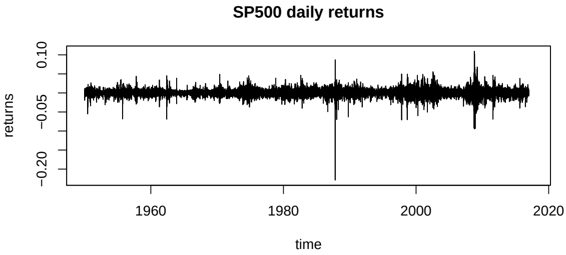
Risk Measurement with Heteroscedasticity

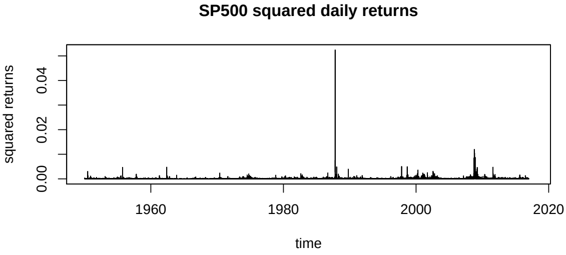
Carlo Favero and Celso Brunetti

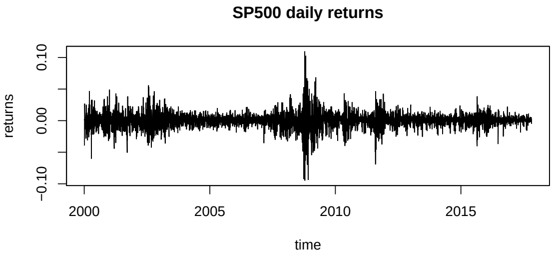
November 13, 2017

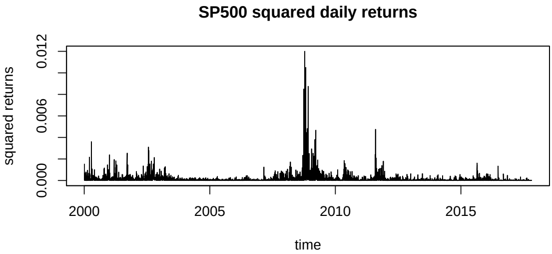
The Evidence form high frequency data

- Data at high-frequency (monthly or higher) show:
  - very little or no persistence in first moments
  - persistence in the variance
  - non-normality
- These features of the data can be used to measure VaR, using appropriate models for hetroscedasticity and non-normality


SP500 daily returns
returns
0.10
-0.05
-0.20
1960
1980
2000
2020
time


SP500 squared daily returns
squared returns
0.00
0.02
0.04
1960
1980
2000
2020
time





- Visibly, volatility "clusters" in time: high (low) volatility tends to be followed by high (low) volatility

Auto-Correlation

- The estimates from a standard ACF can be used to test the hypothesis that the process generating observed returns is a series of independent and identically distributed (IID) variables. The asymptotic (also called Bartlett's) standard error of the autocorrelation estimator is approximately $1/\sqrt{T}$, where T is the sample size
- The no-autocorrelation hypothesis can more formally be tested using the Portmanteau Q-statistic of Box and Pierce (1970), $\hat{Q}_k$, calculated from the first k autocorrelations of returns as:

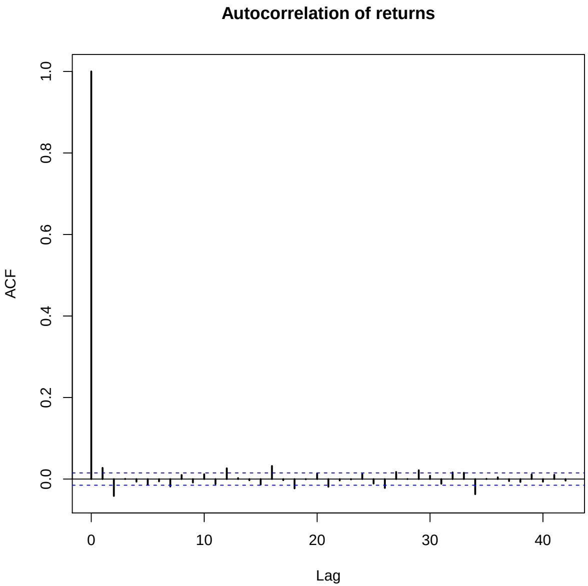
$$\hat{Q}_k \equiv T \sum_{\tau=1}^{k} \hat{\rho}_\tau^2 \overset{a}{\sim} \chi_k^2 \quad \text{where} \quad \hat{\rho}_\tau \equiv \frac{\sum_{t=1}^{T-\tau} (R_t - \bar{R})(R_{t+\tau} - \bar{R})}{\sum_{t=1}^{T-\tau} (R_t - \bar{R})^2}$$

Auto-Correlation

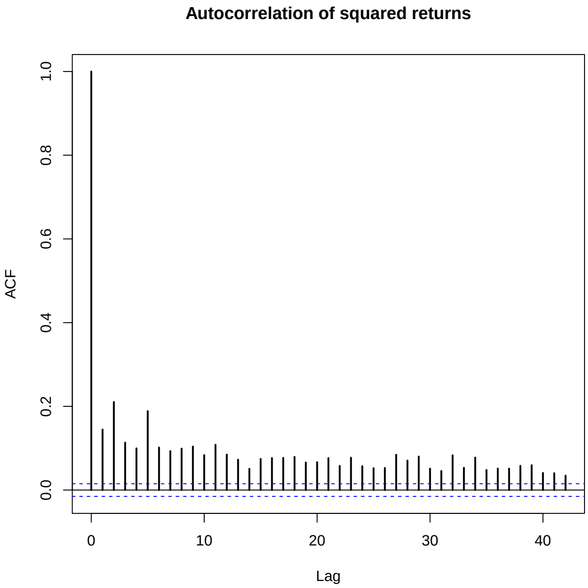
- Note that the null of no-autocorrelation is not rejected for returns but it for squared returns therefore the independence hypothesis of returns is rejected as there is persistence in the second moments
- The definition of independence of a time series process has the following characterization:

$$
\begin{aligned}
R_t \text{ is IID} \quad &\Longleftrightarrow \quad \hat{Q}_k^g \simeq 0 \text{ for all } k \geq 1 \\
\hat{Q}_k^g \quad &\equiv \quad T \sum_{\tau=1}^{k} (\hat{\rho}_\tau^g)^2 \overset{a}{\sim} \chi_k^2 \\
\hat{\rho}_\tau^g \quad &\equiv \quad \frac{\sum_{t=1}^{T-\tau} \left(g(R_t) - \overline{g(R_t)}\right)\left(g(R_{t+\tau}) - \overline{g(R_t)}\right)}{\sum_{t=1}^{T-\tau} \left(g(R_t) - \overline{g(R_t)}\right)}
\end{aligned}
$$

Auto-Correlation



Auto-Correlation



Normality

Normality of a distribution can be assessed via

- Jarque-Bera test
- Comparison of the empirical density with the normal density
- QQ-plot

Jarque-Bera Test

The key tool to perform statistical tests of normality is Jarque and Bera's (1980) test. if $X_t \sim \mathcal{N}(\mu, \sigma^2)$, then the distribution of X_t is symmetric—therefore it has zero skewness—and it has a kurtosis of 3. In particular, if we define the unconditional mean $\mu \equiv E[X_t]$ and the variance $\sigma^2 \equiv Var[X_t]$, then skewness and kurtosis are defined as:

$$\begin{aligned} Skew[X_t] &\equiv \frac{E[(X_t-\mu)^3]}{(Var[X_t])^{3/2}} = \frac{E[(X_t-\mu)^3]}{\sigma^3}, \\ Kurt[X_t] &\equiv \frac{E[(X_t-\mu)^4]}{(Var[X_t])^2} = \frac{E[(X_t-\mu)^4]}{\sigma^4} \geq 0 \end{aligned}$$

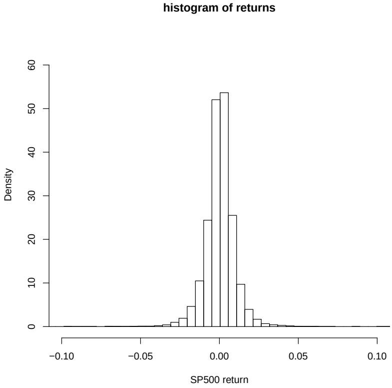
Jarque and Bera's test statistic is:

$$\widehat{JB} \equiv \frac{T}{6}\left\{\widehat{Skew}[X_t]\right\}^2 + \frac{T}{24}\left\{\widehat{Kurt}[X_t] - 3\right\}^2 \overset{a}{\sim} \chi^2_2$$

Jarque-Bera Test

```
Jarque Bera Test

data:  sp_ret
X-squared = 519710, df = 2, p-value < 2.2e-16
```



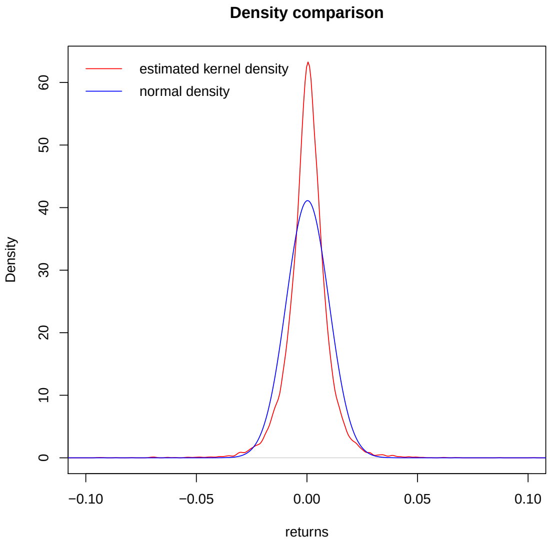
Density Comparison

Compute kernel density estimator of our variable and compare it with the normal distribution.

A kernel density estimator is an empirical density "smoother" based on the choice of two objects: (i) the *kernel function* $K(x)$, and (ii) the *bandwidth parameter*, h. The kernel function is defined as some smooth function (read, continuous and sometimes also differentiable) that integrates to 1. The *kernel density estimator*:

$$\hat{f}_X^{\text{ker}}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$$

Density Comparison

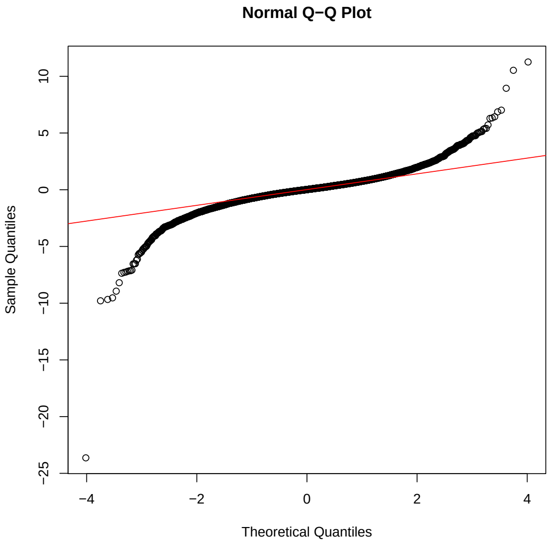


QQ-plot

The idea is to plot in a standard Cartesian reference graph:

- the quantiles of the series under consideration, X_t,against the quantiles of any given distribution. If the returns were truly normal, then the graph should look like a straight line with a 45-degree angle.
  - first, sort all (standardized) returns in ascending order, and call the ith sorted value x_i;
  - second, compute the empirical probability of getting a value below the actual as $(i - 0.5) / T$, where T is number of observations available in the sample.
  - Finally, we calculate the quantiles of the benchmark distribution quantiles as $\Phi^{-1}((i - 0.5) / T)$, where $\Phi^{-1}(\cdot)$ denotes the inverse of the benchmark density.
  - Represent on a scatter plot the (standardized) returns and sort the data on the Y-axis against the standard disribution quantiles on the X-axis.

QQ-plot



Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
−5
−10
−15
−20
−25
−4
−2
0
2
4

GARCH

A parsimonious model capable of capturing all the features of high-frequency returns described in the previous section:

$$
\begin{aligned}
R_{t+1} &= \mu_t + \sigma_{t+1} z_{t+1} \quad z_{t+1} \sim \text{IID } \mathcal{N}(0,1), \\
\sigma_{t+1}^2 &= \omega + \alpha \left(R_t - \mu_t\right)^2 + \beta \sigma_t^2 \\
\alpha + \beta &< 1
\end{aligned}
$$

where returns have a constant mean (that is usually zero) and a time varying GARCH(1,1) structure.
In a model like this the innovation $\epsilon_t \equiv \sigma_t z_t$ has zero mean and is serially uncorrelated at all lags $j \geqslant 1$.

GARCH Properties

R_{t+1} has a finite unconditional long-run variance of $\frac{\omega}{1-\alpha-\beta}$

$$
\begin{aligned}
\sigma^2 &= E\left(\sigma_{t+1}^2\right) = \omega + \alpha E\left(R_t - \mu\right)^2 + \beta\sigma^2 \\
&= \omega + \alpha\sigma^2 + \beta\sigma^2 \\
&= \frac{\omega}{1-\alpha-\beta}
\end{aligned}
$$

Substituting ω out of the GARCH expression:

$$
\begin{aligned}
\sigma_{t+1}^2 &= \left(1-\alpha-\beta\right)\sigma^2 + \alpha R_t^2 + \beta\sigma_t^2 \\
&= \sigma^2 + \alpha\left(\left(R_t-\mu\right)^2 - \sigma^2\right) + \beta\left(\sigma_t^2 - \sigma^2\right)
\end{aligned}
$$

which illustrates the relation between predicted variance and long-run variance in a GARCH model.

GARCH Properties

A GARCH(1,1) model can be considered as the equivalent of an ARMA(1,1) model for the variance, More generally, in the ARMA(q,p) case, we have:

$$\sigma_{t+1}^2 = \omega + \sum_{i=1}^{q} \alpha_i \left(R_{t+1-i}^2 - \mu\right)^2 + \sum_{j=1}^{p} \beta_j \sigma_{t+1-j}^2. \quad (1)$$

$$\bar{\sigma}^2 = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j} \quad (2)$$

Because unconditional variance exists only if $\bar{\sigma}^2 > 0$, the equation above implies that when $\omega > 0$, the condition

$$1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j > 0 \Longrightarrow \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$$

must hold. When the long-run variance of a GARCH process exists, we say that the GARCH process is stationary and we refer to the condition $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ as a stationarity condition.

Testing for GARCH

A (Lagrange multiplier) test for (G)ARCH in returns/disturbances has been proposed by Engle (1982). The methodology involves the following two steps:

- First, use simple OLS to estimate the most appropriate regression equation or ARMA model on asset returns and let $\{\hat{z}_t^2\}$ denote the squares of the standardized returns (residuals), for instance coming from a homoskedastic model, $\hat{z}_t^2 = R_t^2/\hat{\sigma}^2$;
- Second, regress these squared residuals on a constant and on q lagged values $\hat{z}_{t-1}^2$, $\hat{z}_{t+2}^2$, ..., $\hat{z}_{t-q}^2$ (e_t is a white noise shock):

$$\hat{z}_t^2 = \xi_0 + \xi_1 \hat{z}_{t-1}^2 + \xi_2 \hat{z}_{t-2}^2 + ... + \xi_q \hat{z}_{t-q}^2 + e_t.$$

If there are no ARCH effects, the estimated values of ξ_1 through ξ_q should be zero, $\xi_1 = \xi_2 = ... = \xi_q$.

GARCH Forecasting

$$
\begin{aligned}
\sigma^2_{t+1|t} &= \bar{\sigma}^2 + \alpha \left[(R_t - \mu_t)^2 - \bar{\sigma}^2 \right] + \beta \left(\sigma^2_t - \bar{\sigma}^2 \right), \\
\sigma^2_{t+2|t} &= \bar{\sigma}^2 + (\alpha + \beta)\, \sigma^2_{t+1|t} \\
\sigma^2_{t+n+1|t} &= \bar{\sigma}^2 + (\alpha + \beta)^n\, \sigma^2_{t+1|t}
\end{aligned}
$$

MLE Estimation

The assumption of IID normal shocks (z_t),

$$R_{t+1} = \sigma_{t+1} z_{t+1} \qquad z_{t+1} \sim \text{IID}\, \mathcal{N}(0,1),$$

implies (from normality and identical distribution of z_{t+1}) that the density of the time t observation is:

$$l_t \equiv \Pr(R_t; \boldsymbol{\theta}) = \frac{1}{\sigma_t(\boldsymbol{\theta})\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{R_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right),$$

where the notation $\sigma_t^2(\boldsymbol{\theta})$ emphasizes that conditional variance depends on $\boldsymbol{\theta} \in \Theta$.

MLE Estimation

Because each shock is independent of the others (from independence over time of z_{t+1}), the total probability density function (PDF) of the entire sample is then the product of T such densities:

$$L(R_1, R_2, ..., R_T; \boldsymbol{\theta}) \equiv \prod_{t=1}^{T} l_t = \prod_{t=1}^{T} \frac{1}{\sigma_t(\boldsymbol{\theta})\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{R_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right).$$

taking logs

$$\mathcal{L}(R_1, R_2, ..., R_T; \boldsymbol{\theta}) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log\sigma_t^2(\boldsymbol{\theta}) - \frac{1}{2}\sum_{t=1}^{T}\frac{R_t^2}{\sigma_t^2(\boldsymbol{\theta})}$$

MLE Estimation

Substituting an expression for $\sigma_t^2(\boldsymbol{\theta})$ (given by the chosen GARCH specification) given the observations on the returns and given an initial observation for variance

$$\begin{aligned}
\mathcal{L}(R_1, R_2, ..., R_T; \boldsymbol{\theta}) &= -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log\left[\omega + \alpha R_{t-1}^2 + \beta\sigma_{t-1}^2\right] \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}\frac{R_t^2}{\omega + \alpha R_{t-1}^2 + \beta\sigma_{t-1}^2}, \\
\sigma_0^2 &= \frac{\omega}{1-\alpha-\beta}
\end{aligned}$$

maximizing the log-likelihood to select the unknown parameters will deliver the MLE, denoted as $\hat{\boldsymbol{\theta}}_T^{ML}$

Quasi MLE Estimation

- The QMLE result says that we can still use MLE estimation *based on normality assumptions* even when the shocks are not normally distributed, if our choices of conditional mean and variance function are defendable, at least in empirical terms (i.e. conditional mean and conditional variance are correctly specified).
- However, because the maintained model still has that $R_{t+1} = \sigma_{t+1} z_{t+1}$ with $z_{t+1} \sim \text{IID } \mathcal{D}(0,1)$, the shocks will have to be anyway IID: you can just do without normality, but the convenience of $z_{t+1} \sim \text{IID } \mathcal{D}(0,1)$ needs to be preserved.

An example

Because we know that the long-run (ergodic) variance from a GARCH(1,1) is $\bar{\sigma}^2 = \omega/(1-\alpha-\beta)$, instead of jointly estimating ω, α, and β, you simply set

$$\tilde{\omega} = (1-\alpha-\beta)\left[\frac{1}{T}\sum_{t=1}^{T} R_t^2\right]$$

for whatever values of α and β. Note that (i) you impose the long-run variance estimate on the GARCH model directly and avoid that the model may yield nonsensical estimates;(ii) you have reduced the number of parameters to be estimated in the model by one. These benefits must be carefully contrasted with the well-known costs, the loss of efficiency caused by QMLE.

Threshold GARCH

GARCH models can be extended A number of empirical papers have emphasized that for many assets and sample periods, a negative return increases conditional variance by more than a positive return of the same magnitude does, the so-called *leverage effect*.
A way of capturing the leverage effect is to directly build a model that exploits the possibility to define an indicator variable, I_t , to take on the value 1 if on day t the return is negative and zero otherwise. For concreteness, in the simple (1,1) case, variance dynamics can now be specified as:

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \alpha\theta I_t R_t^2 + \beta\sigma_t^2 \qquad I_t \equiv \begin{cases} 1 & \text{if } R_t < 0 \\ 0 & \text{if } R_t \geq 0 \end{cases} \quad \text{or}$$

$$\sigma_{t+1}^2 = \begin{cases} \omega + \alpha(1+\theta)R_t^2 + \beta\sigma_t^2 & \text{if } R_t < 0 \\ \omega + \alpha R_t^2 + \beta\sigma_t^2 & \text{if } R_t \geq 0 \end{cases}.$$

A $\theta > 0$ will capture the leverage effect.

Threshold GARCH

This model is sometimes referred to as the GJR-GARCH model—from Glosten, Jagannathan, and Runkle's (1993) paper—or threshold GARCH (TGARCH) model.
In this model, because when 50% of the shocks are assumed to be negative and the other 50% positive, so that $E[I_t] = 1/2$, the long-run variance equals:

$$
\begin{aligned}
\bar{\sigma}^2 &\equiv E[\sigma_{t+1}^2] = \omega + \alpha E[R_t^2] + \alpha\theta E[I_t R_t^2] + \beta E[\sigma_t^2] \\
&= \omega + \alpha\bar{\sigma}^2 + \alpha\theta E[I_t]\bar{\sigma}^2 + \beta\bar{\sigma}^2 \\
&= \omega + \alpha\bar{\sigma}^2 + \frac{1}{2}\alpha\theta\bar{\sigma}^2 + \beta\bar{\sigma}^2 \Longrightarrow \bar{\sigma}^2 = \frac{\omega}{1 - \alpha(1 + 0.5\theta) - \beta}.
\end{aligned}
$$

Visibly, in this case the persistence index is $\alpha(1 + 0.5\theta) + \beta$.

From GARCH to VaR

After estimation a GARCH model can be simulated using bootstrap or Monte-Carlo to derive the distribution of returns and the relevant VaR

$$
\begin{aligned}
R_{t+1} &= \mu + \sigma_{t+1} z_{t+1} \quad z_{t+1} \sim \text{IID}\, \mathcal{N}(0,1), \\
\sigma_{t+1}^2 &= \omega + \alpha \left(R_t - \mu_t\right)^2 + \beta \sigma_t^2 \\
\alpha + \beta &< 1
\end{aligned}
$$

Given estimation, derive $\hat{z}_t = \frac{R_t}{\hat{\sigma}_t}$. At time t you can now predict σ_{t+1}^2 and the distribution of R_{t+1} can now be simulated via the preferred method.

Recursion can then be applied to derive the distribution of R_{t+n} with n > 1.

Backtesting VaR

How do we test the validity of a VaR model ? The relevant evidence to judge a VaR model are violations:

$$Min(R_{t+1} - VaR^{p}_{t+1}, 0)$$

(a) A good VaR model should not feature neither too few nor too many violations.
(b) We have to few violations when a VaR at the confidence level of alpha shows less then 100*alpha violations in a sample of 100 observations. In this case the VaR model is too conservative.
(c) when we have violations there are two interesting aspects of that: their number and their timing. A five per cent VaR that feature 5 violations in five successive periods cannot be taken as a valid VaR model as violations are not independent. Clustering of violations is a problem that should lead to reject specific VaR models.

Unconditional Coverage Testing

Given a time-series of VaR and observed returns the "hit sequence" of VaR violations is defined as follows:

$$\begin{aligned} I_{t+1} &= 1, \textit{if } R_{t+1} > VaR_{t+1}^{p} \\ I_{t+1} &= 0, \textit{if } R_{t+1} < VaR_{t+1}^{p} \end{aligned}$$

In the null hypothesis that the VaR is a valid model violations should not be predictable: the forecast of the probability of a VaR violation should be p every day. The hit sequence in this case should be distributed over time as a Bernoulli variable that takes the value 1 with probability p and the value 0 with probability $1-p$. So

$$\begin{aligned} H_0 &: I_{t+1} \sim i.i.d. \text{ Bernoulli}\,(p) \\ f\,(I,p) &= (1-p)^{1-I_{t+1}}\, p^{I_{t+1}} \end{aligned}$$

Unconditional Coverage Testing

The first test of validity of a VaR is therefore constructed as follows. Take a Bernoulli distribution (I_{t+1}, x) for the that the number of violations, derive a maximum likelihood estimator $\hat{x}$ of x, and test using a likelihood ratio test that $\hat{x}$ is not statistically different from p.

$$\begin{aligned} L\left(I_{t+1}, x\right) &= \prod_{i=1}^{T} (1-x)^{1-I_{t+i}} \, x^{I_{t+i}} \\ &= (1-x)^{T_0} \, x^{T_1} \end{aligned}$$

where T_1 is the number of violations of the VaR observed in the sample, and $T_0 = T - T_1$.
The maximum likelihood estimator $\hat{x} = \frac{T_1}{T}$.

Unconditional Coverage Testing

A likelihood ratio test of the null hypothesis $\hat{x} = p$, can then be constructed as follows:

$$LR_{uc} = -2\ln\left[\frac{L(p)}{L(\hat{x})}\right]$$

which is distributed as a χ^2 with one degree of freedom.
Note that usually the number of violations and the number of observations available will not be large, so rather than relying upon the χ^2 distribution, it is adviceble to use Monte-Carlo simulations to build the relevant distribution to conduct the test. In this case the simulated P-values would be obtained by drawing an artificial sample of the relevant size from the null, and using as a P-value the share of simulated test that are larger than the observed one.

Independence Testing

We concentrate now on a test able to reject a VaR with clustered violations. In this case the hit sequence is dependent over time and its evolution over time can be described by a so called Markov sequence where the transition from the relevant states (violation and no violation) can be described by the following transition probability matrix

$$X_1 = \begin{bmatrix} x_{00} & 1 - x_{00} \\ 1 - x_{11} & x_{11} \end{bmatrix}$$

where:

$$\begin{aligned} x_{00} &= \Pr\left(I_{t+1} = 0 \mid I_t = 0\right) \\ 1 - x_{00} &= \Pr\left(I_{t+1} = 1 \mid I_t = 0\right) \\ x_{11} &= \Pr\left(I_{t+1} = 1 \mid I_t = 1\right) \\ 1 - x_{11} &= \Pr\left(I_{t+1} = 0 \mid I_t = 1\right) \end{aligned}$$

Independence Testing

If we observe a sample of T observations the likelihood function of a first order Markov process can be written as follows:

$$L\left(X_1, I_{t+1}\right) = x_{00}^{T_{00}}\left(1 - x_{00}\right)^{T_{01}}\left(1 - x_{11}\right)^{T_{10}} x_{11}^{T_{11}}$$

The maximum likelihood estimates of the relevant parameters are then

$$\begin{aligned} \hat{x}_{00} &= \frac{T_{00}}{T_{00}+T_{01}} \\ \hat{x}_{11} &= \frac{T_{11}}{T_{10}+T_{11}} \end{aligned}$$

and so

$$\hat{X}_1 = \begin{bmatrix} \frac{T_{00}}{T_{00}+T_{01}} & \frac{T_{01}}{T_{00}+T_{01}} \\ \frac{T_{10}}{T_{10}+T_{11}} & \frac{T_{11}}{T_{10}+T_{11}} \end{bmatrix}$$

Independence Testing

Under independence

$$\hat{X}_1^{id} = \begin{bmatrix} 1-\hat{x} & \hat{x} \\ 1-\hat{x} & \hat{x} \end{bmatrix}$$

and therefore the independence hypothesis $(1-\hat{x}_{00}) = \hat{x}_{11}$ can be tested using a likelihood ratio test

$$LR_{ind} = -2\ln\left[\frac{L\left(\hat{X}_1^{id}\right)}{L\left(\hat{X}_1\right)}\right] \sim \chi_1^2$$

Independence Testing

As for the unconditional coverage test small sample problems can be fixed by Monte Carlo simulation of the critical values, moreover samples in which $\mathrm{T}_{11} = 0$ are often observed. In this cases the likelihood function is computed as

$$L\left(X_1, I_{t+1}\right) = x_{00}^{T_{00}}\left(1 - x_{00}\right)^{T_{01}}$$

Conditional Coverage Testing

Having constructed the test for independence we can test jointly the hypothesis of conditional coverage and independence via the following likelihood ratio test:

$$LR_{cc} = -2\ln\left[\frac{L(p)}{L(\hat{X}_1)}\right] \sim \chi^2_2$$

note that

$$LR_{cc} = LR_{uc} + LR_{ind}$$