1. Empirical Finance & Portfolio Theory

J.P Bouchaud

CAPITAL FUND MANAGEMENT

http://www.cfm.fr

Single asset returns: Stylized facts

- Returns statistics depend on observation frequency: $r_t^{(\tau)} = \ln(P_{t+\tau}/P_t)$

- High frequency returns: very fat tails $P(r) \approx_{r\to\infty} |r|^{-1-\mu}$, $\mu \sim 3$

- Small linear correlations and small predictability

- Low frequency returns are more Gaussian, but slow convergence because of long memory in volatility fluct.; Slow vol. relaxation after jumps ('aftershocks')

- Leverage effect: $\sigma_{t'}$ negatively correlated with r_t for $t' \geq t$

Single asset returns: Stylized facts

- Complete description: multivariate distribution of successive returns:

$$P(...,r_{t-1}^{(\tau)},r_t^{(\tau)},r_{t+1}^{(\tau)},r_{t+2}^{(\tau)},....)$$

- Simplifying assumptions:

$$r_t^{(\tau)}=\sigma_t\xi_t \qquad \langle\xi_t\xi_{t'}\rangle\sim\delta_{t,t'}$$

where

  – σ_t is $\sim$ log-normal or inverse Gamma, and long-range correlated (eg multifractal model)

  – ξ_t still has fat-tails (jumps)

Single asset returns: Stylized facts

- Note: Simplest model is $\sigma_t = \sigma_0$, ξ_t Gaussian $\longrightarrow r_t^{(\tau)}$ Gaussian $\forall \tau$

Multivariate asset returns

- Complete description of *simultaneous* returns:

$$P(r_{1t}^{(\tau)}, r_{2t}^{(\tau)}, ... r_{it}^{(\tau)}, .., r_{Nt}^{(\tau)})$$

- Must describe correlations of the ξ_i's and of the σ_i's

- The simplest case: Gaussian multivariate

$$P(\{r_i\}) \propto \exp\left[-\frac{1}{2}\sum_{ij}\sigma_i r_i C_{ij}^{-1}\sigma_j r_j\right] \qquad (\langle r\rangle \approx 0)$$

Maximum likelihood estimator of $\mathbf{C}$ from empirical data:

$$E_{ij} = \frac{1}{T}\sum_t \hat{r}_{it}\hat{r}_{jt}$$

Multivariate asset returns

- A more realistic description: on a given day, all vols. are proportional $\rightarrow$ Elliptic distribution:

$$P(\{r_i\}) \propto \int \mathrm{d}s P(s) \exp\left[-\frac{s}{2}\sum_{ij}\sigma_i r_i C_{ij}^{-1}\sigma_j r_j\right] \qquad (\langle r\rangle \approx 0)$$

- Example: Student multivariate: $P(s) = s^{\mu/2-1}e^{-s}/\Gamma(\mu/2)$ Maximum likelihood estimator of $\mathbf{C}$ from empirical data:

$$E^*_{ij} = \frac{T+\mu}{N}\sum_t \frac{\hat{r}_{it}\hat{r}_{jt}}{\mu + \sum_{mn}\hat{r}_{mt}(E^{*-1})_{mn}\hat{r}_{nt}}$$

- When $\mu \rightarrow \infty$ for fixed T, Student becomes Gaussian and $\mathbf{E}^* = \mathbf{E}$

The large NT problem

- Determining $\mathbf{C}$ requires knowing $N(N-1)/2$ correlation coefficients. Size of data: N series of length T/τ

- For $NT/\tau \gg N^2/2$, this should work – but if $NT/\tau \ll N^2/2$ there is a problem even when $T/\tau \gg 1$!

- Actually, when $T/\tau < N$, $\mathbf{E}$ has $N-T/\tau$ exact zero eigenvalues

- For $Q = T/N\tau = O(1)$, the correlation matrix is very noisy

- Going to high frequency ($\tau \to 0$): Beware the Epps effect – $\mathbf{C}$ depends on τ!

The Epps effect

- Epps effect: Correlations grow with time lag: [FTSE, 1994-2003]

  $\langle \rho_{i \neq j}(5') \rangle = 0.06; \qquad \langle \rho_{i \neq j}(1h) \rangle = 0.19; \qquad \langle \rho_{i \neq j}(1d) \rangle = 0.29$

- Change of structure:

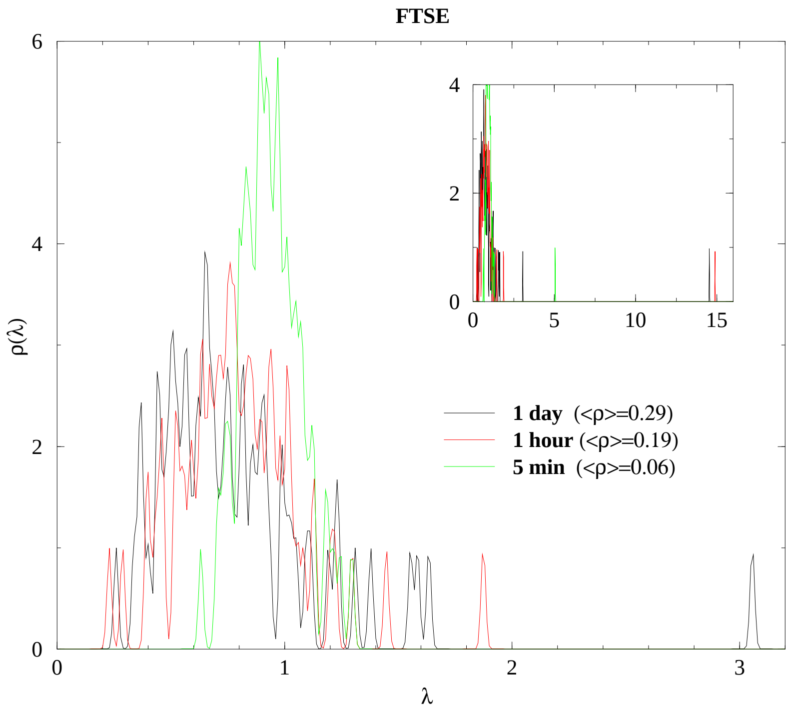
  – Modification of the eigenvalue distribution

  – Emergence of more special eigenvalues ('sectors') with time

  – Modification of the Mantegna correlation tree – market as an embryo with progressive differenciation

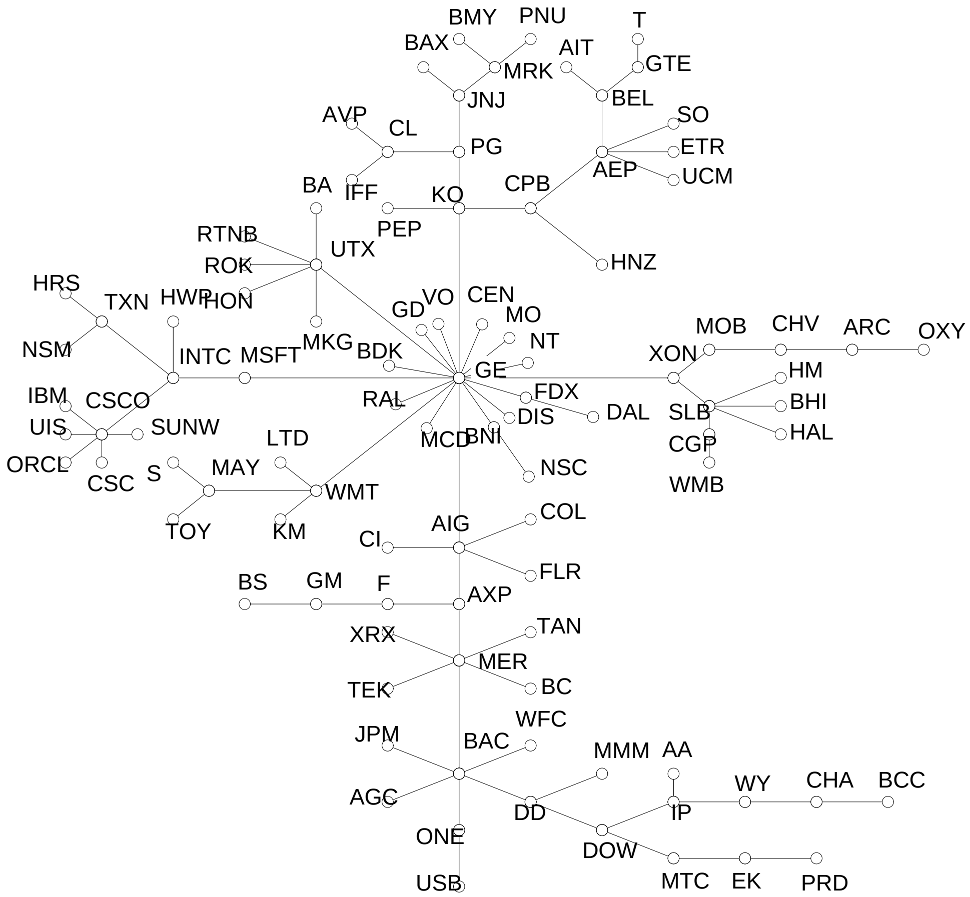
  – Weaker and shifted to higher frequencies since ~ 2000

The eigenvalue distribution on different time scales



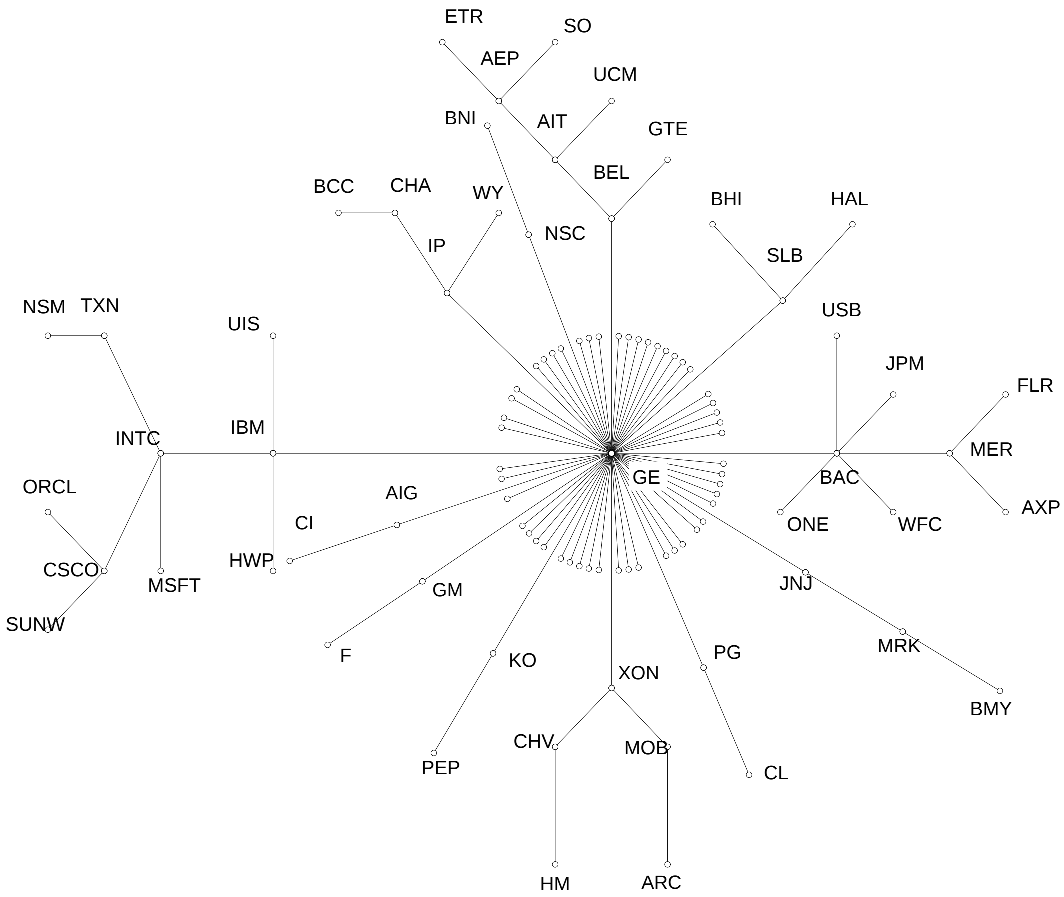
Eigenvalue distribution at different time scales for the FTSE.

The daily correlation tree



Correlation tree constructed from the correlation matrix (From Mantegna et al.)

The high frequency correlation tree



Correlation tree constructed from the high frequency correlation matrix (From Mantegna et al.)

CAPITAL FUND MANAGEMENT

J.Ph. Bouchaud

The Marcenko-Pastur distribution

- Assume $\mathbf{C} \equiv \mathbf{1}$: no 'true' correlations and Gaussian returns

- What is the spectrum of $\mathbf{E}$?

- Marcenko-Pastur $q = 1/Q$

$$\rho(\lambda) = (1-Q)^{+}\delta(\lambda) + \frac{\sqrt{4\lambda q - (\lambda + q - 1)^2}}{2\pi\lambda q} \qquad \lambda \in [(1-\sqrt{q})^2, (1+\sqrt{q})^2]$$

- Two sharp edges ! (when $N \to \infty$)

- Results also known for $\mathbf{E}$ and $\mathbf{E}^*$ in the Student ensemble

Portfolio theory: Basics

- Portfolio weights w_i,

- If predicted gains are g_i then the expected gain of the portfolio is $G = \sum w_i g_i$.

- Risk: variance of the portfolio returns

$$R^2 = \sum_{ij} w_i \sigma_i C_{ij} \sigma_j w_j$$

where σ_i^2 is the variance of asset i and C_{ij} is the correlation matrix.

Markowitz Optimization

- Find the portfolio with maximum expected return for a given risk or equivalently, minimum risk for a given return (G)

- In matrix notation:

$$\mathbf{w}_C = G \frac{\mathbf{C}^{-1}\mathbf{g}}{\mathbf{g}^T \mathbf{C}^{-1}\mathbf{g}}$$

- Where all returns are measured with respect to the risk-free rate and $\sigma_i = 1$ (absorbed in g_i).

- Non-linear problem: $\sum_i |w_i| \le A$ – a spin-glass problem!

- Related problem: find the idiosyncratic part of a stock

Risk of Optimized Portfolios

- Let $\mathbf{E}$ be an noisy estimator of $\mathbf{C}$ such that $\langle \mathbf{E} \rangle = \mathbf{C}$

- "In-sample" risk

$$R^2_{\text{in}} = \mathbf{w}_E^T \mathbf{E} \mathbf{w}_E = \frac{G^2}{\mathbf{g}^T \mathbf{E}^{-1} \mathbf{g}}$$

- True minimal risk

$$R^2_{\text{true}} = \mathbf{w}_C^T \mathbf{C} \mathbf{w}_C = \frac{G^2}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}$$

- "Out-of-sample" risk

$$R^2_{\text{out}} = \mathbf{w}_E^T \mathbf{C} \mathbf{w}_E = \frac{G^2 \mathbf{g}^T \mathbf{E}^{-1} \mathbf{C} \mathbf{E}^{-1} \mathbf{g}}{(\mathbf{g}^T \mathbf{E}^{-1} \mathbf{g})^2}$$

Risk of Optimized Portfolios

- Using convexity arguments, and for large enough matrices: $R^2_{\text{in}} \leq R^2_{\text{true}} \leq R^2_{\text{out}}$

- Importance of eigenvalue cleaning:

$$w_i \propto \sum_{kj} \lambda_k^{-1} V_i^k V_j^k g_j = g_i + \sum_{kj} (\lambda_k^{-1} - 1) V_i^k V_j^k g_j$$

  – Eigenvectors with $\lambda > 1$ are suppressed,

  – Eigenvectors with $\lambda < 1$ are enhanced. Potentially very large weight on small eigenvalues.

  – Must determine which eigenvalues to keep and which one to correct

Quality Test

- Out of Sample quality of the cleaning: $R^2_{\text{in}}/R^2_{\text{out}}$ as close to unity as possible for a random choice of g.

- For example, when g is a random vector on the unit sphere,

$$R^2_{\text{in}} = \frac{G^2}{\text{Tr}\mathbf{E}^{-1}} \qquad R^2_{\text{out}} = \frac{G^2\text{Tr}\mathbf{E}^{-1}\mathbf{C}\mathbf{E}^{-1}}{(\text{Tr}\mathbf{E}^{-1})^2}$$

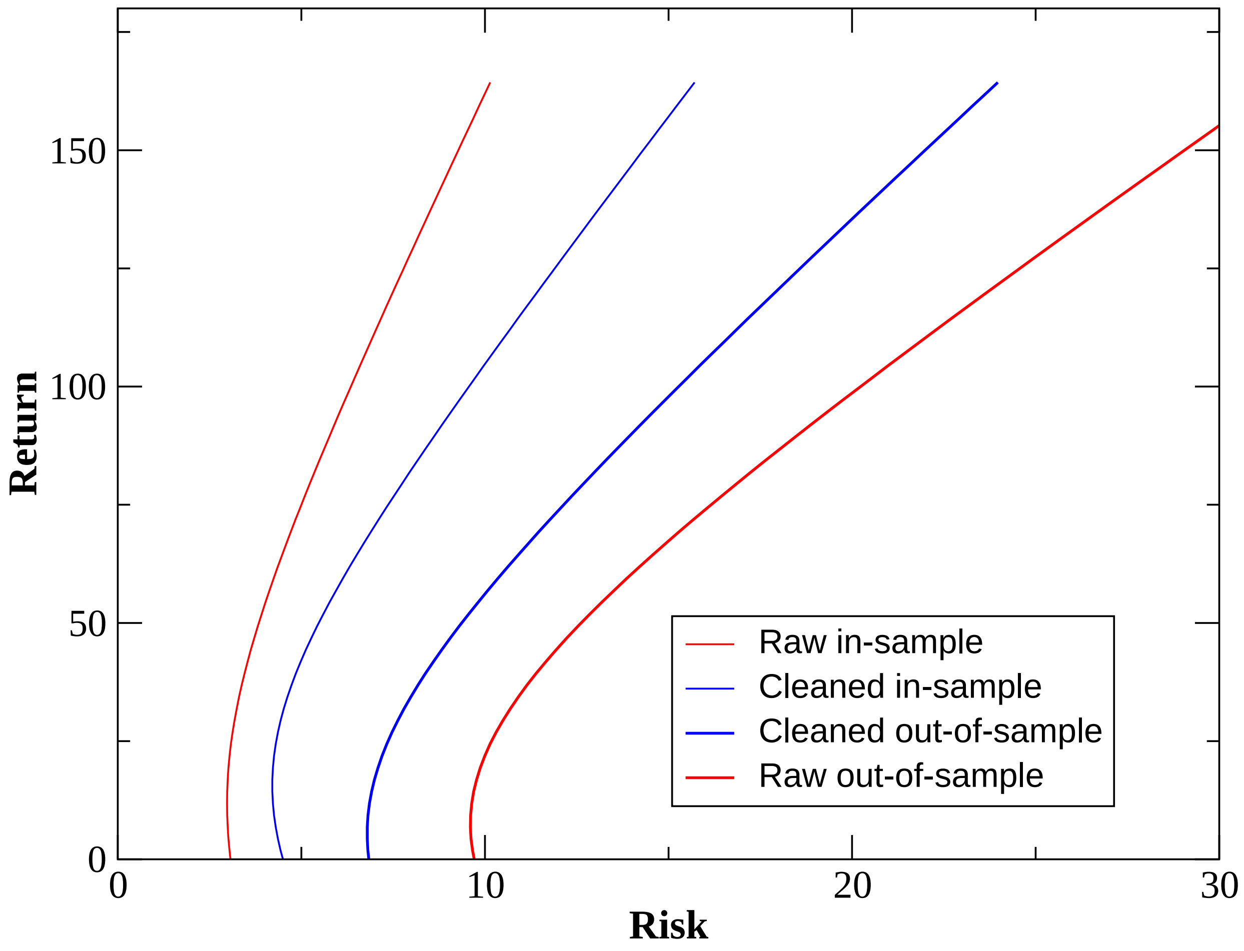
- Example: In the MP case,

$$R^2_{\text{in}} = R^2_{\text{true}}(1-q) \qquad R^2_{\text{out}} = \frac{R^2_{\text{true}}}{1-q}$$

(from:

$$G_{MP}(z \to 0) \approx \frac{1}{1-q} + \frac{z}{(1-q)^3} \equiv -\text{Tr}\mathbf{E}^{-1} - z\text{Tr}\mathbf{E}^{-2})$$

Matrix Cleaning



Cleaning Algorithms

- Shrinkage estimator

$$\mathbf{E}_c = \alpha \mathbf{E} + (1-\alpha)\mathbf{1} \qquad \text{so} \qquad \lambda_c^k = 1 + \alpha(\lambda^k - 1)$$

- Eigenvector cleaning

$$\lambda_c^k = 1 - \delta \qquad \text{if} \qquad k < k_{\text{min}}$$

$$\lambda_c^k = \lambda_E^k \qquad \text{if} \qquad k \geq k_{\text{min}}$$

Effective Number of Assets

- Definition: (Hirfindahl index)

$$N_\mathrm{e} = \left(\sum_{i=1}^{N} w_i^2 \right)^{-1}$$

  - measure the diversification of a portfolio
  - equals N iff $w_i \equiv 1/N$

- Optimization

$$\max \left\{ \sum_{i,j=1}^{N} w_i w_j C_{ij} + \zeta_1 \sum_{i=1}^{N} p_i w_i + \zeta_2 \sum_{i=1}^{N} w_i^2 \right\}$$

  - same as replacing C_{ij} by $C_{ij} + \zeta_2 \delta_{ij}$.

RMT Clipping Estimator Revisited

- Where is the edge? Finite size effects, bleeding.

- In practice non trivial on financial data:
  - Fat tails ($\mu = 3$?),
  - Correlated volatility fluctuations,
  - Time dependence.

- Is there information below the lower edge?
  - Inverse participation ratio is high (localized),
  - Pairs at high frequency.

Other measures of risks

- Risk of an hedged option portfolio:

$$\delta\Pi = \frac{1}{2}\sum_i \Gamma_i r_i^2 + \sum_i \Upsilon_i \delta\sigma_i$$

- Correlation matrices for squared returns and for change of implied vols.

- Extreme Tail correlations:

$$\mathcal{C}_{ij}(p) = P(|r|_i > R_{ip}||r|_j > R_{jp}) \quad \text{with} \quad P(|r|_i > R_{ip}) = p, \forall i$$

- For Gaussian RV, $\mathcal{C}_{ij}(p \to 0) = 0$

Other measures of risks

- For Student RV (or any elliptic power-law), $\mathcal{C}_{ij}(p \to 0) = Z(\theta)/Z(\pi/2)$ with:

$$\rho = \sin\theta; \qquad Z(\theta) = \int_{\pi/4-\theta/2}^{\pi/2} du \cos^{\mu}(u)$$

- Empirically, all these non-linear correlation matrices have a very similar structure to E_{ij}

More General Correlation matrices

- Non equal time correlation matrices

$$E_{ij}^{\tau} = \frac{1}{T} \sum_t \frac{X_i^t X_j^{t+\tau}}{\sigma_i \sigma_j}$$

$N \times N$ but not symmetrical: 'leader-lagger' relations

- General rectangular correlation matrices

$$G_{\alpha i} = \frac{1}{T} \sum_{t=1}^{T} Y_\alpha^t X_i^t$$

N 'input' factors X; M 'output' factors Y

  – Example: $Y_\alpha^t = X_j^{t+\tau}$, $N = M$

- The large N-M-T problem ! Sunspots and generalisation of Marcenko-Pastur – See later