Quantitative Research & Trading Internship - Long-Short Hedge Fund

Loïc Cantin

SU/ Ecole Polytechnique (Master Probability Finance)

September, 18th 2023

Overview of the Internship

- Quantitative Research in a major multi-strategy American hedge fund in London
- Within a team of a dozen of people with different backgrounds: fundamental approach, quant, data science
- Goal: contribute to optimize the absolute performance of the team with **risk management** and **alpha** generation ideas

Long-Short Strategy

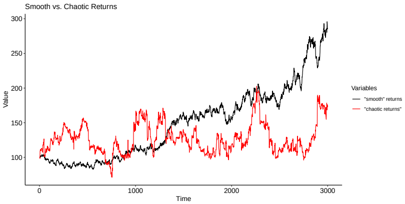
Long-Short funds aim at achieving an optimal **absolute performance** that is decorrelated from the markets (i.e. **beta-neutral**). The goal is to produce a small return with very low volatility to leverage the positions and boost the performance. To do so, they often rely on the use of pair trading.

$\rightarrow$ **General goal**: maximizing returns and limiting maximum drawdowns

Return-Risk Maximization?

Smoothness of daily log-returns is important for long-term performance → we want to limit drawdowns.

$$\boxed{SR = \frac{r_p - r_f}{\sigma_p}}$$



- $\mathbb{E}(\epsilon_{1t}) = \mathbb{E}(\epsilon_{2t}) = 0.04\%$
- $\sigma_1 \approx 0.010$, $\sigma_2 \approx 0.025$
- $\boldsymbol{CAGR_1 = 9.2\%}$, $\boldsymbol{CAGR_2 = 4.7\%}$, with $CAGR = \left(\frac{p_T}{p_1}\right)^{\left(\frac{250}{T}\right)} - 1$

Goal of the Internship

1. Contribute to **Risk Management**: Study the effect of a shock on a major financial series on the portfolio (e.g. oil prices, currency exchange, index etc.) $\rightarrow$ propose a risk measure and an asymptotic theory to measure tail co-movement (implementation of the paper Cantin, Francq, Zakoian (2023)
2. Contribute to improve the absolute performance via **factor analysis** and help **discretionary** anticipations for regime switches
3. Benefit from **mean-reversion** in pair trading strategies

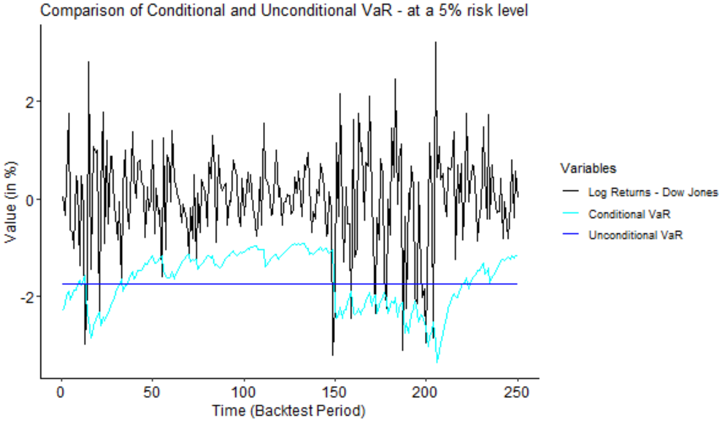
Motivation for the Research

- **Focus**: The first section focuses on **market risk measures** and we will discuss the various risks associated with modelling and estimating the defined measures.
- **Model**: To construct conditional risk measures, we model the log-returns of different financial series using **conditional volatility models** (here GARCH models).
- **Motivation**: Most of the models used by practitioners today face two limitations:
  - Many practitioners still use **unconditional models** (which do not take into account conditional volatility) and which are less efficient (see next slide).
  - Popular risk measures (Value at Risk, Expected Shortfall) do not take into account **co-movements** between assets and are therefore not good indicators for measuring systemic risk.

(Conditional) Value at Risk - Definition

(Conditional) VaR Characterization

$$\mathbb{P}_{t-1}\left[\epsilon_t < -VaR_{t-1}(\alpha)\right] = \alpha \tag{1}$$



(Conditional) Value at Risk / CoVaR - Definitions (1/2)

VaR

$$VaR_{t-1}(\alpha) = -q_\alpha\left(\epsilon_t | \mathcal{F}_{t-1}\right)$$

ES

$$ES_{t-1}(\alpha) = \mathbb{E}_{t-1}\left(\epsilon_t \,|\, \epsilon_t < -VaR_{t-1}(\alpha)\right)$$

(Conditional) Value at Risk / CoVaR - Definitions (2/2)

CoVaR

$$\begin{aligned} CoVaR^{(i|j)}(\alpha, \alpha') &= -inf\{\epsilon_i \in \Omega_{\epsilon_i} : \\ &\qquad F_{\Upsilon_i|\{\epsilon_j < -VaR^{(j)}(\alpha')\}} > \alpha\} \\ &= -F^{-1}_{\Upsilon_i|\{\epsilon_j < -VaR^{(j)}(\alpha')\}}(\alpha) \end{aligned}$$

- With ϵ_i a log-return of the series i in the set of taken values $\Omega_{\epsilon i}$.
- Υ_i is the distribution of the ϵ_{it} for t $\in$ 1, T
- F_{Υ_i} is the cdf of the distribution Υ_i.
- $F^{-1}_{\Upsilon_i}$ is the is the inverse function of F_{Υ_i}.
- α and α' are the risk levels under consideration.

Univariate Model

GARCH(1,1) Conditional Volatility Model

$$\begin{cases} \epsilon_t = \sqrt{h_t}(\theta_0)\,\eta_t \\ h_t(\theta_0) = \omega + \alpha\epsilon_{t-1}^2 + \beta\, h_{t-1}(\theta_0) \end{cases}$$

- $\sqrt{h_t}$ represents the conditional volatility at time t ($\mathcal{F}_{t-1}$ measurable) of the series under consideration.
- η_t is the innovation at time t verifying $\mathbb{E}[\eta_t] = 0$, $\mathbb{V}[\eta_t] = 1$ and the sequence (η_t) is iid, but its distribution is supposed unknown.
- True parameter: $\theta_0 = (\omega_0, \alpha_0, \beta_0)$ / Estimator: $\hat{\theta}_T = (\hat{\omega}_T, \hat{\alpha}_T, \hat{\beta}_T)$, with T the in-sample size.

VaR Estimation via GARCH Model

$$VaR_{t-1}(\alpha) = \sqrt{h_t(\theta_0)}\,\xi_\alpha,\ with\ -\xi_\alpha\ the\ \alpha-quantile\ of\ the\ \eta_t$$

$$\widehat{VaR}_{t-1}(\alpha) = \sqrt{h_t(\hat{\theta}_T)}\,\hat{\xi}_\alpha,\ with\ -\hat{\xi}_\alpha\ the\ empirical\ quantile$$

GARCH Model - QML Estimation

To understand the problems involved in estimating the stochastic volatility model, it may be appropriate to compare it with the **QML estimation** of a GARCH [1] model. It has been proved that:

QML Estimator of GARCH(p,q) process

$$\hat{\theta}_T = argmax_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{n} \left(log(h_t(\theta)) + \frac{\epsilon_t^2}{h_t(\theta)} \right)$$
$$= argmax_{\theta \in \Theta} log(\mathcal{L}_T(\theta))$$

- With $h_t(\theta) = \omega + \sum_{i=1}^{p} \alpha_i \, \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \, h_{t-j}(\theta)$
- $\boldsymbol{\theta_0} = (\omega, \boldsymbol{\alpha}, \boldsymbol{\beta})'$, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_q)$

is a CAN estimator of $\boldsymbol{\theta_0}$.

[1] same notations as in the previous definition

Multivariate CCC-GARCH Model

Definition CCC(m)-GARCH(p,q)

$$\begin{cases} \underline{\epsilon_t} & = H_t^{1/2} \underline{\eta_t} \\ H_t & = D_t R_0 D_t \\ \underline{h_t} & = \underline{\omega_0} + \sum_{i=1}^{q} \mathbf{A}_{0i} \underline{\epsilon}_{t-i}^2 + \sum_{j=1}^{p} \mathbf{B}_{0j} \underline{h}_{t-j} \end{cases}$$

- $D_t = diag\left(\sqrt{h_{11,t}}, ..., \sqrt{h_{mm,t}}\right)$
- $\underline{\epsilon}_t = (\epsilon_{1t}, ..., \epsilon_{mt})', \mathbb{V}[\eta_t] = Id_m$
- $\underline{h}_t = (h_{11t}, ..., h_{mmt})'$ and $\sqrt{h_{iit}}$ is the conditional volatility of series i.
- R_0 is a correlation matrix of the m assets, such that: $R_0 = (\rho_{ij})_{(i,j)\in 1,m^2}$, with $\rho_{ii} = 1$ and $\rho_{ij} \leq 1$ for (i,j) $\in 1, m^2$.
- $\mathbf{A}_{0i} = (\alpha_{k,j})_{(k,j)\in 1,m^2} \in \mathcal{M}_m(\mathbb{R})$, $\mathbf{B}_{0i} = (\beta_{k,j})_{(k,j)\in 1,m^2} \in \mathcal{M}_m(\mathbb{R})$ is assumed diagonal to allow equation by equation estimation.

But can we use the same estimation approach as for VaR?

- Recall that to estimate VaR we used the fact that - with our GARCH-type model - it could be written under the multiplicative form: $VaR_{t-1}(\alpha) = \sqrt{h_t(\theta_0)}\,\xi_\alpha$
- However with a "standard" CCC-GARCH model written under its traditional form (as on previous slide) this is not the case:

Normal Bivariate CCC-GARCH

$$\begin{cases} \epsilon_{1t} &= \sqrt{h_{11,t}(\theta_0)}\,.\,\eta_{1t} \\ \epsilon_{2t} &= \rho_{1,2}\,.\,\sqrt{h_{22,t}(\theta_0)}\,.\,\eta_{1t} + \sqrt{(1-\rho_{1,2}^2)\,h_{22,t}(\theta_0)}\,.\,\eta_{2t} \end{cases}$$

We cannot write $\epsilon_{2t} = K_t\,.\,\eta_{2t}$

- We showed that any CCC-GARCH model can be written under a "diagonal" form (defined next slide).

CCC-GARCH Model under its "Diagonal" Form - and equation-by-equation estimation

Definition (Diagonal Form of the CCC-GARCH Model)

$$\begin{cases} \underline{\epsilon_t} & = D_t \underline{\tilde{\eta}}_t \\ \underline{h}_t & = \underline{\omega}_0 + \sum_{i=1}^{q} \mathbf{A}_{0i} \underline{\epsilon}_{t-i}^2 + \sum_{j=1}^{p} \mathbf{B}_{0j} \underline{h}_{t-j} \end{cases}$$

- $\underline{\epsilon}_t = (\epsilon_{1t}, ..., \epsilon_{mt})'$, $\underline{\tilde{\eta}}_t = R^* \, \underline{\eta}_t$.
- $R^{*2} = \mathbb{V}[\tilde{\eta}_t]$ is a correlation matrix.

Bivariate CCC-GARCH Model used to Estimate the CoVaR

$$\begin{cases} \epsilon_{1t} & = \sqrt{h_{11,t}(\theta_0)} \, . \, \tilde{\eta}_{1t} \\ \epsilon_{2t} & = \sqrt{h_{22,t}(\theta_0)} \, . \, \tilde{\eta}_{2t} \end{cases}$$

Conditional Value at Risk (CoVaR)

CoVaR - Characterisation (GARCH Model)

$$\begin{aligned}
&\mathbb{P}_{t-1}\left[\,\epsilon_{it} < -CoVaR^{(i|j)}_{t-1}(\alpha,\alpha')\,|\,\epsilon_{jt} < -VaR^{(j)}_{t-1}(\alpha')\,\right] \qquad (2)\\
&= \mathbb{P}_{t-1}\left[\,\epsilon_{it} < -CoVaR^{(i|j)}_{t-1}(\alpha,\alpha')\,|\,\tilde{\eta}_{jt} < -\xi^{(j)}_{\alpha'}\,\right]\\
&= \mathbb{P}_{t-1}\left[\,\epsilon_{it} < -\sqrt{h_{ii,t}(\theta_0)}\,u^{(i|j)}(\alpha,\alpha')\,|\,\tilde{\eta}_{jt} < -\xi^{(j)}_{\alpha'}\,\right]\\
&= \mathbb{P}\left[\,\tilde{\eta}_{it} < -u^{(i|j)}(\alpha,\alpha')\,|\,\tilde{\eta}_{jt} < -\xi^{(j)}_{\alpha'}\,\right]
\end{aligned}$$

CoVaR Estimation

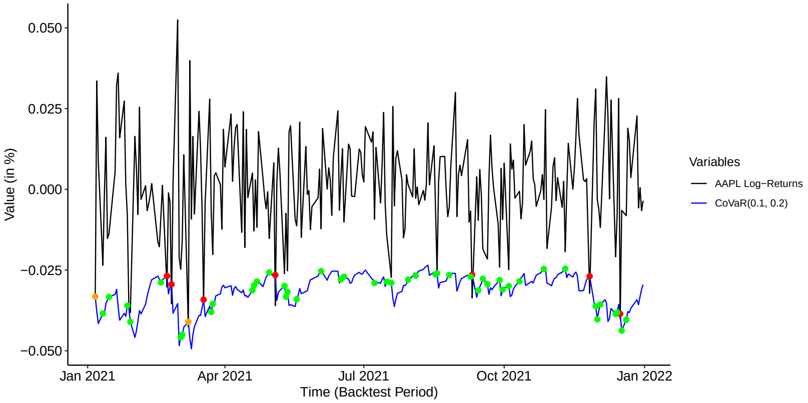
$$CoVaR^{(i|j)}_{t-1}(\alpha,\alpha') = \sqrt{h_{ii,t}(\theta_0)}\,u^{(i|j)}(\alpha,\alpha')$$

$$\widehat{CoVaR}^{(i|j)}_{t-1}(\alpha,\alpha') = \sqrt{h_{ii,t}(\hat{\theta}_T)}\,\hat{u}^{(i|j)}(\alpha,\alpha')$$

The function u(.,.) can be seen as a conditional quantile function, except that this function is conditioned by an inequality.

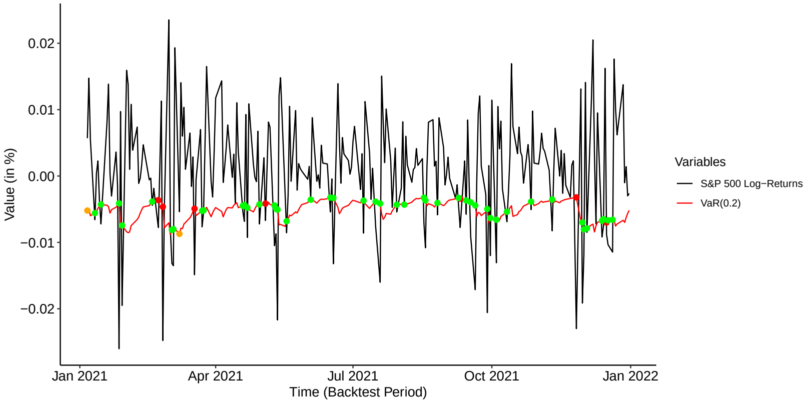
VaR Violations vs. CoVaR violations (1/2)

Figure: Apple CoVaR (5%,10%) vs. VaR(5%) of S&P 500 log-returns



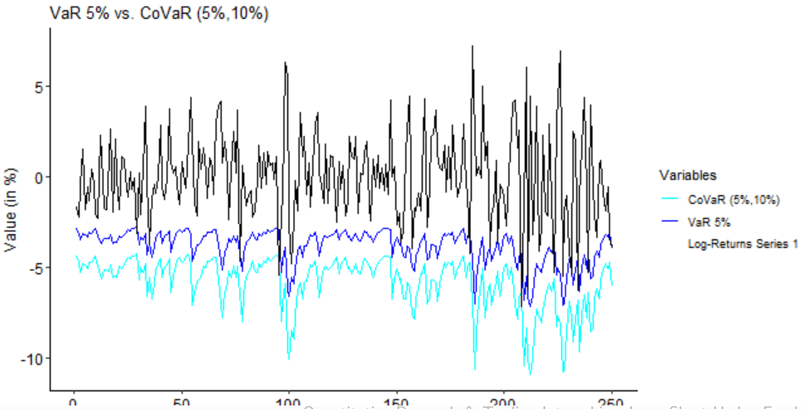
VaR Violations vs. CoVaR violations (2/2)

Figure: VaR(5%) of S&P 500 log-returns



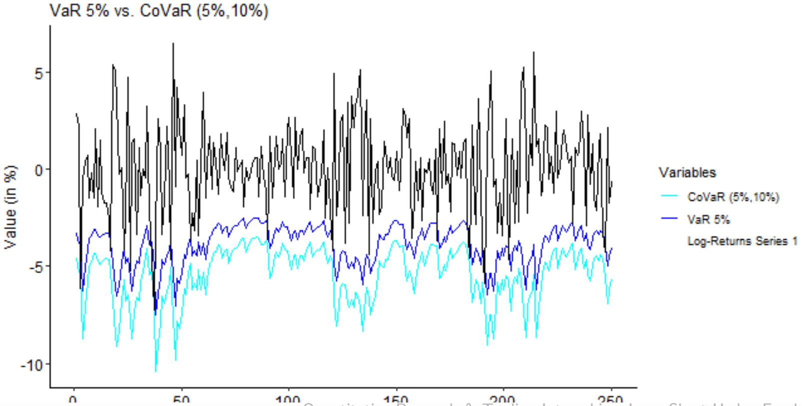
CoVaR Sensitivity to Different Correlation Matrices (1/5)

Figure: CoVaR vs. VaR with $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ as correlation matrix



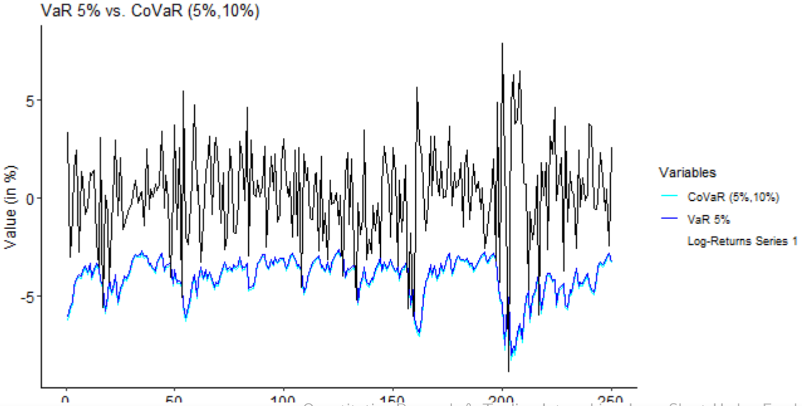
CoVaR Sensitivity to Different Correlation Matrices (2/5)

Figure: CoVaR vs. VaR with $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ as correlation matrix



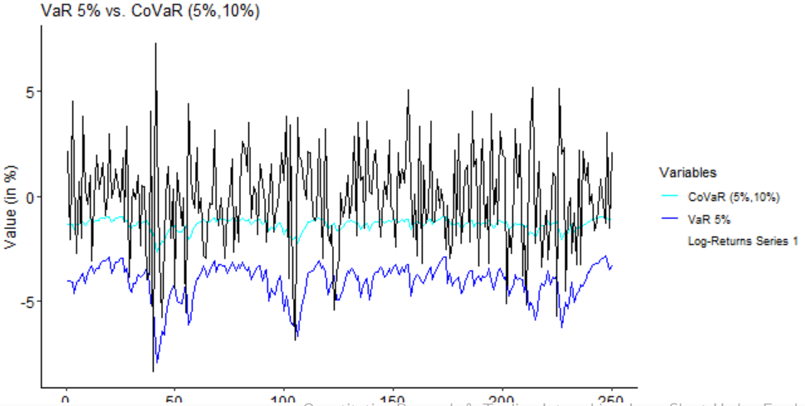
CoVaR Sensitivity to Different Correlation Matrices (3/5)

Figure: CoVaR vs. VaR with $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ as correlation matrix



CoVaR Sensitivity to Different Correlation Matrices (4/5)

Figure: CoVaR vs. VaR with $\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ as correlation matrix



Tests based on Violations

- We can divide the main existing tests in 3 categories:
  - The tests based on the **number of violations** and their **random** appearance.
  - The tests based on **multiple risk levels** (see Multivariate Portmanteau Test from Hurlin (2007)
  - The tests based on the **distance between the VaR and the log-returns** (see what we call the "α-criterion").
- In this presentation we will focus on the most traditional Kupiec, Christoffersen tests.

Hit Variable

$$Hit_t(\alpha) = \mathbb{1}\{\epsilon_t < -VaR_{t-1}(\alpha)\}$$

- Christoffersen has established 3 different tests:
  - The **unconditional coverage** test, testing the consistency between the frequency of violations and the theoretical value α.
  - The **independence** test, testing the random appearance of the violations.

Christoffersen's Tests

- **Unconditional Coverage Test**:
  - Null hypothesis: $H_0^{UC} : \mathbb{P}[Hit_t = 1] = \alpha$
  - Test Statistics: $LR_{UC} = 2log \frac{\pi_{exp}^{n_1} (1-\pi_{exp})^{n_0}}{\pi_{obs}^{n_1} (1-\pi_{obs})^{n_0}}$
- **Independence Test**:
  - Null Hypothesis: $H_0^{Ind} : \mathbb{P}[Hit_t = 1 | Hit_{t-1} = 0] = \mathbb{P}[Hit_t = 1 | Hit_{t-1} = 1]$
  - Test Statistics: $LR_{Ind} = 2log \frac{\pi_{obs}^{n_1} (1-\pi_{obs})^{n_0}}{\pi_{01}^{n_{01}} (1-\pi_{01})^{n_{00}} \pi_{11}^{n_{11}} (1-\pi_{11})^{n_{10}}}$
- **Conditional Coverage Test**:
  - Test Statistics: $LR_{cc} = 2log \frac{\pi_{exp}^{n_1} (1-\pi_{exp})^{n_0}}{\pi_{01}^{n_{01}} (1-\pi_{01})^{n_{00}} \pi_{11}^{n_{11}} (1-\pi_{11})^{n_{10}}}$
- π_{exp} is the expected proportion of violations.
- π_{obs} is the observed proportion of violations.
- n_1 is the number of violations and $n_0 = n - n_1$ is the sample size.

"α - Criterion" for VaR Models Comparison

Definition ("α - Comparison Criterion")

$\mathbb{E}[(1-\alpha).[r-q]^{-}+\alpha.[r-q]^{+}]$

Definition ("Empirical α Comparison Criterion")

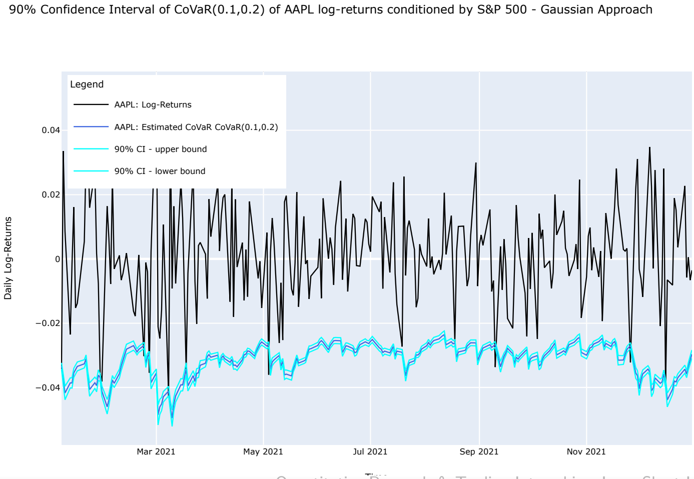
$T^{-1}.\sum_{t=1}^{T}(1-\alpha).[r_t-VaR_t]^{-}+\alpha[r_t-VaR_t]^{+}$

Confidence Interval for Conditional VaR Model

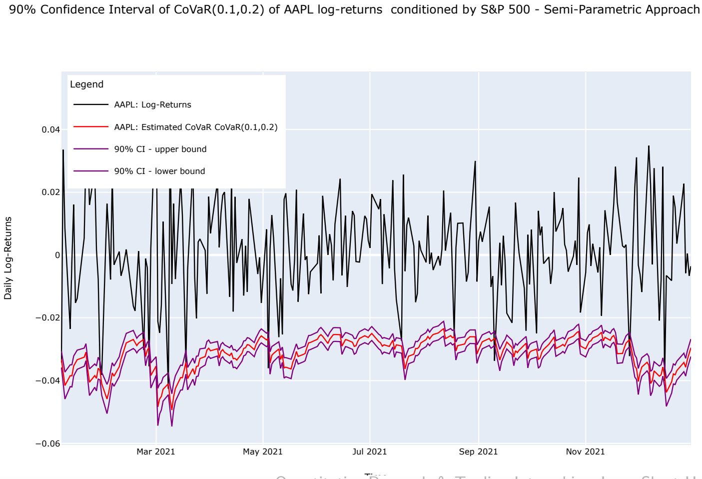
Asymptotic Convergence - With Estimation Risk

- $\sqrt{T}\begin{bmatrix}(\hat{\theta}_{1,T}-\theta_{1,0})\\(\hat{u}_T-u_{\alpha,\alpha'})\end{bmatrix} \sim \mathcal{N}(0,\Sigma(\alpha,\alpha'))$
    - $CoVaR_t(\alpha,\alpha') = \sigma_t(\theta_{1,0})\, u_{\alpha,\alpha'} \; = \; F_{t-1}(\theta_{1,0}, u_{\alpha,\alpha'})$
    - $\widehat{CoVaR}_t(\alpha,\alpha') = \hat{\sigma}_t(\hat{\theta}_{1,T})\, \hat{u}_T \; = \; \widehat{F_{t-1}}(\hat{\theta}_{1,T}, \hat{u}_T)$
- $\sqrt{T}\,[\widehat{F_{t-1}}(\hat{\theta}_{1,T}, \hat{u}_T) - F_{t-1}(\theta_{1,0}, u_{\alpha,\alpha'})] \sim \mathcal{N}(0, c'_{t-1}\Sigma(\alpha,\alpha')c_{t-1})$
- $CoVaR_t(\alpha,\alpha') \simeq \underbrace{\widehat{F_{t-1}}(\hat{\theta}_{1,T}, \hat{u}_T)}_{\widehat{CoVaR}_t(\alpha,\alpha')} \pm \frac{1.96}{\sqrt{T}}\,(c'_{t-1}\widehat{\Sigma(\alpha,\alpha')}c_{t-1})^{1/2}$
    - with $c_{t-1} := \frac{\partial F_{t-1}(\theta_{1,0}, u_{\alpha,\alpha'})}{\partial(\theta_{1,0}, u_{\alpha,\alpha'})}$

95% Confidence Interval vs Real VaR - Simulated Data - An Example



95% Confidence Interval vs Real VaR - Simulated Data - An Example



Full Bootstrap Approach

2. We decompose the sample of size N=3,250 in **two consecutive sub-samples**:
   - The **estimation sample**: of size T=3,000 observations on which we estimate the parameters $\theta_0^{(1)}$ and $\theta_0^{(2)}$ as well as $\xi_{\alpha'}^{(2)}$ and $u(\alpha, \alpha')$.
   - The **"backtest sample"**: of size n=250 on which we will consider the confidence interval via the "bootstrap" approach.

3. Once the model is calibrated, we can simulate S=80 "boostrapped CoVaR" at each time of the backtest period, i.e. we get S different $\widehat{CoVaR}^{(s)}$ processes of size n. The detailed method is as follows:

Method (2/3)

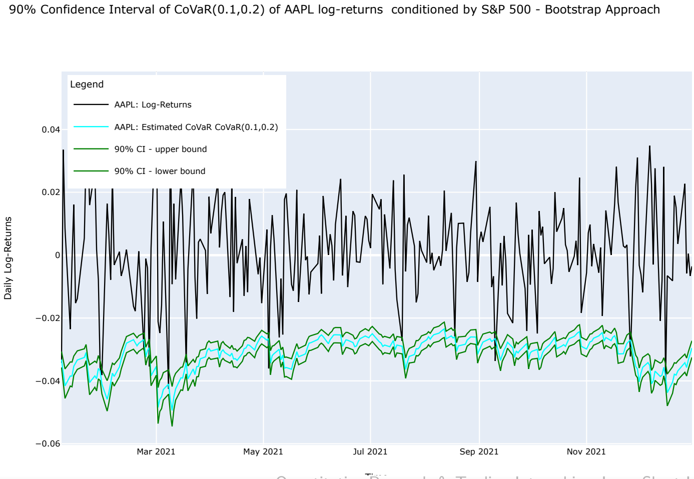
- We draw S different vectors of couple of residuals $(\hat{\eta}_{1t}, \hat{\eta}_{2t})^{(s)}_{1 \leq t \leq T}$ among the T couples obtained after the estimation of the model, **with replacement** (i.e. we can have several times the same couple).
- We re-simulate S different GARCH models using, for each $s \in S$, the vector of couple of residuals $(\hat{\eta}_{1t}, \hat{\eta}_{2t})^{(s)}_{1 \leq t \leq T}$ as innovations and the parameter $\widehat{\theta_T}$.
- For each $s \in S$, we re-estimate the newly-simulated model by QML and obtain estimated parameters $\widehat{\theta^*_T}^{(s)}$ as well as a new vector of couple of residuals $(\widehat{\eta}^*_{1t}, \widehat{\eta}^*_{2t})^{(s)}_{1 \leq t \leq T}$.
- We then compute $\widehat{\xi^*_{\alpha'}}^{(2,s)}$ and $\widehat{u^{*(s)}}(\alpha, \alpha')$ on these new residuals.

Method (3/3)

- We use $\widehat{\theta_T^*}^{(s)}$ and the **observed** ϵ_t to compute recursively $\widetilde{\sigma}_t(\widehat{\theta_{1,T}^*}^{(s)})$ and the residuals $(\widetilde{\eta}_{1t}^*, \widetilde{\eta}_{2t}^*)_{1\leq t\leq T}^{(s)}$ such that $\widetilde{\eta}_{it}^{*(s)} := \epsilon_{it}/\widetilde{\sigma}_t(\widehat{\theta_{i,T}^*}^{(s)})$ for $i \in 1, 2$.
- We then compute $\widetilde{\xi_{\alpha'}^*}^{(2,s)}$ and $\widetilde{u^{*(s)}}(\alpha, \alpha')$ on these new residuals.
- Eventually, we get $\widehat{CoVaR}_t^{(s)} = \widetilde{\sigma}_t(\widehat{\theta_{1,T}^*}^{(s)})\ \widehat{u^{*(s)}}(\alpha, \alpha')$.

4. We then create a 95% confidence interval:
$$I = \left\{ q_{2.5\%}\left(\widehat{CoVaR}_t^{(s)}, s \in [1, S]\right), q_{97.5\%}\left(\widehat{CoVaR}_t^{(s)}, s \in [1, S]\right)\right\}$$
5. Eventually, we check if the true $CoVaR_0$ violates on average (MC exp.) the confidence region with a frequency of 5%.

90% Confidence Interval vs Real CoVaR - Simulated Data - Bootstrap



Portfolio VaR - Model

- In order to consider VaR for portfolio we slightly modified the model in order to allow for positive (or negative) trends. This is mainly used to find optimal weights when we consider dynamic portfolio rebalancing.

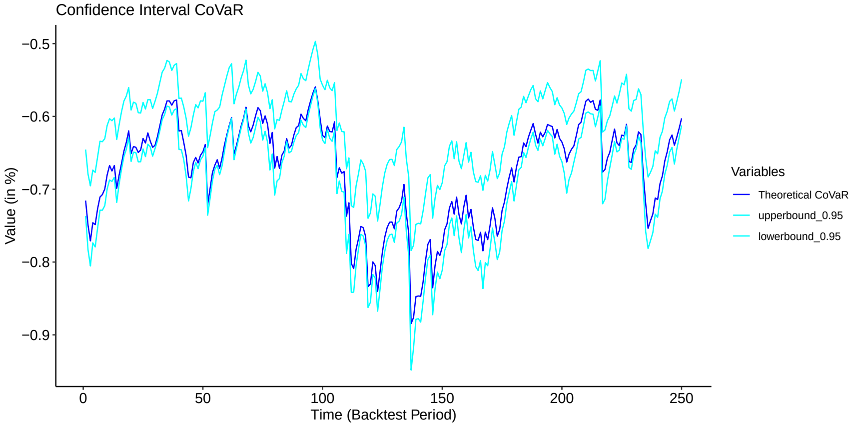
AR(1)-CCC(m)-GARCH(1,1) Model

$$\underline{r_t} = \underline{\mu} + \underline{\delta} + \underline{\psi}\,\underline{r_{t-1}} + \underline{\epsilon_t}$$

- $\underline{r_t} = (r_{1t}, ..., r_{mt})$ represents the log-returns fo the different series.
- $\underline{\mu} = (\mu_1, ..., \mu_m)$ is the vector of the mean of the different series.
- $\underline{\delta} = (\delta_1, ..., \delta_m)$ is the vector of the intercepts of the AR process (of the demeaned returns).
- $\underline{\psi} = (\psi_1, ..., \psi_m)$ is the vector of the auto-regressive coefficients.
- $\underline{\epsilon_t} = (\epsilon_{1t}, ..., \epsilon_{mt})$ is the vector of residuals following a CCC-GARCH model.

Illustration Confidence Interval

True CoVaR & 95% Confidence Interval



Filtered Historical Simulation (FHS) Approach

- The idea of the FHS approach is to simulate a large number of log-returns $\underline{\epsilon}_t^{(s)}$ (that we call a scenario) for time t at time t-1, by drawing innovation vectors $\underline{\eta}_t^{(s)}$ among the observed in-sample innovation vectors $\underline{\eta}_t^{(s)}$ for $s \in 1, T$ or by using all of them.

- We then take the opposite of the α-quantile of the $\underline{\epsilon}_t^{(s)}$ for $s \in 1, T$ as the $VaR^{(\alpha)}_{FHS,t-1}(\epsilon_t^{(p)})$.

(Conditional) VaR of Portfolio Return - FHS

$$VaR^{(\alpha)}_{FHS,t-1}(\epsilon_t^{(p)}) = -q_\alpha(\{\underline{w}_t\, D_t(\theta_0)\, \underline{\tilde{\eta}}_t^{(s)}\}, 1 \leq s \leq T)$$

$$\widehat{VaR}^{(\alpha)}_{FHS,t-1}(\epsilon_t^{(p)}) = -q_\alpha(\{\underline{w}_t\, D_t(\hat{\theta}_T)\, \underline{\hat{\tilde{\eta}}}_t^{(s)}\}, 1 \leq s \leq T)$$

- $\epsilon_t^{(p)}$ refers to the log-returns of the portfolio at time t with weights chosen at time t-1.

Filtered Historical Simulation - Multivariate Method

- **Step 1: Generate the Dataset**
  We consider daily log-returns for an in-sample estimation period of size T = 3,000 on which we calibrate our models (parameter estimation by QMLE) and we keep 500 out-of-sample observations for backtesting.

- **Step 2: Estimate the Parameter of the Model**
  We use a **AR(1)-CCC(m)-GARCH(1,1)** model that we will estimate by Gaussian QMLE.

- **Step 3 :Correlation Matrix Estimation**
  With our CCC-GARCH model under its diagonal form we can retrieve the correlation matrix by taking the empirical Covariance matrix of the $(\tilde{\eta}_t)$.

Filtered Historical Simulation - Multivariate Method

- **Step 4: Choose Dynamic Weights for each Asset at each instant t.**
  We will re-balance the portfolio at the end of day t-1 for day t, considering the optimal weights of the GMVP. To find the coefficient $\underline{w}_{t-1}$ we consider the expected volatility $\hat{D}_t$ estimated with the AR-CCC-GARCH model.

- **Step 5: Estimate the VaR**

$$VaR^{(\alpha)}_{FHS,t-1}(\epsilon^{(p)}_t) = -q_\alpha(\{\underline{w}_t\, D_t(\theta_0)\, \underline{\tilde{\eta}}^{(s)}_t\}, 1 \leq s \leq T)$$

Virtual Historical Simulation (VHS) Approach

- The simple idea of the VHS approach is to retrieve all the past returns of the (virtual) portfolio $\epsilon^{(p)}_{v,s,t}$ for $s \in 1, T$ that we construct with weigths $\underline{w}_t$ chosen at the end of t-1.
- Once we have these returns we just fit a univariate GARCH(1,1) model (as already seen in this presentation) and find the (conditional VaR).

Estimator of the Conditional VaR of Portfolio Return - GARCH

$$\widehat{VaR}^{(\alpha)}_{VHS,t-1}(\epsilon^{(p)}_{v,t,t}) = \sqrt{h^{(p)}_{v,t}(\hat{\theta_T})}\,\hat{\xi}^{(p)}_{\alpha} \tag{3}$$

With $\epsilon^{(p)}_{v,s,t} = \sum_{i=1}^{m} w_{i,t-1}\epsilon_s$ for s$\in 1, t-1$.

Virtual Historical Simulation (VHS) - Univariate Method

- **Step 1: Generate the Dataset** (same data as in the FHS case)
  We consider daily log-returns for an in-sample estimation period of size T = 3,000 on which we calibrate our models (parameter estimation by QMLE) and we keep 500 out-of-sample observations for backtesting.

- **Step 2: Estimate the Parameter of the Model** (same as in the FHS case)
  We will use a **AR(1)-CCC(m)-GARCH(1,1)** model that we will estimate by Gaussian QMLE.

- **Step 3 :Correlation Matrix Estimation** (same as in the FHS case)
  With our CCC-GARCH model under its diagonal form we can retrieve the correlation matrix by taking the empirical Covariance matrix of the $(\tilde{\eta}_t)$.

Virtual Historical Simulation (VHS) - Univariate Method

- **Step 4: Choose Dynamic Weights for each Asset at each instant t.** (same data as in the FHS case)
  We will re-balance the portfolio at the end of day t-1 for day t, considering the optimal weights of the MVP given the expected return of 5% that we consider here. To find the coefficient $\underline{w}_{t-1}$ we consider the expected returns $\hat{\underline{\epsilon}}_t$ and the expected volatility $\hat{D}_t$ estimated with the AR-CCC-GARCH model.

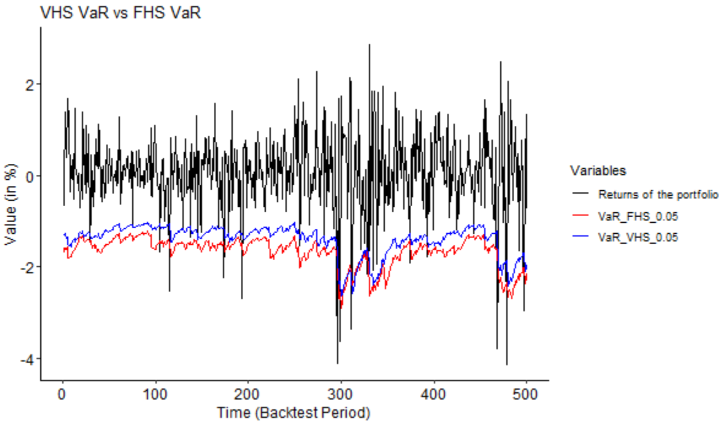
- **Step 5: Estimate the Univariate AR(1)-GARCH(1,1) model for portfolio return with crystalized weigths $\underline{w}_{t-1}$**
  We reconstruct the "virtual" portfolio returns for $1 \leq s \leq t-1$ with weight $\underline{w}_{t-1}$ and then we estimate the coefficients of a univariate AR(1)-GARCH(1,1) model.

Virtual Historical Simulation (VHS) - Univariate Method

- **Step 6: VaR Estimation** We eventually proceed to the VaR estimation as we do in a standard univariate conditional model except here we base our estimation on $\tilde{\eta}_{v,t}$, which are estimated by QML using the $\epsilon_{v,s,t}$.

Portfolio VaR - An Example - GMV Rebalancing

Figure: Portfolio VaR - GMV Optimization Re-balancing



A few Factors (1/2) - Based on Fundamentals

- **Value factor**: Price-to-Earnings (P/E) Ratio that compares a stock's current price to its earnings per share. Price-to-Book (P/B) Ratio that compares a stock's market value to its book value (assets - liabilities). Price-to-Sales (P/S) Ratio that compares a stock's market value to its revenue.
- **Quality Factor**: Return on Equity (ROE) that measures a company's profitability in relation to shareholder equity. Return on Assets (ROA) that measures a company's profitability in relation to its total assets. Profit Margin that measures the percentage of revenue that turns into profit.
- **Quality Factor**: Return on Equity (ROE) that measures a company's profitability in relation to shareholder equity. Return on Assets (ROA) that measures a company's profitability in relation to its total assets. Profit Margin that measures the percentage of revenue that turns into profit.

A few Factors (2/2) - Based on Fundamentals

- **Size Factor**: Market Capitalization refers to the total market value of a company's outstanding shares. Total Assets refers to the total assets of a company.
- **Momentum**: The price momentum measures the recent price performance of a stock and earnings momentum measures the recent growth in earnings.
- **Volatility**: log-returns standard deviation measures the historical volatility of a stock's returns.
- **Growth Factor**: earnings growth measures the rate at which a company's earnings are growing and revenue growth measures the rate at which a company's revenue is growing.

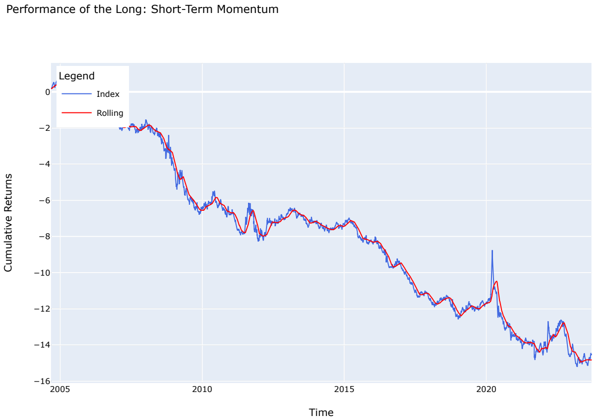
A few Factors (1/2) - A few Stylized Facts

- **Value Factor**:
  - **Expansionary/Recovery**: Value factors often perform well during economic recoveries when interest rates are low, and companies with lower valuations become attractive.
  - **Contractionary/Recession**: On the contrary, value may struggle during recessions when investors flock to safer assets, and companies with higher quality and growth characteristics are favored.
- **Quality Factor**:
  - **Expansionary/Recovery**: Quality factors tend to perform consistently well during economic expansions when investors seek financially stable and well-managed companies.
  - **Contractionary/Recession**: In addition hand, quality may also perform relatively well during recessions as investors prioritize companies with strong balance sheets and low debt.

A few Factors (2/2) - A few Stylized Facts

- **Size Factor**:
  - **Expansionary/Recovery**: Smaller-cap stocks often outperform during economic recoveries as they have more room for growth.
  - **Contractionary/Recession**: Larger-cap stocks may be favored during economic downturns as they are often perceived as more stable.
  - **Momentum Factor**:
  - **Expansionary/Recovery**: Momentum factors can perform well during expansions when trends are strong and investors follow winners.
  - **Contractionary/Recession**: Momentum factors may lose steam during market downturns when trends become less clear, and investors seek safety.

Fama-French Factor Analysis: Short-term Momentum



PCA Approach 1/3

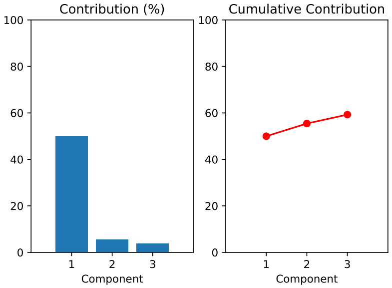


Figure: Cumulative Conrtibution of Principal Components

PCA Approach 2/3

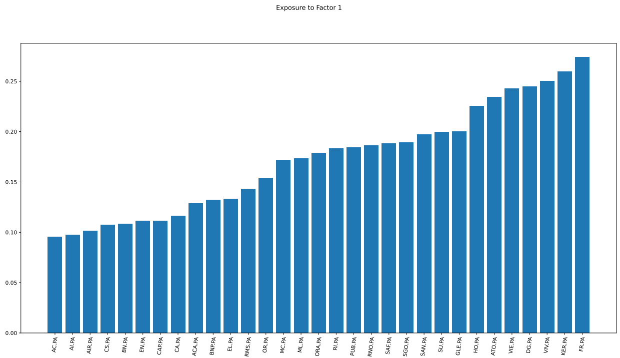


Figure: Stocks Exposure to PC1

PCA Approach 3/3

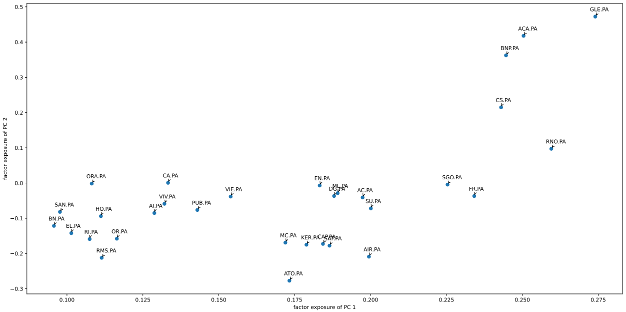


Figure: PC1/ PC2 mapping

$$log(index_t) = c_0 + \sum_{k=1}^{N_a} c_k log(p_{k,t}) + \epsilon_t \text{ for } t \in 1, ..., n \qquad (4)$$

$$\pi_{k,t} = \frac{c_k pf_{t-1}}{p_{k,t}} \text{ for } t \in 1, ..., n \qquad (5)$$

with pf_0 being the initial capital which is available and pf_t the price of the replicating portfolio at date t. We then get the prices of our replicating portfolio given by equation:

$$pf_t = \sum_{k=1}^{N_a} \pi_{k,t} p_{k,t} \text{ for } t \in 1, ..., n \qquad (6)$$

Index Replication - Implementation (Dimitriu, Alexander (2002))

Concerning the practical implementation of the procedure, we rely on the one well described in the paper, namely:

1. Choose a calibration period length to perform the estimation of the number of shares $\pi_{k,t}$ (e.g. 1,2,3,4,5 years)
2. Fix a number of days (e.g. 10 days) between two portfolio rebalancing[2].
3. Choose the number of assets in the replicating portfolio (we followed the simple method given by the authors to take the assets whose true weights in the index are the largest and which is highly debatable[3]). Before each rebalancing date the cointegration relationship is checked and new weights are computed over the rolling period.
4. Before each rebalancing we compute the implied transaction costs by the updating the portfolio composition and spread them uniformly over the period between two rebalancing (to avoid jumps

Index Rebalancing - Illustration - CAC 40 replication

Figure: Replication of the CAC 40 Index using 5 stocks with daily 10-day rebalancing

OU Process for Pair Trading

As previously mentioned, the long-short trading approach is well suited to pair trading and the use of statistical arbitrage, notably through an Ornstein Uhlenbenck process. To ensure the model's validity, we can perform an ADF (Augmented Dickey Fuller) test to verify the absence of unit root and therefore the model's stationarity.
We recall the natural formulation of an Ornstein-Uhlenbeck process:

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

The parameters are:

- $\kappa > 0$: mean reversion coefficient
- $\theta \in \mathbb{R}$: long term mean
- $\sigma > 0$: volatility coefficient

and whose well-known solution takes the following form:

$$X_t = \theta + (X_0 - \theta)e^{-\kappa t} + \int_0^t \sigma \, e^{\kappa(s-t)} dW_s.$$

Basic Potential Strategy

The basic idea of a simple mean-reversion trading strategy based on an Ornstein-Uhlenbeck process could consist in calculating the spread between the current process and its long-term mean parameter: θ. Then, the trader can open a position when the spread significantly deviates from zero, indicating that the asset's price has moved away from its mean. We can use a threshold to trigger an entry. For example, if the spread exceeds 2σ, we can consider it as a good entry signal. Then, we could consider closing the position when the spread approaches zero or starts to move back towards zero, indicating mean reversion. This can be determined using another threshold, such as when the spread falls below 0.5σ.

$\rightarrow$ We can rely on the expertise of the analyst to assess if a potential change in companies' fundamentals can perturb the stationarity of the spread