MASTER
Econométrie et Statistique Appliquée
Université d'Orléans

Advanced Financial Econometrics

Gilles de Truchis

Master 1 ESA

January 19, 2025

Table of contents

Reminders
Stylized Facts
GARCH models
GARCH-based Portfolio Optimization
Multivariate GARCH models
Market Efficiency
Spot and Integrated Volatility
Continuous-time models
Continuous-time GARCH
Realized volatility measures
Multivariate realized models

Introduction

Reminders and Stylized Facts

Back to basics

- Standard time series analysis rests on important concepts such as stationarity, spherical errors, and on a central family of models, the autoregressive moving average (ARMA) models.
- But these concepts are insufficient for the analysis of financial time series

1. First recall standard time series properties
2. Then discuss the main stylized facts of financial series

Objective

- Modeling financial time series is a complex problem (see Mandelbrot 1963)
  - There is a variety of the series in use (stocks, exchange rates, interest rates, etc.)
  - The availability of very large data sets at different frequencies
  - Mainly because of statistical regularities (**stylized facts**) which are common to a large number of financial series and are difficult to reproduce artificially using stochastic models

Definitions

Definition 1

*A **stochastic process**, $\{Y_t(\omega), \omega \in \Omega, t \in \mathbb{R}\}$, is an ordered sequence or random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with*

- *Ω the set of events*
- *$\mathcal{F}$ a σ-field representing the events*
- *$\mathbb{P}$ a probability mesure such that $\mathbb{P}(A)$ is the probability of event A*

$\Rightarrow$ In the following $\{Y_t(\omega), \omega \in \Omega, t \in \mathbb{R}\}$ is denoted $\{Y_t\}_{t\in\mathbb{R}}$ or Y_t

Definitions

Definition 2

A ***time series*** *denoted* $\{y_t\}_{t\in\mathbb{T}}$ *or* y_t *is a set of realizations of a stochastic process* $\{Y_t\}_{t\in\mathbb{Z}}$ *with* $\mathbb{Z} \supseteq \mathbb{T}$

Definition 3

An ***infinite time series*** *denoted* $\{Y_t\}_{t=-\infty}^{\infty}$ *is an infinite set of realizations of a stochastic process* $\{Y_t\}_{t\in\mathbb{Z}}$

Definitions

Definition 4

*The **unconditional central moments** of Y_t may be written as the expected value of $h(Y_t)$, which is a continuous function of Y_t*

$$\mathbb{E}(h(Y_t)) = \int h(Y_t) f(Y_t) dY_t$$

with $f(Y_t)$ the unconditional density function of Y_t

- To compute the expected value of Y_t we choose $h(Y_t) = Y_t$

$$\mathbb{E}(Y_t) = \mu_t$$

- To compute the variance of Y_t we choose $h(Y_t) = (Y_t - \mathbb{E}(Y_t))^2$

$$\mathbb{V}(Y_t) = \sigma_t^2$$

Definitions

Definition 5

The ***autocovariance*** *function of* Y_t *is obtained from the joint density of* $(Y_t, Yy_{t-1}, \dots, Y_{t-h})$ *and is denoted by*

$$\begin{aligned}\gamma(h) &= Cov(Y_t, Y_{t-h})\\ &= \mathbb{E}\Big((Y_t - \mu_t)(Y_{t-h} - \mu_{t-h})\Big)\\ &= \int \dots \int (Y_t - \mu_t)(Y_{t-h} - \mu_{t-h}) f(Y_t, \dots, Y_{t-h}) dY_t \dots dy_{t-h}\end{aligned}$$

with $f(Yt, \dots, Yt - h)$ *the unconditional density function of* Y_t

White noises

Definition 6

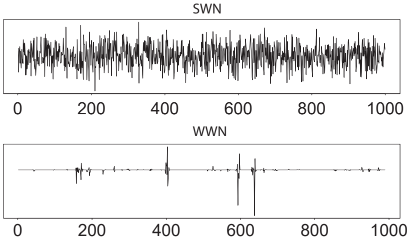
A Gaussian White Noise ε_t *is a sequence of* $\mathrm{i.i.d.}$ *random variable with* $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$

Definition 7

A Strong White Noise ε_t *is a sequence of* $\mathrm{i.i.d.}$ *random variable with* $\mathbb{E}(\varepsilon_t) = 0$ *and* $\mathbb{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$

Definition 8

A Weak White Noise $\tilde{\varepsilon}_t$ *is a sequence of uncorrelated random variable with* $\mathbb{E}(\tilde{\varepsilon}_t) = 0$ *and* $\mathbb{E}(\tilde{\varepsilon}_t^2) = \sigma_{\tilde{\varepsilon}}^2$. *For instance* $\tilde{\varepsilon}_t = u_t u_{t+1} \cdots u_{t+k}$ *is a Weak White Noise.*



Second order stationarity

- Let $\{Y_t\}_{t \in \mathbb{Z}}$, denoted Y_t, be a sequence of random variables

Definition 9

Y_t *is* ***second order stationary*** *if*

$$\forall t \in \mathbb{Z}, \quad \mathbb{E}(Y_t) = \mu < \infty$$

$$\forall t, h \in \mathbb{Z}, \quad Cov(Y_t, Y_{t+h}) = \gamma(h) < \infty$$

$$\forall t \in \mathbb{Z}, \quad \mathbb{V}(Y_t) = \sigma^2 < \infty \text{ as } Cov(Y_t, Y_{t+h}) = \mathbb{V}(Y_t) \text{ for } h = 0$$

- We summarize here the stability in distribution of the process Y_t only through its first two moments
  - Relevant in the Gaussian case but restrictive in general
  - This **weak** form of stationarity is simple to test

Strict Stationarity

- Let Y_t be a sequence of random variables

Definition 10

*Y_t is **strictly stationary** if the joint distribution of Y_t and Y_{t+h}, $\forall t, h$ does not depend on t but only on h*

$$(Y_t, \ldots, Y_{t-h}) \stackrel{d}{=} (Y_\tau, \ldots, Y_{\tau-h})$$

with t $\neq \tau$

- The joint distribution of the process Y_t should hence be **shift-invariant in time**
  - Relevant in non-Gaussian and Gaussian cases
  - This **strong** form of stationarity is difficult to test

Theorem : Wold decomposition

Theorem 1

Let $\{Y_t\}_{t=0}^{n}$ be a second order stationary process. We can show that Y_t can always be written as a weighted sum of innovations of Y_t and a deterministic component μ_t

$$Y_t = \mu_t + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$$

with $\sum_{j=0}^{\infty} a_j^2 < \infty$ and $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2 < \infty)$

- The mean-square convergence of a_j is important as

$$\mathbb{E}(Y_t^2) = \sum_{j=0}^{\infty} a_j^2 \sigma_\varepsilon^2 < \infty$$

if for $m > n$, $\mathbb{E}(\sum_{j=0}^{m} a_j - \sum_{j=0}^{n} a_j)^2 < c \Rightarrow \sum_{j=0}^{\infty} a_j^2 < \infty$

Wold Theorem and autocovariance

- The rate of decrease of the coefficients a_j determines also the shape of the autocovariance of y_t as

$$
\begin{aligned}
\gamma(h) &= \mathbb{E}\Big((Y_t - \mu_t)(Y_{t-h} - \mu_t)\Big) \\
&= \mathbb{E}\Big(\sum_{m=0}^{\infty} a_m \varepsilon_{t-m} \sum_{s=0}^{\infty} a_{s+h} \varepsilon_{t-s+h}\Big) \\
&= \sum_{m,s=0}^{\infty} a_m a_{s+h} \gamma_\varepsilon(m - s + h) \\
&= \sigma_\varepsilon^2 \sum_{m,s=0}^{\infty} a_m a_{s+h}
\end{aligned}
$$

Stochastic trend and unit root

- Suppose an AR(1): $Y_t = \rho Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.\ (0, \sigma_\varepsilon^2 < \infty)$
- If $\rho = 1$ we have $Y_t = Y_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j}$ and hence

$$Cov(Y_t, Y_{t-j}) = (t-j)\sigma_\varepsilon^2 \text{ et } \mathbb{V}(Y_t) = \mathbb{V}\left(\sum_{j=0}^{t-1} \varepsilon_{t-j}\right) = t\sigma_\varepsilon^2$$

$\Rightarrow$ The variance of Y_t depends of $t \Rightarrow Y_t$ is **non-stationary**

$\Rightarrow$ $\rho = 1$ places a solution of the lag polynomial on the unit circle, so Y_t is a **unit root** process

$\Rightarrow$ Y_t is a **random walk** of conditional expectation

$$\mathbb{E}(Y_t | Y_{t-1}, Y_{t-2}, \cdots, Y_0) = Y_{t-1} = \sum_{j=1}^{t-1} \varepsilon_{t-j} = \text{stochastic trend}$$

$\Rightarrow$ Y_t is a **discrete martingale** of **stochastic trend** $\sum_j \varepsilon_{t-j}$

Stochastic trend and differenciation

- Suppose an AR(1): $Y_t = \mu + \rho Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.\ (0, \sigma_\varepsilon^2 < \infty)$
- If $\rho = 1$ and $Y_0 = 0$, the $MA(\infty)$ representation gives

$$Y_t = \rho^t Y_0 + \mu \sum_{j=0}^{t} \rho^j + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j} = \mu t + \sum_{j=0}^{t} \varepsilon_{t-j}$$

- We see that

$$\mathbb{E}(Y_t) = \mu t$$

- The **variance** and **expected value** are not independent of t

$\Rightarrow$ Y_t is an **explosive** processus of **random walk** type and **derivative** μ

- Y_t is first difference **stationary** as

$$\Delta Y_t = (1 - L) Y_t = Y_t - Y_{t-1} = \mu + \varepsilon_t$$

Financial Series

Notations

S_t : asset (or portfolio) price at time t

p_t : asset (or portfolio) log-price at time t

r_t : the continuously compounded or log-return of a financial asset (or portfolio) at time t

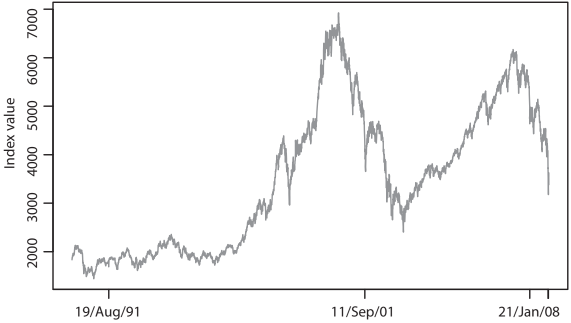
$r_t = p_t - p_{t-1}$

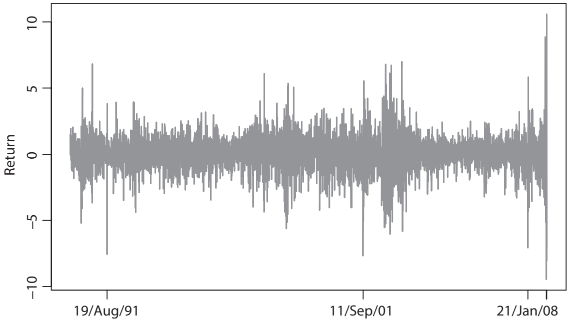
$r_t = log(1 + R_t)$ with $R_t = \frac{S_t - S_{t-1}}{S_{t-1}}$

- Their properties have been amply commented upon in the financial literature
- These stylized facts are mainly concerned with daily stock prices

- Nonstationarity of price series
  - The stochastic process S_t is generally non-stationary in the sense of second-order stationarity

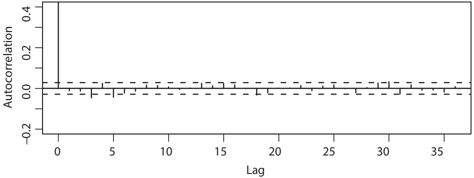
- Stationarity of return series
  - The stochastic process r_t is compatible with the second-order stationarity property



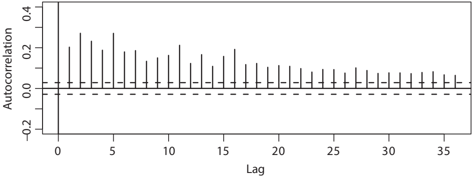


Autocorrelation

- Absence of autocorrelation for the price variations : (a)
  - — The series of price variations generally displays small autocorrelations, making it close to a white noise
- Autocorrelations of the squared price returns : (b)
  - — Squared returns (r_t^2) or absolute returns ($|r_t|$) are generally strongly autocorrelated



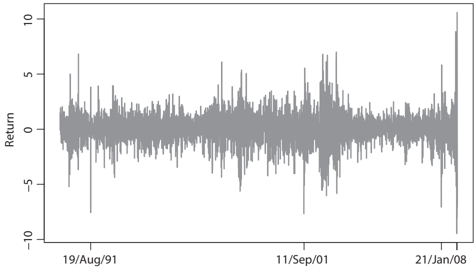
(a)



(b)

Volatility clustering

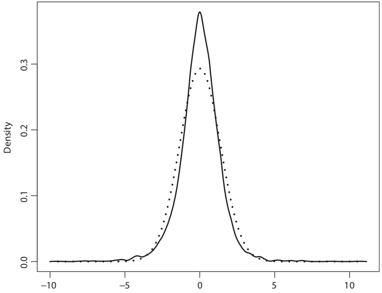
- Large absolute returns $|r_t|$ tend to appear in clusters
- Turbulent (high-volatility) sub-periods are followed by quiet (low-volatility) periods. These sub-periods are recurrent but do not appear in a periodic way (which might contradict the stationarity assumption)
- In other words, volatility clustering is not incompatible with a homoscedastic (i.e. with a constant variance) marginal distribution for the returns



Fat-tailed distributions

- The empirical distribution of daily returns does not resemble a Gaussian one
- The densities have fat tails and are sharply peaked at zero: they are called leptokurtic
- When the time interval over which the returns are computed increases, leptokurticity tends to vanish and the empirical distributions get closer to a Gaussian (Aggregational Gaussianity property)

Note : Below is represented the Kernel estimator of the CAC 40 returns density (solid line) and density of a Gaussian with mean and variance equal to the sample mean and variance of the returns (dotted line).



Conditional fat tails

- Even after accounting for volatility clustering, (by using for example ARCH / GARCH models as we will see in the next section), the distribution of the residuals is leptokurtic
- Its kurtosis is however smaller than in the case of a residuals of a simple ARMA model

Leverage effects

- Asymmetry in the response of volatility to positive and negative past returns, respectively
- A diminishing price generates an increase in volatility larger than a price increase of the same amount

Example: In the table below, $r_t^+ = \max(r_t, 0)$ and $r_t^- = \min(r_t, 0)$

Table: Various return autocorrelation

h	$\hat{\rho}_r(h)$	$\hat{\rho}_{\lvert r \rvert}(h)$	$\hat{\rho}(r^+_{t-h'}, \lvert r_t \rvert)$	$\hat{\rho}(-r^-_{t-h'}, \lvert r_t \rvert)$
1	-0.012	0.175	0.038	0.160
2	-0.014	0.229	0.059	0.200
3	-0.047	0.235	0.051	0.215
4	0.025	0.200	0.055	0.173
5	-0.043	0.218	0.059	0.190
6	-0.023	0.212	0.109	0.136
7	-0.014	0.203	0.061	0.173

Seasonality

- Calendar effects: the day of the week, the proximity of holidays, among other seasonalities, may have significant effects on returns
- Following a period of market closure, volatility tends to increase, reflecting the information cumulated during this break
- The seasonal effect is also very present for intraday series (beyond the scope of this course)

Table: January effect

Average return (monthly %)		
Period	January	Other months
1904-1928	1.3	0.44
1929-1940	6.63	-0.6
1940-1974	3.91	0.7
1904-1974	3.84	0.42

Seasonality

- Calendar effects: the day of the week, the proximity of holidays, among other seasonalities, may have significant effects on returns
- Following a period of market closure, volatility tends to increase, reflecting the information cumulated during this break
- The seasonal effect is also very present for intraday series (beyond the scope of this course)

Table: Week-end effect

		Monday	Tuesday	Wednesday	Thursday	Friday
French (1980)	1953-1977	-0.17	0.02	0.1	0.04	0.09
Gibbons and Hess (1981)	1962-1978	-0.13	0	0.1	0.03	0.08

Summary

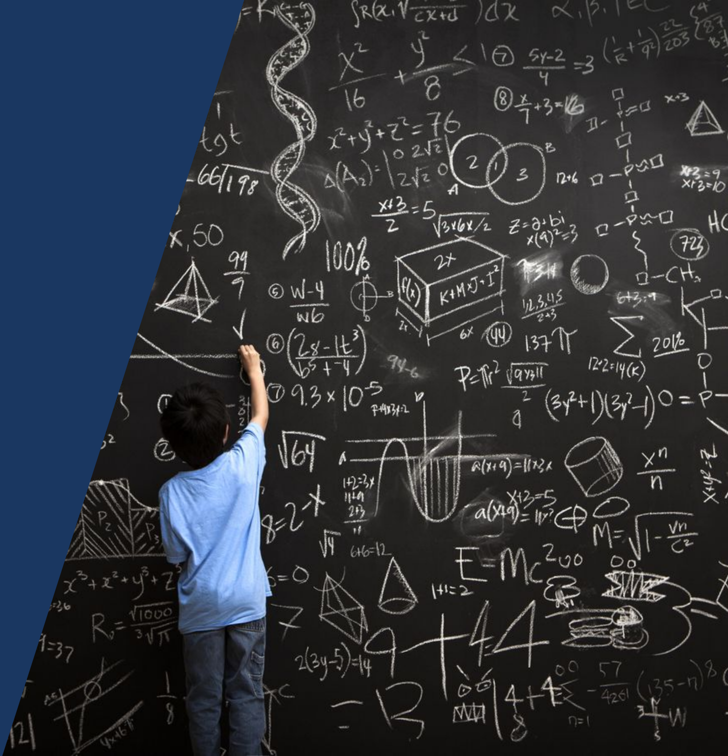
- Any satisfactory statistical model for daily returns must be able to capture these main stylized facts, mainly leptokurticity, the unpredictability of returns, and the existence of positive autocorrelations in the squared and absolute returns
- Classical models (such as ARMA models) centered on the second-order structure are inappropriate
- There is evidence of conditional heterokedasticity (time-varying volatility):

$$\mathbb{V}(r_t|r_{t-1}, r_{t-2}, \dots) \not\equiv const$$

- Conditional heteroscedasticity is perfectly compatible with stationarity, just as the existence of a non-constant conditional mean is compatible with stationarity

Chapter 1

GARCH Type Models



Modelling Approaches

- Objective: account for the very specific nature of financial series
- Example: Stationary AR(1) provides a model specification for the conditional mean

$$r_t = \theta r_{t-1} + \varepsilon_t$$

with $\varepsilon_t \ \mathrm{i.i.d.} \ N(0, \sigma_\varepsilon^2)$

- We hence have

$$\mathbb{E}(r_{t+1}) = 0$$

and

$$\mathbb{E}(r_{t+1} | r_t, r_{t-1}, \ldots) = \theta r_t$$

Modelling Approaches

- Engle (1982)'s idea : account for other conditional moments of the return processus
- But, for an AR(1) process

$$\mathbb{E}(r_{t+1}^2) = \sigma_\varepsilon^2/(1-\theta^2)$$

$$\mathbb{E}(r_{t+1}^2 | r_t, r_{t-1}, \dots) = \sigma_\varepsilon^2$$

  are constants

- Such models are unable to measure changes in forecast error variance although we want them to be impacted by their past evolution

Modelling Approaches

Solution

- Models that capture time-varying volatility are written in the multiplicative form

$$r_t = \sigma_t z_t$$

- where (z_t) and (σ_t) are real processes such that:
  - σ_t is measurable with respect to a σ-field, denoted $\mathcal{I}_{t-1}$;
  - z_t is a strong white noise process with unit variance, z_t being independent of $\mathcal{I}_{t-1}$ and $\sigma(r_u; u < t)$;
  - $\sigma_t > 0$

- This formulation implies that the sign of the current price variation (that is, the sign of r_t) is that of z_t, and is independent of past price variations

- Most importantly, if the first two conditional moments of r_t exist, they are given by

$$\mathbb{E}(r_t|\mathcal{I}_{t-1}) = 0, \ \ \mathbb{E}(r_t^2|\mathcal{I}_{t-1}) = \sigma_t^2$$

- The random variable σ_t is called the volatility of r_t

Modelling Approaches

- As $Cov(r_t, r_{t-h}) = \mathbb{E}(z_t)\mathbb{E}(\sigma_t r_{t-h}) = 0$, r_t^2, generally have nonzero autocovariances

$\Rightarrow$ r_t is a weak white noise

- The kurtosis coefficient of r_t, if it exists, is related to that of z_t

$$\frac{\mathbb{E}(r_t^4)}{\mathbb{E}(r_t^2)^2} = k_z[1 + \frac{Var(\sigma^2)}{\mathbb{E}(\sigma^2)^2}]$$

- Hence, the leptokurticity of financial time series can be taken into account in two different ways:
  - either by using a leptokurtic distribution for the weak white noise sequence (z_t),
  - or by specifying a process (σ_t^2) with a great variability

Modelling Approaches

Different classes of models can be distinguished depending on the specification adopted for σ_t

- Conditionally heteroscedastic (or GARCH-type) processes
  - Here $\mathcal{I}_{t-1} = \sigma(r_s; s < t)$ is the σ-field generated by the past of r_t
  - The volatility is here a deterministic function of the past of r_t
  - Processes of this class differ by the choice of a specification for this function
  - The GARCH model is characterized by a volatility specified as a linear function of the past values of r_t^2
- Stochastic volatility processes
  - Here $\mathcal{I}_{t-1}$ is a σ-field generated by $v_t, v_{t-1}, \cdots$, where (v_t) is a strong white noise and is independent of (z_t)
  - volatility is a latent process
  - a popular specification is the one where the process $\log \sigma_t$ follows an AR(1)

Conditionally heteroscedastic processes

- In these models, the key concept is the conditional variance: the variance conditional on the past
- We can reproduce the autocorrelation empirically seen in conditional volatility by using the information in the previous value(s) of r_t^2

$\Rightarrow$ in an ARCH(q) specification, perturbations follow an AR process of order q

$\Rightarrow$ ARCH(q) are autoregressive conditionally heteroskedastic models

$$\mathbb{V}(r_t) = const$$

$$\mathbb{V}(r_t|\mathcal{I}_{t-1}) = f(r_{t-1}, r_{t-2}, \cdots ; \theta)$$

ARCH test

Usual (Ljung-Box) autocorrelation test on squared returns

- H0: $\rho_1 = \rho_2 = \ldots = \rho_K = 0$

$$Q_{LB}(K) = T(T+2)\sum_{k=1}^{K}\frac{\hat{\rho}_k^2}{T-k}\xrightarrow[T\to\infty]{d}\chi^2(p),$$

where $\hat{\rho}_k$ is the empirical autocorrelation

ARCH-LM test

- Auxiliary regression

$$\hat{\varepsilon}_t^2 = \phi_0 + \phi_1\hat{\varepsilon}_{t-1}^2 + \cdots + \phi_p\hat{\varepsilon}_{t-p}^2 + \eta_t$$

- H0′: $\phi_1 = \cdots = \phi_p = 0$
- Test-statistic: $LM(p) = T \times R^2 \xrightarrow[T\to\infty]{d}\chi^2(p)$

Family of ARCH-type models

- Linear models

$\Rightarrow$ ARCH(q), GARCH(p, q), IGARCH(p, q), etc.

- Non-linear models (i.e. asymmetric models)

$\Rightarrow$ EGARCH(p, q), GJRGARCH(p,q), TGARCH(p, q), etc.

ARCH models Engle (1982)

Model 1

r_t follows an ARCH(1) if

$$r_t = z_t\sqrt{(\sigma_t^2)}, \quad \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2$$

and z_t is a strong white noise with σ_t^2 deterministic and positive process conditionally on the σ-field

- For the ARCH we have :

$$\begin{aligned}\mathbb{V}(r_t|\mathcal{I}_{t-1}) &= \mathbb{V}(z_t\sqrt{(\sigma_t^2)}|\mathcal{I}_{t-1}) \\ &= \sigma_t^2\mathbb{V}(z_t|\mathcal{I}_{t-1}) \\ &= \sigma_t^2\end{aligned}$$

- If $\mathbb{V}(z_t|\mathcal{I}_{t-1})$ is normalized to 1 we can see that σ_t^2 is the conditional variance of r_t

Moments of ARCH(1) process

- Regarding the conditional mean :

$$\begin{aligned}\mathbb{E}(r_t|\mathcal{I}_{t-1}) &= \mathbb{E}(z_t\sigma_t|\mathcal{I}_{t-1}) \\ &= \sigma_t\mathbb{E}(z_t|\mathcal{I}_{t-1}) = 0 \text{ if } z_t \text{ is weak white noise}\end{aligned}$$

- Regarding the unconditional mean :

$$\mathbb{E}(r_t) = \mathbb{E}(\mathbb{E}(r_t)|\mathcal{I}_{t-1}) = 0$$

- Regarding the conditional variance :

$$\begin{aligned}\mathbb{V}(r_t|\mathcal{I}_{t-1}) &= \sigma_t^2\mathbb{V}(z_t|\mathcal{I}_{t-1}) = \sigma_t^2\mathbb{V}(z_t) \\ &= \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2.\end{aligned}$$

- Regarding the unconditional variance :

$$\mathbb{V}(r_t) = \mathbb{E}\left((r_t - \mathbb{E}(r_t))^2\right) = \mathbb{E}(r_t^2)$$

- Under stationarity assumption, we hence have $\mathbb{E}(r_t^2) = \alpha_0 + \alpha_1\mathbb{E}(r_t^2)$ and

$$\mathbb{E}(r_t^2) = \mathbb{V}(r_t) = \frac{\alpha_0}{1-\alpha_1}$$

ARCH(q) Models

Model 2

r_t follows an ARCH(q) process if $r_t = z_t\sigma_t$ with

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2$$

and where z_t is a strong white noise such that $\mathbb{E}(z_t) = 0$ et $\mathbb{E}(z_t^2) = \sigma_z^2$.

This model fulfills the martingale difference and time-varying conditional variance properties

$$\mathbb{E}(r_t|r_{t-1}) = 0 \text{ and } \mathbb{V}(r_t|r_{t-1}) = \alpha_0 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2$$

ARCH-errors model

Model 3

Consider $r_t = \mathbb{E}(r_t | r_{t-1}) + \varepsilon_t$ *with* ε_t *a weak white noise satisfying the martingale difference hypothesis :*

$$\mathbb{E}(\varepsilon) = 0 \text{ and } \mathbb{E}(\varepsilon_t \varepsilon_s) = 0.$$

Then, r_t *follows an ARCH-errors model if* $\varepsilon_t = z_t \sigma_t$ *with*

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2$$

and where z_t *is a strong white noise*

Example of ARCH-errors model

- **AR(1) - ARCH(1)**

$$Y_t = \mu + \rho Y_{t-1} + \varepsilon_t, \varepsilon_t = z_t \sigma_t$$

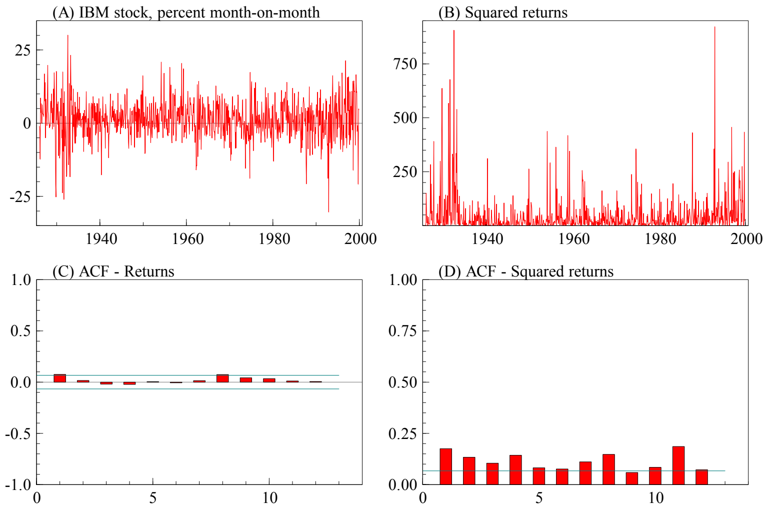
with $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$ and $|\rho| < 1$

The model describes the evolution of both the conditional mean and the conditional variance of Y_t through time

- ε_t : residuals
- z_t : standardized residuals

Residuals ε_t satisfy the properties of an ARCH process:
martingale difference; time-varying conditional variance; zero conditional auto-covariances; leptokurtic distribution

IBM example


(A) IBM stock, percent month-on-month
25
0
-25
1940
1960
1980
2000
(B) Squared returns
750
500
250
1940
1960
1980
2000
(C) ACF - Returns
1.0
0.5
0.0
-0.5
-1.0
0
5
10
(D) ACF - Squared returns
1.00
0.75
0.50
0.25
0.00
0
5
10

Non-negativity constraints and Limits of ARCH

- Non negativity constraints
  - For a ARCH(1) model , $\alpha_0 \geq 0$; $\alpha_1 \geq 0$
  - For an ARCH(q) model, $\alpha_i \geq 0, \ \forall i = 0, 1, \dots, q$
- Limitations of ARCH(q) models
  - q, number of lags of the squared residuals, is potentially very large
  - Non negativity constraints might be violated

IBM example - ARCH(10)

Coeff.	Estimate	Std. Error	t-stat
ω	0.2605	0.0155	16.785
α_1	0.0366	0.0099	3.700
α_2	0.0809	0.0123	6.575
α_3	0.0657	0.0118	5.585
α_4	0.0866	0.0133	6.525
α_5	0.1035	0.0140	7.420
α_6	0.0746	0.0125	5.943
α_7	0.0780	0.0130	6.002
α_8	0.0892	0.0135	6.452
α_9	0.0875	0.0134	6.530
α_{10}	0.0789	0.0130	6.074

Generalised ARCH(1,1) or GARCH(1,1) Models

- Bollerslev (1986): the conditional variance depends on its own past values and on past values of ε_t^2
- GARCH(1,1) model containing 3 parameters is a very parsimonious infinite ARCH model

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

with $\alpha_0 > 0, \alpha_1 > 0$ and $\beta > 0$

The GARCH(p, q) model

Model 4

A process ε_t satisfies a GARCH(p, q) representation if $\varepsilon_t = z_t \sigma_t$ and

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

where z_t is a weak white noise and where $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, q$ and $\beta_i \geq 0$, $i = 1, \ldots, p$

- The conditional variance of the error term depends on own p past values and on q past values of the squared residuals
- But in general a GARCH(1,1) model will be sufficient to capture the volatility clustering in the data

Moments of GARCH

Conditional Moments

- $\mathbb{E}(\varepsilon_t|\varepsilon_{t-1}) = 0$
- $\mathbb{V}(\varepsilon_t|\varepsilon_{t-1}) = \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$

Unconditional Variance

- $\mathbb{V}(r_t) = \mathbb{E}(r_t - \mathbb{E}(r_t)^2) = \mathbb{E}(r_t^2)$
- $\mathbb{V}(r_t) = \mathbb{E}(r_t^2) = \alpha_0 \times \left(1 - \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i\right)^{-1}$

The GARCH model meets financial stylized facts

- Under additional assumptions (second-order stationarity of ε_t^2), we can state that if ε_t is GARCH(p, q), then ε_t^2 is an ARMA(p, q) process

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i)\varepsilon_{t-i}^2 + v_t - \sum_{i=1}^{p} \beta_i v_{t-i}$$

  where $v_t = \varepsilon_t^2 - \sigma_t^2$ are the innovations of the process

- GARCH processes are hence able to capture one important characteristic of financial series: squared returns are autocorrelated

- The sum $\alpha + \beta$ is referred to as the persistence of the conditional variance process

The GARCH model meets financial stylized facts

- Contrary to standard time series models (ARMA), the GARCH structure allows the magnitude of the noise ε_t^2 to be a function of its past values.
- Thus, periods with high volatility level (corresponding to large values of ε_{t-i}^2) will be followed by periods where the fluctuations have a smaller amplitude.

GARCH Stationarity

Theorem 2

A process ε_t satisfies a GARCH(p, q) representation if $\varepsilon_t = z_t\sigma_t$,

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

where z_t is a weak white noise and where $\alpha_0 > 0$, $\alpha_i \geq 0$,$i = 1, \ldots, q$ and $\beta_i \geq 0$, $i = 1, \ldots, p$.

Also, ε_t is **asymptotically second-order stationary** *if and only if*

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j \leq 1$$

GARCH

Drost & Nijman, 1993 define 3 types of GARCH

- The strong GARCH where z_t is a Strong White Noise
- The semi-strong GARCH where z_t is a Weak White Noise
- The weak GARCH where only projections of the conditional variance are considered

$\Rightarrow$ We focus on the simplest case : the strong GARCH

Examples of GARCH(1,1)

Table: Estimation results

		S&P500		DAX	
		statistic	std error	statistic	std error
Daily returns	α_0	0.0074	0.0012	0.0248	0.0031
None	α_1	0.0513	0.0039	0.0910	0.0065
None	β_1	0.9422	0.0042	0.8954	0.0069
weekly returns	α_0	0.0829	0.0292	0.2369	0.0634
None	α_1	0.1015	0.0165	0.1091	0.0165
None	β_1	0.8872	0.0174	0.8642	0.0195
Monthly returns	α_0	0.6531	0.4497	3.4344	1.8789
None	α_1	0.1297	0.0419	0.1276	0.0487
None	β_1	0.8444	0.0505	0.7837	0.0817

IGARCH(p,q)

- When

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j = 1$$

  the model is called an integrated GARCH(p, q) or IGARCH(p, q) model (see Engle and Bollerslev, 1986)

- There is a unit root in the autoregressive part of the ARMA representation of ε_t^2 representation

- Returns are strictly stationary with an infinite variance

Maximum Likelihood (ML)

- Easy to implement once the density function of z_t is specified
- Let us call θ the vector of the parameters to be estimated
- If z_t are assumed to be normally distributed, then the log likelihood function for a sample of T observations is:

$$\begin{aligned}\ell(\varepsilon_1, \varepsilon_2, ..., \varepsilon_T; \theta) &= \sum_{t=1}^{T} \log f\left(\varepsilon_t | \mathcal{I}_{t-1}\right) \\ &= -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log\left(\sigma_t^2(\theta)\right) - \frac{1}{2}\sum_{t=1}^{T}\frac{\varepsilon_t^2(\theta)}{\sigma_t^2(\theta)},\end{aligned}$$

where $\frac{\varepsilon_t^2(\theta)}{\sigma_t^2(\theta)} = z_t^2$

Maximum Likelihood (ML)

$$z_t^2 = z_t^2(\theta) = \frac{r_t - \mathbb{E}(r_t|\mathcal{I}_{t-1})}{\sigma_t^2}$$

and

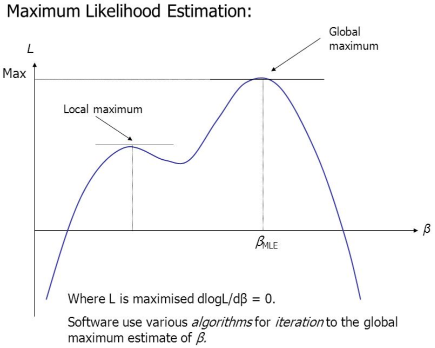
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

- Note that σ_t^2 is not observed for $t = 0, -1, \ldots, -p+1$
- To initialize the process, the unobserved squared residuals are
  1. set to their sample mean
  2. set to the unconditional variance
  3. obtained using a pre-sample
  4. or considered as additional parameters to be estimated
- Under the regularity assumptions, this estimator

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_T^{-1}(\theta_0))$$

Maximum Likelihood (ML)

- $\widehat{\theta}$ does not have a closed-form formula and numerical optimization methods are used
- We need
  1. Initial condition
  2. Moving rule
  3. Stopping rule



Maximum Likelihood (ML) Example

Residuals GARCH(1,1)				
Observation	u_i	$u_i \times u_i$	variance, σ_t^2	$-ln(\sigma_t^2) - \frac{u_t^2}{\sigma_t^2}$
30/04/2004	-0.01195	0.00014	0.00014	7.85
31/05/2004	-0.01082	0.00012	0.00014	8.03
30/06/2004	-0.00015	0.00000	0.00014	8.91
31/07/2004	0.00719	0.00005	0.00010	8.70
31/08/2004	-0.00272	0.00001	0.00009	9.26
30/09/2004	0.01046	0.00011	0.00007	7.99
...	...	...	...	...
30/11/2013	-0.00038	0.00000	0.00004	10.18
31/12/2013	0.00254	0.00001	0.00003	10.11
31/01/2014	0.00594	0.00004	0.00003	9.25
28/02/2014	0.00367	0.00001	0.00004	9.79
31/03/2014	-0.00150	0.00000	0.00004	10.09

α_0	α_1	β
0.00	0.28	0.66

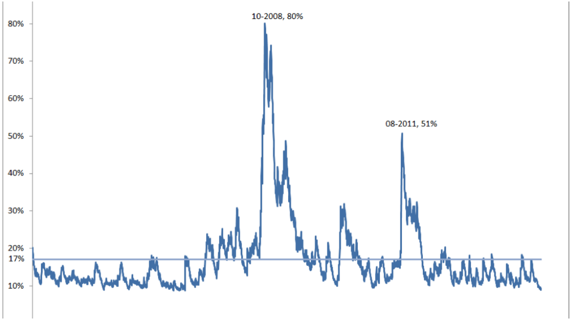
$$LLF = \sum_{i=1}^{m} \left[-ln(\sigma_t^2) - \frac{u_t^2}{\sigma_t^2} \right]^2$$

Maximum Likelihood (ML) Example 2

- Annualized GARCH(1,1) volatility fitted to daily US market returns

$$\sigma_t^2 = \underset{(4.46)}{0.0} + \underset{(8.59)}{0.09}\varepsilon_{t-1}^2 + \underset{(80.19)}{0.9}\ \sigma_{t-1}^2$$

- Volatility is : time varying, persistent, mean-reverting



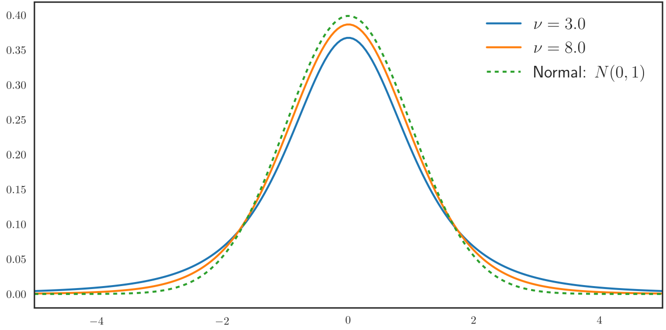
GARCH Extensions

- GARCH models are doing better than ARCH models (e.g., persistence in volatility) but there are still some issues (e.g., tails are not enough fat, etc)
- Improvements in various directions:
  - Non-normality of the conditional distribution: e.g. GARCH-t model
  - Asymmetric GARCH models : e.g. Exponential GARCH model (EGARCH), Threshold GARCH model (TGARCH), GJR model
  - Trade-off mean vs variance : e.g. GARCH-in-mean model

GARCH Extensions

In case of non-normality:

- Student distribution
- Skewed Student distribution
- Generalized error distribution (GED)



GARCH Extensions

The GJR-GARCH(1,1) Model

- Due to Glosten, Jaganathan and Runkle

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma \varepsilon_{t-1}^2 I_{t-1}$$

where $I_{t-1} = \begin{cases} 1 \text{ if } \varepsilon_{t-1} < 0 \\ 0 \text{ otherwise} \end{cases}$

- For a leverage effect, we would see $\gamma > 0$
- We require $\alpha_1 + \gamma \geq 0$ and $\alpha_1 \geq 0$ for non-negativity
- We require $\alpha_1 + 0.5\gamma + \beta < 1$ for stationarity

GARCH Extensions

The GJR-GARCH(1,1) Model

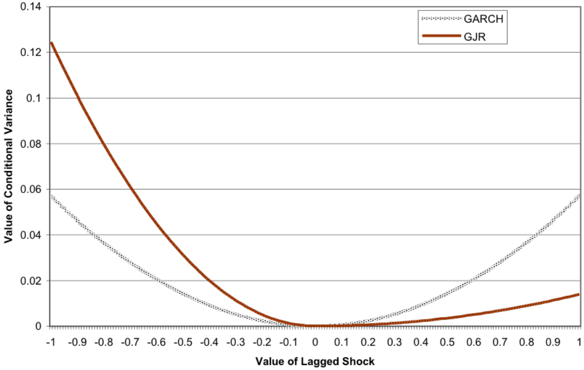
- Using monthly S&P 500 returns, December 1979- June 1998
- Estimating a GJR model, we obtain the following results

$$r_t = \underset{(3.198)}{0.172}$$

$$\sigma_t^2 = \underset{(16.372)}{1.243} + \underset{(0.437)}{0.015}\varepsilon_{t-1}^2 + \underset{(14.999)}{0.498}\,\sigma_{t-1}^2 + \underset{(5.772)}{0.604}\varepsilon_{t-1}^2 I_{t-1}$$

News Impact Curves

- The news impact curve plots the next period volatility (σ_t) that would arise from various positive and negative values of ε_{t-1}, given an estimated model
- News Impact Curves for S&P 500 Returns using Coefficients from GARCH and GJR Model Estimates



EGARCH

- Suggested by Nelson (1991)
- The variance equation is given by

$$\log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{q} a_i z_{t-i} + \sum_{i=1}^{q} b_i(|z_{t-i}| - \mathrm{E}[|z_{t-i}|]) + \sum_{i=1}^{p} \beta_i log(\sigma_{t-i}^2)$$

- Advantages of the model
  — Since we model the $\log \sigma_t^2$, then even if the parameters are negative, σ_t^2 will be positive

GARCH-in-mean or GARCH-M

- Asset pricing models suppose that higher risks should be rewarded by higher returns
- The GARCH-in-mean model lets the mean of an asset's returns to be determined by its lagged conditional volatility
  $$r_t = \mu + \delta G(\sigma_t^2) + \varepsilon_t, \ \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2)$$
  with $G(\sigma_t^2)$ a linear or square-root function and
  $$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta \sigma_{t-1}^2,$$
- The parameter δ can be interpreted as the price of risk and can thus be assumed to be positive
- Hence, if $\delta > 0$, increases in risk (given by increases in conditional volatility) lead to higher mean returns

Portfolio: optimization problem

- Consider two assets and their log-returns $r_t = (r_{1,t}, r_{2,t})'$, denoted $\boldsymbol{r}$ when stacked over time.
- If the market is efficient, returns are unpredictable with mean μ
- The conditional covariance matrix

$$H_t = \begin{pmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{pmatrix}$$

  is however predictable if one can fit a stationary GARCH-type model for $r_{1,t}$ and $r_{2,t}$
- Let define Σ_τ, the forecast of H_t at date τ and the expected (targeted) return for the portfolio, $\tilde{\mu}$
- What is of interest for the portfolio manager is to find a minimum risk portfolio (efficient) subjected to provide the targeted return :

$$\min_{\omega} \sigma_p^2(\omega) = \omega' \Sigma_\tau \omega$$
$$\text{s.c.} \quad \omega' \mu = \tilde{\mu} \quad \text{and} \quad \omega' \iota = 1$$

  with ι a unit vector and ω the allocation weights

Portfolio: resolution

- You should be able to derive the following results (based here on the Lagrangien).

$$\mathcal{L}(\omega, \lambda_1, \lambda_2) = \omega' \Sigma_\tau \omega + \lambda_1(\omega'\mu - \tilde{\mu}) + \lambda_2(\omega'\iota - 1)$$

- The partial derivatives lead to the following system

$$2\Sigma_\tau \omega + \lambda_1 \mu + \lambda_2 \iota = 0 \Rightarrow \omega = \frac{1}{2}\Sigma_\tau^{-1}(-\lambda_1 \mu - \lambda_2 \boldsymbol{u}) \quad (1)$$

$$\omega'\mu - \tilde{\mu} = 0 \Rightarrow \frac{1}{2}\lambda_1 \mu' \Sigma_\tau^{-1} \mu + \frac{1}{2}\lambda_2 \iota' \Sigma_\tau = \tilde{\mu} \quad (2)$$

$$\omega'\iota - 1 = 0 \Rightarrow \frac{1}{2}\lambda_1 \mu' \Sigma_\tau^{-1} \iota + \frac{1}{2}\lambda_2 \iota' \Sigma_\tau = 1 \quad (3)$$

- Denoting $A = \iota' \Sigma_\tau^{-1} \mu = \mu' \Sigma_\tau^{-1} \iota$, $B = \mu' \Sigma_\tau^{-1} \mu$ and $C = \iota' \Sigma^{-1} \iota$ we have

$$A\lambda_1 + B\lambda_2 = 2\tilde{\mu}$$
$$A\lambda_1 + C\lambda_2 = 2$$

- Solving this system we obtain

$$\lambda_1 = 2(C\tilde{\mu} - A)(BC - A^2)^{-1} \text{ et } \lambda_2 = 2(B - A\tilde{\mu})(BC - A^2)^{-1}$$

Portfolio: resolution

- Substituting in the weight equation we finally get

$$
\begin{aligned}
\omega^*(\tilde{\mu}) &= \frac{1}{2}\Sigma_\tau^{-1}\left[\lambda_1\mu + \lambda_2\iota\right] \\
&= \frac{1}{2}\Sigma_\tau^{-1}\left[2\left(\frac{C\tilde{\mu}-A}{D}\right)\mu + 2\left(\frac{B-A\tilde{\mu}}{D}\right)\iota\right] = \frac{1}{D}\Sigma_\tau^{-1}\Big[(C\tilde{\mu}-A)\mu + (B-A\tilde{\mu})\iota\Big] \\
&= \frac{1}{D}\Sigma_\tau^{-1}\Big[(B\iota - A\mu) + (C\mu - A\iota)\tilde{\mu}\Big] = \frac{1}{D}\Sigma_\tau^{-1}(B\iota - A\mu) + \frac{1}{D}\Sigma_\tau^{-1}(C\mu - A\iota)\tilde{\mu} \\
&= E\tilde{\mu} + F
\end{aligned}
$$

- The minimal risk (efficient) portfolio is hence given by

$$
\sigma_p^2(\tilde{\mu}) = \omega^*(\tilde{\mu})'\Sigma_\tau^{-1}\omega^*(\tilde{\mu})
$$

- If we omit the investor preferences ($\tilde{\mu}$) in the minimization problem the solution simplifies to

$$
\omega^* = \frac{\Sigma_\tau^{-1}\iota}{\iota'\Sigma_\tau^{-1}\iota}
$$

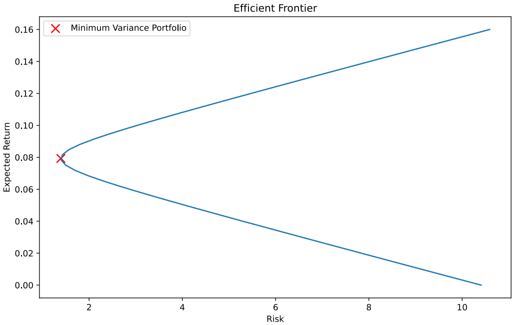
and we get the so-called minimum-variance portfolio

Portfolio: implementation

- After estimating individual GARCH models, collect conditional variance $h_{t,ii}$, $i = 1, 2$
- Forecast $\widehat{h}_{ii,t}$ at horizon k, k being the investment date or next portfolio rebalancing
- As here we assume r_t unpredictable, we just use the unconditionnal mean as a predictor
- In a univariate framework, we need a constant correlation assumption to go further because :

$$\rho_{ij,t} = \frac{h_{ij,t}}{\sqrt{h_{ii,t}h_{jj,t}}}$$

$\Rightarrow$ If $\rho_{ij,t} = \rho_{ij}$ we can use the Pearson correlation and $h_{ij,t} = \sqrt{h_{ii,t}h_{jj,t}}\rho_{ij}$ to reconstruct $\widehat{\Sigma}_\tau$



Multivariate GARCH models

- While the volatility of univariate series has been the focus of the previous chapters, modeling the comovements of several series is of great practical importance
- The standard linear modeling of real time series has a natural multivariate extension through the framework of the vector ARMA (VARMA) models
- Similarly, here we introduce the concept of multivariate GARCH model
- Essential for asset pricing and risk management crucially depend on the conditional covariance structure of the assets of a portfolio

Multivariate GARCH models

Let denote by r_t a column vector of k asset returns and the vector of their conditional expectations by μ_t

- Returns' equation implies a **conditional covariance** matrix H_t:

$$r_t - \mu_t = \varepsilon_t = H_t^{1/2} z_t, \quad H_t^{1/2}(H_t^{1/2})' = H_t$$

  — ε_t is a vector, not a scalar as previously

  — where H_t is a matrix $k \times k$ with elements h_{ijt}

  — and z_t is i.i.d Gaussian such that $\mathbb{E}(z_t) = 0$ and $\mathbb{E}(z_t z_t') = I$ with I a $k \times k$ identity matrix

- The **conditional covariance** matrix H_t takes the form

$$H_t = f(H_{t-1}, H_{t-2}, \cdots, \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots)$$

- If $H_t^{1/2}$ exists, H_t is positive definite

  $\Rightarrow$ the transformation $f(.)$ ought to insure that H_t is symmetric and positive definite (strictly positive eigen-values)

- But $f(H_{t-1}, H_{t-2}, \cdots, \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots)$ is complex...

Multivariate GARCH models

Choosing a specification for H_t is obviously more delicate than in the univariate framework because:

(i) H_t should be (almost surely) symmetric, and positive definite for all t

(ii) the specification should be simple enough to be amenable to probabilistic study (existence of solutions, stationarity, ...), while being of sufficient generality

(iii) the specification should be parsimonious enough to enable feasible estimation

(iv) but, the model should not be too simple to be able to capture the - possibly sophisticated - dynamics in the covariance structure

Multivariate GARCH models

- Moreover, it may be useful to have the so-called **stability by aggregation property**
- If $\varepsilon_t = H_t^{1/2} z_t$ is satisfied, the process $\tilde{\varepsilon}_t = P\varepsilon_t$, where P is an invertible square matrix, is such that

$$\mathbb{E}(\tilde{\varepsilon}_t | \tilde{\varepsilon}_u, u < t) = 0, \;\; \mathbb{V}(\tilde{\varepsilon}_t | \tilde{\varepsilon}_u, u < t) = \tilde{H}_t = P H_t P$$

- The stability by aggregation of a class of specifications for H_t requires that the conditional variance matrices $\tilde{H}_t$ belong to the same class for any choice of P
- Relevance: if the components of the vector ε_t are asset returns, $\tilde{\varepsilon}_t$ is a vector of portfolios of the same assets, each of its components consisting of amounts (coefficients of the corresponding row of P) of the initial assets

Distribution of multivariate GARCH models

- Generally z_t is assumed to follow the multivariate Gaussian distribution, $z_t \sim N(0, I)$, since it provides the basis of QML estimation as in the univariate case
- Another choice of density for z_t is the multivariate t
- Multivariate skewed distributions can also be used (e.g. the skewed-t of Bauwens and Laurent, 2005)
- As in the univariate case, distributions with fat-tails and skewness are usually better fitting data than the Gaussian

Representation of multivariate GARCH models

- Unlike the ARMA models, however, the GARCH model specification does not suggest a natural extension to the multivariate framework
- Indeed, the (conditional) expectation of a vector of size k is a vector of size k, but the (conditional) variance is a $k \times k$ matrix
- Important milestones are
  - the BEKK model of Engle and Kroner (1995)
  - the constant conditional correlation (CCC) model of Bollerslev (1990)
  - the dynamic correlation model (DCC) of Engle (2002a)
  - the time-varying correlation (TVC) model of Tse and Tsui (2002)
- Earlier models had too many parameters to be useful for modeling more than two asset returns jointly (e.g. VEC model)

VEC models

- Take the case of a bivariate ($k = 2$) GARCH(p, q):

$$H_t = \begin{pmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{pmatrix} \Rightarrow h_t = vec(H_t) = \begin{pmatrix} h_{11t} \\ h_{12t} \\ h_{21t} \\ h_{22t} \end{pmatrix}$$

Note : the operator $vec(.)$ consists in vectorizing a matrix by stacking the columns of the matrix on top of one another

- Using this operator, Engle et Kroner (1995) propose the VEC model:

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i vec(\varepsilon_{t-i}\varepsilon'_{t-i}) + \sum_{i=1}^{p} \beta_i h_{t-i}$$

with ω a $k \times 1$ vector, and α_i and β_i $k \times k$ matrices

- **Problem:**
  - the model is big and some equations are redundant: $h_{12t} = h_{21t}$ as H_t is a covariance matrix
  - it will not in general produce positive definite covariance matrices H_t

VECH models

- Apply a $vech(.)$ operator now to the previous $GARCH(p,q)$:

$$H_t = \begin{pmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{pmatrix} \Rightarrow h_t = vech(H_t) = \begin{pmatrix} h_{11t} \\ h_{21t} \\ h_{22t} \end{pmatrix}$$

Note : The operator $vech(.)$ consists in vectorizing a matrix by stacking the columns of the lower triangular part of its argument square matrix

- One obtains the VECH model where $h_t = vech(H_t)$

Model 5

The process ε_t is said to admit a VEC-GARCH(p,q) representation (relative to the i.i.d sequence z_t) if it satisfies

$$\varepsilon_t = H_t^{1/2} z_t, \text{ where } H_t \text{ is positive definite such that}$$

$$vech(H_t) = \omega + \sum_{i=1}^{q} A^{(i)} vech(\varepsilon_{t-1}\varepsilon'_{t-1}) + \sum_{j=1}^{p} B^{(j)} vech(H_{t-j}),$$

where ω is a vector of size $\{k(k+1)/2\} \times 1$, and the $A^{(i)}$ and $B^{(j)}$ are matrices of dimension $k(k+1)/2 \times k(k+1)/2$.

VEC models

- In particular, for a bivariate VECH-GARCH(1,1)

$$h_t = \begin{pmatrix} h_{11t} \\ h_{21t} \\ h_{22t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \begin{pmatrix} \varepsilon^2_{1,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon^2_{2,t-1} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{21,t-1} \\ h_{22,t-1} \end{pmatrix}$$

- Every conditional covariance is a function of lagged conditional variances as well as lagged cross-products of all components
- More parsimonious model than the VEC-GARCH
- But the VECH-GARCH still implies a big number of coefficients
- **Problem:** VEC and VECH are not able to generally insure that H_t is positive definite

Diagonal VECH-GARCH models

- To further simplify the model and its estimation, one may assume that volatilities and covariances depend only on their past values (Bollerslev, Engle, and Wooldridge, 1988)

  $\Rightarrow$ Non-diagonal coefficients of A_i and B_i are null

- For instance, the diagonal VECH-GARCH(1,1) gives

$$h_t = \begin{pmatrix} h_{11t} \\ h_{21t} \\ h_{22t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{pmatrix} \begin{pmatrix} \varepsilon^2_{1,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon^2_{2,t-1} \end{pmatrix} + \begin{pmatrix} \beta_{11} & 0 & 0 \\ 0 & \beta_{22} & 0 \\ 0 & 0 & \beta_{33} \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{21,t-1} \\ h_{22,t-1} \end{pmatrix}$$

- More parsimonious than the VEC-GARCH

- The diagonal VEC-GARCH is stable by aggregation

- In this case it is possible to obtain conditions for positive definiteness of H_t for all t

BEKK model

- Developed by Baba, Engle, Kraft and Kroner, in a preliminary version of Engle and Kroner (1995)

Model 6

Let (z_t) denote an i.i.d. sequence with common distribution. The process (ε_t) is called a BEKK-GARCH(p, q), with respect to the sequence (z_t), if it satisfies

$$\varepsilon_t = H_t^{1/2} z_t$$

$$H_t = C'C + \sum_{n=1}^{N}\sum_{i=1}^{q} A_{in}\varepsilon_{t-i}\varepsilon_{t-i}'A_{in}' + \sum_{n=1}^{N}\sum_{i=1}^{p} B_{in}'H_{t-i}B_{in}'$$

with A_{in}, B_{in}, $n \in \{1, \dots, N\}$, and C matrices of dimension $k \times k$

- Each BEKK model implies a unique VECH model, while the converse implication is not true
- The BEKK class contains the diagonal models by choosing diagonal matrices A_{ik} and B_{jk}

Note: The sum over N introduces a complex generalization so we consider only the case $N = 1$ hereafter

BEKK model

Theorem 3

H_t **is positive definite** *if matrices* $H_{t-i},\ i = 1, \ldots, p$*, are almost surely positive definite and*

$$ker\{C\} \bigcap_{j=1}^{p} \bigcap_{n=1}^{N} ker\{B_{jn}\} = \{0\}$$

- This is a weak condition, requiring only that C and B_{jn} are full rank (e.g. triangular C with positive diagonal elements)
- an identifiability restriction is needed, $H_{jj,t}$ being invariant to a change of sign of the j-th row of any matrix A_i
- BEKK-GARCH$(1, 1)$ in the bivariate case ($k = 2$) with $N = 1$

$$H_t = C'C + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}' \begin{pmatrix} \varepsilon^2_{1,t-1} & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon^2_{2,t-1} \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}$$
$$+ \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}' \begin{pmatrix} h^2_{11,t-1} & h_{12,t-1} \\ h_{21,t-1} & h^2_{22,t-1} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}$$

Stationarity of the BEKK model

Definition 11

Let C be an upper triangular $n \times n$ matrix and A_{in}, B_{in} be $n \times n$ parameter matrices. Let z_t be an i.i.d. process with mean zero and unit variance. Hence z_t is independent of $\mathcal{I}_{t-1}$, and $cov(z_t|\mathcal{I}_{t-1}) = cov(z_t) = I$.

There exists a covariance stationary BEKK process ε_t, such that $\varepsilon_t = H_t^{1/2} z_t$, where $H_t = cov(\varepsilon_t|\mathcal{I}_{t-1})$ and $\mathcal{I}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ if and only if all the eigenvalues of

$$\sum_{i=1}^{q} A_{in} \otimes A_{in} + \sum_{n=1}^{N} \sum_{i=1}^{p} B_{in} \otimes B_{in}$$

are less than one in modulus.

Estimation of the BEKK model

- Under the assumption that z_t are i.i.d. conditionally on initial values, the quasi log-likelihood function of the BEKK model is given by

$$L_n(\theta) = L_n(\theta; \varepsilon_1, \ldots, \varepsilon_n) = \sum_{t=1}^{n} -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|H_t| - \frac{1}{2}\varepsilon_t' H_t^{-1}\varepsilon_t,$$

where

$$\varepsilon_t = H_t^{1/2} z_t$$

$$H_t = C'C + \sum_{n=1}^{N}\sum_{i=1}^{q} A_{in}\varepsilon_{t-i}\varepsilon_{t-i}' A_{in}' + \sum_{n=1}^{N}\sum_{i=1}^{p} B_{in}' H_{t-i} B_{in}'$$

Estimation of the BEKK model

- Comte and Lieberman (2003) provide conditions for strong consistency and asymptotic normality of the quasi maximum likelihood estimator
  - Strong consistency

$$\hat{\theta}_n \rightarrow \theta_0 \textit{ almost surely when } n \rightarrow \infty$$

  - Asymptotic normality

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, J^{-1}IJ^{-1}),$$

where J is a positive definite matrix and I is a positive semi-definite matrix, defined by

$$I = \mathbb{E}\Big(\frac{\partial l_t(\theta_0)}{\partial \theta}\frac{\partial l_t(\theta_0)}{\partial \theta'}\Big), \quad J = \mathbb{E}\Big(\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'}\Big)$$

Conditional correlations

- Multivariate GARCH models allow one to compute conditional **variances and covariances**
- **Conditional correlations** can hence be reconstructed

$$\rho_{ij,t} = \frac{h_{ij,t}}{\sqrt{h_{ii,t}h_{jj,t}}}$$

with $\{i,j\} = 1, \cdots, k$ and $i \neq j$

Constant Conditional Correlations models

- Suppose that for a multivariate GARCH process of the form

$$\varepsilon_t = H_t^{1/2} \tilde{z}_t$$

all the past information on ε_{it}, involving all the variables $\varepsilon_{j,t-i}$, is summarized in the conditional variance

- The standardized innovations $z_{it} = h_{ii,t}^{-1/2} \varepsilon_{it}$ are sequences of i.i.d (0,1) variables generally correlated
- Denote the covariance matrix $R = \mathbb{V}(z_t) = (\rho_{i,j})$, with $z_t = (z_{1t}, \ldots, z_{kt})$
- In CCC models the conditional covariances $h_{ij,t}$ are obtained as $h_{ij,t} = \rho_{i,j} \sqrt{(h_{ii,t} h_{jj,t})}$ for $i \neq j$ and they are time varying although the correlations are constant
- In matrix notations,

$$H_t = D_t R D_t = \rho_{i,j} \sqrt{(h_{ii,t} h_{jj,t})}$$

with D_t a $k \times k$ diagonal matrix with $\sqrt{h_{11,t}}, \cdots, \sqrt{h_{kk,t}}$ on its main diagonal

$\Rightarrow$ H_t is positive-definite if $h_{ii,t}$ is positive for all i and R_t is positive-definite

Constant Conditional Correlations models

Definition 12

Let $\tilde{z}_t$ be a sequence of i.i.d. variables. A process ε_t is called CCC-GARCH(p, q) if it satisfies

$$\varepsilon_t = H_t^{1/2}\tilde{z}_t$$
$$H_t = D_t R D_t$$
$$\boldsymbol{h}_t = \boldsymbol{\omega} + \sum_{s=1}^{q} \boldsymbol{A}_s \boldsymbol{\varepsilon}_{t-s} + \sum_{v=1}^{p} \boldsymbol{B}_v \boldsymbol{h}_{t-v},$$

where $R = cov(z_t z_t')$ is a correlation matrix, $D_t = diag(\sqrt{\boldsymbol{h}_t})$, $\boldsymbol{h}_t$ is the vector of k conditional variances with elements $(h_{ii,t})$, $\boldsymbol{\varepsilon}_t$ is the vector of k squared innovations (non-standardized), ω is a $m \times 1$ vector with positive coefficients, $\boldsymbol{A}_s$ and $\boldsymbol{B}_v$ are $k \times k$ matrices with nonnegative coefficients

- Note that $\varepsilon_t = D_t z_t$, where $z_t = R^{1/2}\tilde{z}_t$ is a centered vector with covariance matrix R such that

$$\varepsilon_{i,t} = h_{ii,t}^{1/2} z_{i,t}$$

- Note that $h_{ii,t}$ may depend on the past of all the components of ε_t

Strict stationarity of the CCC model

Definition 13

The CCC-GARCH(p,q) model admits a second-order stationary solution if the vector of parameters is such that the roots of the polynomial $\det(I - \sum_{i=1}^{s}(A_i + B_i)\lambda)$ *with* $s = sup(p, q)$*, are outside the unit cercle. This solution is unique and ergodic.*

Estimation of CCC models

- By the quasi-maximum likelihood method
- Overall there are $k + k^2(p+q) + k(k-1)/2$ parameters to estimate $\theta = (\boldsymbol{\omega}', \alpha', \beta', \rho')'$
- Let $(\varepsilon_1, \ldots, \varepsilon_n)$ be a sample of length n of the unique nonanticipative and strictly stationary variable ε_t of the CCC model
- Conditionally on nonnegative initial values $\varepsilon_0, \ldots, \varepsilon_{1-q}, \boldsymbol{h}_0, \ldots, \boldsymbol{h}_{1-p}$, the Gaussian quasi-likelihood is written as

$$L_n(\theta) = L_n(\theta; \varepsilon_1, \ldots, \varepsilon_n) = \prod_{t=1}^{n} \frac{1}{(2\pi)^{k/2}|H_t|^{1/2}} \exp\left(-\frac{1}{2}\varepsilon_t' H_t^{-1} \varepsilon_t\right),$$

where H_t are recursively defined, for $t \geq 1$, by

$$H_t = D_t R D_t, \quad D_t = \{\mathrm{diag}(\boldsymbol{h_t})\}^{1/2} \tag{4}$$

$$\boldsymbol{h_t} = \boldsymbol{h_t}(\theta) = \boldsymbol{\omega} + \sum_{s=1}^{q} \boldsymbol{A}_s \boldsymbol{\varepsilon}_{t-s} + \sum_{v=1}^{p} \boldsymbol{B}_v \boldsymbol{h}_{t-v} \tag{5}$$

Estimation of CCC models

- Under the assumption that each conditional variance is specified as a function of its own lags and the i^{th} element of ε_t (denoted by ε_{it}), for example, by a GARCH(1,1) equation, an important simplification is obtained in QML estimation

- This assumption splits the log-likelihood function into two parts

$$
\begin{aligned}
l_n(\theta) = \log L_n(\theta) &= -\frac{1}{2}\sum_{t=1}^{n}(2\log|D_t| + \log|R| + z_t' R z_t) \\
&= -\frac{1}{2}\sum_{t=1}^{n}(2\log|D_t| + z_t' z_t) \\
&\quad -\frac{1}{2}\sum_{t=1}^{n}(\log|R| + z_t' R z_t - z_t' z_t)
\end{aligned}
$$

- The parameters of the conditional variances appear only in D_t (first term), while the parameters of the conditional correlation matrix R_t appear only in the second term

Estimation of CCC models

- So the estimation can be performed in two steps
  - Estimate univariate GARCH models for each asset $i = 1, \dots, k$ and construct standardized residuals
    $$z_t = D_t^{-1}\varepsilon_t$$
  - In a second step, estimate the correlation model (i.e. the constant conditional correlations) based on
    $$\mathbb{E}(z_t z_t') = D_t^{-1} H_t D_t^{-1} = R,$$
    where R is symmetric and positive definite

Remark: The separate estimation of each conditional variance model and of the correlation model is the key to enable estimation of MGARCH models of conditional correlations when k is large, where large means more than, say, 5

Remark: The price to pay for this is the impossibility of including spillover terms in the conditional variance equations, i.e. terms involving $\varepsilon_{t-1,j}$ or $h_{t-1,j}$ for $j \neq i$

Estimation of CCC models

- A QMLE of θ is defined as a measurable solution $\hat{\theta}_n = \underset{\theta \in \Theta}{\arg\max}\, L_n(\theta)$
- Under several assumption the following asymptotic properties of the QMLE estimator can be established (Francq and Zakoïan, 2010)
  — Strong consistency

$$\hat{\theta}_n \to \theta_0 \textit{ almost surely when } n \to \infty$$

  — Asymptotic normality

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, J^{-1} I J^{-1}),$$

  where J is a positive definite matrix and I is a positive semi-definite matrix, defined by

$$I = \mathbb{E}\Big(\frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'}\Big), \;\; J = \mathbb{E}\Big(\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'}\Big)$$

CCC models

- The hypothesis of CCCs is not tenable except for specific cases and short periods
- Several tests of the null hypothesis of constant correlations exist: see e.g. Silvennoinen and Terasvirta (2005)
- Indeed, many empirical work show that the matrix R is time-varying

$$H_t = D_t R_t D_t,$$

  with R_t measurable with respect to the past variables $\{\varepsilon_u, u < t\}$
- Dynamic conditional correlations GARCH (DCC-GARCH) of Engle et Sheppard (2001) is the most well known multivariate approach introducing dynamics for the conditional correlation
- For reasons of parsimony, it seems reasonable to choose diagonal matrices $\boldsymbol{A}_s$ and $\boldsymbol{B}_v$ as discussed on slide 88 regarding the definition of CCC models (on slide 85), corresponding to univariate GARCH models for each component

DCC models

- Dynamic conditional correlations GARCH models are an extension of CCC-GARCH, obtained by introducing a dynamic for the conditional correlation (Engle 2002)

Definition 14

The DCC process is a martingale difference sequence ε_t relative to a given filtration $\mathcal{I}_t$, whose conditional covariance matrix $H_t = cov(\varepsilon_t|\mathcal{I}_{t-1})$ satisfies

$$H_t = D_t R_t D_t$$

where $D_t = diag(h_{11,t}^{1/2} \dots h_{kk,t}^{1/2})$ and R_t is a $k \times k$ **time varying** *correlation matrix of z_t.*
Besides, $h_{ii,t}$ is defined as univariate GARCH(p, q) model where the usual restrictions for non-negativity and stationarity are imposed.

- The univariate GARCH models can have different orders

- The number of parameters to be estimated is quite large when k is large (e.g. equal to $(k+1)(k+4)/2$ in bivariate case for a DCC(1,1))

DCC models

Different DCC(1,1) models are obtained depending on the specification of R_t

- Simple GARCH-like formulation

$$R_t = \theta_0 R + \theta_1 \Psi_{t-1} + \theta_2 R_{t-1},$$

  with R a constant correlation matrix, and Ψ_{t-1} the empirical correlation matrix of $z_{t-1}, \ldots, z_{t-M}$ and

$$R_t = \mathrm{diag}(Q_t)^{-1/2} Q_t \,\mathrm{diag}(Q_t)^{-1/2},$$

  where

$$Q_t = \theta_0 \bar{Q} + \theta_1 z_{t-1} z'_{t-1} + \theta_2 Q_{t-1}$$

  with $\theta_1 > 0$, $\theta_2 > 0$, $\theta_1 + \theta_2 < 1$, $\theta_0 = 1 - \theta_1 - \theta_2$, and $\bar{Q} = cov(z_t z'_t)$

- One can test the assumption of constant conditional covariance matrix through the restriction $\theta_2 = \theta_3 = 0$

- Both ensure that H_t is positive definite if R_t is positive definite with elements in the unit cercle. For this, Q_t and its initial value have to be positive definite.

Estimation of the DCC model

Suppose that the process z_t is multivariate Gaussian distributed such that $\mathbb{E}(z_t) = 0$ and $\mathbb{E}(z_t z_t') = I$.

- The DCC model can be estimated by a two-step procedure as the conditional variance $H_t = D_t R_t D_t$ can be divided into volatility part and correlation part (Engle 2002)
- The method is thought to produce consistent but not efficient estimators
- The log-likelihood takes the form of

$$
\begin{aligned}
l_n(\theta) &= -\frac{1}{2}\sum_{t=1}^{n}\left(\log(|H_t|) + \varepsilon_t' H_t^{-1}\varepsilon_t\right) \\
&= -\frac{1}{2}\sum_{t=1}^{n}\left(2\log(|D_t|) + \log(|R_t|) + \varepsilon_t' D_t^{-1} R_t^{-1} D_t^{-1}\varepsilon_t\right)
\end{aligned}
$$

Estimation of the DCC model

- In the first step the likelihood involves replacing R_t with the identity matrix I

$$l_{1,n}(\theta_a) = -\frac{1}{2}\sum_{t=1}^{n}\left(2\log(|D_t|) + \log(|I|) + \varepsilon_t^{'} D_t^{-1} I^{-1} D_t^{-1}\varepsilon_t\right)$$
$$= -\frac{1}{2}\sum_{i=1}^{k}\sum_{t=1}^{n}\left(\log(h_{ii,t}) + \frac{\varepsilon_{ii,t}^2}{h_{ii,t}}\right),$$

  where θ_a corresponds to the vector of parameters of the univariate GARCH model for all series

- Once θ_a is estimated, $h_{ii,t}$ is estimated such that z_t and $\bar{Q}$ can be estimated as well
- In the second step, $\theta_b = (\theta_1, \theta_2)$ is estimated, given the estimated parameters from step one

$$l_{2,n}(\theta_b|\hat{\theta}_a) = -\frac{1}{2}\sum_{t=1}^{n}\left(2\log(|D_t|) + \log(|R_t|) + \varepsilon_t^{'} D_t^{-1} R_t^{-1} D_t^{-1}\varepsilon_t\right)$$
$$= -\frac{1}{2}\sum_{t=1}^{n}\left(2\log(|D_t|) + \log(|R_t|) + z_t' R_t^{-1} z_t\right)$$

Estimation of the DCC model

- Asymptotic properties of the two-step estimation procedure have been studied in Engle and Sheppard (2001)
- However, Aielli (2009) showed that the estimation of Q by $\hat{R}$ is **inconsistent** since

$$\mathbb{E}(z_t z_t) = \mathbb{E}(\mathbb{E}(z_t z_t' | \mathcal{I}_{t-1})) = \mathbb{E}(R_t) \neq \mathbb{E}(Q_t)$$

- The consistent DCC (cDCC) relies on a consistent specification of Q_t

$$Q_t = (1 - \theta_1 - \theta_2)\bar{Q} + \theta_1 \operatorname{diag}(Q_{t-1}^{1/2}) z_{t-1} z_{t-1}' \operatorname{diag}(Q_{t-1}^{1/2}) + \theta_2 Q_{t-1},$$

such that $\bar{Q}$ is the unconditional covariance matrix of $\operatorname{diag}(Q_{t-1}^{1/2}) z_t$

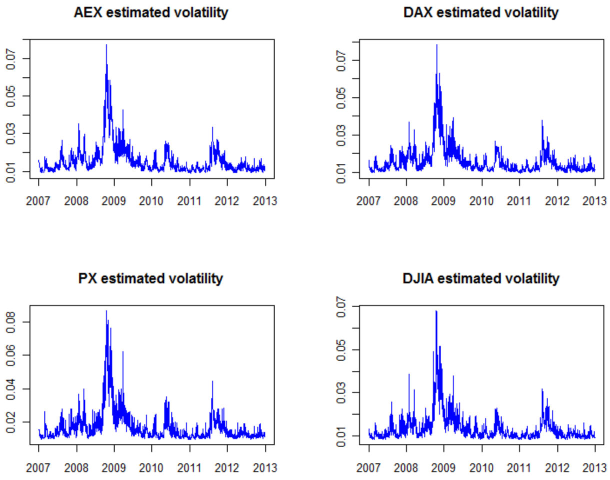
Multivariate GARCH models: example

- Returns on 4 stock market indices: AEX, DAX, PX and DJIA from January 2007 to December 2012

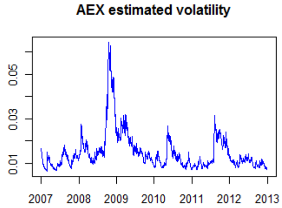
Unconditional Correlation coefficients of the returns series

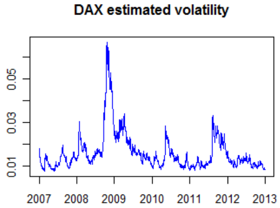
	AEX	DAX	PX	DJIA
AEX	1			
DAX	0.8568444	1		
P	0.5330840	0.4924072	1	
DJIA	0.5630591	0.6086716	0.3260289	1

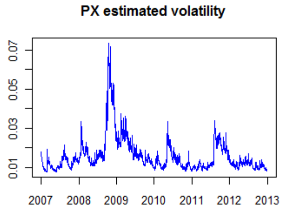
Multivariate GARCH models: BEKK model

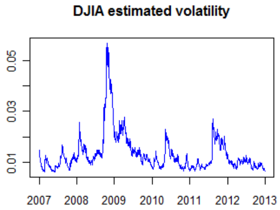


Multivariate GARCH models: DCC model (smoother volatilities)

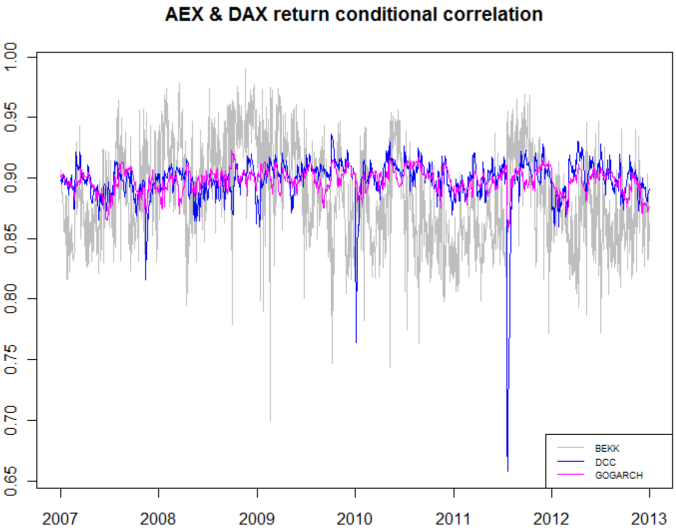




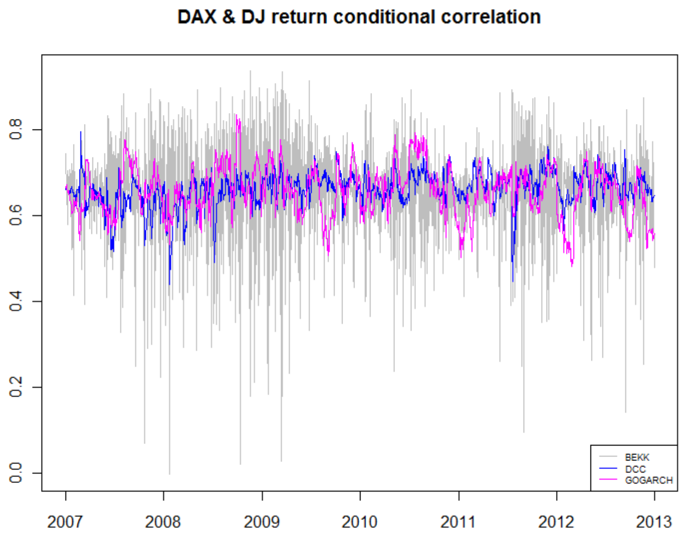




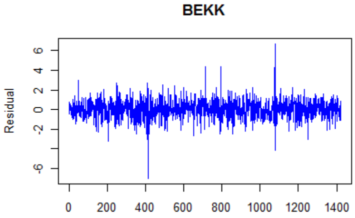
Multivariate GARCH models: example

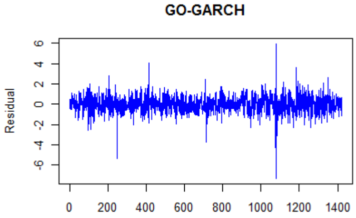


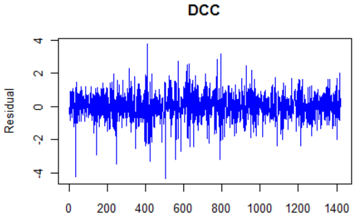
Multivariate GARCH models: example



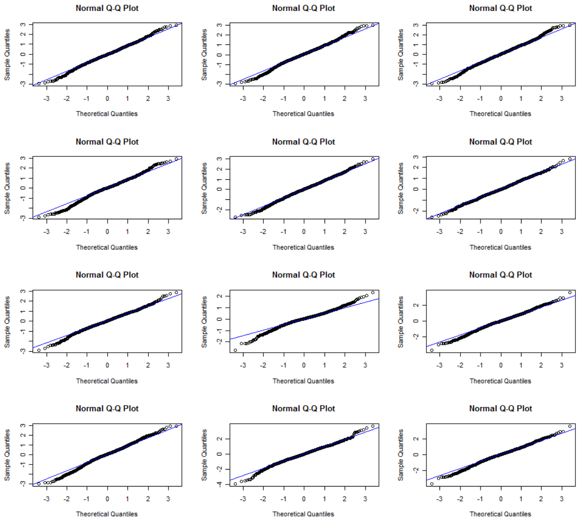
Multivariate GARCH models: example







Multivariate GARCH models: Residual QQ plots



Chapter 2

Finance Theory and High Frequency Modelling

Efficiant Market Hypothesis (EMH)

- The EMH applies rational expectations to asset pricing
  - Let S_t be an asset price and $\mathcal{I}_t$ the information available up to t
  - EMH : if investors use all available information in forming expectations

$$\mathbb{E}(S_{t+1}|\mathcal{I}_t) = S_t$$

  ⇒ future prices are impossible to forecast

- As we also assume that

$$\mathbb{E}(|S_t|) < \infty \text{ and } \mathbb{E}(S_{t+1} - S_t|\mathcal{I}_t) = 0,$$

  S_t is defined as a martingale

Brownian motion (or Wiener process)

- The martingale property is crucial in finance

⇒ it is the cornerstone of most of asset pricing theories

- A particular type of martingale is the Wiener process $W_t \in \mathcal{M}$
  - $W_0 = 0$
  - $W_{t+u} - W_t, \forall u \geq 0$ is independent of W_s, $s \leq t$
  - $W_{t+u} - W_t \sim \mathcal{N}(0, u)$
  - W_t is continuous in t

⇒ Asset prices are often defined as Brownian martingales

$S_t = \mu t + \sigma W_t$, with μ a drift and σ a finite constant volatility

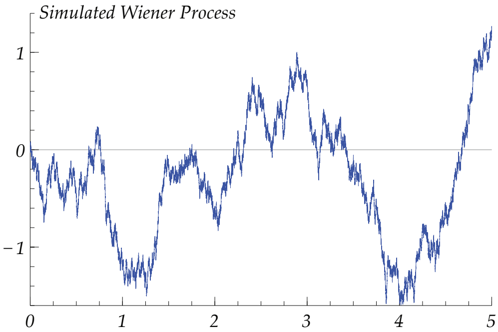


Figure: AI generated picture

Semimartingale

- A stochastic process M_t can satisfy locally the martingale property

$\Rightarrow$ M_t is a local martingale ($\mathcal{LM}$) : $M_t \in \mathcal{LM}$

- M_t enters a more general class of processes : the semimartingale ($\mathcal{SM}$)
- If a stochastic process $S_t \in \mathcal{SM}$, then it can be decomposed as

$$S_t = A_t + M_t$$

  where A_t is a càdlàg adapted process with locally bounded variation

$\Rightarrow$ "càdlàg" means "continue à droit et limite à gauche"

$\Rightarrow$ "adapted" means "that cannot see into the future"

$\Rightarrow$ "bounded variation" means "a function whose total variation is finite"

- M_t often describes the continuous part of the asset price dynamics
- A_t often describes a drift term or predictable path

σ-martingale

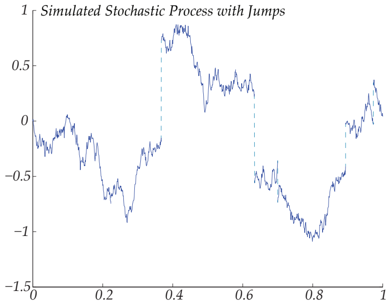
- Many asset prices are represented by semimartingales

$\Rightarrow$ e.g. Brownian martingales are semimartingales

$\Rightarrow$ e.g. stochastic volatility models can be semimartingales

$\Rightarrow$ e.g. stochastic processes with jumps can be semimartingales

$\Rightarrow$ if S_t has an integral representation it is a σ-martingale (more general)



Fundamental Theorem of Asset Pricing (FTAP)

- Asset pricing requires market equilibrium to exist
- Under the EMH the market exists but risk free profit is impossible

$\Rightarrow$ FTAP ensures that for an $\mathbb{R}^d$-valued semimartingale $S = (S_t)_{0 \leq t \leq T}$:

$\Rightarrow$ if there exists a risk-neutral probability measure $\mathbb{Q}$ equivalent to the

... original probability measure $\mathbb{P}$ under which S is a σ-martingale, then

... S does not permit free lunch with vanishing risk

$\Rightarrow$ any arbitrage is mathematically prohibited

A trivial model

- Let's define $\mathbb{P}$, $\mathbb{Q}$ and explain that theorem with a trivial model
  a. $(B_t)_{t=0,1} = 1 \; \forall \, t$, the price of a risk free bond with null interest rate
  b. $(S_t)_{t=0,1}$, the price of a risky asset with $S_0 = 1$
- S_1 being uncertain, it is a random variable defined on a probability space

$$(\Omega, (\mathcal{F}_t)_{t=0,1}, \mathbb{P})$$

  $\Rightarrow$ Ω is the set of all possible outcomes

  $\Rightarrow$ $\mathcal{F}_t$ is the set of events

  $\Rightarrow$ $\mathbb{P}$ is a probability function assigning a probability to each event
- To simplify, let Ω consist of only two elements g and b

  $\Rightarrow$ any random element $\omega \in \Omega$ has outcomes g or b

  $\Rightarrow$ g and b stand for "good" and "bad"' and occur with probability

$$\mathbb{P}(g) = \mathbb{P}(b) = 1/2$$

  $\Rightarrow$ At time $t = 1$, we finally define that

$$S_1(\omega) = \begin{cases} 2 & \text{for} \quad \omega = g \\ 1/2 & \text{for} \quad \omega = b \end{cases}$$

Option pricing

- An interesting strategy to prevent b or g is to buy an Option

$\Rightarrow$ the investor buys the right but not the obligation to buy S_t

... at time $t = 1$ at the pre-defined strike price $K = 1$

- At Option expiration (i.e. at time $t = 1$), the payoff is simply

$$C_1(\omega) = (S_1(\omega) - K)_+ = \begin{cases} 1 & \text{for} \quad \omega = g \\ 0 & \text{for} \quad \omega = b \end{cases}$$

- Knowing $C_1(\omega)$, the Option pricing puzzle is to determine C_0

$\Rightarrow$ What is the price at which the investor will buy the Option today ?

— In M1 we have studied: for risky assets, “expectation” is a bad criterion

$$C_0 := \mathbb{E}_{\mathbb{P}}(C_1) = 1/2 < \mathbb{E}_{\mathbb{P}}(B_1) = 1 < \mathbb{E}_{\mathbb{P}}(S_1) = 1.25$$

Reminder : expectation operator is valid only for risk neutral investors

$\Rightarrow$ The binomial model (Cox, Ross et Rubinstein) provides a solution

Reminder : binomial model

- The investor going long in the Option can try to cover its position

$\Rightarrow$ invests in a portfolio that replicates the Option profit $(S_1(\omega) - K)_+$

$$P_1 = \alpha S_1 + \beta B_1 = C_1$$

where α and β are the quantities invested in the stock and the bond

- Solving that system (composed by each state of the nature) gives

$$\alpha^* = 2/3 \text{ and } \beta^* = -1/3$$

- Indeed, the system is

$$C_1 = 1 = 2\alpha + \beta \text{ for } \omega = g$$
$$C_1 = 0 = \alpha/2 + \beta \text{ for } \omega = b$$

Reminder : binomial model

- P_0 equals the Option price C_0 by “no arbitrage argument”

$$P_0 = 2/3S_0 - 1/3B_0 = 1/3 = C_0$$

- Proof by contradiction : suppose that $C_0 = \mathbb{E}_{\mathbb{P}}(C_1) = 1/2 \neq P_0$

$\Rightarrow$ As the portfolio replicates the Option

... free lunch would be possible by going long in P_0 and going short in C_0

... and getting back arbitrage profit $C_0 - P_0 = 1/6$

$\Rightarrow$ Arbitrage opportunities vanish only for $C_0 \to P_0$

Risk-neutral measure (or equivalent martingale measure)

- Suppose that the world is governed by a new probability measure $\mathbb{Q}$

$\Rightarrow$ $\mathbb{Q}$ assigns new weights to g and b such that $\mathbb{E}_{\mathbb{Q}}(S_1) = \mathbb{E}_{\mathbb{P}}(B_1)$

- As the bond is free of risk, $\mathbb{Q}$ is called risk-neutral probability measure

$\Rightarrow$ recall that $\mathbb{E}_{\mathbb{P}}(B_1) = 1$ and the unique solution ensuring $\mathbb{E}_{\mathbb{Q}}(S_1) = 1$ is

$$\mathbb{Q}(g) = 1/3 \text{ and } \mathbb{Q}(b) = 2/3$$

as $\mathbb{E}_{\mathbb{Q}}(S_1) = 2 \times 1/3 + 1/2 \times 2/3 = 1$

- Strictly speaking, $\mathbb{Q}$ is a martingale measure for S

... or equivalently, S is a martingale under $\mathbb{Q}$

- The now valid expectation criterion can be applied to determine C_0

$$C_0 = \mathbb{E}_{\mathbb{Q}}(C_1) = 1 \times 1/3 + 0 \times 2/3 = 1/3$$

$\Rightarrow$ C_0 is compatible with the “arbitrage-free” value obtained under $\mathbb{P}$

Risk-neutral measure in continuous time models

- This trivial example can be extended to more general processes

$\Rightarrow$ let $(S_t)_{0 \leq t \leq T}$ be a continuous-time Brownian martingale

$$S_t = \sigma W_t$$

modeled on a filtered probability space $(\Omega, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$

- For C_T any contingent claim, i.e. a $\mathcal{F}_T$-measurable random variable

$\Rightarrow$ $C_0 := \mathbb{E}_{\mathbb{Q}}(C_T)$ yields the arbitrage-free prices for C_T when $\mathbb{Q}$ runs

... through the probability measures on $\mathcal{F}_T$ which are equivalent to $\mathbb{P}$

... under which the stochastic process S is a martingale

No free lunch with vanishing risk

- When the risk-neutral measure is unique, one may replicate C_T as

$$C_T = C_0 + \int_0^T P_t dS_t$$

where P_t is a predictable trading strategy (e.g. replication portfolio)

$\Rightarrow$ P_t models the holding in S during the infinitesimal interval $[t, t + dt]$

- Many models like the one of Black & Scholes use that result to find C_0

- We now get more intuition on the importance of the FTAP

$\Rightarrow$ roughly speaking, the absence of arbitrage possibilities for a stochastic process S is equivalent to the existence of an equivalent martingale measure for S

Stochastic differential equation (SDE)

- Financial observations are often sampled at high frequency

⇒ to the limit, infinitesimal differences can be considered :

$$dt = (t + h) - t, \quad h \to 0$$

- Applied to S_t, infinitesimal calculus comes down to the differential dS_t
- Assume the following Itô drift-diffusion for the returns of S_t

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t \Rightarrow dS_t = S_t(\mu dt + \sigma dW_t)$$

- As S_t is stochastic, solving the SDE dS_t is not immediate

⇒ Itô's lemma is a powerful alternative to chain rule derivatives and states

$$d\big(f(S_t, t)\big) = \frac{\partial f}{\partial t}(S_t, t)dt + \frac{\partial f}{\partial S_t}(S_t, t)dS_t + \frac{1}{2}\frac{\partial^2 f}{\partial S_t^2}(S_t, t)\sigma^2 dt$$

Solution of Itô drift-diffusion

- Define $f(S_t, t) = \log(S_t)$ and apply Itô's lemma

$$
\begin{aligned}
d\big(\log(S_t)\big) &= 0dt + \frac{1}{S_t}dS_t + \frac{1}{2}\left(-\frac{1}{S_t^2}\right)\sigma^2 S_t^2 dt \\
&= \frac{1}{S_t}(\mu S_t dt + \sigma S_t dW_t) - \frac{1}{2}\sigma^2 dt \\
&= \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t
\end{aligned}
$$

- Now, we can integrate and obtain the solution of the SDE

$$
\begin{aligned}
\log(S_t) &= \log(S_0) + \int_0^t \mu ds - \frac{1}{2}\int_0^t \sigma^2 ds + \int_0^t \sigma dW_s \\
&= \log(S_0) + \mu t - \frac{\sigma^2}{2}t + \sigma W_t \\
S_t &= S_0 \exp\left(\mu t - \frac{\sigma^2}{2}t + \sigma W_t\right)
\end{aligned}
$$

$\Rightarrow$ Modeling log-price ensures the stochastic path to be positive

Martingale property of the Wiener integral under $\mathbb{P}$

- Any integrable process X_t
  - whose increments are independent
  - and centered under probability measure $\mathbb{P}$: $\mathbb{E}(X_t - X_s) = 0$

… is a martingale with respect to the filtration $\mathcal{F}_t$ as for $0 \leq s \leq t$ we have

$$
\begin{aligned}
\mathbb{E}(X_t|\mathcal{F}_s) &= \mathbb{E}(X_t - X_s + X_s|\mathcal{F}_s) \\
&= \mathbb{E}(X_t - X_s|\mathcal{F}_s) + \mathbb{E}(X_s|\mathcal{F}_s) \\
&= \mathbb{E}(X_t - X_s) + X_s \\
&= X_s
\end{aligned}
$$

- The Wiener process W_t, whose stochastic integral representation is

$$
W_t = \int_0^t dW_s,
$$

has centered and independent increments and hence $W_t \in \mathcal{M}$

Martingale property of the Itô integral under $\mathbb{P}$

- The Wiener integral is a particular type of Itô integral

… which is a stochastic generalization of the Riemann–Stieltjes integral (limit of Riemann sums) where the integrands and the integrators are stochastic

$$\int_0^t u_\tau dX_\tau = \lim_{M\to\infty} \sum_{i=1}^{M} u_{t_i}(X_{t_i} - X_{t_{i-1}})$$

with $t-1 = t_0 < t_1 < \ldots < t_M = t$ and X_τ a semi-martingale

- This more general representation, where X_τ is not necessarily a Wiener process, can help to generate more realistic price dynamics

- If $X_t = W_t$, the Itô integral of any square integrable adapted process u_t is a martingale as

$$\mathbb{E}\left(\int_0^t u_\tau dW_\tau | \mathcal{F}_s\right) = \int_0^s u_\tau dW_\tau$$

No arbitrage in continuous time

- Under a risk-neutral probability measure $\mathbb{Q}$, the return of the risky asset

... equals the return of the risk-less asset if it is discounted by r

$$\tilde{S}_t = e^{-rt} S_t = \frac{S_t}{B_t / B_0}$$

with r the risk-free rate of a zero-coupon bond B_t

$\Rightarrow$ The discounted process $\tilde{S}_t$ satisfies

$$\begin{aligned} d\tilde{S}_t &= d(e^{-rt} S_t) \\ &= S_t d(e^{-rt}) + e^{-rt} dS_t + (de^{-rt}) dS_t \\ &= -re^{-rt} S_t dt + e^{-rt} dS_t + (-re^{-rt} dt) \bullet dS_t \\ \text{Itô rules} : &\; dt \bullet dt = 0 \mid dt \bullet dW_t = 0 \mid dW_t \bullet dW_t = dt \\ &= -re^{-rt} S_t dt + \mu e^{-rt} S_t dt + \sigma e^{-rt} S_t dW_t \\ &= (\mu - r) \tilde{S}_t dt + \sigma \tilde{S}_t dW_t \end{aligned}$$

and $\tilde{S}_t \in \mathcal{M}$ under $\mathbb{Q}$ when $\mu = r \Rightarrow$ no arbitrage by FTAP theorem

- When $\mu \neq r$, one can define a martingale under $\mathbb{Q}$ but Girsanov Theorem is necessary

Time-varying instantaneous variance

- In previous model, S_t as well as $\tilde{S}_t$ have constant volatility σ but

... Itô's lemma remains valid if volatility changes in time

- At this stage assume σ_t is càdlàg, a.s. positive and consider $p_t = \log(S_t)$

$$dp_t = m_t dt + \sigma_t dW_t, \quad m_t = \mu_t - \sigma_t^2/2, \quad t \geq 0$$

- It turns out that log returns are

$$r_t = p_t - p_{t-1} = \int_{t-1}^{t} m_s ds + \int_{t-1}^{t} \sigma_s dW_s$$

... and distributed as follows

$$r_t \sim \mathcal{N}\left(\int_{t-1}^{t} m_s ds, \int_{t-1}^{t} \sigma_s^2 dW_s \right)$$

$\Rightarrow$ σ_t^2 (resp. σ_t) is called instantaneous or spot variance (volatility)

Integrated Variance (IV)

- Over an interval of time $[t-1, t]$ the log-return variance is hence

$$IV_t = \int_{t-1}^{t} \sigma_s^2 ds$$

also called Integrated Variance

- Assuming that the time length of one day is $t-(t-1)=1$

... $\sqrt{IV}$ represents the daily log-return volatility

- For any sequence of partitions $t-1=t_0<t_1<\ldots<t_M=t$ with

... $\sup_j\{t_{j+1}-t_j\} \to 0$ for $M \to \infty$, the Quadratic Variation (QV) can be defined as

$$[p,p]_t = \operatorname*{plim}_{M\to\infty} \sum_{j=0}^{M-1} (p_{t_{j+1}} - p_{t_j})(p_{t_{j+1}} - p_{t_j}) = IV_t$$

Note In the particular constant volatility case, $IV_t = \sigma^2 t$

Integrated Variance and semi-martingale

- Now, consider a more general representation with $p_t \in \mathcal{SM}$

$$p_t = A_t + M_t$$

  where A_t has finite variation and M_t is a local martingale

- If A_t is continous, it can be show that the QV

$$[p,p]_t = M_t$$

  even if M_t is contaminated by discontinuities (jumps)

- However, one cannot ensure that QV converges to IV

Simple continuous-time models

- Under the assumption that $S_t \in \mathcal{SM}$, many models can be considered
- The simplest one is

$$S_t = S_0 + \int_0^t \sigma_s dW_s$$

  for which we have $IV_t = \int_0^t \sigma_s ds$ over interval $[0, t]$

- It appears as the solution of the SDE

$$dS_t = \sigma_t dW_t$$

- However,

$\Rightarrow$ more general integrators ($\mathcal{LI}$) can be considered in place of W_t

$\Rightarrow$ many models exist for the stochastic dynamics of price ($\mathcal{SP}$)

$\Rightarrow$ many models exist for the stochastic dynamics of volatility ($\mathcal{SV}$)

$\mathcal{LI}$: Lévy processes

Definition 15

A càdlàg process $(X_t)_{t\geq 0}$ *defined on* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a Lévy process if* $X_0 = 0$ *and*

- *For any increasing sequence* $t_0, \cdots, t_n$*, the random variables* $X_{t_0}, X_{t_1} - X_{t_0}, \cdots, X_{t_n} - X_{t_{n-1}}$ *are independent*
- *The law of* $X_{t+h} - X_t$ *does not depend on t such that increments are stationary*
- X_t *is stochastically continuous, i.e.* $\lim_{h\to 0} \mathbb{P}(|X_{t+h} - X_t| \geq \varepsilon) = 0$

Note The last point just rules out nonrandom occurrence of discontinuities

$\Rightarrow$ $X_t \sim \mathcal{L}(.)$ can be discontinuous as we will see in a few slides

- Lévy processes are more general than Brownian motions and possibly have non-Gaussian increments

$\mathcal{LI}$: Infinite divisibility of Lévy processes

- Let $X_t \sim \mathcal{L}(.)$ be sampled at regular time intervals $0, \Delta, 2\Delta, \cdots, n\Delta$

$\Rightarrow$ the resulting process is still a random walk as

$$X_{n\Delta} = \sum_{k=0}^{n-1} (X_{(k+1)\Delta} - X_{k\Delta}) = \sum_{k=0}^{n-1} \Delta_n X_k$$

has i.i.d. increments $\Delta_n X_k$

- Moreover, their distribution is the same as the one of X_Δ whatever the sampling frequency
- For $n\Delta = t > 0$ and $n \geq 1$, X_t and $X_{\Delta=t/n}$ have the same distribution

$\Rightarrow$ X_t can be divided into n i.i.d. parts : it is infinitely divisible

- A random walk can have arbitrary distribution whereas the distribution of increments of $X_t \sim \mathcal{L}(.)$ has to be infinitely divisible

$\Rightarrow$ the most common eligible distributions are the Gaussian, the α-stable and the Poisson distributions

$\mathcal{SP}$: Ornstein-Uhlenbeck (OU) model

- The OU model appears as the counterpart of an AR(1) process

$$dX_t = \alpha(\theta - X_t)dt + \sigma dW_t$$

  with $\theta > 0$ and $\sigma > 0$

- The solution of the Ornstein-Uhlenbeck SDE is

$$X_t = X_0 e^{-\alpha t} + \theta \left(1 - e^{-\alpha t}\right) + \sigma \int_0^t e^{-\alpha(t-s)}\, dW_s \tag{6}$$

- In the financial literature, the OU coefficients can be interpreted as
  — θ represents the asset equilibrium value
  — σ the volatility of shocks coming from the diffusion
  — α the rate at which these shocks vanish & X_t reverts towards the mean
  — when (6) is used to model interest rates, it is also known as Vasicek model

$\mathcal{SP}$: Hull-White (HW) model

- The HW model extends the Vasicek model

$$dX_t = \alpha(\theta(t) - X_t)dt + \sigma(t)dW_t$$

with $\theta(t)$ a time-dependent coefficient

- The solution of the SDE is

$$X_t = X_0 e^{-\alpha t} + \alpha \int_0^t e^{\alpha(s-t)} \theta_s ds + \sigma e^{-\theta t} \int_0^t e^{\theta s}\, dW_s \tag{7}$$

- As θ is time-dependent, one can use the Yield Curve for calibration
- Unfortunately, the HW model cannot ensure the positivity of X_t

$\Rightarrow$ not suitable for interest rate (in normal periods) nor for volatility

$\mathcal{SP}$: Cox-Ingersoll-Ross (CIR) model

- The CIR model is also designed to mimic the evolution of interest rates

$$dX_t = \theta(\mu - X_t)dt + \sigma\sqrt{X_t}\, dW_t$$

  with
  - $\theta(\mu - X_t)$ the short term dynamics of X_t
  - σ the volatility of X_t
  - Interestingly, as long as $2\theta\mu \geq \sigma^2$, X_t will never reach 0 (Feller condition)

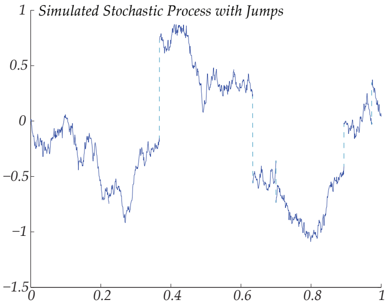
- This type of square-root SDE is hence useful to model volatility

$\Rightarrow$ However, this SDE has no closed-form solution

$\mathcal{SP}$: Jump augmented models

- Abrupt changes can occur in price dynamics

$\Rightarrow$ such jumps in price cannot be modeled by Gaussian increments



- Counting processes $\mathcal{CM}$ are good candidates
- A counting process is a stochastic non-decreasing process $N_t \in \mathbb{N}$, $t \geq 0$

$\Rightarrow$ As $N_t \in \mathbb{N}$, for $s \leq t$, $N_t - N_s$ is the number of events occurred in $(s, t]$

$\mathcal{SP}$: Poisson model

- A Poisson point process $N_t \in \mathcal{CM}$ is defined as follows
  - $N_0 = 0$ and for $s \leq t$ where increments $N_t - N_s$ are independent
  - the probability mass function of N_t is

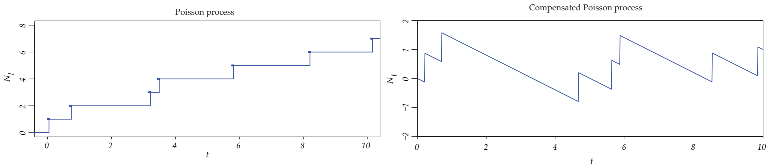
$$\mathbb{P}(N_t = c) = \frac{(\lambda t)^c}{c!} \exp(-\lambda t)$$

with λ the arrival intensity of events and $\mathbb{E}(N_t) = \lambda t$

$\Rightarrow$ N_t will represent the jumps occurrence irrespective of the jumps size

- Consider $(N_t - \lambda t)_t$ a compensated process satisfying $\mathbb{E}(N_t - \lambda t) = 0$

$\Rightarrow$ Since it has centered and independent increments, $\tilde{N}_t = (N_t - \lambda t)_t \in \mathcal{M}$



$\mathcal{SP}$: Compound Poisson model

- Let's define J_t a compound Poisson process

$$J_t = \sum_{i=1}^{N_t} Y_i,$$

where Y_i is i.i.d, square-integrable and generally Gaussian

- Defining $Y_{t-} := \lim_{s \to t} Y_s$, the jump size is hence

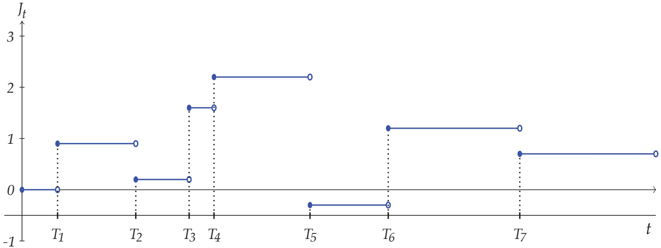
$$\Delta J_t = Y_t - Y_{t-} = dJ_t$$

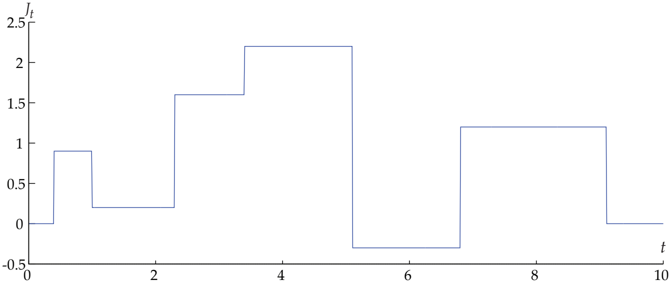
- ΔJ_t will be used latter to obtain an integral representation of J_t
- Notice that $\mathbb{E}(J_t) = \lambda t \, \mathbb{E}(Y)$ and hence the compensated process

$$\tilde{J}_t = \big(J_t - \lambda t \, \mathbb{E}(Y)\big)_t$$

has centered and independent increments $\Rightarrow \tilde{J}_t \in \mathcal{M}$

$\mathcal{SP}$: Illustration of compound Poisson paths





Stochastic integral representation of J_t

- Let $f_Y(.) \in \mathbb{R}$ be the jump size distribution function and observe that

$$\nu(dy) = \lambda f_Y(dy)$$

  is the density of J_t in the jump size interval $[y, y + dy]$

- $\nu(dy)$ is called Lévy or intensity measure but is not a probability measure since

$$\int_{\mathbb{R}} \nu(dy) = \lambda \neq 1$$

- Now we call Poisson random measure, for $B = [t_1, t_2]$,

$$\mu(B, A) = \#\{(J_{t_2} - J_{t_1}) \in A\}$$

  a measure that counts the jumps in B such that their sizes are in A

- Then, J_t has a stochastic integral representation given by

$$J_t = \sum_{s \in [0,t]} \Delta J_s = \int_{[0,t] \times \mathbb{R}} y \times \mu(ds \times dy)$$

Lévy-Itô decomposition

Model 7

Let $(X_t)_{t\geq 0}$ be a Lévy process and $\nu(.)$ its Lévy measure verifying

$$\int_{|x|\leq\kappa} |x|^2 \nu(dx) < \infty \text{ and } \int_{|x|\geq\kappa} \nu(dx) < \infty$$

such that the intensity of jumps larger than $\kappa > \varepsilon > 0$ is finite and denote $\mu(.)$ its Poisson random measure on $[0,\infty) \times \mathbb{R}$. Then, there exists a drift α and a Brownian motion $(W_t) \geq 0$ with variance β such that

$$X_t = \alpha t + W_t + J_t^l + \lim_{\varepsilon\to 0} \tilde{J}_t^{\varepsilon}$$

where large jumps with finite activity are modeled by

$$J_t^l = \int_0^t \int_{|x|\geq\kappa} x\mu(ds \times dx)$$

and small jumps with possibly infinite activity by

$$\tilde{J}_t^{\varepsilon} = \int_0^t \int_{\varepsilon\leq|x|<\kappa} x(\mu(ds \times dx) - \nu(dx)ds)$$

Lévy characteristic triplet

- The Lévy-Itô decomposition entails that every Lévy process

1. is a sum of (Brownian) continuous and discontinuous paths
2. is characterized by the triplet $X_t = \mathcal{L}(\alpha, \beta, \nu)$
3. has large jumps J_t^l that follow a finite activity compound Poisson
4. has small jumps J_t^ε that follow an infinite activity compound Poisson

- As $\nu < \infty$ is imposed only for $|x| \geq \kappa$, the Lévy process can diverge

$\Rightarrow$ J_t^ε needs to be compensated $\Rightarrow \tilde{J}_t^\varepsilon$

- One can define a "pure-jump" Lévy process by setting $X_t = \mathcal{L}(0, 0, \nu)$

$\Rightarrow$ generally unrealistic in finance as it rules out the diffusion

$\mathcal{SP}$: Jump diffusion model and Exponential Lévy

- A jump-diffusion process combines a Brownian diffusion with jumps

$$dS_t = \mu S_{t-}\,dt + \sigma S_{t-}\,dW_t + S_{t-}\,dJ_t$$

$\Rightarrow$ it is an Exponential Lévy if $S_t = \exp(X_t)$ with $X_t \sim \mathcal{L}(\mu, \sigma^2, \nu)$

- Applying an Itô formula for jump, the SDE solution is $S_t = A_t + M_t$
- The martingale part is

$$M_t = 1 + \int_0^t S_{t-}\sigma dW_s + \int_{[0,t]\times\mathbb{R}} S_{t-}(e^y - 1) \times \tilde{\mu}(ds \times dy)$$

and the continuous finite variation drift part is given by

$$A_t = \int_0^t S_{t-}\left(\mu - \sigma^2/2 + \int_{\mathbb{R}} (e^y - 1 - y\mathbb{1}_{|y|\leq\kappa}) \times \nu(dy)\right)ds$$

- S_t will be a martingale if $\mathbb{E}(S_t|S_0) = S_0$ and hence if

$$\mu - \sigma^2/2 + \int_{\mathbb{R}} (e^y - 1 - y\mathbb{1}_{|y|\leq\kappa}) \times \nu(dy) = 0$$

Note S_t could be defined directly as a Lévy but without positivity constraint

$\mathcal{SP} + \mathcal{SV}$: Heston model

- Previous models are unrealistic because σ is constant over time
- Heston (1993) suggests the following stochastic volatility model

$$dS_t = \mu S_t dt + S_t \sigma_t dW_{1,t}$$
$$d\sigma_t^2 = -\theta(\sigma_t^2 - \eta)dt + \gamma \sigma_t dW_{2,t}$$

with $W_{1,t}$ and $W_{2,t}$ two Brownian motions and $\theta, \eta, \gamma > 0$

$\Rightarrow$ The stochastic volatility is modeled as a CIR process

- The Heston model is a particular case of two-factor stochastic volatility model

$$dS_t = m_t S_t dt + \sigma_t S_t dW_{1,t}$$
$$d\sigma_t^2 = \alpha(t, \sigma_t^2)dt + \beta(t, \sigma_t^2)dW_{2,t}$$

where $Cov(W_{1,t}, W_{2,t}) = \rho t$ is possibly non-null

- In practice $d\sigma_t$ tends to be negatively correlated with $dS_t \Rightarrow \rho < 0$

$\mathcal{SP} + \mathcal{SV}$: Infinite activity jump diffusion stochastic volatility model

- Let $p_t \in \mathcal{SM}$ be stated by the following general $\mathcal{SV}$ model

$$p_t = p_0 + \underbrace{\int_0^t m_s ds + \int_0^t \sigma_s dW_s}_{\text{Continuous (1)}} + \underbrace{\int_0^t dJ_s}_{\text{Discontinuous}}$$

$$\int_0^t dJ_s = \underbrace{\int_0^t \int_{|x|\geq\kappa} x\mu(ds \times dx)}_{\text{Big jumps (2)}} + \underbrace{\int_0^t \int_{\varepsilon\leq|x|<\kappa} x(\mu(ds \times dx) - \nu(dx)ds)}_{\text{Small jumps (3)}}$$

- As seen before, p_t will always generate a finite number of big jumps

... but it may give rise to either a finite or infinite number of small jumps

- Each component of p_t can be mapped into an economic source of risk

Part 1 capture the normal risk of the asset, which is hedgeable

Part 2 capture default risk, or more generally big news-related events

Part 3 represent price moves which are large on a time scale of a few seconds, but generally not significant on a daily frequency

$\mathcal{SP} + \mathcal{SV}$: Finite activity jump diffusion stochastic volatility model

- Let $p_t \in \mathcal{SM}$ be stated as previously but impose $\kappa = 0$

$\Rightarrow$ all jumps are qualified as being big jumps

$\Rightarrow$ hence p_t has finite activity and jumps are summable

$$\sum_{s \leq t} |\Delta p_s| < \infty, \text{ where } \Delta p_s = p_s - p_{s-}$$

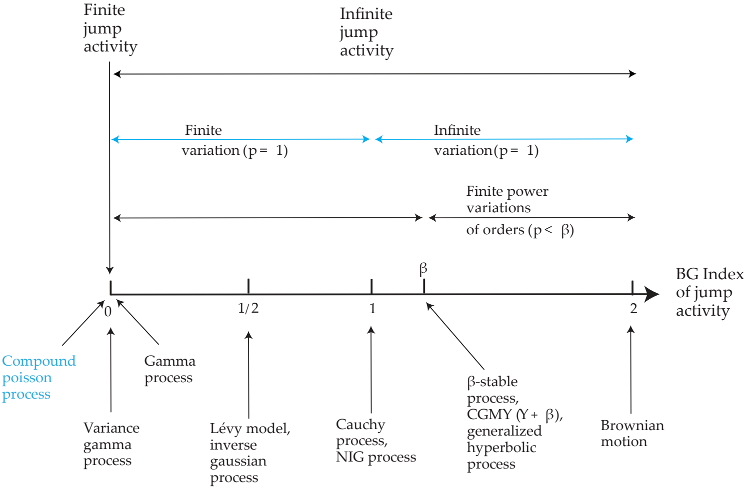
- In such a case, on can rewrite p_t as

$$p_t = p_0 + \underbrace{\int_0^t \tilde{m}_s ds + \int_0^t \sigma_s dW_s}_{\text{Continuous}} + \underbrace{\sum_{s \leq t} \Delta p_s}_{\text{Discontinuous}}$$

where $\tilde{m}_s = m_s - \int_{\mathbb{R}} y f(dy)$

- In practice, the jump activity is difficult to measure but index such as Blumenthal-Getoor (BG or β) index exist

Finite versus Infinite activity : the Blumenthal-Getoor index



- Finite varition of order p means that for $\Delta X_k = X_{k\Delta} - X_{(k-1)\Delta}$: $\sum_{k=1}^{[t/\Delta]} |\Delta X_k|^p < \infty$ as $\Delta \to 0$

$\mathcal{SP} + \mathcal{SV}$: Bates model

- Bates (1996) extends the Heston model to Poisson jump-diffusion

$$
\begin{aligned}
dS_t &= \mu S_{t-} dt + S_{t-} \sigma_t dW_{1,t} + S_{t-} dJ_t \\
d\sigma_t^2 &= -\theta(\sigma_t^2 - \eta)dt + \gamma \sigma_t dW_{2,t}
\end{aligned}
$$

- In this model, the QV of $p_t = \log(S_t)$ no-longer converges to IV

$\Rightarrow$ For any sequence of partitions $t - 1 = t_0 < t_1 < \ldots < t_M = t$ with

... $\sup_j \{t_{j+1} - t_j\} \to 0$ for $M \to \infty$, the Quadratic Variation (QV) is

$$[p, p]_t = \underset{M \to \infty}{\text{plim}} \sum_{j=0}^{M-1} (p_{t_{j+1}} - p_{t_j})(p_{t_{j+1}} - p_{t_j}) = IV_t + JV_t$$

where

$$IV_t = \int_{t-1}^{t} \sigma_s^2 ds \text{ and } JV_t = \sum_{i=N_{t-1}+1}^{N_t} \xi_i^2 = \sum_{t-1 \leq s \leq t} |\Delta p_s|^2$$

with $\xi_i = \log(1 + Y_i)$ and Y_i the jump magnitude of the count i

Quadratic variation of various processes

- Overall, depending on the stochastic process we have :

$$[p,p]_t = IV_t = \sigma^2 t$$

if p_t is a standard Brownian motion,

$$[p,p]_t = IV_t = \int_{t-1}^{t} \sigma_s^2 ds$$

if p_t is a Brownian motion with stochastic volatility

$$[p,p]_t = IV_t + JV_t = \int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1 \leq s \leq t} |\Delta p_s|^2$$

if p_t is a finite activity jump diffusion process

$$[p,p]_t = IV_t + JV_t = \int_{t-1}^{t} \sigma_s^2 ds + \int_0^t \int_{\mathbb{R}} y^2 \mu(ds \times dy)$$

if p_t is a general Lévy process

Why continuous time models?

- Observations are quite often irregularly spaced
- Observations quite often come in at a very high frequency
- Then a continuous time model may provide a better approximation to the discrete data than a discrete model

At the same time

- Continuous-time models are central to mathematical finance
- Most theoretical results on derivative pricing rely on continuous-time processes, obtained as solutions of diffusion equations

Aim: Construct continuous time models with features of GARCH (it is possible to establish connections between the two approaches: discrete-time and continuous-time models)

Note: In continuous time it is natural to model the logarithm of the asset price itself, that is p_t, rather than its increments r_t as in discrete time

Continuous time GARCH approximations

- A continuous time model may serve as an approximation to a GARCH process
- Main question: starting from the continuous model, how close would a process be to a GARCH process when sampled at discrete times?
- An optimal situation would be that the process itself is a GARCH process, whenever sampled at equidistant times $(kh)_{k\in\mathbb{N}_0}$, for each $h > 0$
- **Issue**: GARCH processes are not closed under temporal aggregation (Drost and Nijman, 1993)
- A continuous time process $(Y_t)_{t\geq 0}$ which happens to be a GARCH(1,1) process when sampled at $0, h, 2h, \ldots$ for some frequency h will not be GARCH when sampled at $0, 2h, 4h, \ldots$

The diffusion approximation of Nelson (1990)

Model 8

The GARCH(1,1) diffusion limit satisfies

$$dp_t = \sigma_t dW_t^{(1)},$$
$$d\sigma_t^2 = (\omega - \theta\sigma_t^2)dt + \lambda\sigma_t^2 dW_t^{(2)}, \ t \geq 0$$

- Although the GARCH process is driven by a single noise sequence, the diffusion limit is driven by two independent Brownian motions $(W_t^{(1)})_{t\geq 0}$ and $(W_t^{(2)})_{t\geq 0}$
- The behavior of this diffusion limit is therefore rather different from that of the GARCH process itself since the volatility process $(\sigma_t^2)_{t\geq 0}$ evolves independently of the driving process $(W_t^{(1)})_{t\geq 0}$ in the first of the equations above

The diffusion limit of Nelson

Remarks:

- The equation of $d\sigma_t^2$ in the Theorem has a **strictly stationary solution** $(\sigma_t^2)_{t\geq 0}$ if

  $$2\theta/\lambda^2 > -1 \text{ and } \omega > 0,$$

  in which case the marginal stationary distribution of σ_0^2 is inverse Gamma distributed with parameters $1 + 2\theta/\lambda^2$ and $2\omega/\lambda^2$
- The stationary limiting process $d\sigma_t^2$ has Pareto like tails
- The limit (p_t, σ_t^2) is driven by **two independent** Brownian motions
- The processes p_t and σ_t^2 are continuous. But empirical volatility can exhibit jumps

A continuous time GARCH model designed for option pricing (Kallsen and Taqqu, 1998)

- Option pricing for Nelson's model (designed as limit of discrete time GARCH processes) may be demanding since the model gives rise to incomplete markets
- Kallsen and Taqqu (1998) developed a continuous time process which is a GARCH process when sampled at integer times
- This process is driven by a single Brownian motion only

A continuous time GARCH model designed for option pricing (Kallsen and Taqqu, 1998)

Definition 16

Let $\omega, \lambda > 0$, $\delta \geq 0$ *and* $(B_t)_{t\geq 0}$ *be a standard Brownian motion. For some starting random variable* σ_0^2, *define the volatility process* $(\sigma_t)_{t\geq 0}$ *by* $\sigma_t^2 = \sigma_0^2$ *for* $t \in [0, 1)$ *and*

$$\sigma_t^2 = \omega + \lambda\Big(\int_{\lfloor t\rfloor -1}^{\lfloor t\rfloor} \sigma_{s-}\, dB_s\Big)^2 + \delta\sigma_{\lfloor t\rfloor -1}^2,\ t \geq 1$$

The continuous-time GARCH process $(p_t)_{t\geq 0}$ *then models the log-price process, and is given by*

$$p_t = p_0 + \int_0^t (\mu(\sigma_{s-}) - \sigma_{s-}^2/2)ds + \int_0^t \sigma_s dB_s$$

A continuous time GARCH model designed for option pricing (Kallsen and Taqqu, 1998)

- the drift function μ is assumed to have continuous derivatives
- the volatility process $(\sigma_t)_{t\geq 0}$ is constant on intervals $[n, n+1)$ for $n \in \mathbb{N}_0$
- the process $(p_t - p_{t-1}, \sigma_{t-1})_{t\geq 1}$, when sampled at integer times, gives rise to a discrete time GARCH(1,1)-M process

$$p_n - p_{n-1} = \mu(\sigma_{n-1}) - \sigma_{n-1}^2/2 + \sigma_{n-1}(B_n - B_{n-1}),\ n \in \mathbb{N},$$
$$\sigma_n^2 = \omega + \lambda\sigma_{n-1}^2(B_n - B_{n-1})^2 + \delta\sigma_{n-1}^2,\ n \in \mathbb{N}$$

- This differs from a usual GARCH(1,1) process only by the term $\mu(\sigma_{n-1}) - \sigma_{n-1}^2/2$, which vanishes if the function μ is chosen as $\mu(x) = x^2/2$
- If we are not in the classical GARCH situation but rather have

$$\limsup_{x\to\infty} \mu(x)/x < \infty,$$

then Kallsen and Taqqu (1998) show that the continuous time model is arbitrage free and complete. This is then used to derive pricing formulas for contingent claims such as European options.

GARCH is Slow: A Thought Experiment

- Suppose a sudden change in the true latent volatility

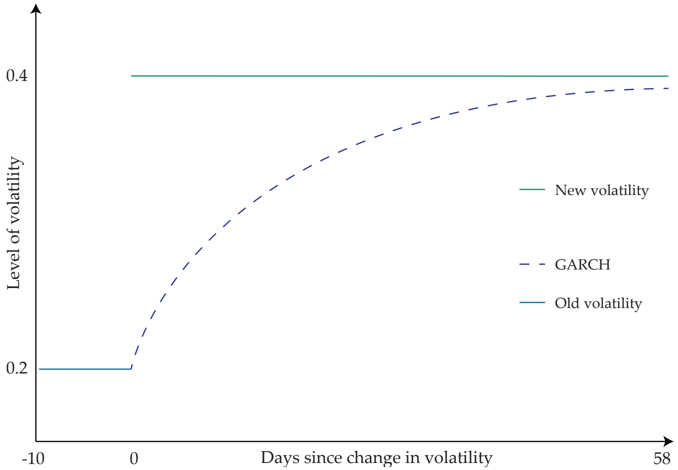
$$\sqrt{\text{var}(r_t|\mathcal{F}_{t-1})} = \begin{cases} 0.2 & t \leq T \\ 0.4 & t > T \end{cases}$$

- With $\alpha + \beta = 1$ (e.g. $\alpha = 0.05$ and $\beta = 0.95$), GARCH(1,1) implies

$$\mathbb{E}(\sigma^2_{T+k}) = \alpha \sum_{j=0}^{\infty} \beta^j \mathbb{E}(r^2_{T+k-j}) = \alpha \sum_{j=0}^{k-1} \beta^j (0.4)^2 + \alpha \sum_{j=k}^{\infty} \beta^j (0.2)^2$$

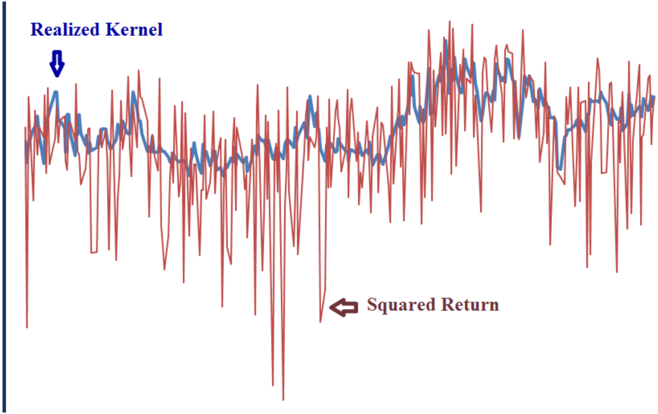
- How quickly does σ_t^2 converge to the new level of $\text{var}(r_t|\mathcal{F}_{t-1})$?

GARCH is Slow



Squared Return is a Noisy Signal of Volatility

- r_t^2 is a very noisy measure of variance
- High-frequency data $\longrightarrow$ Realized Measures $\longrightarrow$ far more accurate signals



Realized Variance

- In practice IV is unobservable and one only observe asset prices
- Moreover, log price data are only available in discrete time at frequency

$$\Delta\tau = t_i - t_{i-1}$$

where sampling rates $\Delta\tau$ can be seconds or minutes

$\Rightarrow$ For a given sampling rate, $r_{t_i} = p_{t_i} - p_{t_{i-1}}$ are intradaily log-returns

- We assume M intradaily observations $t - 1 = t_0 < t_1 < \ldots < t_M = t$

$\Rightarrow$ the daily Realized Variance (RV) is given by

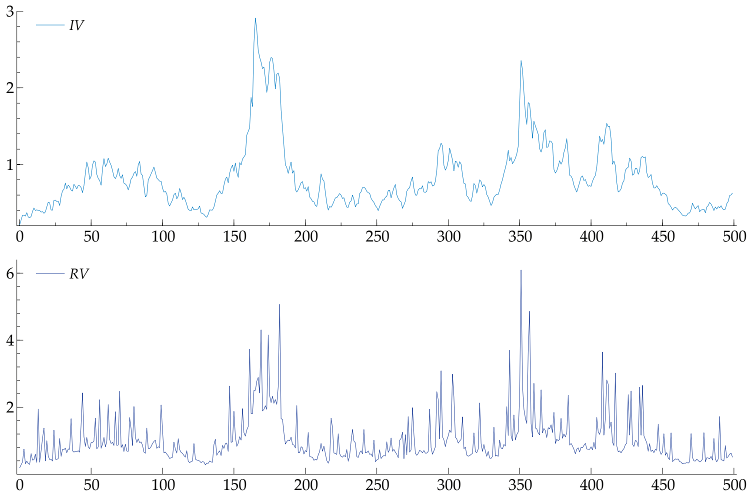
$$RV_t = \sum_{i=1}^{M} r_{t_i}^2$$

and one can show that in absence of jumps

$$\sqrt{n}\left(\sum_{i=1}^{M} r_{t_i}^2 - \int_{t-1}^{t} \sigma_s^2 ds\right)\left(2\underbrace{\int_{t-1}^{t} \sigma_s^4 ds}_{IQ}\right)^{-1/2} \xrightarrow{d} \mathcal{N}(0,1)$$

with IQ the Integrated Quarticity (the limit behavior of the vol-of-vol)

Realized Variance



Realized Semi-Variance

- The behavior of r_t in the lower tail of their distribution is of interest

... as it often differs from the behavior in the upper tail

$\Rightarrow$ requires an asymmetric treatment of the downside and upside risks

- Barndorff-Nielsen et al. (2010) suggest the Realized Semi-variance

$$RS_t^+ = \sum_{i=1}^{M} r_{t_i}^2 \mathbb{1}_{r_{t_i} \geq 0} \text{ and } RS_t^- = \sum_{i=1}^{M} r_{t_i}^2 \mathbb{1}_{r_{t_i} \leq 0}$$

- As for RV_t in presence of finite activity jumps, they no-longer converge to IV_t as

$$RS_t^+ \xrightarrow{p} \frac{1}{2}\int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1<s\leq t} |\Delta p_s|^2 \mathbb{1}_{r_{t_i}} \geq 0$$

and

$$RS_t^- \xrightarrow{p} \frac{1}{2}\int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1<s\leq t} |\Delta p_s|^2 \mathbb{1}_{r_{t_i}} \leq 0$$

Realized Multipower Variation

- In presence of jumps, $RV_t \xrightarrow{p} QV_t$ instead of the IV_t as $M \to \infty$
- As jumps cannot be easily distinguished from the continuous part,

... if IV_t is of interest, robust realized measures are needed

$\Rightarrow$ Barndorff-Nielsen et al. (2006) introduce the Multipower Variation

$$MV_t^{\{\gamma_1,\dots,\gamma_m\}} = \left(\frac{1}{M}\right)^{1-\sum \gamma_k/2} \sum_{j=m}^{M} \prod_{i=1}^{m} |r_{t_{j-i+1}}|^{\gamma_i}$$

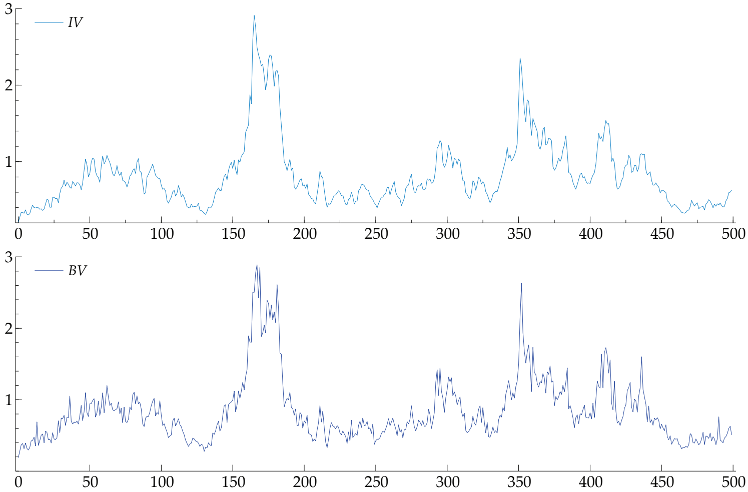
- The simplest case is the Bipower Variation (BV) with $\gamma_1 = \gamma_2 = 1$

$$BV_t^{\{1,1\}} = BV_t = \frac{\pi}{2} \sum_{j=2}^{M} |r_{t_j}||r_{t_{j-1}}|$$

- Conversely to RV_t, this measure is jump robust as $M \to \infty$:

$$BV_t \xrightarrow{p} \int_{t-1}^{t} \sigma_s^2 ds$$

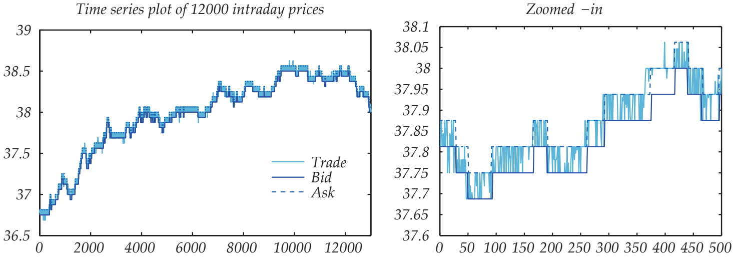
Bipower Variation



Microstructure noise

- The most striking feature of the empirical prices is their discreteness

$\Rightarrow$ price movement cannot be smaller than one tick



- At the opposite, theoretical $\mathcal{M}$-based models are continuous in time

$\Rightarrow$ efficient log-price are unobservable due to microstructure noise, i.e.

  — frictions in the trading process : tick limit, rounding error

  — informational effects : gradual response of prices to a block trade

  — recording errors : prices entered as zero or misplaced decimal points

Realized Kernel

- Let $p_t^* = p_t + \varepsilon_t$ be the observed price

$\Rightarrow$ $\varepsilon_t \perp p_t$ acts as microstructure noise and is defined as a white noise

- Barndorff-Nielsen et al. (2008) suggest the Realized Kernel (RK)

$$RK_t = \underbrace{\gamma_0(p_t^*)}_{RV} + \underbrace{\sum_{h=1}^{H} \mathcal{K}\left(\frac{h}{H}\right)(\gamma_h(p_t^*) + \gamma_{-h}(p_t^*))}_{\text{noise correction}} \xrightarrow{p} \int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1<s\leq t} |\Delta p_s|^2$$

with $\gamma_h(p_t^*) = \sum_{j=1}^{M}(p_{t_j}^* - p_{t_{j-1}}^*)(p_{t_{j-h}}^* - p_{t_{j-1-h}}^*)$ and $\mathcal{K}(.)$ a kernel

- The recommended kernel $\mathcal{K}(.)$ is Parzen's kernel

$$\mathcal{K}(x) = \begin{cases} 1 - 6x^2 + 6x^3, & 0 > x \geq \frac{1}{2} \\ 2(1-x)^3, & \frac{1}{2} > x \geq 1 \\ 0, & x > 1 \end{cases}$$

- The preferred choice for the bandwidth is $H = 3.5134 \times \hat{\xi}^{4/5} M^{3/5}$ with

$$\hat{\xi} = q^{-1} \sum_{i=1}^{q} \hat{\omega}_{(i)}^2 / RV_{\text{sparse}} \text{ and } \hat{\omega}_{(i)}^2 = RV_{\text{dense}}^{(i)} / (2M_{(i)})$$

Jumps identification and Realized Jumps (RJ)

- It is hard to disentangle the continuous and jump parts of the QV
- Consider for instance $RK_t \xrightarrow{p} IV_t$ and $BV_t \xrightarrow{p} QV_t$ as $M \to \infty$

$\Rightarrow$ Hence, we have $RJ_t = RK_t - BV_t \xrightarrow{p} -J_t = -\sum_{t-1<s\leq t} |\Delta p_s|^2$

- Based on that result, several test statistics can be defined ...

... e.g. the one of Barndorff-Nielsen and Shephard (2006) :

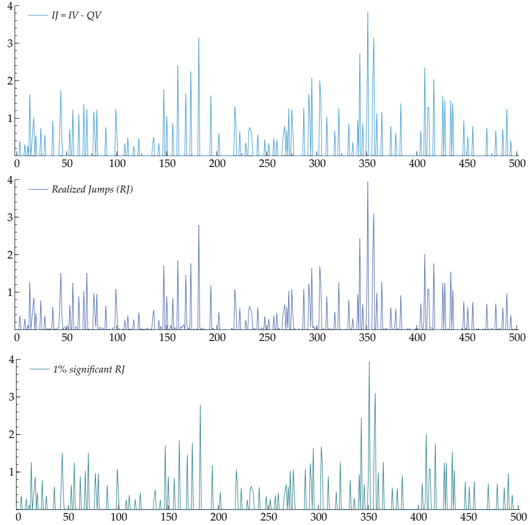
$$T_M = \left(1 - \frac{BV_{t,M}}{RV_{t,M}}\right)\left(\frac{1}{M}\frac{\vartheta_{BV} MV_{t,M}^{\{1,1,1,1\}}}{BV_{t,M}^2}\right)^{-1/2} \xrightarrow{d} \mathcal{N}(0,1)$$

under the null H_0 : no jump, with $\vartheta_{BV} \approx 2.609$ and as $M \to \infty$

$\Rightarrow$ The rejection of the null at 5% is thus given by $T_M > 1.64$

... and allows to identify insignificant Realized Jumps

Realized Jumps



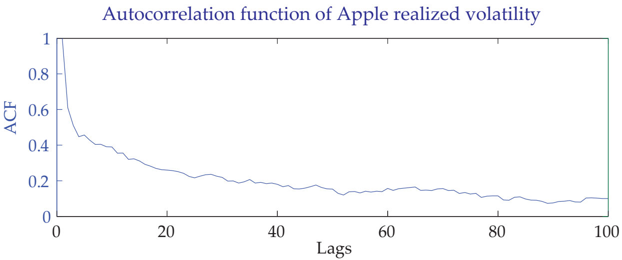
Nonparametric realized measure

- RV_t, RS_t, BV_t and RK_t are ex-post realized measures ($\mathcal{RM}$) of volatility

... at the opposite GARCH-type models estimate ex-ante volatility

- Realized measures are nonparametric estimators of IV_t or QV_t

$\Rightarrow$ without conditional models, parametric forecasts are infeasible



- $\mathcal{RM}$ are highly persistent with hyperbolic decay of autocorrelations

$\Rightarrow$ one could use this stylized fact to forecast volatility

Modelling realized measures with ARFIMA

- Autocorrelations of ARMA models decay at exponential rate

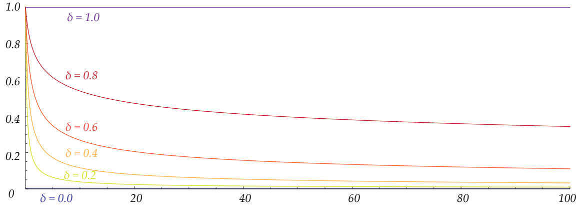
$\Rightarrow$ only ARMA(∞) can mimic an hyperbolic decay

- Such a slow decay reflects the presence of Long Memory ($\mathcal{LM}$)

$\Rightarrow$ the simplest parametric model of $\mathcal{LM}$ is the ARFIMA(p, δ, q)

$$(1-L)^{\delta} \underbrace{\Phi(L)(x_t - \mu) = \Theta(L)}_{\text{ARMA}(p,q)} \varepsilon_t \Rightarrow \varepsilon_t = \sum_{i=0}^{\infty} \Xi_i x_{t-i}$$

with $\delta \in (-1/2, 1)$ and $x_t \in \log(\mathcal{RM})$ or $x_t \in (\mathcal{RM})^{1/2}$



Estimating and forecasting with ARFIMA models

- Under Gaussian assumption (reasonable if $x_t \in \log(\mathcal{RM})$), the exact log-likelihood is

$$\mathcal{L}_E(\vartheta; \boldsymbol{x}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma(\vartheta)| - \frac{1}{2}\mathbf{x}'\Sigma(\vartheta)^{-1}\mathbf{x}$$

  with $\boldsymbol{x} = (x_1, \cdots, x_n)'$, $\vartheta = (\delta, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_p)$ and

$$\Sigma(\vartheta) = \gamma_x(r - s; \vartheta), \text{ for } r, s = 1, \cdots, n$$

- As the fractional polynomial $(1-L)^\delta$ implies an infinite dependence

$$(1-L)^\delta = \sum_{j=0}^{\infty} \frac{\Gamma(j+\delta)}{\Gamma(j+1)\Gamma(\delta)} L^j$$

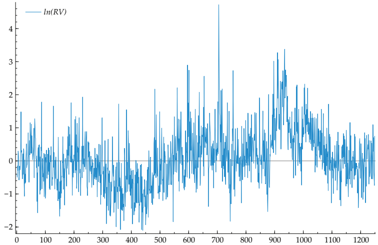
... the MLE $\widehat{\vartheta} = \arg\min_{\vartheta \in \Theta} \mathcal{L}_E(\vartheta; \boldsymbol{x})$ is time consuming

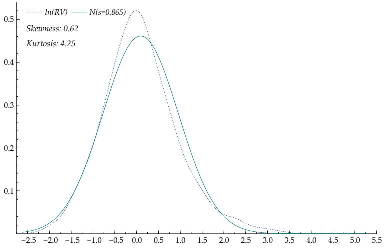
- The log-variance forecasts are then obtained by

$$\hat{x}_{t+h} \equiv \mathbb{E}_t(x_{t+h}) = \sum_{i=0}^{\infty} \Xi_i x_{t+h-i} \approx \sum_{i=0}^{r} \Xi_{h+i} x_{t-i}$$

  where in practice the infinite sum is truncated at lag r

Normality and Realized Variance





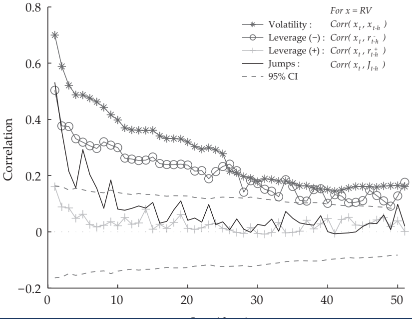
Modelling realized measures with HAR

- Corsi (2008) suggests an Heterogeneous AR model that mimics $\mathcal{LM}$

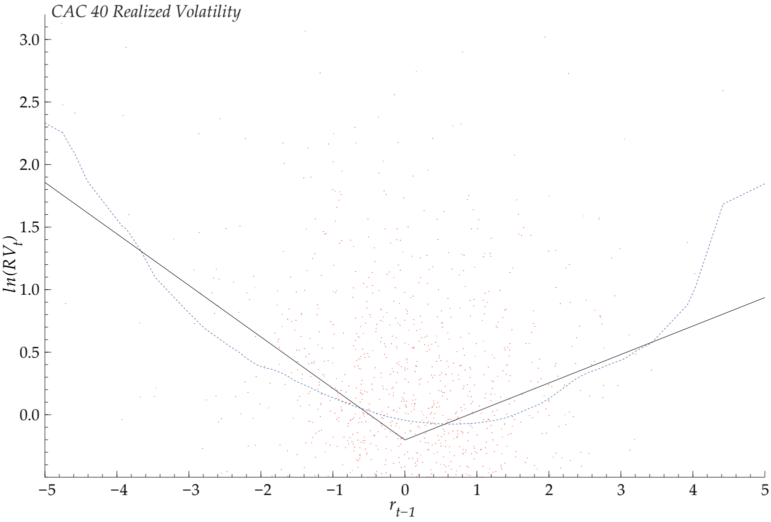
$$x_{t+1} = c + \underbrace{\beta_d x_t}_{\text{day}} + \underbrace{\beta_w x_t^w}_{\text{week}} + \underbrace{\beta_m x_t^m}_{\text{month}} + \varepsilon_{t+1}, \ \varepsilon_t \sim \text{ i. i. d. } ,$$

where $x_t^w = \frac{1}{5}\sum_{j=1}^{4} x_{t-j}$ and $x_t^m = \frac{1}{22}\sum_{j=1}^{21} x_{t-j}$ and $x_t \in (\mathcal{RM})^{1/2}$

- A log-HAR version is possible if $x_t \in \log(\mathcal{RM})$
- The HAR is simple and flexible $\Rightarrow$ can model additional stylized facts



Leverage in Realized Variance



Extensions of the HAR model : LHAR

- Define aggregated negative and positive returns at given frequency s

$$(r_t^s)^- = \frac{1}{s}\sum_{j=1}^{s} r_{t-j}\mathbb{1}_{\sum_{j=1}^{s} r_{t-j}<0} \text{ and } (r_t^s)^+ = \frac{1}{s}\sum_{j=1}^{s} r_{t-j}\mathbb{1}_{\sum_{j=1}^{s} r_{t-j}\geq 0}$$

- We allow for the leverage effect to impact each market component

$$\begin{aligned} x_{t+1} &= c + \beta_d x_t + \beta_w x_t^w + \beta_m x_t^m \\ &+ \gamma_d^- (r_t)^- + \gamma_w^- (r_t^w)^- + \gamma_m^- (r_t^m)^- \\ &+ \gamma_d^+ (r_t)^+ + \gamma_w^+ (r_t^w)^+ + \gamma_m^+ (r_t^m)^+ + \varepsilon_{t+1} \end{aligned}$$

$\Rightarrow$ The Leverage HAR (LHAR) fits particularly well the asymmetric relationship between r_t and x_{t+1} with $x_t \in (\mathcal{RM})^{1/2}$

Extensions of the HAR model : LHAR-CJ

- Assume now that J_{t,T_M} is the sequence of significant realized jumps
- In the spirit of the HAR let's define at given frequency s

$$J^s_{t,T_M} = \frac{1}{s}\sum_{j=1}^{s} J_{t-j,T_M}$$

- The Leverage HAR with Continuous volatility and Jumps (LHAR-CJ) is given by

$$\begin{aligned} x_{t+1} &= c + \beta_d x_t + \beta_w x_t^w + \beta_m x_t^m \\ &+ \alpha^d J_{t,T_M} + \alpha^w J^w_{t,T_M} + \alpha^m J^m_{t,T_M} \\ &+ \gamma_d^- (r_t)^- + \gamma_w^- (r_t^w)^- + \gamma_m^- (r_t^m)^- \\ &+ \gamma_d^+ (r_t)^+ + \gamma_w^+ (r_t^w)^+ + \gamma_m^+ (r_t^m)^+ + \varepsilon_{t+1} \end{aligned}$$

- As the model is linear, one can use OLS to estimate all parameters
- This general model offers good in-sample and out-of-sample performances

GARCH-X with realized measures

- Engle (2002) (and many others)

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma x_{t-1}$$

$x_t \in \mathcal{RM}$ is a realized measure of the QV (not IV)

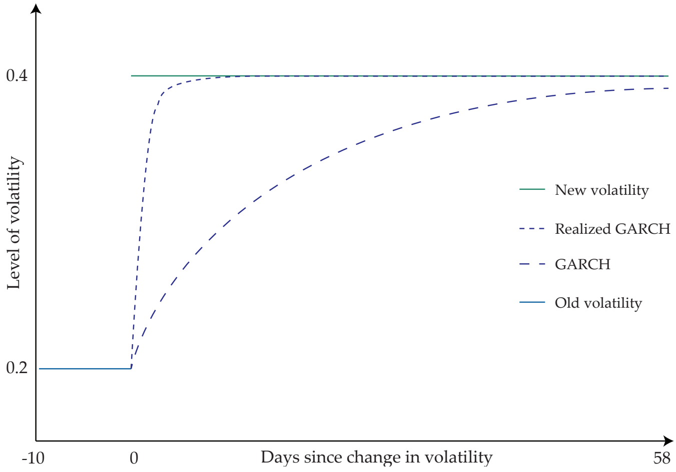
- Leads to substantial empirical gains

$\Rightarrow$ Typically

  - $\hat{\gamma} \simeq 0.5$
  - $\hat{\alpha} \simeq 0$ (ARCH parameter becomes insignificant)

- However, GARCH-X models are partial (incomplete) models that have nothing to say about returns and volatility beyond a single period into the future.

GARCH with a Realized Measure is Fast



Realized GARCH (Hansen, Huang, Shek, 2012)

- The Realized GARCH is a complete model, contrary to GARCH-X

$\Rightarrow$ It fully specifies the dynamic properties of both r_t and x_t :

$$\begin{aligned} r_t &= \overbrace{\mu + \sqrt{\sigma_t^2} z_t}^{\text{Return Equation}} \\ \log \sigma_t^2 &= \alpha + \beta \log \sigma_{t-1}^2 + \gamma \log x_{t-1} \Big\} \text{ GARCH Equation} \\ \log x_t &= \underbrace{\xi + \psi \log \sigma_t^2 + \tau(z_t) + u_t}_{\text{Measurement Equation}} \end{aligned}$$

- $x_t \in \mathcal{RM}$ is a realized measure of the QV (not IV)
- $\tau(z) = \tau_1 z + \tau_2 (z^2 - 1)$ models an asymmetric response in volatility to return shocks (leverage effect)
- $z_t \sim \text{ i.i.d. } (0, 1)$, $u_t \sim \text{ i.i.d. } (0, \sigma_u^2)$
- $\tau(z_t) + u_t$ are the volatility shocks

Realized GARCH estimation

- Assume $z_t \sim \text{i.i.d. } \mathcal{N}(0,1)$ and $u_t \sim \text{i.i.d. } \mathcal{N}(0,\sigma_u^2)$
- The quasi-log likelihood function is constructed based on

$$\overbrace{f(r_t, x_t | \mathcal{F}_{t-1})}^{\text{Joint density}} = f(r_t|\mathcal{F}_{t-1}) f(x_t|r_t, \mathcal{F}_{t-1}),$$

$$l(r,x;\theta) = -\frac{1}{2}\sum_{t=1}^{n}\left(\log(\sigma_t^2) + r_t^2/\sigma_t^2\right) - \frac{1}{2}\sum_{t=1}^{n}\left(\log(\sigma_u^2) + u_t^2/\sigma_u^2\right)$$

- Based on Straumann and Mikosch (2006) the authors conjecture that

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0,\Omega),$$

with Ω the usual QMLE covariance matrix

Realized GARCH multi-period forecast

- Predicts both the conditional return variance and the realized measure

e.g. RGARCH(1,1) forecasts through the VARMA(1,1) structure

$$\begin{pmatrix} \log \sigma_t^2 \\ \log x_t \end{pmatrix} = \begin{pmatrix} \beta & \gamma \\ \psi\beta & \psi\gamma \end{pmatrix} \begin{pmatrix} \log \sigma_{t-1}^2 \\ \log x_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha \\ \psi\alpha + \xi \end{pmatrix} + \begin{pmatrix} 0 \\ \tau(z_t) + u_t \end{pmatrix}$$

- h-step ahead forecasts are obtained from

$$\begin{pmatrix} \widehat{\log \sigma_{t+h}^2} \\ \widehat{\log x_{t+h}} \end{pmatrix} = \begin{pmatrix} \beta & \gamma \\ \psi\beta & \psi\gamma \end{pmatrix}^h \begin{pmatrix} \log \sigma_{t-1}^2 \\ \log x_{t-1} \end{pmatrix} + \sum_{j=0}^{h-1} \begin{pmatrix} \beta & \gamma \\ \psi\beta & \psi\gamma \end{pmatrix}^j \left(\begin{pmatrix} \alpha \\ \psi\alpha + \xi \end{pmatrix} + \begin{pmatrix} 0 \\ \tau(z_{t+h-j}) + u_{t+h-j} \end{pmatrix} \right)$$

- One would have to account for distributional aspects of $\log \sigma_{t+h}^2$ in order to produce an unbiased forecast of it since $\mathbb{E}(\log x) \neq \log(\mathbb{E}(x))$

Chapter 3

Multivariate Realized Mesures

- There is growing theoretical and empirical interest in extending the results of the univariate processes discussed previously to a multivariate framework
- Multivariate volatility modelling is of particular importance in the areas of risk management, portfolio management and asset pricing
- We will discuss:
  - Multivariate realized measures
  - Modelling and forecasting multivariate realized measures

Multivariate realized measures

1. Realized Covariance (Barndorff-Nielsen and Shephard, 2004)

- Suppose that along day t the log-prices of financial assets follow a k-variate continuous time diffusion process

$$d\boldsymbol{p}_{t+\tau} = \boldsymbol{\mu}_{t+\tau} + \boldsymbol{\Sigma}_{t+\tau}^{1/2} d\boldsymbol{W}_{t+\tau}, \;\; 0 \leq \tau \leq 1, \; t = 1, 2, \ldots,$$

  where
  - $\boldsymbol{\mu}_{t+\tau}$ is the multivariate drift component
  - $\boldsymbol{\Sigma}_{t+\tau}^{1/2}$ is the instantaneous $k \times k$ co-volatility matrix
  - $\boldsymbol{W}_{t+\tau}$ is the standard multivariate Brownian motion
- $\boldsymbol{\Sigma}_{t+\tau}^{1/2}$ is orthogonal to $\boldsymbol{W}_{t+\tau}$
- A generic element of $\boldsymbol{\Sigma}_{t+\tau}$ is given by $\Sigma_{t+\tau}^{(u)(s)}$

Multivariate realized measures

1. Realized Covariance (Barndorff-Nielsen and Shephard, 2004)

Definition 17

The realized covariance over the arbitrary interval between 0 and 1 (representing day t) is computed using the outer-product of high-frequency returns

$$\mathbf{RCov}_t^{(ALL)} = \sum_{j=1}^{M} \boldsymbol{r}_{t_j} \boldsymbol{r}_{t_j}',$$

where $\boldsymbol{r}_{t_j}$ is the j-th return on day t and it is a consistent estimator for the sum of the integrated covariance matrix and the realized jump variability $\int_0^1 \boldsymbol{\Sigma}_{t+\tau} d\tau + \int_0^1 \xi_\tau \xi_\tau' dN(\tau)$

- Barndorff-Nielsen and Shephard (2004) showed that when the price process is Brownian motion with drift, as $M \to \infty$

$$M^{1/2}\left[vech(\mathbf{RCov}_t^{(ALL)}) - vech(\int_0^1 \boldsymbol{\Sigma}_{t+\tau} d\tau)\right] \xrightarrow{d} \mathcal{N}_k(0, \boldsymbol{\Pi}_t)$$

- In principle prices should be sampled as frequently as possible to maximize the precision of the realized covariance estimator

Multivariate realized measures

1. Realized Covariance (Barndorff-Nielsen and Shephard, 2004)

In practice:

- Prices, especially transaction prices (trades), are contaminated by noise (e.g. bid-ask spread, non-trading, price discreteness, trades occurring on different markets or networks, rounding errors)
- Bandi and Russell (2005) showed that in the presence of microstructure noise the realized covariation estimator is not consistent

Solutions:

- The standard method to address these concerns is to sample relatively infrequently, for example every 5 minutes
- An improved method is to use modified realized covariance estimators (e.g. using subsampling)
- Difficult to avoid singular covariance matrices issue when k is large

Multivariate realized measures

2. Realized Semi-Covariance (Bollerslev, Patton, Quaedvlieg, 2020)
   - Let $r_{t,j,i}$ denote the return over the j-th intra-daily time on day t for asset i and $\boldsymbol{r}_{t,j}$ the vector returns
   - With $\mathbb{I}\{.\}$ is the indicator function and $\odot$ the Hadamard product, the vectors of k signed returns are

$$\boldsymbol{r}_{t,j}^{+} = \boldsymbol{r}_{t,j} \odot \mathbb{I}\{\boldsymbol{r}_{t,j} > 0\}, \quad \boldsymbol{r}_{t,j}^{-} = \boldsymbol{r}_{t,j} \odot \mathbb{I}\{\boldsymbol{r}_{t,j} \leq 0\},$$

Definition 18

Therefore, the standard realized covariance matrix can be decomposed in four realized semi-covariance matrices:

$$\boldsymbol{P}_t^{(M)} \equiv \sum_{j=1}^{M} \boldsymbol{r}_{t,j}^{+} \boldsymbol{r}_{t,j}^{+\prime}, \quad \boldsymbol{Q}_t^{(M)+} \equiv \sum_{j=1}^{M} \boldsymbol{r}_{t,j}^{+} \boldsymbol{r}_{t,j}^{-\prime},$$

$$\boldsymbol{Q}_t^{(M)-} \equiv \sum_{j=1}^{M} \boldsymbol{r}_{t,j}^{-} \boldsymbol{r}_{t,j}^{+\prime}, \quad \boldsymbol{N}_t^{(M)} \equiv \sum_{j=1}^{M} \boldsymbol{r}_{t,j}^{-} \boldsymbol{r}_{t,j}^{-\prime},$$

where $\boldsymbol{P}$, $\boldsymbol{N}$ *and* $\boldsymbol{Q}$ *correspond to "positive", "negative" and "mixed" vector signs, such that the realized covariance matrix is given by*

$$RCov_t^{M} = \boldsymbol{P}_t^{(M)} + \boldsymbol{N}_t^{(M)} + \boldsymbol{Q}_t^{(M+)} + \boldsymbol{Q}_t^{(M-)}$$

Multivariate realized measures

2. Realized Semi-Covariance (Bollerslev, Patton, Quaedvlieg, 2020)

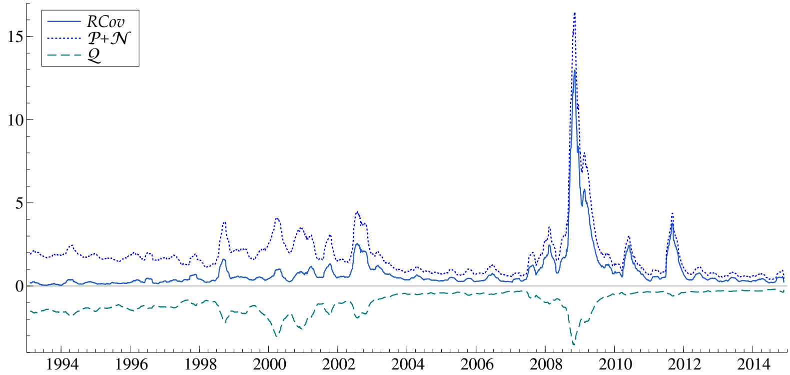
- Since $RCov_t^M$, $\boldsymbol{P}_t^{(M)}$ and $\boldsymbol{N}_t^{(M)}$ are all defined as sums of vector outerproducts, these matrices are all positive semidefinite
- Since the diagonal elements of $\boldsymbol{Q}_t^{(M)+}$ and $\boldsymbol{Q}_t^{(M)-}$ are identically zero by construction, these matrices are necessarily indefinite
- If the assets have a clear ordering, the two realized semicovariances may have different economic interpretations
- If the ordering is arbitrary, they may convey the same information, combined in a single semi-covariance matrix $\boldsymbol{Q}_t^{(M)} = \boldsymbol{Q}_t^{(M)+} + \boldsymbol{Q}_t^{(M)-}$
- E.g. In a bivariate case

$$\boldsymbol{P}_t^{(M)} = \begin{pmatrix} \mathcal{P}_{1,t}^{(M)+} & \mathcal{P}_{12,t}^{(M)} \\ \bullet & \mathcal{P}_{2,t}^{(M)+} \end{pmatrix}, \quad \boldsymbol{N}_t^{(M)} = \begin{pmatrix} \mathcal{N}_{1,t}^{(M)-} & \mathcal{N}_{12,t}^{(M)} \\ \bullet & \mathcal{N}_{2,t}^{(M)-} \end{pmatrix}$$

$$\boldsymbol{Q}_t^{(M)} = \begin{pmatrix} 0 & \mathcal{Q}_{12,t}^{(M)} \\ \bullet & 0 \end{pmatrix}$$

Multivariate realized measures

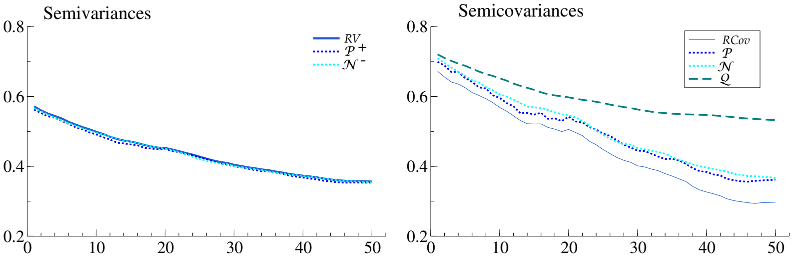
2. Realized Semi-Covariance (Bollerslev, Patton, Quaedvlieg, 2020)



- Daily realized semicovariances averaged across 500 randomly-selected pairs of S&P 500 stocks over the 1993-2014 period (smoothed by a moving average of 50 obs)

Multivariate realized measures

2. Realized Semi-Covariance (Bollerslev, Patton, Quaedvlieg, 2020)



- Autocorrelations functions for the different realized semicovariance elements averaged across 1000 randomly-selected pairs of S&P 500 stocks over the 1993-2014 period (smoothed by a moving average of 50 obs)

Multivariate realized measures

2. Realized Semi-Covariance (Bollerslev, Patton, Quaedvlieg, 2020) asymptotic theory ($k = 2$)

Definition 19

Assume that the bivariate log-price process evolves continuously through time according to the semimartingale $\boldsymbol{p}_t = \boldsymbol{p}_0 + \int_0^t \boldsymbol{m}_s ds + \int_0^t \boldsymbol{\sigma}_s d\boldsymbol{W}_s, \quad 0 \leq t \leq 1$, where $\boldsymbol{W}_s$ denotes a 2-dimensional Brownian motion, $\boldsymbol{m}_s$ is a 2-dimensional locally bounded predictable drift process, and $\boldsymbol{\sigma}$ is a $\mathbb{R}^{2\times 2}$-valued càdlàg volatility process. Then for $\boldsymbol{r}_{j,i} \equiv \boldsymbol{p}_{j/M,i} - \boldsymbol{p}_{(j-1)/M,i}$, $M \to \infty$, with $\sigma^2_{s,i}$ and ρ_s the spot variance and correlation respectively,

$$\boldsymbol{V}^{(M)} = \begin{bmatrix} \mathcal{P}_1^{(M)+} \\ \mathcal{P}_2^{(M)+} \\ \mathcal{N}_1^{(M)-} \\ \mathcal{N}_2^{(M)-} \\ \mathcal{P}_{12}^{(M)} \\ \mathcal{N}_{12}^{(M)} \\ \mathcal{Q}_{12}^{(M)+} \\ \mathcal{Q}_{12}^{(M)-} \end{bmatrix} \xrightarrow{p} \int_0^1 \begin{bmatrix} \sigma^2_{s,1}/2 \\ \sigma^2_{s,1}/2 \\ \sigma^2_{s,2}/2 \\ \sigma^2_{s,2}/2 \\ \sigma^2_{s,1}\sigma^2_{s,2}\left(\rho_s \arccos(-\rho_s) + \sqrt{1-\rho_s^2}\right)/(2\pi) \\ \sigma^2_{s,1}\sigma^2_{s,2}\left(\rho_s \arccos(-\rho_s) + \sqrt{1-\rho_s^2}\right)/(2\pi) \\ \sigma^2_{s,1}\sigma^2_{s,2}\left(\rho_s \arccos \rho_s - \sqrt{1-\rho_s^2}\right)/(2\pi) \\ \sigma^2_{s,1}\sigma^2_{s,2}\left(\rho_s \arccos \rho_s - \sqrt{1-\rho_s^2}\right)/(2\pi) \end{bmatrix} ds \equiv \boldsymbol{V}$$

Multivariate realized measures

2. Realized Semi-Covariance (Bollerslev, Patton, Quaedvlieg, 2020)

Asymptotic theory ($k = 2$ and a unit time interval)

Definition 20 (CLT and feasible CLT)

Assume that the bivariate log-price process evolves continuously through time according to slide (186), with the $\boldsymbol{\sigma}_t$ volatility process determined by $\boldsymbol{\sigma}_t = \boldsymbol{\sigma}_0 + \int_0^t \nu_s' d\boldsymbol{W}_s^*$, with ν_s' an adapted càdlàg process, and $\boldsymbol{W}_s^*$ a 2×2-dimensional Brownian motion independent of $\boldsymbol{W}_s$. Then for $M \to \infty$,

$$\sqrt{M}(\boldsymbol{V}^{(M)} - \boldsymbol{V}) \xrightarrow{d_{st}} \int_0^1 \boldsymbol{\alpha}_s d\boldsymbol{W}_s + \int_0^1 \boldsymbol{\beta}_s d\tilde{\boldsymbol{W}}_s \equiv \boldsymbol{U},$$

where d_{st} denotes stable convergence in distribution, $\boldsymbol{\alpha}_s$ and $\boldsymbol{\beta}_s$ are 8×2-dimensional processes. With

$$\boldsymbol{\Pi} \equiv Var(\boldsymbol{U}) = \int_0^1 (\boldsymbol{\alpha}_s \boldsymbol{\alpha}_s' + \boldsymbol{\beta}_s \boldsymbol{\beta}_s') ds,$$

$$\boldsymbol{\Pi}^{(M)} \xrightarrow{p} \boldsymbol{\Pi},$$

$$\{\boldsymbol{\Pi}^{(M)}\}^{-1/2} \sqrt{M}(\boldsymbol{V}^{(M)} - \boldsymbol{V}) \xrightarrow{d} \mathcal{N}_2(0, \boldsymbol{I})$$

Multivariate realized measures

3. Realized Bipower Covariation (Barndorff-Nielsen and Shephard, 2004b)

Definition 21

The realized bipower covariation over the arbitrary interval between 0 and 1 (representing day t) computed as

$$RBPCov_t = \frac{1}{4}\Big(\sum_{j=2}^{M} |r_{(v),t_j} + r_{(l),t_j}||r_{(v),t_{j-1}} + r_{(l),t_{j-1}}| - |r_{(v),t_j} - r_{(l),t_j}||r_{(v),t_{j-1}} - r_{(l),t_{j-1}}|\Big)$$

where $r_{(v),t_j}$ is the v-th component of the return vector $\mathbf{r}_{t_j}$, is an estimator of the integrated covariance matrix (robust to jumps)

$$RBPCov_t \xrightarrow[\Delta\to 0]{p} \int_0^1 \boldsymbol{\Sigma}_{t+\tau} d\tau$$

Multivariate realized measures: Asynchronous trading

- Compared to the univariate case, the additional issue of **synchronicity** arises for multivariate measures
- The asynchronous nature of intraday prices biases realized covariances toward 0, unless an appropriate adjustment is made
- The downward bias occurs because when trading is infrequent, news that affect a pair of assets will be incorporated at different times simply as a result of asynchronous trading

Definition 22 (Epps effect, Epps, 1979)

Information arrives at different frequencies for different assets, therefore introducing additional microstructure effects that are related to the nonsynchronicity in the process of price formation.
Even when there is no microstructure friction as previously discussed, nonsynchronous trading introduces a downward bias in the realized covariance estimates when sampling returns in calendar time at high frequencies.

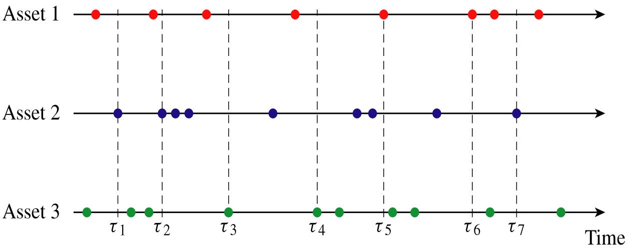
Multivariate realized measures: Asynchronous trading

An approach to tackle the problem of asynchronous trading :

- **previous tick aggregation :** forces prices to an equispaced grid by taking the last price realized before each grid point or, alternatively, the interpolation of the 1st and last price in the interval
- But, at least one quote should be available for both assets in the chosen time interval for this algorithm to be applicable
- Empirical work use heuristically chosen 5 or 30 minutes return interval to try to avoid the bias and market microstructure effects
- But this type of correction will increase the variance of the realized covariance estimator and no rule to choose an optimal frequency exists

Multivariate realized measures: Asynchronous trading

A better approach: **Refresh-time sampling (Harris and Wood, 1995)**



- Each $\{\tau_j\}$ with $j = 1, \ldots, M$ is the time it has taken for all the assets to (re)-trade.
- This procedure forces the time series to synchronize but not necessarily on equispaced time grid
- Some realized measures are based on this time clock $\{\tau_j\}$ while showing that stale pricing errors have no impact on the asymptotic distribution of the measures

Multivariate realized measures : Realized Kernel

4. Realized Kernel (Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2011)

Definition 23

The multivariate realized kernel over the arbitrary interval between 0 and 1 (representing day t) defined as

$$RK_t = \Gamma_0 + \sum_{h=1}^{H} \mathcal{K}\left(\frac{h}{H}\right)(\Gamma_h + \Gamma_h'),$$

$$\Gamma_h = \sum_{j=h+1}^{\tilde{M}} \tilde{\boldsymbol{r}}_{t_j} \tilde{\boldsymbol{r}}_{t_{j-h}},$$

with $\tilde{\boldsymbol{r}}_{t_j}$ refresh-time returns, $\mathcal{K}$ a kernel weighting function (e.g. Parzen), $\tilde{M}$ is the refresh-time sample size after jittering (averaging observations on the boundaries of the sample) and H is a parameter which controls the bandwidth.

- Simultaneously guarantees consistency, positive semi-definiteness and robustness to microstructure noise
- Accounts for nonsynchroneity of observations by using refresh-time returns

Multivariate realized measures : Realized Kernel

Theorem 4 (Asymptotic theory)

$$\frac{H^2}{\tilde{M}} RK_t \xrightarrow{p} |\mathcal{K}''(0)|\Omega, \; \textit{if } \eta < 1/2,$$

$$RK_t = \int_0^1 \Sigma(s)ds + c_0^{-2}|\mathcal{K}''(0)|\Omega + O_p(1), \; \textit{if } \eta = 1/2,$$

$$RK_t \xrightarrow{p} \int_0^1 \Sigma(s)ds \textit{ if } \eta > 1/2,$$

where $H = c_0 \tilde{M}^{\eta}$, $c_0 > 0$, $\eta \in (0,1)$, *and* Ω*: average long run variance of the noise*

$$\tilde{M}^{1/5}\left(RK_t - \int_0^1 \Sigma(s)ds\right) \xrightarrow{d_{st}} \mathcal{MN}(c_0^{-2}|\mathcal{K}''(0)|\Omega, 4c_0\mathcal{K}_\bullet^{0,0})IQ),$$

with $\mathcal{K}_\bullet^{0,0} = \int_0^\infty \mathcal{K}(s)^2 ds$ *and IQs the multivariate integrated quarticity*

- The bandwidth H plays a crucial role, as it has to increase with $\tilde{M}$ quite quickly to remove the influence on the estimator of the noise

Multivariate realized measures : Realized Kernel

Bandwidth choice

- It must grow with $\tilde{M}$ at rate $\tilde{M}^{3/5}$
- Solutions for estimating a good constant of proportionality in this multivariate case:

i) Apply the univariate optimal mean square error bandwidth selection to each asset price individually, i.e. $\hat{H}_{(i)} = c_0 \hat{\xi}_{(i)}^{4/5} \tilde{M}^{3/5}$, for $i = 1, 2, \ldots, k$, with $c_0 = 3.5134$ for Parzen kernel
One then gets k bandwidths and constructs some ad hoc rules for choosing the global $\hat{H}$, such as Min(.), Max(.) or Average(.)

ii) Construct a sort of equally weighted "market portfolio"
Once prices are converted into Refresh Time, one computes the market return and then carries out a univariate analysis on it, choosing an optimal $\hat{H}$ for the market
This single $\hat{H}$ is then applied to the multivariate problem

Multivariate realized measures

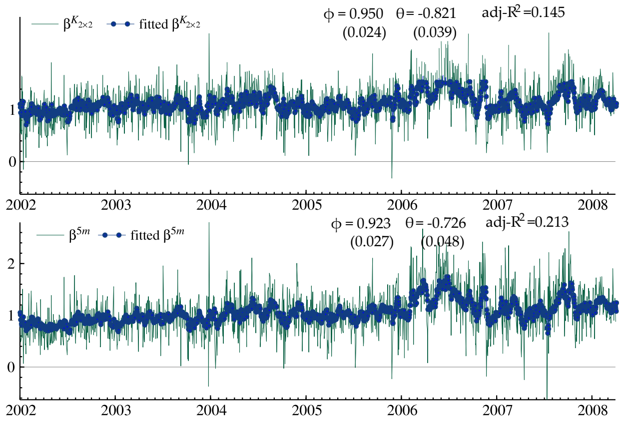
2. Realized correlation and realized beta
   - Realized Correlation is the realized analogue of the usual correlation estimator, but defined in terms of realized covariance (RC)
   $$RCorr = \frac{RC_{vl}}{\sqrt{RC_{vv} RC_{ll}}}$$
   - Suppose
   $$RC = \begin{pmatrix} RV_v & RC'_{fv} \\ RC_{fv} & RC_{ff} \end{pmatrix}$$
   is the realized covariance matrix of an asset with a set of observable factors. The realized beta is defined as
   $$R\beta = RC_{ff}^{-1} RC_{fv}$$
   - Realized betas are similar to other realized measures in that they are model free and, as long as prices can be sampled frequently and have little market microstructure noise, is an accurate measure of the current exposure to changes in the market

Multivariate realized measures

2. Realized correlation and realized beta

Stylized facts (AA <Alcoa Inc.> - SPY analysis)

- Estimate the realised kernel beta and 5 min-based realized betas
- Model them as ARMA(1,1) processes



Modelling and forecasting multivariate realized measures

- An alternative to multivariate GARCH models is based on **realized covariance measures**
- Problem: the matrix constructed from the variance and correlation **forecasts** obtained from **disjoint (univariate) models** is not guaranteed to be positive definite
- Solutions:
  - VARFIMA approach (Chiriac and Voev, 2010)
  - Heterogeneous Autoregressive approach (Chiriac and Voev, 2011; Cech and Barunik, 2016)
  - Multivariate Realized GARCH approach (Hansen et al. 2014)

Modelling and forecasting multivariate realized measures

1. VARFIMA approach (Chiriac and Voev, 2010)

- Let r_t denote a $k \times 1$ vector of asset returns
- Let Y_t denote the realized covariance matrix, i.e. a non-parametric estimator of QV_t associated with r_t

... and P_t the upper triangular matrix of the Choleski decomposition of Y_t such that $P_t' P_t = Y_t$

- Let $X_t = vech(P_t)$ be the $g \times 1$ vector obtained by stacking the upper triangular components of P_t, where $g = k(k+1)/2$

Modelling and forecasting multivariate realized measures

1. VARFIMA approach (Chiriac and Voev, 2010)

- X_t is modeled as a vector autoregressive fractionally integrated moving average (VARFIMA(p, δ, q)) process

$$\Phi(L)D(L)[X_t - BZ_t] = \Theta(L)\varepsilon_t, \; \varepsilon_t \sim \mathcal{N}(0, \Omega)$$

- where Z_t is a $n \times 1$ vector of exogenous variables
- B is a $g \times n$ matrix of coefficients
- $\Phi(L) = I_g - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p$ and $\Theta(L) = I_g - \Theta_1 L - \Theta_2 L^2 - \ldots - \Theta_q L^q$ are matrices of lag AR and MA polynomial, respectively
- $D(L) = diag\{(1-L)^{\delta_1}, \ldots, (1-L)^{\delta_g}\}$ with $d_1, \ldots, \delta_g$ the degrees of fractional integration of each of the g elements of X_t
- We assume that the roots of $\Phi(L)$ and $\Theta(L)$ lie outside the unit circle
- Z_t could include exogenous variables (trading volume, corporate bond returns, short-term interest rates, etc.)

Modelling and forecasting multivariate realized measures

1. VARFIMA approach (Chiriac and Voev, 2010)
   - Advantages of Choleski factors VARFIMA
     - No need to impose parameter restrictions on the model
     - The out-of-sample covariance matrice is always positive definite by the 'reverse' Cholesky transformation

$$Y_{ij,t} = \sum_{l=1+i(i-1)/2}^{i(i-1)/2} X_{l,t} X_{l+j(j-1)/2-i(i-1)/2,t}, \; i,j = 1,\ldots,k, \; j \geq i,$$

     where $X_{l,t}$ is the l-th element of X_t
   - Parsimonious VARFIMA (1,d,1) model for forecasting

$$(1-\phi L)D(L)[X_t - c] = (1-\theta L)\varepsilon_t, \; \varepsilon_t \sim \mathcal{N}(0,\Omega),$$

     where c is a $g \times 1$ vectors of constants
   - Estimation under the assumption of normally distributed errors by the approximate ML approach in the spirit of Beran (1995), which is also applicable for non-stationary processes ($\delta > 0.5$)

Modelling and forecasting multivariate realized measures

1. Forecasting with the VARFIMA model for the Cholesky factors

- From the VAR(∞) representation of the model

$$\Phi(L)D(L)\Theta(L)^{-1}X_t = \sum_{i=0}^{\infty} \Xi_i X_{t-i},$$

one obtains multi-step-ahead forecasts by truncation at r lags

$$\hat{X}_{t+s} = \mathbb{E}_t(X_{t+s}) = \sum_{i=0}^{\infty} \Xi_i X_{t+s-i} \approx \sum_{i=0}^{r} \Xi_{i+s} X_{t-i}$$

- The forecast is unbiased
- Having obtained $\hat{X}_{t+s}$, we construct the forecast $\hat{Y}_{t+s}$ by applying the 'reverse' Cholesky transformation on the previous slide
- Since $\hat{Y}_{ij,t+s}$ is a quadratic transformation of $\hat{X}_{t+s}$, it is biased by $\sigma^*_{s,ij} = \sum_{l=1+i(i-1)/2}^{i(i+1)/2} \sigma_{s(l,l+j(j-1)/2-i(i-1)/2)}$ where $j \geq i$, $i = 1, \ldots, k$ and $\sigma_{s(u,v)}$ is the (u, v)-element of $\Omega_s = \sum_{i=0}^{\infty} \Psi_i \Omega \Psi_i'$ with Ψ_i the coefficients of the VMA(∞) representation of the model

Modelling and forecasting multivariate realized measures

2. Heterogeneous Autoregressive approach (Chiriac and Voev, 2011; Cech and Barunik, 2016)

   - Extensions of Corsi (2009) to a multivariate framework
   - Let r_t denote a $k \times 1$ vector of asset returns
   - Let Y_t denote the realized covariance matrix, i.e. a non-parametric estimator of QV_t associated with r_t

   ... and P_t the upper triangular matrix of the Choleski decomposition of Y_t such that $P_t' P_t = Y_t$

   - Let $X_t = vech(P_t)$ be the $g \times 1$ vector obtained by stacking the upper triangular components of P_t, where $g = k(k+1)/2$

Modelling and forecasting multivariate realized measures

2. Heterogeneous Autoregressive approach (Chiriac and Voev, 2011; Cech and Barunik, 2016)

- Chiriac and Voev (2011) use the HAR representation for the vector of Cholesky factors X_t

$$X_{t+1} = c + \beta_d X_t + \beta_w X_t^w + \beta_m X_t^m + \varepsilon_t, \ \varepsilon_t \sim \text{ i.i.d. } ,$$

where $X_t^w = \frac{1}{5}\sum_{j=1}^{4} X_{t-j}$ and $X_t^m = \frac{1}{22}\sum_{j=1}^{21} X_{t-j}$

  — Estimation by OLS

  — Assumes the same structure for all elements of the factors in X_t

  — Assumes homoscedasticity and no cross-correlation of the error term

Modelling and forecasting multivariate realized measures

Example:

- Data: tick-by-tick bid and ask NYSE quotes 01.01.2000–30.07.2008 ($n = 2156$ trading days)
- Use six highly liquid stocks: American Express Inc. (AXP), Citigroup (C), General Electric (GE), Home Depot Inc. (HD), International Business Machines (IBM) and JPMorgan Chase & Co. (JPM)
- Through previous-tick interpolation obtain 78 intraday returns by sampling every 5 minutes (and subsampling at 300 seconds) and construct daily realized covariance matrices
- In-sample from 01.01.2000 to 31.12.2005 and out-of-sample from 01.01.2006 to 30.07.2008

Modelling and forecasting multivariate realized measures

RMSE based on the Frobenius norm of the forecasting error

Model	1 day	Iterated		Direct	
		5 days	10 day	5 days	10 days
VARFIMA-Cholesky	3.897[a]	3.388[a]	3.515[a]	3.540[a]	3.716[a]
VARFIMA-Log	3.937[a]	3.498	3.610[a]	3.525[a]	3.700[a]
HAR-Cholesky	3.940	3.459[a]	3.628	3.652[a]	3.919
HAR-Log	3.943	3.492	3.627	3.585[a]	3.871
Diagonal WAR	4.990	6.198	7.055	4.673	4.608
Diagonal WAR-HAR	4.598	4.995	5.752	4.489	4.659
DCC	5.195	4.727	4.851	5.252	4.945
FIDCC	5.613	4.613	4.767	5.435	5.224

[a] Model belongs to the 5% MCS of Hansen et al.(2009).

- MCS : Model Confidence Set, i.e. the set of models with significantly better forecasting abilities

Modelling and forecasting multivariate realized measures

Economic value of volatility forecasts : portfolio optimization

- Suppose the return distribution is completely characterized by its first two moments

$\Rightarrow$ Portfolio optimization reduces to finding asset weights which minimize the portfolio volatility for a given expected return (Markowitz, 1952)

- Denote by μ_p the annualized expected return
- The optimal portfolio is given by the solution to the following quadratic problem

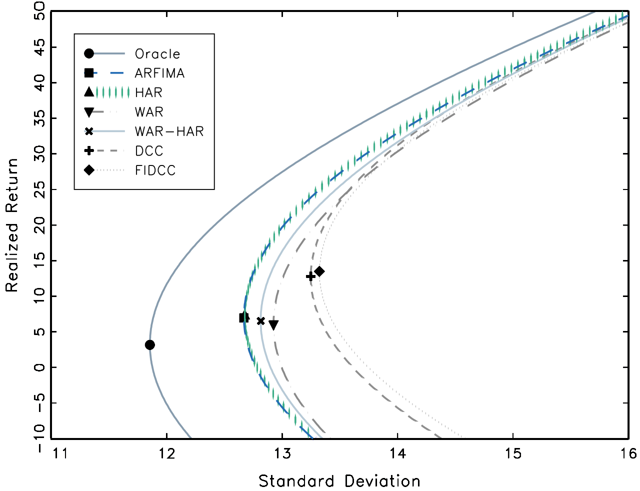
$$\min_{w_{t+s|t}} w'_{t+s|t}\hat{H}_{t:t+s}w'_{t+s|t}, \text{ s.t. } w'_{t+s|t}\mathbb{E}_t[r_{t:t+s}] = \frac{s\mu_p}{250}, \ w'_{t+s|t}\mathbb{1}_k = 1,$$

with

- $w_{t+s|t}$ the $k \times 1$ vector of portfolio weights chosen at t and held until $t+s$,
- $\hat{H}_{t:t+s}$ a covariance matrix forecast
- $\frac{s\mu_p}{250}$ the target return scaled to the investment horizon s

Modelling and forecasting multivariate realized measures

Mean-variance plots for the ex-post realized conditional mean against realized conditional standard deviation



All plots are averages across the 648 out-of-sample periods (days)

Modelling and forecasting multivariate realized measures

Annualized realized conditional standard deviations of the ex-post global minimum variance portfolio (%)

Model	1 day	Iterated		Direct	
		5 days	10 days	5 days	10 days
VARFIMA-Cholesky	12.669[a]	12.931[a]	13.016[a]	12.924[a]	13.001
HAR-Cholesky	12.676	12.934[a]	13.023[a]	12.934	13.018
Diagonal WAR	12.925	13.462	13.786	13.219	13.143
Diagonal WAR-HAR	12.814	13.186	13.398	13.178	13.150[a]
DCC	13.248	13.501	13.588	13.805	15.202
FIDCC	13.323	13.552	13.530	14.134	15.279

[a]Model belongs to the 5% MCS of Hansen et al. (2009).
All numbers are averages across the out-of-sample periods

Modelling and forecasting multivariate realized measures

3. Multivariate Realized GARCH approach (Hansen et al. 2014)

- They propose a Hierarchical Realized GARCH Framework

... which ties all individual return series to the market return

  — It models the the conditional distribution of a vector of returns as well as realized measures of volatility and correlation

  ⇒ parsimonious and simple to estimate model

  ⇒ relates key variables in the model to dynamic market betas

- It includes i) a marginal model for the market return and its realized measure of volatility

... and ii) conditional (on the market) models for individual asset returns, variance, and correlation (with the market)

Modelling and forecasting multivariate realized measures

3. Multivariate Realized GARCH approach (Hansen et al. 2014)

- Let $r_{i,t}$ and $x_{i,t}$ denote the returns and a corresponding realized measure of variance, where $i = 0$ corresponds to the market and $i = 1, \ldots, k$ the assets
- Let $\varrho_{i,t}\ in(-1, 1)$ denote a realized measure of correlation between asset i and the market
- Let $h_{i,t}|\mathcal{I}_{t-1}$, $x_{i,t}|\mathcal{I}_{t-1}$ and $\rho_{i,t}|\mathcal{I}_{t-1}$ denote the conditional equivalents

$\Rightarrow$ $\beta_{i,t} = \rho_{i,t}\sqrt{h_{i,t}/h_{0,t}}$ for $i \geq 1$ denotes market beta whose dynamics is of interest

Modelling and forecasting multivariate realized measures

3. Multivariate Realized GARCH approach (Hansen et al. 2014)

Marginal model for the market return

$$\begin{aligned} r_{0,t} &= \mu_0 + \sqrt{\sigma^2_{0,t}} z_{0,t} \\ \log \sigma^2_{0,t} &= a_0 + b_0 \log \sigma^2_{0,t-1} + c_0 \log x_{0,t-1} \\ \log x_{0,t} &= \xi_0 + \psi_0 \log \sigma^2_{0,t} + \tau_0(z_{0,t}) + u_{0,t} \end{aligned}$$

- $z_{0,t} \sim \text{ i.i.d. } (0,1)$, $u_{0,t} \sim \text{ i.i.d. } (0, \sigma^2_{u_0})$ for the estimation
- A simple second-order polynomial is used for the leverage effect function

Modelling and forecasting multivariate realized measures

3. Multivariate Realized GARCH approach (Hansen et al. 2014)

Conditional model for an individual asset return and its realized measures ($i > 0$)

$$r_{i,t} = \mu_i + \sqrt{\sigma_{i,t}^2} z_{i,t} \quad \text{Return Eq.}$$

- where the dependence on $(r_{0,t}, x_{0,t})$ operates through $\rho_{i,t} = \text{cov}(z_{0,t}, z_{i,t} | I_{t-1})$

$\Rightarrow$ i.e. there is a "factor" structure $z_{i,t} = \rho_{i,t} z_{0,t} + \sqrt{1 - \rho_{i,t}^2 w_{i,t}}$, where $w_{i,t}$ has mean zero, unit variance and is uncorrelated with $z_{0,t}$

$$\left.\begin{array}{c} \log \sigma_{i,t}^2 = a_i + b_i \log \sigma_{i,t-1}^2 + c_i \log x_{i,t-1} + d_i \log \sigma_{0,t-1}^2 \\ F(\rho_{i,t}) = a_{i,0} + b_{i,0} F(\rho_{i,t-1}) + c_{i,0} F(\varrho_{i,t-1}) \end{array}\right\} \text{GARCH Eq.}$$

$$\left.\begin{array}{c} \log x_{i,t} = \xi_i + \psi_i \log \sigma_{i,t}^2 + \tau_i(z_{i,t}) + u_{i,t} \\ F(\varrho_{i,t}) = \xi_{i,0} + \psi_{i,0} F(\rho_{i,t}) + v_{i,t} \end{array}\right\} \text{Measurement Eq.}$$

Fisher transform, F, is a 1 to 1 mapping from (-1,1) to $\mathbb{R}$

Modelling and forecasting multivariate realized measures

3. Multivariate Realized GARCH approach (Hansen et al. 2014)

Conditional model for an individual asset return and its realized measures ($i > 0$)

- The measurement errors will be assumed to be independent of the studentized innovations
- The measurement errors are allowed to be correlated

$$\mathbb{V}\begin{pmatrix} u_{0,t} \\ u_{i,t} \\ v_{i,t} \end{pmatrix} = \begin{pmatrix} \sigma^2_{u_0} & \sigma_{u_0,u_i} & \sigma_{u_0,v_i} \\ \bullet & \sigma^2_{u_i} & \sigma_{u_1,v_i} \\ \bullet & \bullet & \sigma^2_{v_i} \end{pmatrix}$$

Modelling and forecasting multivariate realized measures

3. Multivariate Realized GARCH approach (Hansen et al. 2014)

Estimation

- The joint density of the observables conditional on the information set can be decomposed

$$
\begin{aligned}
f(r_{0,t}, x_{0,t}, r_{i,t}, x_{i,t}, y_{i,t}|\mathcal{F}_{t-1}) &= f(r_{0,t}, x_{0,t}|\mathcal{F}_{t-1}) f(r_{i,t}, x_{i,t}, y_{i,t}|r_{0,t}, x_{0,t}, \mathcal{F}_{t-1}) \\
&= f(r_{0,t}|\mathcal{F}_{t-1}) f(x_{0,t}|r_{0,t}, \mathcal{F}_{t-1}) \times \\
&\quad f(r_{i,t}|r_{0,t}, x_{0,t}, \mathcal{F}_{t-1}) f(x_{i,t}, y_{i,t}|r_{i,t}, r_{0,t}, x_{0,t}, \mathcal{F}_{t-1})
\end{aligned}
$$

- As $z_{0,t} \sim \text{ i.i.d. } (0,1)$, $u_{0,t} \sim \text{ i.i.d. } (0, \sigma^2_{u_0})$, the four components of the likelihood function corresponding to the four conditional densities above are easy to explicitate
- The parameters of the model, generally denoted by θ are hence estimated by

$$
\hat{\theta} = \arg\max_{\theta \in \Theta} \log L(r_{0,t}, x_{0,t}, r_{i,t}, x_{i,t}, y_{i,t}|\theta)
$$

Thank you!