) + a_t \quad a_t = \sigma_t \epsilon_t \\
\log(\sigma_t^2) &= \alpha_0 + \sum_{i=1}^{m} \alpha_i \log(\sigma_{t-i}^2) + v_t
\end{aligned}
$$

where $\{\epsilon_t\}$ are i.i.d. $N(0,1)$, $\{v_t\}$ are i.i.d. $N(0,\sigma_v^2)$, $\{\epsilon_t\}$ and $\{v_t\}$ are independent.

- The logarithm of an unobserved variance process is modeled by a linear stochastic specification, such as an autoregressive model.

Stochastic Volatility (SV) Models

- The primary advantage of stochastic volatility models is that they can be viewed as discrete time approximations to the continuous time models employed in options pricing frameworks.
- Such models are hard to estimate, so they have not been popular in empirical discrete-time financial applications

Use of High-Frequency Data

Suppose we like to estimate the monthly volatility of a stock return

- Data: Daily returns
- Let r_t^m be the t^{th} month log return.
- Let $\{r_{t,i}\}_{i=1}^n$ be the daily log returns within month t.
- Using properties of log returns, we have $r_t^m = \sum_{i=1}^n r_{t,i}$.
- Assuming conditional variance and covariance exist, we have

$$Var(r_t^m|F_{t-1}) = \sum_{i=1}^n Var(r_{t,i}|F_{t-1}) + 2\sum_{i<j} Cov[(r_{t,i}, r_{t,j})|F_{t-1}]$$

where F_{t-1} = the information available at month $t-1$

Use of High-Frequency Data

Further simplication is possible under additional assumptions.

- If $\{r_{t,i}\}$ is a white noise series, then

$$Var(r_t^m|F_{t-1}) = nVar(r_{t,1})$$

where $Var(r_{t,1})$ can be estimated from the daily returns $\{r_{t,i}\}_{i=1}^n$ by

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (r_{t,i} - \bar{r}_t)^2}{n-1}, \quad \bar{r}_t = \frac{\sum_{i=1}^n r_{t,i}}{n}$$

The estimated monthly volatility is then

$$\hat{\sigma}_m^2 = \frac{n}{n-1}\sum_{i=1}^n (r_{t,i} - \bar{r}_t)^2 \approx \sum_{i=1}^n (r_{t,i} - \bar{r}_t)^2$$

Use of High-Frequency Data

- If $\{r_{t,i}\}$ follows an MA(1) model, then

$$Var(r_t^m | F_{t-1}) = nVar(r_{t,1}) + 2(n-1)Cov(r_{t,1}, r_{t,2})$$

which can be estimated by

$$\hat{\sigma}_m^2 = \frac{n}{n-1} \sum_{i=1}^{n} (r_{t,i} - \bar{r}_t)^2 + 2 \sum_{i=1}^{n-1} (r_{t,i} - \bar{r}_t)(r_{t,i+1} - \bar{r}_t)$$

Use of High-Frequency Data

- Advantage: Simple
- Weaknesses:
  1. Models for daily returns $\{r_{t,i}\}$ are unknown.
  2. Typically, 21 or 22 trading days in a month, resulting in a small sample size.

Use of High-Frequency Data : Realized Integrated Volatility

- If the sample mean $\bar{r}_t$ is zero, then $\hat{\sigma}_m^2 \approx \sum_{i=1}^{n} r_{t,i}^2$
  $\Rightarrow$ Use cumulative sum of squares of daily log returns within a month as an estimate of monthly volatility.
- Apply the idea to intradaily log returns and obtain (daily) realized integrated volatility of an asset.

Realized Integrated Volatility

- Let r_t be the daily log return of an asset.
- Suppose that there are n equally spaced intradaily log returns available such that $r_t = \sum_{i=1}^{n} r_{t,i}$.
- The quantity

$$RV_t = \sum_{i=1}^{n} r_{t,i}^2$$

is called the realized volatility of r_t.

Realized Integrated Volatility

- Advantages: simplicity and using intraday information
- Weaknesses:
  1. Effects of market microstructure (noises)
     The problem of choosing an optimal time interval for constructing realized volatility has attracted much research lately. For heavily traded assets in the United States, a time interval of 4-15 minutes is often used.
  2. Overlook overnight volatilities
     The overnight return, which is the return from the closing price of day $t-1$ to the opening price of t , tends to be substantial. Ignoring overnight returns can seriously underestimate the volatility.

What We Know

- Volatilities and correlations vary over time, sometimes abruptly
- Risk management, asset allocation, derivative pricing and hedging strategies all depend upon up to date correlations and volatilities

Multivariate Volatility Models

- Can we anticipate future correlations?
- How and why do correlations change over time?
- How can we get the best estimates of correlations for financial decision making?

Multivariate Volatility Models

How do the correlations between asset returns change over time?

Focus on two series (Bivariate)
Two asset return series:

$$\mathbf{r}_t = \begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix}$$

Data : $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_T$.

Multivariate Volatility Models

Basic concept

- Let F_{t-1} denote the information available at time $t-1$.
- Partition the return as

$$\mathbf{r}_t = \mu_t + \mathbf{a}_t, \quad \mathbf{a}_t = \mathbf{\Sigma}_t^{1/2}\epsilon_t$$

where $\mu_t = E(\mathbf{r}_t|F_{t-1})$ is the predictable component, and

$$Cov(\mathbf{a}_t|F_{t-1}) = \mathbf{\Sigma}_t = \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{22,t} \end{bmatrix}$$

$\{\epsilon_t\}$ are i.i.d. 2-dimensional random vectors with mean zero and identity covariance matrix.

Multivariate Volatility Modeling

Study time evolution of $\{\mathbf{\Sigma}_t\}$

- $\mathbf{\Sigma}_t$ is symmetric, i.e. $\sigma_{12,t} = \sigma_{21,t}$
- There are 3 variables in $\mathbf{\Sigma}_t$
- For k asset returns, $\mathbf{\Sigma}_t$ has $k(k+1)/2$ variables

Multivariate Volatility Modeling

Requirement

- $\mathbf{\Sigma}_t$ must be positive definite for all t,

$$\sigma_{11,t} > 0, \quad \sigma_{22,t} > 0, \quad \sigma_{11,t}\sigma_{22,t} - \sigma_{12,t}^2 > 0.$$

- The time-varying conditional correlation between r_{1t} and r_{2t} is

$$\rho_{12,t} = \frac{\sigma_{12,t}}{\sqrt{\sigma_{11,t}\sigma_{22,t}}}$$

Some Complications

- Positiveness requirement is not easy to meet
- Too many series to consider

Multivariate Volatility Modeling

Some simple models available

- Exponentially weighted covariance
- Diagonal VEC model
- BEKK model
- Dynamic correlation models

Exponentially Weighted Model

Given the innovations $F_{t-1} = \{\mathbf{a}_1, \cdots, \mathbf{a}_{t-1}\}$,

- the (unconditional) covariance matrix of the innovation can be estimated by

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{t-1} \sum_{j=1}^{t-1} \mathbf{a}_j \mathbf{a}_j'$$

- To allow for a time-varying covariance matrix and to emphasize that recent innovations are more relevant, we estimate the conditional covariance matrix of $\mathbf{a}_t$ by

$$\widehat{\boldsymbol{\Sigma}}_t = \frac{1-\lambda}{1-\lambda^{t-1}} \sum_{j=1}^{t-1} \lambda^{j-1} \mathbf{a}_{t-j} \mathbf{a}_{t-j}' \quad \lambda \in (0,1)$$

Exponentially Weighted Model

- For a sufficiently large t such that $\lambda^{t-1} \approx 0$,

$$\widehat{\mathbf{\Sigma}}_t = (1-\lambda)\mathbf{a}_{t-1}\mathbf{a}'_{t-1} + \lambda\widehat{\mathbf{\Sigma}}_{t-1}$$

- Just one parameter to calibrate for memory decay for all volatilities and correlations
- Pros : positive definite
- Let μ_t be a function of parameter Θ . λ and Θ can be estimated jointly by the maximum-likelihood method.
- Example 10.1.

Diagonal Vectorization (VEC) Model

- Model elements of $\boldsymbol{\Sigma}_t$ separately
- For instance, the bivariate DVEC(1,1) model

$$
\begin{aligned}
\sigma_{11,t} &= c_{11} + \alpha_{11} a_{1,t-1}^2 + \beta_{11}\sigma_{11,t-1} \\
\sigma_{12,t} &= c_{12} + \alpha_{12} a_{1,t-1} a_{2,t-1} + \beta_{12}\sigma_{12,t-1} \\
\sigma_{22,t} &= c_{22} + \alpha_{22} a_{2,t-1}^2 + \beta_{22}\sigma_{22,t-1}
\end{aligned}
$$

where each element of $\boldsymbol{\Sigma}_t$ depends only on its own past value and the corresponding product term in $\mathbf{a}_{t-1}\mathbf{a}_{t-1}'$.

Diagonal Vectorization (VEC) Model

- That is, each element of a DVEC model follows a GARCH(1,1)-type model.
- May not produce a positive-definite covariance matrix. Restrictions on parameters are needed to guarantee positive definiteness.
- Example 10.2.

BEKK model of Engle and Kroner (1995)

Simple BEKK(1,1) model

$$\mathbf{\Sigma}_t = \mathbf{A}_0\mathbf{A}_0' + \mathbf{A}_1(\mathbf{a}_{t-1}\mathbf{a}_{t-1}')\mathbf{A}_1' + \mathbf{B}_1\mathbf{\Sigma}_{t-1}\mathbf{B}_1'$$

where $\mathbf{A}_0$ is a lower triangular matrix, $\mathbf{A}_1$ and $\mathbf{B}_1$ are square matrices without restrictions.

- Pros: models with guaranteed positive definite structure
- Cons: many parameters, dynamic relations require further study

Dynamic Conditional Correlation (DCC) Models

Modeling steps involved:

- Obtain univariate GARCH models for each asset return (k models for volatilities σ_{it}^2 for $i = 1, \cdots, k$)
- Model the correlation matrix ρ_t, where $\rho_t = \mathbf{D}_t^{-1}\mathbf{\Sigma}_t\mathbf{D}_t^{-1}$ with $\mathbf{D}_t = diag\{\sigma_{1t}, \cdots, \sigma_{kt}\}$.

Dynamic Conditional Correlation (DCC) Models

- Tse and Tsui (2002) model:

$$\rho_t = (1-\theta_1-\theta_2)\rho + \theta_1\rho_{t-1} + \theta_2\psi_{t-1}$$

where $0 \leq \theta_1, \theta_2 < 1$ and $\theta_1 + \theta_2 < 1$, and ψ_{t-1} is the sample correlation matrix of the past m returns.
In other words , ψ_{t-1} is the sample correlation matrix of $\{\mathbf{r}_{t-1}, \cdots, \mathbf{r}_{t-m}\}$ for some pre-specified m with $m > k$.

Dynamic Conditional Correlation (DCC) Models

- Engle (2002) model: $\rho_t = \mathbf{J}_t \mathbf{Q}_t \mathbf{J}_t$, where $\mathbf{Q}_t = (q_{ij,t})_{k \times k}$ is a positive-definite matrix, $\mathbf{J}_t = diag\{q_{11,t}^{-1/2}, \cdots, q_{kk,t}^{-1/2}\}$

  $\mathbf{Q}_t$ satisfies $\mathbf{Q}_t = (1-\theta_1-\theta_2)\bar{\mathbf{Q}} + \theta_1 \epsilon_{t-1}\epsilon'_{t-1} + \theta_2 \mathbf{Q}_{t-1}$, where $0 \le \theta_1, \theta_2 < 1$ and $\theta_1 + \theta_2 < 1$, $\bar{\mathbf{Q}}$ is the sample correlation matrix of the data, and ϵ_t is the standardized innovation vector with elements $\epsilon_{it} = a_{it}/\sigma_{it}$.

  The $\mathbf{J}_t$ matrix is a normalization matrix to guarantee that ρ_t is a correlation matrix.

Dynamic Conditional Correlation (DCC) Models

Discussion: Volatility matrix is always positive definite, but the model is restrictive. All correlations evolve in the same way.

Assignment 4

- Tsay : 3.1; 3.6 ; 3.7 ; 3.9; 3.11; 3.12; 3.13;
- Due Nov. 30 in class