Empirical properties of large covariance matrices in finance

Gilles Zumbach

Ex: RiskMetrics Group, Geneva
Since 2010: Swissquote, Gland

December 2009

Covariance and large random matrices

Many problems in finance require to estimate the covariance matrix Σ in a given “universe” of assets.

Σ measures the volatilities and co-dependencies between time series.

Today portfolios are very large, ranging from $N = 100$ to 100’000 positions.

At first order, prices are random walks.

With such large sizes, an approach with random matrices seems appropriate.

Is such an approach valid, when information should be present in a financial universe?

Extracting information from prices

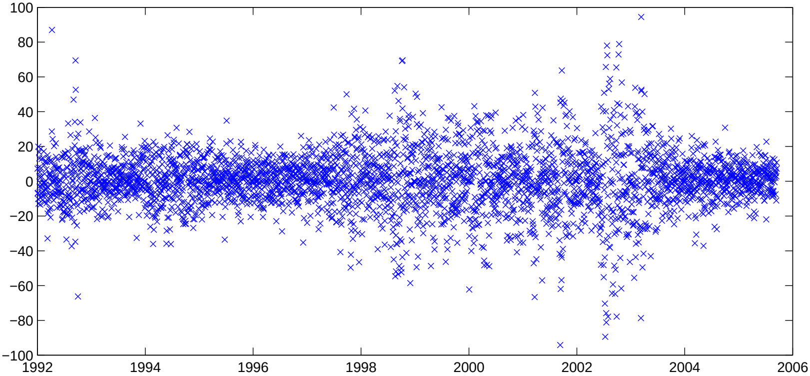
From a given asset α, the prices $p_\alpha(t)$ are the known informations from the trades.

The returns (profit and loss), at the scale δt (1 day), are

$$r_\alpha(t) = \ln(p_\alpha(t)/p_\alpha(t-\delta t))$$

The returns are dimentionless, but their magnitudes depend on δt.

The key specificity in finance: heteroskedasticity



Time series for the annualized returns $r[\delta t]$ for the FTSE100.

The dominant features of the data are heteroscedasticity and fat tails.

What is the correct measure of volatility?

Goal: measure the volatility $\sigma(t)$.
Problem: many definitions can be given, with different properties.

A process that reproduces the empirical features has to contain a dynamical volatility $\sigma(t)$.
This $\sigma(t)$ is the volatility in the forthcoming time step.
The conditional expectations of quadratic GARCH processes can be computed $\longrightarrow$ forecasts for any horizons.
The forecasts take the form:

$$\begin{aligned}\tilde{\sigma}^2(t, t + n\delta t) &= \sum_{i=0} \lambda(i,n)\, r^2(t - i\,\delta t) \\ \sum_i \lambda(i,n) &= 1\end{aligned}$$

The volatility forecast is such that the recent past carry more information than the distant past.

Definition of the covariance

The covariance matrix $\Sigma(t)$ is a cross product of the past return vectors

$$\begin{aligned}\Sigma_{\alpha,\beta}(t, t + n\delta t) &= \sum_{i=0} \lambda(i,n)\, r_\alpha(t - i\delta t)\, r_\beta(t - i\delta t) \\ \sum_i \lambda(i,n) &= 1\end{aligned}$$

The mean return can also be subtracted.

More general quadratic extensions can be used (see later).

The dependency on n is weak $\longrightarrow$ $n = 1$.

The covariance $\Sigma(t)$ has a dynamical structure.

The covariance's weights

Common choices for $\lambda(i)$:

- **Equal weights in a time window**: $\lambda(i) = 1/T \quad \text{for} \quad i < T$
  A convenient definition for rigorous results.
- **Exponential**: $\lambda(i) \simeq \mu^i = \exp(-i\,\delta t/\tau)$
  A convenient practical definition.
- **Long memory**: $\lambda(i) \simeq 1 - \log(i\,\delta t)/\log(\tau_0)$
  An optimal definition for forecasts.

The correlation matrix

The correlation matrix $\rho(t)$ is defined by

$$\rho_{\alpha,\beta}(t) = \frac{\Sigma_{\alpha,\beta}(t)}{\sqrt{\Sigma_{\alpha,\alpha}(t)\ \Sigma_{\beta,\beta}(t)}}$$

The correlation matrix has simpler properties than the covariance, because it is not depending on the distribution of the volatilities $\Sigma_{\alpha,\alpha}$.

Why covariances are so important in finance?

The covariance is crucial in three areas in finance:

- **Portfolio allocation**: the optimal allocation of assets in a portfolio is depending on the volatilities (diagonal part of Σ) and diversification (off diagonal part of Σ).
- **Risk**: market risks are driven by the possible large losses in a portfolio. The profits and losses are driven by the covariance (more precisely: a forecast for the covariance matrix up to the desired risk horizon).
- **Model building and inferences**: a dynamical description of the market is done by random processes.
  For example, a multivariate Bachelier random walk is

  $$r_\alpha(t) = \Sigma^{1/2}_{\alpha,\beta}\, \varepsilon_\beta(t)$$

  where the innovations $\varepsilon_\beta(t)$ are independent random variables.

The problems with large sizes

- When $N > T$, the covariance contains zero eigenvalues.
  The covariance cannot be inverted!
- Is there actually N^2 (or NT) information in Σ?
  Randomness should be a large part of the covariance.
  Can the "information" be separated from the "noise"?
- Can the dominant structure of Σ be understood in term of a few dominant factors (e.g. stock indexes, industrial sectors, ...)?
  How many factors are needed?

A purely random model for Σ

In 1929, Wishart proposed a purely random model for the covariance

$$\Sigma_{\alpha,\beta} = \frac{1}{T} \sum_{1 \leq t' \leq T} r_\alpha(t')\, r_\beta(t')$$

with $r_\alpha(t)$ independent random variables **with unit variance**.

In finance, the natural structure is a product of random vectors.

Marcenko and Pastur (1967) derived the spectral density in the limit $N \to \infty$ and $T \to \infty$, for a fixed ratio $q = N/T$:

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

The data sets for the empirical study

- **ICM dataset**: 340 time series covering majors asset classes and world geographical areas (19 commodities, 78 foreign exchange rates, 52 equity indices, 127 interest rates, 54 individual stocks)
- **G10 dataset**: 55 time series covering the largest economies (European, Japan, and USA).
- **USA dataset**: 54 times series from USA, mostly large stocks with IR and 2 equity indices.

For constant weights in a window, $T = 260$ days.
Other memory kernels are evaluated with $T = 260$ days.

Spectrum

For symmetric matrices, the eigen-decomposition is

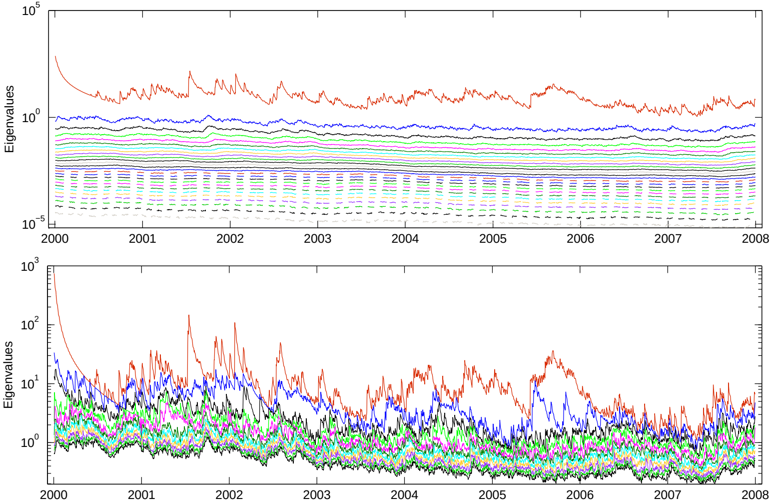
$$\Sigma = \sum_{\alpha=1}^{N} e_\alpha \, \mathbf{v}_\alpha \mathbf{v}'_\alpha$$

and the eigenvectors $\mathbf{v}_\alpha$ are orthogonal.

Similarly for the correlation matrix.

All quantities are time dependent: $\Sigma(t), \rho(t), e_\alpha(t), \mathbf{v}_\alpha(t)$.

Time evolution of the spectrum (ICM data set)



Time evolution of the spectrum

- The bulk of the spectrum spectrum is very rigid.
- Only the first few eigenvalues have an interesting dynamics.
- This goes in the direction of
  a few meaningful eigenvalues + random spectrum.

Spectral density

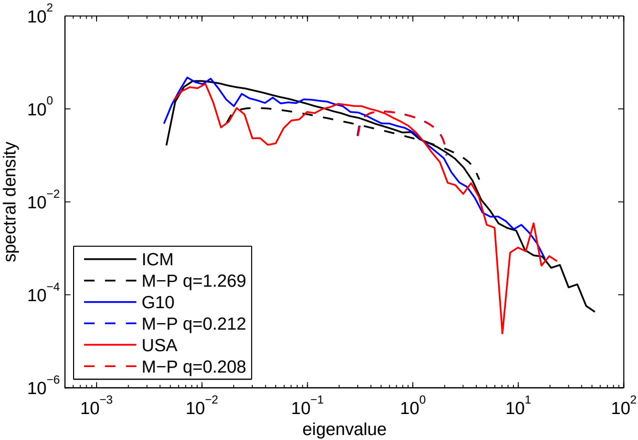
The spectral density in the interval $[\lambda - \delta\lambda, \lambda + \delta\lambda]$ counts the number of eigenvalues in the interval

$$\rho(\lambda) = \frac{1}{2\,\delta\lambda}\,\frac{1}{T}\sum_{t=1}^{T}\,\frac{1}{N}\sum_{\alpha}\,\chi(\lambda - \delta\lambda < e_{\alpha}(t) < \lambda + \delta\lambda).$$

The spectral density $\rho(\lambda)$ measures the time average density of the eigenvalues around λ.

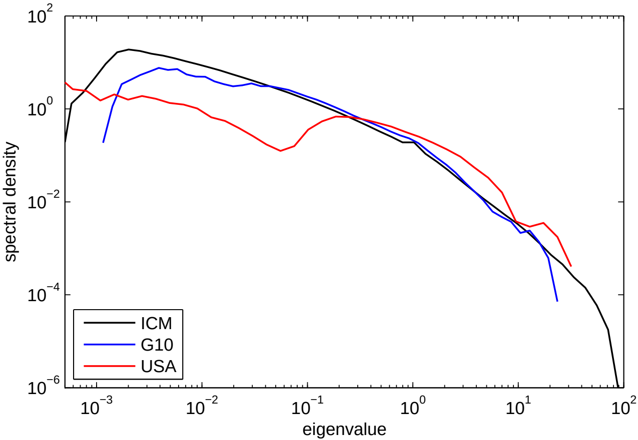
The normalization is defined such that $\int \rho(\lambda)\,d\lambda = 1$.

Spectral density for the correlation (constant weights)



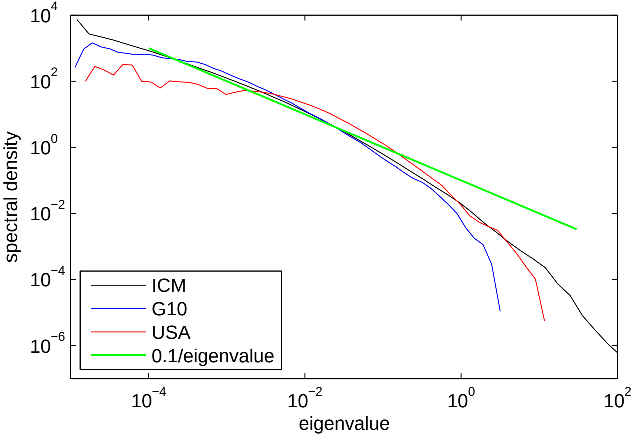
Mean spectral density of the correlation matrix ρ.

Spectral density for the correlation (long memory)



Mean spectral density of the correlation matrix ρ.

Spectral density for the covariance



Mean spectral density of the covariance matrix Σ.

The green curve is a simple Ansatz corresponding to

$$e_\alpha \simeq e^{-a\alpha/N}$$

Remarks on the spectral density

- The eigenvalues decrease exponentially fast toward zero.
  Problems arise even when $N < T$.
- The empirical density (for the correlation) is not close to the Marcenko-Pastur density.
  Marcenko-Pastur: $N \longrightarrow \infty$, $T \longrightarrow \infty$, $q = N/T$ fixed.
  Empirical: $N \longrightarrow \infty$, T fixed

So far, the focus is on the spectrum. What about the eigenvectors?
The relevant objects to study are projectors.

Projector

The projector $\mathrm{P}_k(t)$ on the leading subspace of rank k is

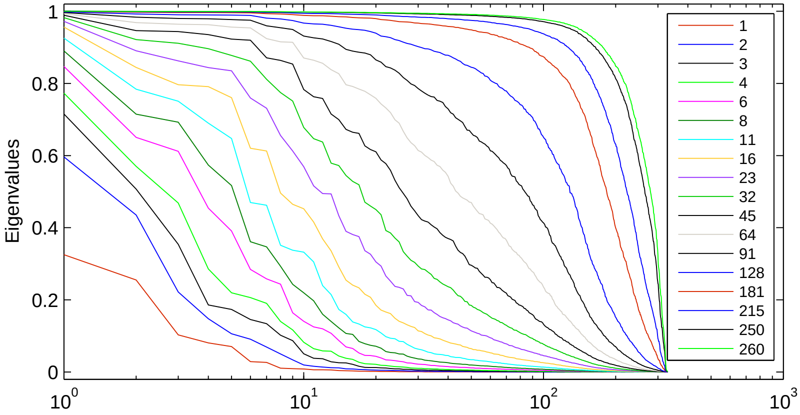
$$\mathrm{P}_k = \sum_{\alpha=1}^{k} \mathbf{v}_\alpha \mathbf{v}'_\alpha$$

The mean projector is defined by the time average

$$\langle \mathrm{P}_k \rangle = \frac{1}{T} \sum_{t=1}^{T} \mathrm{P}_k(t).$$

The trace is preserved by the average $\mathrm{tr}\,\langle \mathrm{P}_k \rangle = k$.
$\langle \mathrm{P}_k \rangle$ is not a projector as its eigenvalues are between 0 and 1.

Empirical spectrum of the mean projector (ICM data set)



No clear invariant subspace!

Projector dynamics and fluctuation index

The fluctuation index γ essentially measures the difference between $\left\langle P^2(t) \right\rangle$ and $\langle P(t) \rangle^2$, with a convenient normalization

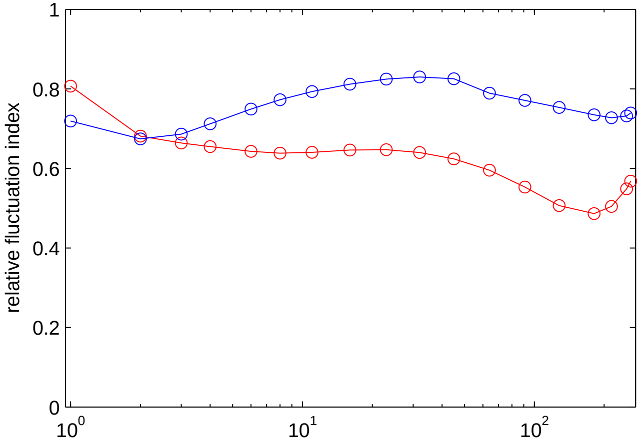
$$\gamma_k = \frac{\mathrm{tr}\left\langle P_k^2 \right\rangle - \mathrm{tr}\left(\langle P_k \rangle^2 \right)}{\mathrm{tr}\left(\langle P_k \rangle \right)} = 1 - \frac{1}{k}\ \mathrm{tr}\left(\langle P \rangle^2 \right).$$

If $\langle P \rangle \simeq P(t)$, then $\langle P \rangle^2 \simeq \langle P \rangle$ and $\gamma \simeq 0$.

If the projector dynamics explores fully the available space $\langle P_k \rangle \simeq k/N\, \mathbb{I}$, then $\gamma_k = \gamma_{k,\max} = 1 - k/N$.

This is a dis-order parameter.

Empirical fluctuation index (ICM data set)



Relative fluctuation index $\gamma_k / \gamma_{k,\max}$ as function of the projector rank k. Projectors of covariance (red) and correlation (blue).

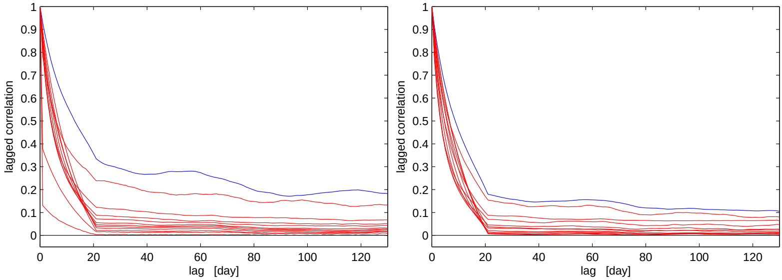
Lagged correlation

For a matrix $X(t)$, the scalar lagged correlation is

$$\rho(\tau) = \frac{\operatorname{tr}\langle (X(t) - \langle X \rangle)(X(t+\tau) - \langle X \rangle)\rangle}{\operatorname{tr}\left\langle (X - \langle X \rangle)^2 \right\rangle}.$$

The matrices $X(t)$ can be the covariance or projector of a given rank, or the correlation or projector of a given rank.

Empirical lagged correlation (USA data set, T = 21 days)



Lagged correlations for the covariance (left) and correlation (right), for full matrix (blue) and the projectors of increasing size (red).

This is the signature of heteroskedasticity and its long memory.

Summarizing the projector dynamics

- Regardless of the rank k, projectors tend to explore a large fraction of the available space.
- Covariance and correlation matrices have a similar dynamics.
- An approximation by constant correlations is inappropriate.

Multivariate ARCH processes

$$
\begin{aligned}
\mathbf{x}(t+\delta t) &= \mathbf{x}(t)+\mathbf{r}(t+\delta t) \\
\mathbf{r}(t+\delta t) &= \Sigma(\gamma,\xi;t)^{1/2}\,\boldsymbol{\varepsilon}(t+\delta t). \\
\Sigma(\gamma,\xi) &= (1-\xi)\left((1-\gamma)\Sigma+\gamma\,\Sigma|_{\text{diag}}\right)+\xi\,\mathrm{tr}(\Sigma)\,\mathbb{I}_N
\end{aligned}
$$

The new terms are quadratic and keep the trace (total variance). For $\xi > 0$, all eigenvalues are positive.

The innovations are iid:

$$
E\left[\,\varepsilon_\alpha(t)\,\varepsilon_\beta(t')\,\right] = \delta_{\alpha,\beta}\,\delta_{t,\,t'}
$$

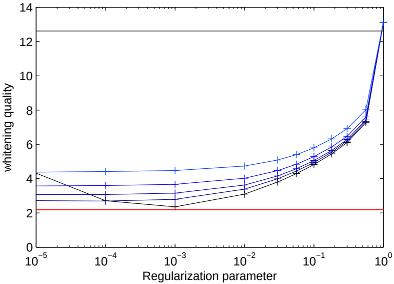
To estimate the process on historical data requires to compute the innovations:

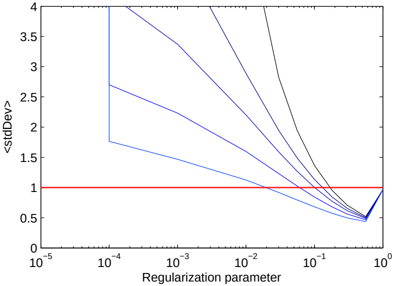
$$
\boldsymbol{\varepsilon}(t+\delta t) = \Sigma^{-1/2}(\gamma,\xi;t)\,\mathbf{r}(t+\delta t).
$$

Statistical properties of the innovations

$\langle \varepsilon_\alpha \, \varepsilon_\beta \rangle = 0$

$\langle \varepsilon_\alpha^2 \rangle = 1$





Shrinkage parameter: $\gamma = 0.0$ (black), 0.05, 0.1, 0.2, 0.4 (blue).

Horizontal black line: return correlation.

Horizontal red lines: uncorrelated Student innovations.

Understanding the causes of the problem

The inability to get white noise innovations is due to the followings.

- Small eigenvalues corresponds to directions (portfolios) without fluctuations (riskfree)
- Fluctuations along these directions do occur out-of-sample
- Large innovations are needed to compensate for the small eigenvalues
- Regularizing the covariance reduces the innovations ...
- ... but washes out the correlation structure

Implication for volatility forecasts: for increasing forecast horizon, the historical correlation decreases for a structureless prior.

Conclusions

- The eigenvalues of the covariance are mostly uninformative, except the first three to ten eigenvalues.
  The eigenvalues of the correlation are even less informative, but the first fews are not constant.
- The transition from significant to noisy eigenvalues is gradual (no gap, no cusp)
- Eigenvectors have informative dynamics deep in the spectrum.
  No clear invariant subspace corresponding to the main market modes, but dynamic subspaces with regularly decreasing importance across the whole spectrum (no universal factors).
- Random matrix theory seems to lack results that can be used in this context ($N \to \infty$ with T fixed, properties of the projectors).
- The inverse covariance is ill defined, even when $N \ll T$.
  A proper regularization should be used to estimate Σ^{-1}.
- No good multivariate processes.

References

The corresponding papers can be downloaded from www.ssrn.com (Social Science Research Network)