1. Empirical Finance & Portfolio Theory

J.P Bouchaud

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 \( t' \geq t \)

\[ J.Ph. \text{ Bouchaud} \]
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:

  \[ r_{t}^{(\tau)} = \sigma_t \xi_t \quad \langle \xi_t \xi_{t'} \rangle \sim \delta_{t,t'} \]

  where

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

  - \( \xi_t \) still has fat-tails (jumps)

  [Logo: Capital Fund Management]

  J.Ph. Bouchaud
Single asset returns: Stylized facts

- Note: Simplest model is $\sigma_t = \sigma_0$, $\xi_t$ Gaussian $\rightarrow r_t^{(\tau)}$ Gaussian $\forall \tau$
Multivariate asset returns

- Complete description of simultaneous returns:
  \[ P(\{r_1^\tau, r_2^\tau, \ldots, r_{it}^\tau, \ldots, r_{Nt}^\tau\}) \]

- Must describe correlations of the \( \xi_i \)'s and of the \( \sigma_i \)'s

- The simplest case: Gaussian multivariate
  \[
P(\{r_i\}) \propto \exp \left[ -\frac{1}{2} \sum_{ij} \sigma_i r_i C^{-1}_{ij} \sigma_j r_j \right] \quad (\langle r \rangle \approx 0)\]
  Maximum likelihood estimator of \( C \) from empirical data:
  \[ E_{ij} = \frac{1}{T} \sum_t \hat{r}_{it} \hat{r}_{jt} \]

*J.Ph. Bouchaud*
Multivariate asset returns

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

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

- Example: Student multivariate: $P(s) = s^{\mu/2-1}e^{-s}/\Gamma(\mu/2)$

  Maximum likelihood estimator of $C$ from empirical data:

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

- When $\mu \rightarrow \infty$ for fixed $T$, Student becomes Gaussian and $E^* = E$
The large $NT$ problem

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

- 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 $T/\tau < N$, $E$ 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 – $C$ depends on $\tau$!

J.Ph. Bouchaud
The Epps effect

• Epps effect: Correlations grow with time lag: [FTSE, 1994-2003]
  \[ \langle \rho_{i\neq j}(5') \rangle = 0.06; \quad \langle \rho_{i\neq j}(1h) \rangle = 0.19; \quad \langle \rho_{i\neq j}(1d) \rangle = 0.29 \]

• 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 \( \sim 2000 \)

J.Ph. Bouchaud
The eigenvalue distribution on different time scales

Eigenvalue distribution at different time scales for the FTSE.

J.Ph. Bouchaud
The daily correlation tree

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

J.Ph. Bouchaud
The high frequency correlation tree

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

J.Ph. Bouchaud
The Marcenko-Pastur distribution

- Assume $C \equiv 1$: no ‘true’ correlations and Gaussian returns

- What is the spectrum of $E$?

- Marcenko-Pastur $q = 1/Q$

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

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

- Results also known for $E$ and $E^*$ in the Student ensemble

J.Ph. Bouchaud
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 $\sigma_i^2$ is the variance of asset $i$ and $C_{ij}$ is the correlation matrix.

J.Ph. Bouchaud
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:

$$w_C = G \frac{C^{-1}g}{g^T C^{-1}g}$$

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

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

- Related problem: find the idiosyncratic part of a stock

J.Ph. Bouchaud
Risk of Optimized Portfolios

• Let $E$ be an noisy estimator of $C$ such that $\langle E \rangle = C$

• “In-sample” risk

$$R_{in}^2 = w_E^T E w_E = \frac{G^2}{g^T E^{-1} g}$$

• True minimal risk

$$R_{true}^2 = w_C^T C w_C = \frac{G^2}{g^T C^{-1} g}$$

• “Out-of-sample” risk

$$R_{out}^2 = w_E^T C w_E = \frac{G^2 g^T E^{-1} C E^{-1} g}{(g^T E^{-1} g)^2}$$

J.Ph. Bouchaud
Risk of Optimized Portfolios

- Using convexity arguments, and for large enough matrices:
  \[ R_{\text{in}}^2 \leq R_{\text{true}}^2 \leq R_{\text{out}}^2 \]

- Importance of eigenvalue cleaning:
  \[ w_i \propto \sum_{kj} \lambda_{kj}^{-1} V_i^k V_j^k g_j = g_i + \sum_{kj} (\lambda_{kj}^{-1} - 1) V_i^k V_j^k g_j \]
  
  - 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

J.Ph. Bouchaud
Quality Test

• Out of Sample quality of the cleaning: $R_{in}^2/R_{out}^2$ as close to unity as possible for a random choice of $g$.

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

$$R_{in}^2 = \frac{G^2}{\text{Tr}E^{-1}} \quad R_{out}^2 = \frac{G^2\text{Tr}E^{-1}CE^{-1}}{(\text{Tr}E^{-1})^2}$$

• Example: In the MP case,

$$R_{in}^2 = R_{true}^2(1 - q) \quad R_{out}^2 = \frac{R_{true}^2}{1 - q}$$

(from:

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

J.Ph. Bouchaud
Matrix Cleaning

![Graph showing the relationship between Return and Risk for different data sets: Raw in-sample, Cleaned in-sample, Cleaned out-of-sample, Raw out-of-sample.]

- **Return** axis ranges from 0 to 150.
- **Risk** axis ranges from 0 to 30.

Legend:
- Red: Raw in-sample
- Blue: Cleaned in-sample
- Light blue: Cleaned out-of-sample
- Red: Raw out-of-sample

**J.Ph. Bouchaud**
Cleaning Algorithms

- **Shrinkage estimator**

$$E_c = \alpha E + (1 - \alpha) I$$

so

$$\lambda_c^k = 1 + \alpha (\lambda^k - 1)$$

- **Eigenvector cleaning**

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

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

\[ J.Ph. \ Bouchaud \]
Effective Number of Assets

• Definition: (Hirfindahl index)

\[ N_e = \left( \sum_{i=1}^{N} w_i^2 \right)^{-1} \]

– 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:
  \[ C_{ij}(p) = P(|r_i > R_{ip}| |r_j > R_{jp}) \quad \text{with} \quad P(|r_i > R_{ip}) = p, \forall i \]

- For Gaussian RV, \( C_{ij}(p \rightarrow 0) = 0 \)
Other measures of risks

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

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

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

---

J.Ph. Bouchaud
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

\[ J.Ph. \ Bouchaud \]