Fiche
High Frequence for Finance

Antoine Falck

December 9, 2017

Contents

Part I

1 Portfolio management – Markovitz and CAPM approach

1.1 Introduction

The goal here is to find the best portfolio combination, *i.e.* with the best returns, given a risk level. The portfolio's performance is measured by its average returns and its risk by the variance. This gives the idea that we are in a normal frame.

1.2 Efficient portfolios

1.2.1 Market assumptions and portfolio property

We have two dates on the market, $t = 0$ is the date of the investment and $t = 1$ is when we look at the returns. We have N risky assets[1] and one non-risky asset (NRA).

The rate r of the NRA is known at $t = 0$; we note $p_{i,t}$ the price of the i-th asset at time t and r_i its rate return between $t = 0$ and $t = 1$,

$$r_i \;=\; \frac{p_{i,1}}{p_{i,0}} \;-\; 1 \;\triangleq\; y_i \;-\; 1.$$

And with $\mathbf{Y} = (y_1, \ldots, y_N)^{\mathrm{T}}$, we suppose that the following parameters are known at time $t = 0$,

$$\begin{aligned} \boldsymbol{\mu} &= \mathbb{E}[\mathbf{Y}] \;; \\ \boldsymbol{\Omega} &= \mathrm{Var}[\mathbf{Y}]. \end{aligned}$$

We also assume that for all $i \in \{1, \ldots, N\}$, $\mu_i > 1 + r$.

Let us consider a portfolio with a quantity a_0 of NRA and a_i units of i-th asset ; its value is

$$V_t \;=\; a_0 p_t^{\mathrm{NRA}} \;+\; \mathbf{a}^{\mathrm{T}}\mathbf{p_t} \;, \tag{1.1}$$

where $\mathbf{a} = (a_1, \ldots, a_N)^{\mathrm{T}}$ and $\mathbf{p_t} = (p_{1,t}, \ldots, p_{N,t})^{\mathrm{T}}$. Especially,

$$\begin{aligned} V_0 &= a_0 \;+\; \mathbf{a}^{\mathrm{T}}\mathbf{p_0} \;; && (1.2) \\ V_1 &= a_0(1+r) \;+\; \mathbf{a}^{\mathrm{T}}\mathrm{Diag}(\mathbf{p_0})\mathbf{Y}. && (1.3) \end{aligned}$$

On a donc,

$$\begin{aligned} \mathbb{E}[V_1] &= a_0(1+r) \;+\; \mathbf{a}^{\mathrm{T}}\mathrm{Diag}(\mathbf{p_0})\boldsymbol{\mu} \;; && (1.4) \\ \mathrm{Var}[V_1] &= \mathbf{a}^{\mathrm{T}}\mathrm{Diag}(\mathbf{p_0})\boldsymbol{\Omega}\mathrm{Diag}(\mathbf{p_0})\mathbf{a}. && (1.5) \end{aligned}$$

1.2.2 Average/variance optimizing

Let's consider that the value of the portfolio at $t = 0$ is v. We are looking for the portfolios of maximum expected return for a variance below a certain level.

Definition 1.1 (Domination)**.** We say that a portfolio A dominates a portfolio B if $\mathbb{E}[V_1^{\mathrm{A}}] > \mathbb{E}[V_1^{\mathrm{B}}]$ and $\mathrm{Var}[V_1^{\mathrm{A}}] \leq \mathrm{Var}[V_1^{\mathrm{B}}]$.

[1] To set the ideas we can take $N = 40$ for the CAC40 index.

Definition 1.2 (Efficient border)**.** We call efficient border the set of portfolios that are not dominated.

To determine the efficient border for each risk level σ^2 we have to solve the problem:

$$\begin{aligned} & \max_{a_0,\mathbf{a}} \mathbb{E}[V_1(a_0,\mathbf{a})] \\ \text{s.t.} \quad & \mathrm{Var}[V_1(a_0,\mathbf{a})] \;\leq\; \sigma^2 \\ & V_0(a_0,\mathbf{a}) \;=\; v \end{aligned} \tag{E}$$

Remark. Solutions of (E) or also solutions of (E') with,

$$\begin{aligned} & \max_{a_0,\mathbf{a}} \mathbb{E}[V_1(a_0,\mathbf{a})] \\ \text{s.t.} \quad & \mathrm{Var}[V_1(a_0,\mathbf{a})] \;=\; \sigma^2 \\ & V_0(a_0,\mathbf{a}) \;=\; v \end{aligned} \tag{E'}$$

Proof. Let's assume A is solution of (E) s.t. $\mathrm{Var}[V_1^{\mathrm{A}}] < \sigma^2$ and $\delta = \sqrt{\frac{\sigma^2}{\mathrm{Var}[V_1^{\mathrm{A}}]}} = 1 + \varepsilon$, with $\varepsilon > 0$. So

$$\begin{aligned} \mathbb{E}[V_1^{\mathrm{A}}] \;=\; & (v - \mathbf{a}^{\mathrm{T}}\mathbf{p_0})(1+r) \\ & + \mathbf{a}^{\mathrm{T}}\mathrm{Diag}(\mathbf{p_0})\boldsymbol{\mu}. \end{aligned} \tag{1.6}$$

Then let's consider the portfolio B $(a_0^{\mathrm{B}}, \delta\mathbf{a}^{\mathrm{T}})$ with a_0^{B} chosen s.t. $V_0^{\mathrm{B}} = v$. Then,

$$\begin{aligned} \mathbb{E}[V_1^{\mathrm{B}}] &= (v - \delta\mathbf{a}^{\mathrm{T}}\mathbf{p_0})(1+r) \\ &\quad + \delta\mathbf{a}^{\mathrm{T}}\mathrm{Diag}(\mathbf{p_0})\boldsymbol{\mu} && (1.7) \\ &= \mathbb{E}[V_1^{\mathrm{A}}] \;+\; \varepsilon\underbrace{(\mathbb{E}[V_1^{\mathrm{A}}] - v(1+r))}_{>0 \text{ if } \sigma\neq 0} \;; && (1.8) \\ \mathrm{Var}[V_1^{\mathrm{B}}] &= \delta^2\mathrm{Var}[V_1^{\mathrm{A}}] && (1.9) \\ &= \sigma^2. && (1.10) \end{aligned}$$

B dominates A which is absurd. □

The problem (E') can be written

$$\begin{aligned} & \max_{a_0,\mathbf{a}} \; a_0(1+r) \;+\; \mathbf{a}^{\mathrm{T}}\mathrm{Diag}(\mathbf{p_0})\boldsymbol{\mu} \\ \text{s.t.} \quad & \mathbf{a}^{\mathrm{T}}\mathrm{Diag}(\mathbf{p_0})\boldsymbol{\Omega}\mathrm{Diag}(\mathbf{p_0})\mathbf{a} \;=\; \sigma^2 \\ & a_0 \;+\; \mathbf{a}^{\mathrm{T}}\mathbf{p_0} \;=\; v \end{aligned}$$

Let's say $\boldsymbol{\omega}_{\mathbf{a}} = \mathrm{Diag}(\mathbf{p_0})\mathbf{a}$ is the composition in \$ of the portfolio in risky assets. One can write (E') as,

$$\begin{aligned} & \max_{a_0,\mathbf{a}} \; (v - \boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\mathbf{1})(1+r) \;+\; \boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\mu} \\ \text{s.t.} \quad & \boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}} \;=\; \sigma^2 \end{aligned}$$

We note $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu} - (1+r)\mathbf{1}$ the extra returns of the risky assets compared to the NRA. By using the Lagrange method,

$$\max_{a_0,\mathbf{a}} \; \mathcal{L}(\boldsymbol{\omega}_{\mathbf{a}}, \lambda) \;, \tag{1.11}$$

with $\mathcal{L}(\boldsymbol{\omega}_{\mathbf{a}}, \lambda) = \boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\tilde{\boldsymbol{\mu}} - \frac{\lambda}{2}(\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}} - \sigma^2)$. Then by deriving

$$\begin{aligned} & \begin{cases} \tilde{\boldsymbol{\mu}} \;+\; \lambda\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}} \;=\; 0 \\ \boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}} - \sigma^2 \;=\; 0 \end{cases} \\ \Leftrightarrow \quad & \begin{cases} \boldsymbol{\omega}_{\mathbf{a}} \;=\; \frac{1}{\lambda}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}} \\ \boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}} \;=\; \sigma^2 \end{cases} \end{aligned}$$

Proposition 1.1. *For portfolios of value v at $t = 0$, the efficient border is made from portfolios s.t.*

$$\begin{aligned}\mathbf{a} &= \frac{1}{\lambda}\text{Diag}(\mathbf{p_0})^{-1}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}} \ ; \\ a_0 &= v - \frac{1}{\lambda}\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{1} \ ; \\ \lambda &= \frac{1}{\sigma}\left(\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}\right)^{\frac{1}{2}} .\end{aligned}$$

1.2.3 Efficient border

For an efficient portfolio,

$$\mathbb{E}[V_1] = v(1+r) + \frac{1}{\lambda}\left(\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}\right) \ ; \quad (1.12)$$

$$\sqrt{\text{VaR}[V_1]} = \frac{\left(\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}\right)^{\frac{1}{2}}}{\lambda}. \quad (1.13)$$

So by setting λ as parameter,

$$\begin{aligned}\mathbb{E}[V_1] &= v(1+r) \\ &\quad + \sqrt{\text{VaR}[V_1]}\left(\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}\right)^{\frac{1}{2}} . \quad (1.14)\end{aligned}$$

We have a line equation, and so we deduce the Proposition 1.2.

Proposition 1.2. *Any efficient portfolio can be written as the linear combination of two particular portfolios P_0 and P^*. Respectively a portfolio of NRA and risky assets.*

Let us talk about the case where there are no NRA on the market. We can then show as previously that the efficient border in the plan (standard deviation, expectation) is a half hyperbola.

We now assume that $\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{1} > 0$. We notice that for a given v we can find a σ s.t. $a_0 = 0$. Let us note $\widetilde{P}^*$ this portfolio. It contains no NRA, he then is also efficient in the case where no NRA is on the market. We then find that the line equation of Proposition 1.2 is necessarily the tangent equation of the hyperbola on the point $\widetilde{P}^*$.

1.2.4 Sharpe ratio

Definition 1.3 (Market Sharpe ratio)**.** We define the market Shape ratio with the quantity:

$$S = \sqrt{\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}}.$$

In the case where $\boldsymbol{\Omega}$ is diagonal we have

$$S = \left(\frac{\sum_{i=1}^{N}\left(\mu_i - (1+r)\right)^2}{\sigma^2}\right)^{\frac{1}{2}} .$$

Definition 1.4 (Sharpe ratio)**.** The Sharpe ratio of the portfolio $(a_0, \mathbf{a})$ is:

$$S(\mathbf{a}) = \frac{\mathbb{E}\left[V_1(a_0,\mathbf{a}) - (1+r)v\right]}{\sqrt{\text{Var}\left[V_1(a_0,\mathbf{a})\right]}}.$$

Recall that $\mathbb{E}[V_1(a_0, \mathbf{a})] = a_0(1 + r) + \boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\mu}$ and $v = a_0 + \boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\mathbf{1}$. And for an efficient porftolio, $\boldsymbol{\omega}_{\mathbf{a}} = \frac{1}{\lambda}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}$ and $\sigma = \frac{1}{\lambda}\left(\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}\right)^{\frac{1}{2}}$. Then,

$$S(\mathbf{a}) = \frac{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\tilde{\boldsymbol{\mu}}}{\left(\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}}\right)^{\frac{1}{2}}}. \quad (1.15)$$

Remark. $S(\mathbf{a})$ is not function of a_0 nor v. All efficient portfolios have the same Sharpe ratio,

$$S(\mathbf{a}) = \frac{1}{\lambda}\frac{\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}}{\sigma} = \left(\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}\right)^{\frac{1}{2}} , \quad (1.16)$$

which is the market Sharpe ratio.

Proposition 1.3. *For all portfolios $(a_0, \mathbf{a})$, we have $S(\mathbf{a}) \leq S$. With equality for the efficient portfolios.*

Proof. Let $\boldsymbol{\Omega} \in \mathcal{S}_n^{++}$, then $\exists \mathbf{R} \in \mathcal{S}_n^{++} \ / \ \boldsymbol{\Omega} = \mathbf{R}^2$.

$$\begin{aligned}S(\mathbf{a}) &\triangleq \frac{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\tilde{\boldsymbol{\mu}}}{\sqrt{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}}}} \\ &= \frac{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Omega}^{-\frac{1}{2}}\tilde{\boldsymbol{\mu}}}{\sqrt{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}}}} \\ &= \frac{\text{Tr}\left((\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\omega}_{\mathbf{a}})^{\mathrm{T}}\boldsymbol{\Omega}^{-\frac{1}{2}}\tilde{\boldsymbol{\mu}}\right)}{\sqrt{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}}}} \\ &\leq \frac{\sqrt{\text{Tr}(\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}})\text{Tr}(\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}})}}{\sqrt{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}}}} \quad \text{(CS)} \\ &\leq \sqrt{\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}}.\end{aligned}$$

□

1.3 Capital Asset Pricing Model

1.3.1 Identification of the portfolio P^*

We add the assumption $\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}} > \mathbf{0}$, *i.e.* we have a long position on all assets. We also assume that there are I investors on the market and they all follow the Markovitz approach. We note λ the risk aversion of the i-th investor. We saw that

$$\mathbf{a}_i = \frac{1}{\lambda_i}\text{Diag}(\mathbf{p_0})^{-1}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}. \quad (1.17)$$

Then the total supply in risky assets is:

$$\sum_{i=1}^{I}\frac{1}{\lambda_i}\text{Diag}(\mathbf{p_0})^{-1}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}} = \frac{1}{\bar{\lambda}}P^* , \quad (1.18)$$

where $\bar{\lambda} = \left(\sum_{i=1}^{I}\frac{1}{\lambda_i}\right)^{-1}$. When the market is at equilibrium the supply is equal to the demand ; and the assets' prices are the parameters that lead to this equilibrium.

Proposition 1.4. *P^* is proportional to the "market portfolio", that is to say the portfolio containing all the risky assets on the market. And the prices adjust themselves so that*

$$\mathbf{a}_m = \frac{1}{\lambda}\mathrm{Diag}(\mathbf{p_0})^{-1}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}} ,$$

$\mathbf{a}_m$ is the vector where its components are the amount of equities for the i-th risky asset.

1.3.2 CAPM equation

We note $\mathbf{z} = \mathbf{y} - (1+r)\mathbf{1}$ the vector of net returns and $z(\mathbf{a})$ the net return of a portfolio with composition $\mathbf{a}$ in risky assets.

$$\begin{aligned} z(\mathbf{a}) &\triangleq \frac{\mathbf{a}^{\mathrm{T}}\mathbf{p_1}}{\mathbf{a}^{\mathrm{T}}\mathbf{p_0}} - (1+r) \\ &= \frac{\mathbf{a}^{\mathrm{T}}\mathrm{Diag}(\mathbf{p_0})(\mathbf{z}+(1+r)\mathbf{1})}{\mathbf{a}^{\mathrm{T}}\mathrm{Diag}(\mathbf{p_0})\mathbf{1}} - (1+r) \\ &= \frac{\mathbf{a}^{\mathrm{T}}\mathrm{Diag}(\mathbf{p_0})\mathbf{z}}{\mathbf{a}^{\mathrm{T}}\mathrm{Diag}(\mathbf{p_0})\mathbf{1}} = \frac{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\mathbf{z}}{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\mathbf{1}} , \end{aligned}$$

and for an efficient portfolio, $\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}} \propto \boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\mu}}$, *i.e.*

$$z(\mathbf{a}) = \frac{\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{z}}{\tilde{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{1}}. \tag{1.19}$$

Lemma 1.5. *Recall that with X and Y two r.v. in L^2, one can find $a, b \in \mathbb{R}$, ε r.v. s.t. $Y = a + bX + \varepsilon$, with $\mathbb{E}[\varepsilon] = 0$ and $\mathrm{Cov}(X, \varepsilon) = 0$.*

Proof. Indeed with $b := \frac{\mathrm{Cov}(Y,X)}{\mathrm{Var}[X]}$, $a := \mathbb{E}[Y] - b\mathbb{E}[X]$ and $\varepsilon := Y - a - bX$. □

Remark. a and b are unique.

Proposition 1.6. *We have the decomposition $\mathbf{z} = \boldsymbol{\beta}z(\mathbf{a}_m) + \boldsymbol{\varepsilon}$, with $\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\varepsilon} \in \mathbb{R}^N$, $z(\mathbf{a}_m) \in \mathbb{R}$ and $\mathbf{a}_m$ the market portfolio. The vector $\boldsymbol{\beta}$ is called the "beta" vector, $\boldsymbol{\beta} = \frac{\mathrm{Cov}(\mathbf{z}, z(\mathbf{a}_m))}{\mathrm{Var}[z(\mathbf{a}_m)]}$, $\mathbb{E}[\boldsymbol{\varepsilon}] = 0$ and $\mathrm{Cov}(\boldsymbol{\varepsilon}, z(\mathbf{a}_m)) = 0$.*

Remark. For all $i \in [\![1, N]\!]$, $z_i = y_i - (1+r) = \beta_i z(\mathbf{a}_m) + \varepsilon_i$ with $\beta_i = \frac{\mathrm{Cov}(z_i, z(\mathbf{a}_m))}{\mathrm{Var}[z(\mathbf{a}_m)]}$.

Proof. □

Corollary. *We have for all $i \in [\![1, N]\!]$,*

$$\mathbb{E}[z_i] = \beta_i \mathbb{E}[z(\mathbf{a}_m)] ; \tag{1.20}$$

$$\mathrm{Var}[z_i] = \beta_i^2 \mathrm{Var}[z(\mathbf{a}_m)] + \sigma_i^2. \tag{1.21}$$

With $\mathbf{a}_m$ the market pottfolio and $\sigma_i^2 = \mathrm{Var}[\varepsilon_i]$.

1.3.3 Comments

We can see that the expectation of return just depends of the Beta and the expected return of the market portfolio.

In (1.21) we call $\beta_i^2\mathrm{Var}[z(\mathbf{a}_m)]$ the systemic risk and σ_i^2 the idiosyncratic risk, which is uncorrelated from the economy.

We say that the idiosyncratic risk can be diversified, contrary to the systemic risk. Indeed let us consider the portfolio of composition $\mathbf{a}$ in risky assets, then $z(\mathbf{a}) = \frac{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\mathbf{z}}{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\mathbf{1}}$, and with what we saw previously,

$$\begin{aligned} z(\mathbf{a}) &= \frac{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\beta}z(\mathbf{a}_m^*)}{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\mathbf{1}} + \frac{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\varepsilon}}{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\mathbf{1}} \\ &= \beta(\mathbf{a})z(\mathbf{a}_m^*) + \varepsilon(\mathbf{a}). \end{aligned}$$

The Beta of a portfolio is the weighted average of the Beta of different assets. Then,

$$\begin{aligned} \mathbb{E}[z(\mathbf{a})] &= \beta(\mathbf{a})\mathbb{E}[z(\mathbf{a}_m^*)] ; \\ \mathrm{Var}[z(\mathbf{a})] &= \beta^2(\mathbf{a})\mathrm{Var}[z(\mathbf{a}_m^*)] + \mathrm{Var}[\varepsilon(\mathbf{a})] \\ &= \beta^2(\mathbf{a})\mathrm{Var}[z(\mathbf{a}_m^*)] + \bar{\boldsymbol{\omega}}_{\mathbf{a}}^{\mathrm{T}}\mathrm{Var}[\boldsymbol{\varepsilon}]\bar{\boldsymbol{\omega}}_{\mathbf{a}} , \end{aligned}$$

where $\bar{\boldsymbol{\omega}}_{\mathbf{a}} = \frac{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}}{\boldsymbol{\omega}_{\mathbf{a}}^{\mathrm{T}}\mathbf{1}}$. So if one wants to reduce the risk of the first member by reducing $\beta(\mathbf{a})$ but then one reduces also the expected return. We say that the systemic risk cannot be diversified.

On the other hand, if there is a lot of different assets in the portfolio, the idiosyncratic risk will be low. Indeed, consider[2] to simplify that $\mathrm{Var}[\boldsymbol{\varepsilon}] = \sigma^2\mathbf{I}_N$ and for all $i \in [\![1, N]\!]$, $\bar{\boldsymbol{\omega}}_{\mathbf{a},i} = \frac{1}{N}$,

$$\begin{aligned} \bar{\boldsymbol{\omega}}_{\mathbf{a}}^{\mathrm{T}}\mathrm{Var}[\boldsymbol{\varepsilon}]\bar{\boldsymbol{\omega}}_{\mathbf{a}} &= \sum_{i=1}^{N} \bar{\boldsymbol{\omega}}_{\mathbf{a},i}^2\sigma^2 \\ &= \frac{\sigma^2}{N} \xrightarrow[N\to\infty]{} 0. \end{aligned}$$

The graph which represents the $(\beta(\mathbf{a}), \tilde{\mu}(\mathbf{a}))$ is a line where the coefficient is the expected return of the market portfolio. We call it the "security market line", which should not be mixed up with "capital market line"[3].

One can show that the rank of $\boldsymbol{\Sigma} := \mathrm{Var}[\boldsymbol{\varepsilon}]$ is $N-1$ and

$$\boldsymbol{\Sigma} = \boldsymbol{\Omega} - \frac{\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}_m^*}\boldsymbol{\omega}_{\mathbf{a}_m^*}^{\mathrm{T}}\boldsymbol{\Omega}}{\boldsymbol{\omega}_{\mathbf{a}_m^*}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega}_{\mathbf{a}_m^*}}.$$

1.3.4 Markovitz and CAPM in practice

To use the previous approaches we have to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$. Assume that the returns are iid and we observe them on n periods, so we have the data $(\mathbf{y}_t)_{1\le t\le n}$,

$$\begin{aligned} \boldsymbol{\mu} &= \mathbb{E}[\mathbf{y}_t] ; \\ \boldsymbol{\Omega} &= \mathrm{Var}[\mathbf{y}_t]. \end{aligned}$$

[2] We will see later that it is in fact impossible.

[3] This is the name of the efficient border.

The estimators associated to the empirical moments are,

$$\begin{aligned}\hat{\boldsymbol{\mu}}_n &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_t \ ; \\ \hat{\boldsymbol{\Omega}}_n &= \frac{1}{n}\sum_{i=1}^{n}(\mathbf{y}_t-\hat{\boldsymbol{\mu}}_n)(\mathbf{y}_t-\hat{\boldsymbol{\mu}}_n)^{\mathrm{T}}.\end{aligned}$$

Proposition 1.7. *We have* $\hat{\boldsymbol{\mu}}_n \xrightarrow[n\to\infty]{a.s.} \boldsymbol{\mu}$ *and* $\hat{\boldsymbol{\Omega}}_n \xrightarrow[n\to\infty]{a.s.} \boldsymbol{\Omega}$ *with the speed* $\sqrt{n}$.

In fact $\hat{\boldsymbol{\Omega}}_n$ is a very noisy estimator and its use can have catastrophic consequences in term of risk. In practice we use the following recipe. First we prefer to work on the empirical correlation matrix $\mathbf{E}$. We can write it as,

$$\mathbf{E} = \sum_{k=1}^{N}\lambda_k\mathbf{V}^k(\mathbf{V}^k)^{\mathrm{T}} \ ,$$

with $\lambda_1 \leq \cdots \leq \lambda_N$ the eigenvalues and $\mathbf{V}^k$ the eigenvectors associated. We would like to apply the following procedure:

$$\begin{aligned}\lambda^k_{clean} &= \begin{cases} a & \text{for } k > k^* \\ \lambda_k & \text{for } k \leq k^* \end{cases} \\ \mathbf{E} &= \sum_{k=1}^{k^*}\lambda_k\mathbf{V}^k(\mathbf{V}^k)^{\mathrm{T}} \ + \ a\mathbf{I}_N.\end{aligned}$$

This comes from the PCA method, we keep the significant values of eigenvalues and replace the others with a constant to keep the trace. Now the question is how to choose k^* ?

Theorem 1.8. *Assume that* $n \to \infty$, $N \to \infty$ *s.t.* $\frac{N}{n} = Q < 1$ *constant. Assume that* $\mathbf{E}_n$ *the empirical correlation matrix of iid returns and the theoretical correlation matrix of returns is* $\mathbf{I}_N$*. So a.s. the convergence of the empirical measure associated to the eigenvalues of* $\mathbf{E}_n$ *goes towards the Marčenko–Pastur distribution:*

$$\rho(\lambda) = \frac{1}{2\pi}\frac{\sqrt{(\lambda^+-\lambda)(\lambda-\lambda^-)}}{\lambda} \ ,$$

where $\lambda^- < \lambda < \lambda^+$, $\lambda^+ := (1+\sqrt{Q})^2$ *and* $\lambda^- := (1-\sqrt{Q})^2$.

So in practice we will keep the eigenvalues greater than λ^+, indeed even in the case where there is no information, a statistical effect makes that eigenvalues are computes in the range $[\lambda^-, \lambda^+]$.

One can also estimate the parameters of the CAPM. For each risky asset we have the regression

$$z_{i,t} = \alpha_i \ + \ \beta_i z_{i,t}(\mathbf{a}^*_m) \ + \ \varepsilon_{i,t}.$$

And with the hypothesis of iid returns we are in the frame of least square regression, so we can estimate α_i and β_i ; and test the nullity of α_i.

However, the CAPM theory being valid for all the risky assets, it is natural to consider the piled regressions

$$\begin{aligned} z_{1,1} &= \alpha_1+\beta_1 z_1(\mathbf{a}^*_m)+\varepsilon_{1,1} \\ &\vdots \\ z_{N,1} &= \alpha_N+\beta_N z_1(\mathbf{a}^*_m)+\varepsilon_{N,1} \\ z_{1,2} &= \alpha_1+\beta_1 z_2(\mathbf{a}^*_m)+\varepsilon_{1,2} \\ &\vdots \end{aligned}$$

where the first subscript correspond to the i-th asset and the second to the time. In this case we have $\varepsilon_{i,t_k} \perp\!\!\!\perp \varepsilon_{j,t_l}$ for $t_k \neq t_l$, but $\mathbb{E}[\varepsilon_{i,t}\varepsilon_{j,t}] \neq 0$, hence $\mathrm{Var}[\boldsymbol{\varepsilon}] \neq \sigma\mathbf{I}_N$. It is then desirable to use the GLS[4] method or the QGLS[5] method instead of the OLS.

But for this structure, as the variable always depends of the market portfolio, one can show that the OLS estimator is the same as the GLS and we have

$$\begin{aligned}\hat{\alpha}_n &= \hat{\mu}-\hat{\beta}_n\hat{\mu}_n(\mathbf{a}^*_m) \ ; \\ \hat{\beta}_n &= \frac{\frac{1}{n}\sum_{t=1}^{n}(z_t-\hat{\mu}_n)(z_t(\mathbf{a}^*_m)-\hat{\mu}_n(\mathbf{a}^*_m))}{\hat{\sigma}^2_n(\mathbf{a}^*_m)} \ ,\end{aligned}$$

where $\hat{\mu}_n = \frac{1}{n}\sum_{t=1}^{n} z_t$, $\hat{\mu}_n(\mathbf{a}^*_m) = \frac{1}{n}\sum_{t=1}^{n} z_t(\mathbf{a}^*_m)$ and $\hat{\sigma}^2_n(\mathbf{a}^*_m) = \frac{1}{n}\sum_{t=1}^{n}(z_t(\mathbf{a}^*_m)-\hat{\mu}_n(\mathbf{a}^*_m))^2$.

We can also estimate $\boldsymbol{\Sigma} = \mathrm{Var}[\boldsymbol{\varepsilon}]$, and we can show that

$$\sqrt{n}\begin{pmatrix}\hat{\alpha}_n-\alpha \\ \hat{\beta}_n-\beta\end{pmatrix} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1+\frac{\mu^2(\mathbf{a}^*_m)}{\sigma^2(\mathbf{a}^*_m)} & -\frac{\mu^2(\mathbf{a}^*_m)}{\sigma^2(\mathbf{a}^*_m)} \\ -\frac{\mu^2(\mathbf{a}^*_m)}{\sigma^2(\mathbf{a}^*_m)} & \frac{1}{\sigma^2(\mathbf{a}^*_m)}\boldsymbol{\Sigma}\end{pmatrix}\right).$$

Recall that if $Y_n \xrightarrow{\mathcal{L}} Y$ with $Y \sim (\mathbf{0},\boldsymbol{\Sigma})$, $\boldsymbol{\sigma} \in \mathcal{S}^{++}_K$, then

$$Y_n^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}Y \xrightarrow{\mathcal{L}} \chi^2(K).$$

So we have in our case

$$\left(1+\frac{\mu^2(\mathbf{a}^*_m)}{\sigma^2(\mathbf{a}^*_m)}\right)^{-1} n\alpha_n^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^*_n\hat{\alpha}_n \xrightarrow{\mathcal{L}} \chi^2(N-1) \ ,$$

where $\hat{\boldsymbol{\Sigma}}^*_n$ is the generalized inverse of $\boldsymbol{\Sigma}$ as it is non inversible ($\mathbf{A}^*$ is the generalized inverse if $\mathbf{A}\mathbf{A}^*\mathbf{A} = \mathbf{A}$). We can easily show that $\boldsymbol{\Omega}^{-1}$ is a generalized inverse of $\boldsymbol{\Sigma}$.

2 Linear regression beyond OLS

2.1 The model

We consider the model

$$Y = X\theta \ + \ \xi \ ,$$

[4]Generalized Least Squares.

[5]Quasi-Generalized Least Squares.

where $Y, \xi \in \mathbb{R}^n$, $\theta \in \mathbb{R}^p$ and $X \in \mathcal{M}_{n,p}(\mathbb{R})$. And $\mathbb{E}[\xi|X] = 0$, $\text{Var}[\xi] = \sigma^2 \mathbf{I}_n$. We observe (X,Y), and the OLS estimator of θ is

$$\hat{\theta} = (X\text{T}X)^{-1}X^{\text{T}}Y.$$

Theorem 2.1 (Gauss–Markov)**.** *The OLS estimator has the minimal variance in the set of unbiased estimator.*

2.2 Limits of the OLS

The first problem is that this estimator favours the bias rather than the variance, which could be a problem.

The OLS doesn't compute null coefficients, there is no variable selection, so the interpretation is difficult.

We need to have $X^{\text{T}}X$ inversible and so $n \geq p$.

To correct the last point we could *e.g.* reject the coefficient that has the highest p-value according to its nullity test, then compute again the regression *etc.* until the distance $\|Y - X\hat{\theta}_{p-k}\|_2^2$ is not stable anymore.

2.3 Ridge regression

The Ridge regression allows us to correct the points 1 and 3 previously detected. The idea is to penalize the distance of OLS and minimize

$$F(\theta) := \|Y - X\theta\|_2^2 + \lambda\|\theta\|_2^2.$$

When $\lambda \to 0$ we have $\hat{\theta}^R = \hat{\theta}^{OLS}$, and when $\lambda \to \infty$ we have $\hat{\theta}^R = 0$. We can show that

$$\nabla F(\theta) = -2X^{\text{T}}(Y - X\theta) + 2\lambda \mathbf{I}_p \theta \ ,$$

so the theta that minimizes this distance is

$$\hat{\theta}^R = (X^{\text{T}}X + \lambda \mathbf{I}_p)^{-1}X^{\text{T}}Y.$$

We have the properties

$$\begin{aligned}\mathbb{E}[\hat{\theta}^R] &= (X^{\text{T}}X + \lambda \mathbf{I}_p)^{-1}X^{\text{T}}X\theta \\ \text{Var}[\hat{\theta}^R] &= \sigma^2(X^{\text{T}}X + \lambda \mathbf{I}_p)^{-1}X^{\text{T}}X(X^{\text{T}}X + \lambda \mathbf{I}_p)^{-1}\end{aligned}$$

Let us now compare $\hat{\theta}^R$ and $\hat{\theta}^{OLS}$ with $p = 1$: $\hat{\theta}^R = \left(\|x\|^2 + \lambda\right)^{-1} X^{\text{T}}Y$ and $\hat{\theta}^{OLS} = \left(\|x\|^2\right)^{-1} X^{\text{T}}Y$. So we have

$$\hat{\theta}^R = \underbrace{\frac{\left(\|x\|^2 - \lambda\right)^{-1}}{\left(\|x\|^2\right)^{-1}}}_{=:c(\lambda)} \hat{\theta}^{OLS}.$$

Let us see the quadratic error:

$$\begin{aligned}\mathbb{E}\left[(\hat{\theta}^R - \theta)^2\right] &= \text{Var}[\hat{\theta}^R] + \left(\mathbb{E}[\hat{\theta}^R] - \theta\right)^2 \\ &= c(\lambda)^2\text{Var}[\hat{\theta}^{OLS}] + ((c(\lambda) - 1)\theta)^2 \\ &= c^2\left(\text{Var}[\hat{\theta}^{OLS}] + \theta^2\right) - 2c\theta^2 + \theta^2\end{aligned}$$

which is a parabole in c that reaches its minimum in $\frac{\theta^2}{\text{Var}[\hat{\theta}^{OLS}]+\theta^2}$ and its value in $c = 1$ is $\text{Var}[\hat{\theta}^{MCO}]$. We can then improve the OLS estimator in the sense of the quadratic error.

The problem is how to choose λ when λ_{opt} is a function of θ which we don't know. In practice we use the cross validation. We divise the data in K sets of equal size, and for each $k \in [\![1, K]\!]$ we compute $\hat{\theta}_{-k}(\lambda)$ on all datas except those of the k-th set. Then we compute the error on the set k:

$$\varepsilon_k^\lambda := \frac{1}{\text{Card}(k\text{-th set})} \sum_{(Y_j, X_j) \in k} \left(Y_j - X_J\hat{\theta}_{-k}(\lambda)\right)^2.$$

The total error is the average over all the k-th sets: $\varepsilon^\lambda := \frac{1}{K}\sum_{k=1}^K \varepsilon_k^\lambda$. And we choose the λ that minimizes this error.

2.4 The LASSO regression (Least Absolute Shrinkage and Selection Operator)

2.4.1 Definition

The only difference with the Ridge regression is that we replace the norme L^2 with the norm L^1:

$$\hat{\beta}^L = \min_\beta \|Y - X\beta\|_2^2 + \lambda\sum_{j=1}^p |\beta_j|.$$

This is equivalent to

$$\begin{aligned} &\min_\beta \|Y - X\beta\|_2^2 \\ u.c. \quad &\sum_{j=1}^p |\beta_j| \leq t.\end{aligned}$$

2.4.2 Interest

In big dimension we usually are looking for sparse solutions *i.e.* with a lot of null coefficient. The idea would be to solde

$$\begin{aligned} &\min_\beta \|Y - X\beta\|_2^2 \\ u.c. \quad &\sum_{j=1}^p |\beta_j|_0 \leq t\,,\end{aligned}$$

where the norme L^0 counts the number of non-zero coefficients. The problem is that the complexity of this is "NP hard". So we replace this norm with the norm L^q with $q = 1$, which is the smallest q for which the constraint region is convex.

3 Principal Component Analysis (PCA)

The goal here is to find a pertinent representation of a scatter plot that are originally in huge dimension.

3.1 Datas

We have a table of n rows and p columns. In row $(x_i^1, \dots, x_i^p)$ is the individual variable i, while in column $(x_1^j, \dots, x_n^j)$ is the variable j.

In our example the individual variables are trading days and variables are the returns of assets. In general we work with the centred and normalized variables:

$$x_i^j = \frac{x_i^j - \bar{x}^j}{\sqrt{\frac{1}{n}\sum_{i=1}^n (x_i^j - \bar{x}^j)^2}},$$

where $\bar{x}^j = \frac{1}{p}\sum_{l=1}^p x_l^j$.

3.2 Space of variables

Let us consider a distance between individuals, typically the euclidean distance:

$$\begin{aligned} d^2(x_i, x_j) &= \|x_i - x_j\|^2 \\ &= \sum_{k=1}^p (x_i^k - x_j^k). \end{aligned}$$

And we have $\bar{x}^k = 0$ for $k \in [\![1, p]\!]$, so the center of mass of the scatter plot is $\vec{0}$.

Definition 3.1 (Inertia)**.** We call inertia of the scatter the dispersion around the center of mass, in the non-normalized case

$$I = \frac{1}{n}\sum_{i=1}^n d^2(\vec{0}, x_i).$$

The bigger I is, the more points are dispersed.

$$\begin{aligned} I &= \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^p \left(x_i^j\right)^2 \\ &= \sum_{j=1}^p \underbrace{\frac{1}{n}\sum_{i=1}^n \left(x_i^j\right)^2}_{=1 \text{ in our case}} = p = \mathrm{Tr}(V). \end{aligned}$$

Where V is the empirical variance matrix between the variables.

The inertia carried buy a linear subspace F is defined by

$$I_F = \frac{1}{n}\sum_{i=1}^n d^2(P_F x_i, \vec{0})$$

with $P_F x_i$ the orthogonal projection of x_i on F.

In the case of a subspace of dimension 1, where the span vector is $\vec{u}$ with $\|\vec{u}\| = 1$,

$$P_F x_i = \langle u, x_i\rangle \vec{u}$$

and $I_{\vec{u}} = \frac{1}{n}\sum_{i=1}^n \langle u, x_i\rangle^2$.

[6] Sub-vector spaces.

3.3 PCA

We are looking for the subspace that F_k of dimension $k < p$ s.t. after projection the scatter is as least deformed as we can.

Theorem 3.1. *Le subspace F_k that caters the highest inertia is spanned by the eigenvectors $\vec{v}_1, \dots, \vec{v}_k$ associated to the eigenvalues $\lambda_1 < \dots < \lambda_k$. And this inertia is $\sum_{j=1}^k \lambda_j$.*

Proof. We are looking for the svs[6] F_k of dimension $k < p$ s.t. after projection the scatter is as less skewed as we can. The projection is decreasing the distance between individual variables, so we are looking for the svs F_k which maximizes

$$\begin{aligned} &\sum_{i=1}^n \sum_{l=1}^n d^2(P_{F_k} x_i, P_{F_k} x_l) \\ &= \sum_{i=1}^n \sum_{l=1}^n \sum_{j=1}^p \left((P_{F_k} x_i)^j\right)^2 + \left((P_{F_k} x_l)^j\right)^2 \\ &= 2n^2 I_{F_k}. \end{aligned}$$

So we are looking for the svs that carries the biggest inertia.

Let us assume that we know F_k, so which is F_{k+1} ? Let E_{k+1} be of dimension $k+1$, $\dim F_k^\perp = p - k$ and

$$\begin{aligned} \underbrace{\dim(E_{k=1} + F_k^\perp)}_{\le p} &= \underbrace{\dim E_{k+1} + \dim F_k^\perp}_{=p+1} \\ &\quad - \dim E_{k+1} \cap F_k^\perp, \end{aligned}$$

and so $\dim E_{k+1} \cap F_k^\perp \geq 1$. Let now $v \in E_{k+1} \cap F_k^\perp$. We can write $E_{k+1} = v \oplus G$, with G orthogonal complement of v in E_{k+1}. So $\dim G = k$.
Let $\tilde{F}_{k+1} = F_k \oplus v$. As $v \perp V$ and $v \perp F_k$,

$$\begin{aligned} I_{E_{k+1}} &= I_G + I_v ; \\ I_{\tilde{F}_{k+1}} &= I_{F_k} + I_v. \end{aligned}$$

By definition $I_{F_k} \geq I_G$ so $\forall E_{k+1}$, $I_{E_{k+1}} \leq I_{\tilde{F}_{k+1}}$ if we choose v s.t. I_v is maximal. And so F_{k+1} is the orthogonal sum of F_k and v with v the axis orthogonal to F_k of maximum inertia.

To find a space F_k we can then look for the axis one after the other.

Let us now look for the axis that carries the maximal inertia. The inertia carried by an axis Δ_u of director vector $\vec{u}$ is

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n \langle x_i, u\rangle^2 &= \frac{1}{n}\sum_{i=1}^n (x_i^{\mathrm{T}} u)^2 \\ &= \frac{1}{n}\sum_{i=1}^n u^{\mathrm{T}} x_i x_i^{\mathrm{T}} u \\ &= u^{\mathrm{T}} \left(\frac{1}{n}\sum_{i=1}^n x_i x_i^{\mathrm{T}}\right) u \\ &= u^{\mathrm{T}} V u. \end{aligned}$$

We will then maximize $u^{\mathrm{T}}Vu$ under the constraint $u^{\mathrm{T}}u = 1$. We have $V = PDP^{\mathrm{T}}$, with $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, with $\lambda_1 \geq \cdots \geq \lambda_p$. P is the orthogonal matrix that contains the eigenvectors of V, (v_j) with $\|v_j = 1$.

$$P^{\mathrm{T}}u = \begin{pmatrix} \langle u, v_1 \rangle \\ \vdots \\ \langle u, v_p \rangle \end{pmatrix}$$

$$\begin{aligned} u^{\mathrm{T}}Vu &= \sum_{j=1}^{p} \lambda_j \langle u, v_j \rangle^2 \\ &\leq \lambda_1 \sum_{j=1}^{p} \langle u, v_j \rangle^2 = \lambda_1. \end{aligned}$$

We then choose $\vec{u} = v_1$ that maximizes the inertia (λ_1 in this case). For the next axis we are looking for $\vec{u} \perp v_1$ that maximizes

$$\sum_{j=2}^{p} \lambda_j \langle u, v_j \rangle^2 \leq \lambda_2.$$

So we choose $\vec{u} = v_2$. □

Part II

4 Modeling high frequency data

4.1 Introduction

The need for models in the derivatives world comes from the problem of pricing and hedging. At the macroscopic scale, for example if we consider one data per day over one year, prices trajectories look like sample paths of models classically used in mathematical finance such as Brownian motion, diffusions, stochastic volatility models

At the microscopic scale, prices are very different from Brownian type sample paths.

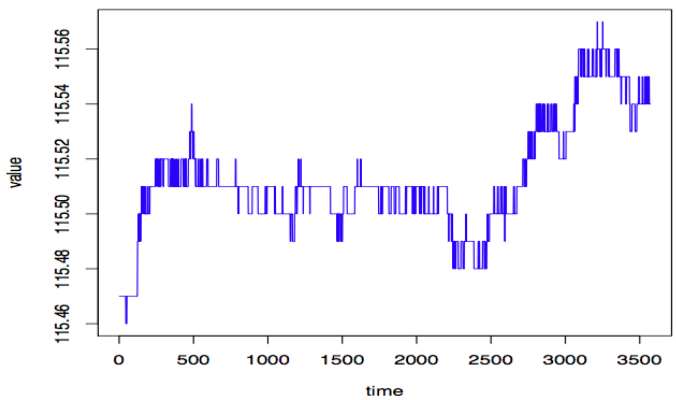

Figure 1: Bund, one hour, one data every second

Essentially, a model is good if it reasonably reproduces the stylized facts of the main quantities of interest and, mostly, if it is useful for market practitioners.

4.2 A first approach

We distinguish four levels of resolution for modeling "through scales". A good probabilistic/statistical model should provide reasonable dynamics across these levels.

L0 The ultimate level of the order book. One proposes a complex stochastic system in continuous time with discrete values in an appropriate state space which describes all the events of a limit order book.

L1 The ultra high frequency level for the price. At this level, one wishes to model all transaction prices and durations between these transactions.

L2 At an intermediate high frequency level. Here one does not focus on durations but essentially on the price, regularly sampled, for example every second or every minute.

L3 The macroscopic level, where the price is viewed as a continuous semi-martingale. It is the dominant and historical approach.

Remark. We will essentially focus on price models here and so only consider levels L1, L2, L3. When is the model valid ? For 5 minutes sampling ? 1 minute sampling ? 1 second sampling ?

Another issue is the question of the modeled price. Is it the last traded ? mid-quote ? best bid ? best ask ? Volume Weighted Average Price ?

4.2.1 From the large scales to the fine scales

It is well known that the no free lunch assumption is essentially only compatible with semi-martingale type dynamics for the price. But various empirical study have shown that over short time periods (order of magnitude of an hour or a day), it is not reasonable to see price observations as observations of a continuous semi-martingale.

Definition 4.1 (Signature plot)**.** Assume we have price observations P_t at times $t = \frac{i}{n}$, $n \in \mathbb{N}$, $i = 0, \ldots, n$, where $t = 1$ represents for example one trading day. The signature plot is the function which to $k = 1, \ldots, n$ associates

$$RV_n(k) \quad = \quad \sum_{i=0}^{\lfloor \frac{n}{k} \rfloor - 1} \left(P_{k\frac{i+1}{n}} - P_{k\frac{i}{n}} \right)^2$$

If P_t is a continuous semi-martingale, as soon as $\frac{n}{k}$ is large enough, $RV_n(k)$ is close to the quadratic variation of the semi-martingale. However, in practice, $RV_n(k)$ is very often a decreasing functional, stabilizing for sampling periods larger than 10 minutes (depending on the asset of course).

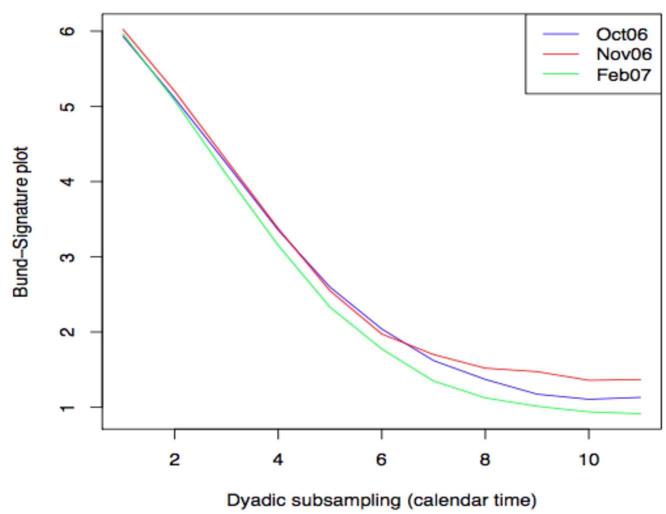

Figure 2: Signature plot for Bund contract

It is clear that high frequency prices time series are not of the same nature as low frequency series. The goal of the coarse to fine approach to microstructure is the following : reconcile these different behaviours across scales starting from the coarse scale, that is the scale where a continuous semi-martingale type dynamics is a relevant model. More precisely, one starts from a continuous semi-martingale, called efficient (latent) price, and apply a stochastic mechanism to it in order to derive the observed prices at higher frequencies.

4.2.2 L3 to L2 models

Additive microstructure noise models In the coarse to fine approach, one can consider the widely used notion of microstructure noise. This noise is defined for all t where the model price exists by

$$\varepsilon_t \;=\; \ln P_t \,-\, \ln X_t\,,$$

where P_t is the model price at time t and X_t is the semi-martingale used in order to build the model. In the so called additive microstructure noise approach, one focuses on modeling the process ε_t.

Some simple considerations enable us to deduce some suitable properties of the observed price and the microstructure noise at level L1:

(i) Observed prices are discrete. Indeed, market prices usually stay on a tick grid. This is not compatible with data from a continuous semi-martingale observed at exogenous times. Therefore, in this approach, this is a "reason" for the presence of microstructure noise.

(ii) For many assets, there are quick oscillations of the transaction prices between two values ("bid-ask bounce").

Since $(P_t)_t$ is a jump process so that the number of jumps is finite over every bounded time interval, we deduce that

(iii) $(\ln P_t)_t$ is almost surely finite.

Additive microstructure noise models are very convenient to carry on computations, see later, and are often reasonable when considering sampling scales larger than about 5 minutes. However, they do not satisfy any of the properties (i), (ii), (iii) and the durations are not modeled. Consequently, they cannot be extended to level L1.

Rounding models Rounding models are a simple way to accommodate properties (i), (ii), (iii) and the assumption of an underlying semi-martingale efficient price. In this model, the efficient price is modeled by a continuous semi-martingale Xt and the observed prices are given by the sample

$$\left(X^{(\alpha_n)}_{\frac{i}{n}} := \alpha_n \left\lfloor \frac{X_{\frac{i}{n}}}{\alpha_n} \right\rfloor,\, i = 0, \ldots, n \right),$$

with rounding error α_n, corresponding to the tick size.

Hence, compare to additive microstructure noise models, rounding models have several nice properties. However, the main drawback of these models is that, as additive microstructure noise models, they remain L3 $\to$ L2 models and cannot be extended to level L1: Assume that for any time t the observed price is given by the rounding value of a semi-martingale. This leads to an observed price with an infinite number of jumps on a finite interval, which is of course hardly acceptable.

4.2.3 L3 to L1 models

It appears that several drawbacks are inherent to L3 $\to$ L2 models. Let us summarize the suitable properties for a L3 $\to$ L1 model:

- A continuous semi-martingale type behaviour at large sampling scales.
- Model for prices and durations (in particular, no notion of sampling frequency is required).
- A clear definition of the price.
- Discrete prices.
- Bid-Ask bounce.
- Usual stylized facts of returns, durations and volatility (in a loose sense here). In particular, inverse relation between durations and volatility.
- An interpretation of the model.
- Finite quadratic variation for the price.
- A testable model.
- A useful model, for example for building statistical procedures.

4.2.4 Fine to coarse models (L1 to L3)

5 Optimal portfolio liquidation

5.1 Introduction

We want to sell a large quantity of a stock (or of several stocks) in one day. How to choose a good way to split this large order in time and volume ?

There are two extreme strategies:

- Sell everything right now, hence huge transaction cost since we need to "eat" a lot in the order book. However this cost is known.
- Sell regularly in the day small amounts of assets, so small transaction costs (volumes are much smaller) but the final profit is unknown because of the daily price fluctuations: volatility risk.

We need to optimize between transaction costs and volatility risk. To do so, we use the Almgren and Chriss framework which takes into account the market impact phenomenon and emphasizes the importance of having good statistical estimators of market parameters.

5.2 Almgren and Chriss model

5.2.1 Setup

We consider we are selling one asset, we have X shares of this assets at $t_0 = 0$, everything has to be sold at $t = T$. The interval $[0, T]$ is split into N intervals of length $\tau = \frac{T}{N}$ and set $t_k = k\tau$, $k \in [\![0, N]\!]$. A trading strategy is a vector $(x_0, \cdot, x_N)$, with x_k the number of shares we still have at time t_k.

So $x_0 = X$, $x_N = 0$ and $n_k = x_{k-1} - x_k$ is the number of assets sold between t_{k-1} and t_k (and not exactly at time t_{k-1} or t_k), decided at time t_{k-1}.

5.2.2 Market impacts

Permanent impact component. Intuitively market participants see us selling large quantities, thus they revise their prices down. Therefore, the "equilibrium price" of the asset is modified in a permanent way. Let S_k be be equilibrium price at time t_k:

$$S_k \;=\; S_{k-1} \;+\; \sigma\sqrt{\tau}\xi_k \;+\; \tau g\left(\frac{n_k}{\tau}\right),$$

with ξ_k i.i.d. r.v. $\sim \mathcal{N}(0, 1)$.

Temporary impact component. This effect is due to the transaction costs: we are liquidity taker since we consume the liquidity available in the order book. If we sell a large amount of shares, our price per share is significantly worse than when selling only one share. We assume this effect is temporary and the liquidity comes back after each period. Let $\tilde{S}_k = \frac{1}{n_k}\sum n_{k,i}p_i$, with $n_{k,i}$ the number of shares sold at price p_i between t_{k-1} and t_k. We set

$$\tilde{S}_k \;=\; S_{k-1} \;-\; h\left(\frac{n_k}{\tau}\right).$$

PnL. The result of the sell of the asset is

$$\begin{aligned}\sum_{k=1}^{N} n_k\tilde{S}_k \;&=\; XS_0 \;+\; \sum_{k=1}^{N} x_k\left(\sigma\sqrt{T}\xi_k - \tau g\left(\frac{n_k}{\tau}\right)\right)\\ &\quad -\; \sum_{k=1}^{N} n_k h\left(\frac{n_k}{\tau}\right).\end{aligned}$$

The trading cost $\mathcal{C} := XS_0 - \sum_{k=1}^{N} n_k\tilde{S}_k$ is equal to the addition of:

- the volatility cost
- the permanent impact cost
- the temporary impact cost.

Consider a static strategy (fully known in t_0), which is in fact optimal in this framework. We have

$$\begin{aligned}\mathbb{E}[\mathcal{C}] \;&=\; \tau\sum_{k=1}^{N} x_k g\left(\frac{n_k}{\tau}\right) \;+\; \sum_{k=1}^{N} n_k h\left(\frac{n_k}{\tau}\right)\;;\\ \mathrm{Var}[\mathcal{C}] \;&=\; \sigma^2\tau\sum_{k=1}^{N} x_k^2.\end{aligned}$$

In order to build optimal trading trajectories, we will look for strategies minimizing

$$\mathbb{E}[\mathcal{C}] \;+\; \lambda\mathrm{Var}[\mathcal{C}]\,,$$

with λ a risk aversion parameter.

5.3 Naive strategies

5.3.1 Assumptions

We assume that the permanent impact is linear: $g(v) = \gamma v$. If we sell n shares, the price per share decreases by γn, thus

$$S_k \;=\; S_0 \;+\; \sigma\sqrt{\tau}\sum_{j=1}^{k}\xi_j \;-\; \gamma(X - x_k).$$

Moreover, in $\mathbb{E}[\mathcal{C}]$, the permanent impact component satisfies

$$\begin{aligned}\tau\sum_{k=1}^{N} x_k g\left(\frac{n_k}{\tau}\right) \;&=\; \gamma\sum_{k=1}^{N} x_k(x_{k-1} - x_k)\\ &=\; \frac{1}{2}\gamma X^2 \;-\; \frac{1}{2}\gamma\sum_{k=1}^{N} n_k^2.\end{aligned}$$

On the other hand we assume affine temporary impact: $h\left(\frac{n_k}{\tau}\right) = \varepsilon + \eta\frac{n_k}{\tau}$. Where ε represents a fixed cost: fees and bid ask spread. Let $\tilde{\eta} := \eta - \frac{1}{2}\gamma\tau$, we get

$$\mathbb{E}[\mathcal{C}] \;=\; \frac{1}{2}\gamma X^2 \;+\; \varepsilon X \;+\; \frac{\tilde{\eta}}{\tau}\sum_{k=1}^{N} n_k^2.$$

5.3.2 Two strategies

Regular liquidation. We take $n_k = \frac{X}{N}$, $x_k = (N-k)\frac{X}{N}$, $k \in [\![1, N]\!]$. And so,

$$\begin{aligned}\mathbb{E}[\mathcal{C}] \;&=\; \frac{1}{2}\gamma X^2 \;+\; \varepsilon X \;+\; \tilde{\eta}\frac{X^2}{T}\;;\\ \mathrm{Var}[\mathcal{C}] \;&=\; \frac{\sigma^2}{3}X^2 T\left(1 - \frac{1}{N}\right)\left(1 - \frac{1}{2N}\right).\end{aligned}$$

We can show this strategy has the smallest expectation. However the variance can be very big if T is large.

Immediate selling. We take $n_1 = X$, $n_2 = \cdots = n_N = 0$, $x_1 = \cdots = x_N = 0$, and we get

$$\begin{aligned}\mathbb{E}[\mathcal{C}] \;&=\; \varepsilon X \;+\; \frac{\eta X^2}{\tau}\;;\\ \mathrm{Var}[\mathcal{C}] \;&=\; 0.\end{aligned}$$

This strategy has the smallest variance. However, if τ is small, the expectation can be very large.

5.4 Optimal strategies

The trader wants to minimize

$$\begin{aligned} U(\mathcal{C}) &= \mathbb{E}[\mathcal{C}] + \lambda \mathrm{Var}[\mathcal{C}] \\ &= \frac{1}{2}\gamma X^2 + \varepsilon X + \frac{\tilde{\eta}}{\tau}\sum_{k=1}^{N}(x_{k-1}-x_k)^2 \\ &\quad + \lambda\sigma^2\tau\sum_{k=1}^{N} x_k^2 . \end{aligned}$$

For $j \in [\![1, N-1]\!]$,

$$\frac{\partial U}{\partial x_j} = 2\tau\left(\lambda\sigma^2 x_j - \tilde{\eta}\frac{x_{j-1}-2x_j+x_{j+1}}{\tau^2}\right),$$

then

$$\begin{aligned} &\frac{\partial U}{\partial x_j} = 0 \\ \Leftrightarrow\quad &\frac{x_{j-1}-2x_j+x_{j+1}}{\tau^2} = \tilde{K} x_j , \end{aligned}$$

with $\tilde{K} = \lambda\frac{\sigma^2}{\tilde{\eta}}$. It is shown that the solution can be written $x_0 = X$ and for $j \in [\![1, N]\!]$:

$$\begin{aligned} x_j &= \frac{\sinh K(T-t_j)}{\sinh KT} X ; \\ n_j &= \frac{2\sinh\frac{K\tau}{2}}{\sinh KT}\cosh\left(K\left(T - j\tau + \frac{\tau}{2}\right)\right) X , \end{aligned}$$

where K satisfies

$$\frac{2}{\tau^2}(\cosh K\tau - 1) = \tilde{K}.$$

If $\lambda = 0$, then $\tilde{K} = K = 0$ and so $n_j = \frac{\tau}{T} = \frac{X}{N}$. We retrieve the strategy with minimal expected cost.

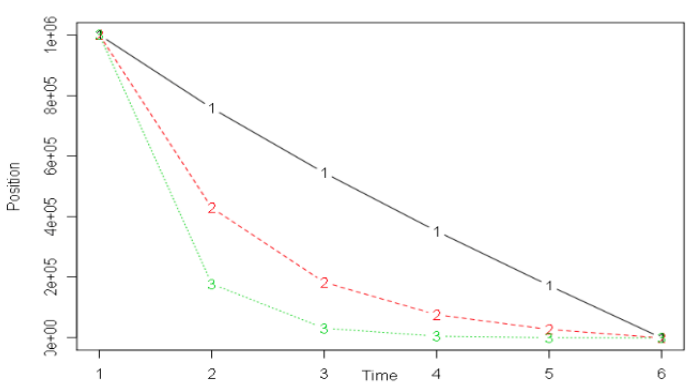

Figure 3: Optimal trajectory for a single-asset portfolio – green: $\lambda = 10^{-5}$, red: $\lambda = 2\cdot 10^{-6}$, black: $\lambda = 10^{-7}$.

Remark. Few remarks:

- It is easy to show that the solution is time homogenous: if we compute the optimal strategy in t_k, we obtain the values between t_k and T of the optimal strategy computed in t_0.
- In this approach, we obtain an efficient frontier of trading.
- The optimal trajectories are very sensitive to the volatility parameter. It is therefore important to obtain accurate volatility estimates.
- The Almgren and Chriss framework can be extended in dimension n (if we sell several assets). In that case, correlation parameters come into the picture.

Only the total quantities which have to be executed in each time window are provided by the Almgren–Chriss approach. The way to deal with them inside each window (the trading tactic) is an intricate issue : Should we use market orders only or a combination with limit orders ? When should we trade inside each window ?

6 Model with uncertainty zones and statistical procedures for volatility, correlation and lead-lag

6.1 Modelling ultra high frequency data

We want the following properties for our model:

- Model for prices and durations, *i.e.* no hesitation about the sampling frequency.
- Discrete prices.
- Bid-Ask bounce.
- Stylized facts of returns, durations and volatility. In particular, inverse relation between durations and volatility.
- A diffusive behaviour at large sampling scales.
- No hesitation about the price.
- Finite quadratic variation for the microstructure noise.
- An interpretation of the model.
- A useful model.

The proposed answer is the model called model with uncertainty zones. In this model, prices and durations are functionals of some hitting times of an underlying continuous semi-martingale.

6.2 Model with uncertainty zones

The added value of this model is to use the aversion for price changes. Indeed in an idealistic framework, transactions would occur when the efficient price crosses the tick grid. In practice, uncertainty about the efficient price and aversion for price changes of market participants. The price changes only when market participants are convinced that the efficient price is far from the last traded price. We introduce a parameter η quantifying this aversion for price changes.

We will use the following notations:

- X_t: efficient price.
- α: tick size.
- t_i: time of the i-th transaction with price change.
- P_{t_i}: transaction price at time t_i.
- $L_i := \frac{|P_{t_{i+1}} - P_{t_i}|}{\alpha}$: size of the i-th price jump.
- $U_k := [0, \infty[\times]d_k, u_k[$: uncertainty zones with

$$\begin{aligned} d_k &= \left(k + \frac{1}{2} - \eta\right)\alpha \ ; \\ u_k &= \left(k + \frac{1}{2} + \eta\right)\alpha. \end{aligned}$$

- τ_i: i-th exit time of an uncertainty zone.

We give the graphical use of the model

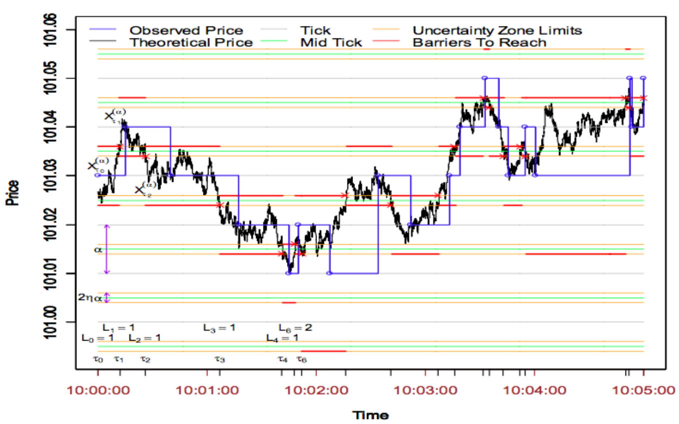

Figure 4: Model with uncertainty zones

The model is good according to the the criterias we gave earlier. Let us say few words about η the only parameter of the model. It quantifies the aversion for price changes (with respect to the tick size) of market participants. In the UHF, the order book can not "follow" the efficient price and is reluctant to price changes. Reluctancy measured by η. $2\eta\alpha$ represents the implicit spread of a large tick asset (see later). A small η ($< \frac{1}{2}$) means that for market participants, the tick size is too big and conversely.

6.3 Volatility estimation

We recall the efficient price and its estimator writes

$$\begin{aligned} X_{\tau_i} &= P_{t_i} - \alpha\left(\frac{1}{2} - \eta\right) \operatorname{sgn}(P_{t_i} - P_{t_{i-1}}) \ ; \\ \hat{X}_{\tau_i} &= P_{t_i} - \alpha\left(\frac{1}{2} - \hat{\eta}\right) \operatorname{sgn}(P_{t_i} - P_{t_{i-1}}). \end{aligned}$$

Let $N_{\alpha,t} = \text{Card}\{t_i,\, t_i \leq t\}$ and define

$$\begin{aligned} N_{\alpha,t}^{(c)} &= \sum_{i=1}^{N_{\alpha,t}} \mathbb{1}_{\{|X_{\tau_i} - X_{\tau_{i-1}}| = \alpha\}} \ ; \\ N_{\alpha,t}^{(a)} &= \sum_{i=1}^{N_{\alpha,t}} \mathbb{1}_{\{|X_{\tau_i} - X_{\tau_{i-1}}| = 2\eta\alpha\}}. \end{aligned}$$

And then the natural estimator

$$\hat{\eta}_t := \frac{1}{2} \frac{N_{\alpha,t}^{(c)}}{N_{\alpha,t}^{(a)}}.$$

Theorem 6.1. *Let*

$$\widehat{\text{RV}}_t = \sum_{i=1}^{N_{\alpha,t}} \left(\ln \hat{X}_{\tau_i} - \ln \hat{X}_{\tau_{i-1}}\right)^2,$$

we have

$$\alpha^{-1}\left(\widehat{\text{RV}}_t - \text{RV}_t\right) \xrightarrow{\mathcal{L}} \gamma_t \int_0^t v_u \, dW_{\theta_u}$$

where W is a Brownian motion independent of B and θ_u, v_u and γ_u depend on X_u, σ_u and explanatory variables, involving for example the order book.

6.4 Covariation estimation

We now consider two assets,

$$\begin{aligned} d\ln X_t &= \mu_t^X \, dt + \sigma_{t-}^X \, dW_t \ ; \\ d\ln Y_t &= \mu_t^Y \, dt + \sigma_{t-}^Y \, dB_t \, , \end{aligned}$$

with

$$d\langle W, B\rangle_t = \rho_t \, dt.$$

We want to estimate $\int_0^t \rho_t \sigma_t^X \sigma_t^Y \, dt$. And we will face two main difficulties:

- asynchronicity of the data,
- microstructure effects.

We start with the usual case without asynchronicity or microstructure effects. We observe $(X_{\frac{i}{n}}, Y_{\frac{i}{n}})$, $i \in [\![0, n]\!]$. Let

$$\Delta_i^n X := \ln X_{\frac{i}{n}} - \ln X_{\frac{i-1}{n}}.$$

An estimator of $\int_0^t \rho_t \sigma_t^X \sigma_t^Y \, dt$ with accuracy $n^{-\frac{1}{2}}$ is

$$\hat{c}_n := \sum_{i=1}^n \Delta_i^n X \Delta_i^n Y.$$

Very often, traders like to think in term of correlation. When the correlation and volatility parameters are supposed to be constant, an estimator of ρ with accuracy $n^{-\frac{1}{2}}$ is given by

$$\frac{\hat{c}_n}{\sqrt{\sum_{i=1}^n (\Delta_i^n X)^2}\sqrt{\sum_{i=1}^n (\Delta_i^n Y)^2}}$$

Proposition 6.2. *In the case where the volatility parameters are no longer constant, one can consider*

$$\hat{\rho}_n = \frac{2}{\pi}\frac{\hat{c}_n}{\hat{a}_n},$$

with

$$\hat{a}_n := \sum_{i=1}^{n-1} \Delta_{i+1}^n X \Delta_i^n Y.$$

Indeed, $\hat{a}_n$ is an estimator of $\frac{2}{\pi}\int_0^1 \sigma_s^X \sigma_s^Y \, \mathrm{d}s$.

Proof. We note that

$$\hat{a}_n \approx \sum_{i=1}^{n-1} \sigma^X_{\frac{i-1}{n}} \sigma^Y_{\frac{i-1}{n}} \left| W_{\frac{i+1}{n}} - W_{\frac{i}{n}} \right| \left| B_{\frac{i}{n}} - B_{\frac{i-1}{n}} \right|,$$

with the Brownian increments in the preceding sum being independent.

$\hat{a}_n$ has the same limit as

$$\begin{aligned} & \sum_{i=1}^{n-1} \mathbb{E}\left[\sigma^X_{\frac{i-1}{n}} \sigma^Y_{\frac{i-1}{n}} \left| W_{\frac{i+1}{n}} - W_{\frac{i}{n}} \right| \left| B_{\frac{i}{n}} - B_{\frac{i-1}{n}} \right| \middle| \mathcal{F}_{\frac{i}{n}} \right] \\ = & \sum_{i=1}^{n-1} \sigma^X_{\frac{i-1}{n}} \sigma^Y_{\frac{i-1}{n}} \left| B_{\frac{i}{n}} - B_{\frac{i-1}{n}} \right| \mathbb{E}\left[\left| W_{\frac{i+1}{n}} - W_{\frac{i}{n}} \right| \right] \\ = & \sqrt{\frac{2}{n}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n-1} \sigma^X_{\frac{i-1}{n}} \sigma^Y_{\frac{i-1}{n}} \left| B_{\frac{i}{n}} - B_{\frac{i-1}{n}} \right|. \end{aligned}$$

This last term has the same limit as

$$\begin{aligned} & \sqrt{\frac{2}{n}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n-1} \sigma^X_{\frac{i-1}{n}} \sigma^Y_{\frac{i-1}{n}} \mathbb{E}\left[\left| B_{\frac{i}{n}} - B_{\frac{i-1}{n}} \right| \right] \\ = & \frac{2}{n}\frac{1}{n} \sum_{i=1}^{n-1} \sigma^X_{\frac{i-1}{n}} \sigma^Y_{\frac{i-1}{n}} \\ \to & \frac{2}{\pi} \int_0^1 \sigma_s^X \sigma_s^Y \, \mathrm{d}s. \end{aligned}$$

□

6.4.1 Hayashi–Yoshida Estimator

Assume now we observe X at times $(T^{X,i})$ and Y at times $(T^{Y,i})$. We build

$$\begin{aligned} \bar{X}_t &:= X_{T^{X,i}} \quad \text{for } t \in [T^{X,i}, T^{X,i+1}[\,; \\ \bar{Y}_t &:= Y_{T^{Y,i}} \quad \text{for } t \in [T^{Y,i}, T^{Y,i+1}[. \end{aligned}$$

For given h, the previous tick covariation estimator is

$$V_h = \sum_{i=1}^m \left(\ln \bar{X}_{ih} - \ln \bar{X}_{(i-1)h} \right) \left(\ln \bar{Y}_{ih} - \ln \bar{Y}_{(i-1)h} \right).$$

But in fact with the Epps effect there is a systematic bias for this estimator.

Definition 6.1 (Hayashi–Yoshida estimator)**.** Let $I_i^X :=]T^{X,i}, T^{X,i+1}]$, and $I_i^Y :=]T^{Y,i}, T^{Y,i+1}]$. The Hayashi–Yoshida estimator is

$$U_n = \sum_{i,j} \Delta X(I_i^X) \Delta Y(I_j^Y) \mathbb{1}_{\{I_i^X \cap I_j^Y \neq \varnothing\}}.$$

Theorem 6.3. *In the model with uncertainty zones, the Hayashi–Yoshida estimator is a consistent estimator of the covariation provided one uses the estimated values of the efficient prices.*

6.5 Lead-Lag estimation

The motivation comes from the observations from practitioners in finance, some assets are leading some other assets. This means that a "lagger" asset may partially reproduce the behaviour of a "leader" asset. This common behaviour is unlikely to be instantaneous. It is subject to some time delay called "lead-lag".

We will work on a Bachelier model, for $t \in [0, 1]$, and $(B^{(1)}, B^{(2)})$ such that $\langle B^{(1)}, B^{(2)} \rangle_t = \rho t$, set

$$\begin{aligned} X_t &:= x_0 + \sigma_1 B_t^{(1)}\,; \\ \tilde{Y}_t &:= y_0 + \sigma_2 B_t^{(2)}. \end{aligned}$$

Define $Y_t := \tilde{Y}_{t-\theta}$, $t \in [\theta, 1]$. Our lead-lag model is given by the bidimensional process (X_t, Y_t). We have

$$\begin{cases} X_t = x_0 + \sigma_1 B_t^{(1)}\,; \\ Y_t = y_0 + \rho\sigma_2 B_{t-\theta}^{(1)} + \sigma_2\sqrt{1-\rho^2} W_{t-\theta}. \end{cases}$$

Assume the data arrive at regular and synchronous time stamps in the Bachelier model, *i.e.* we have data

$$(X_0, Y_0), (X_{\Delta_n}, Y_{\Delta_n}), (X_{2\Delta_n}, Y_{2\Delta_n}), \ldots, (X_1, Y_1)$$

and suppose $\theta = k_0 \Delta_n$, $k_0 \in \mathbb{Z}$. Let

$$\mathcal{C}_n(k) := \sum_i \left(X_{i\Delta_n} - X_{(i-1)\Delta_n} \right) \left(Y_{(i+k)\Delta_n} - Y_{(i+k-1)\Delta_n} \right).$$

Heuristically we have

$$\begin{aligned} \mathcal{C}_n(k) \approx\ & \Delta_n^{-1} \mathbb{E}\left[(X_\cdot - X_{\cdot - \Delta_n})(Y_{\cdot + k\Delta_n} - Y_{\cdot + (k-1)\Delta_n}) \right] \\ & + \Delta_n^{\frac{1}{2}} \xi^n. \end{aligned}$$

Moreover

$$\begin{aligned} & \Delta_n^{-1} \mathbb{E}\left[(X_\cdot - X_{\cdot - \Delta_n})(Y_{\cdot + k\Delta_n} - Y_{\cdot + (k-1)\Delta_n}) \right] \\ = & \begin{cases} 0 & \text{if } k \neq k_0 \\ \rho\sigma_1\sigma_2 & \text{if } k = k_0. \end{cases} \end{aligned}$$

Thus we can (asymptotically) detect the value k_0 that defines θ in the very special case $\theta = k_0\Delta_n$ by maximizing in k the contrast sequence $|\mathcal{C}_n(k)|$.

We set

$$U_n \;=\; \sum_{i,j} \Delta X(I_i^X)\Delta Y((I_j^Y)_{-\theta})\mathbb{1}_{\{I_i^X\cap(I_j^Y)_{-\theta}\neq\varnothing\}}\,,$$

with $(I_j^Y)_{-\theta} =]T^{Y,j}-\theta, T^{Y,j+1}-\theta]$. Then $\hat\theta_n$ is defined as the solution of

$$|U_n(\hat\theta_n)| \;=\; \max_{\theta\in\mathcal{G}^n} |U_n(\theta)|$$

where $\mathcal{G}^n$ is a sufficiently fine grid.

Theorem 6.4. *As $n\to\infty$,*

$$v_n^{-1}\left(\hat\theta_n-\theta\right) \;\to\; 0\,,$$

in probability, on the event $\left\{\langle X^c, \tilde{Y}^c\rangle_T \neq 0\right\}$.

7 Tick values and regulation

7.1 Tick value, tick size and spread

7.1.1 Settings

The tick value is smallest price increment. In practice the tick value is given little consideration. What is important is the tick size, it qualifies the traders' aversion to price movements of one tick. Notion of tick size is ambiguous in general. However, we can identify large tick assets.

Definition 7.1. Large tick stocks are such that the bid-ask spread is almost always equal to one tick, while small tick stocks have spreads that are typically a few ticks.

So for small tick assets spread is a good proxy for the stick size, but for large tick assets, how can we quantify the tick size ?

7.1.2 Madhavan, Richardson, Roomans economic model

Let p_{i+1} be the ex post true or efficient price after the i-th trade (all transactions have the same volume), and ε_i: sign of the i-th trade. The MRR model is defined by:

$$p_{i+1} - p_i \;=\; \xi_i + \theta\varepsilon_i\,,$$

with ξ_i an independent centred shock component (new information, *etc.*) with variance v^2. Market makers cannot guess the surprise of the next trade. So, they post (pre trade) bid and ask prices a_i and b_i given by

$$\begin{aligned} a_i &= p_i + \theta + \phi\,; \\ b_i &= p_i - \theta - \phi\,, \end{aligned}$$

with ϕ an extra compensation claimed by market makers, covering processing costs and the shock component risk. We can compute several quantities, the spread

$$\begin{aligned} S &:= a - b \\ &= 2(\theta-\phi)\,, \end{aligned}$$

the variance per trade of the efficient price

$$\begin{aligned} \sigma_1^2 &\triangleq \mathbb{E}\left[(p_{i+1}-p_i)^2\right] \\ &= \theta^2 + v^2 \\ &\sim \theta^2. \end{aligned}$$

Therefore

$$S \;\sim\; 2\sigma_1 + 2\phi.$$

7.1.3 Market making strategy

Wyart et *al.*: consider a simple market making strategy. Its average P&L per trade is

$$\text{P\&L} \;=\; \frac{S}{2} - \frac{c}{2}\sigma_1\,,$$

with c depending on the assets but of order 1. Moreover, market makers' P&L = 0 (if not so, another market maker comes with a slightly tighter spread), hence

$$S \;\sim\; c\sigma_1.$$

7.2 The model with uncertainty zones

Recall this model from Section 6.2.

Now we can interpret η, it has a direct link with the distribution of high frequency tick returns. Indeed for example if η small, the uncertainty zone is small, then there is a strong mean reversion in the observed price, so the signature plot is decreasing, significant ACV of tick returns, and so finally the tick size is large. We have the opposite for $\eta\sim\frac{1}{2}$.

The distance between Ask Zone and Bid Zone is $2\eta\alpha$. It represents an implicit unobservable spread. Wit M the total number of trades (null returns and not), can we extend $\frac{S}{2}\sim\frac{\sigma}{\sqrt{M}}$, to $\eta\alpha\sim\frac{\sigma}{\sqrt{M}}$?

7.3 Implicit spread and volatility per trade

We want to investigate the relationship

$$\eta\alpha \;\sim\; \frac{\sigma}{\sqrt{M}} + \phi$$

for large tick assets.

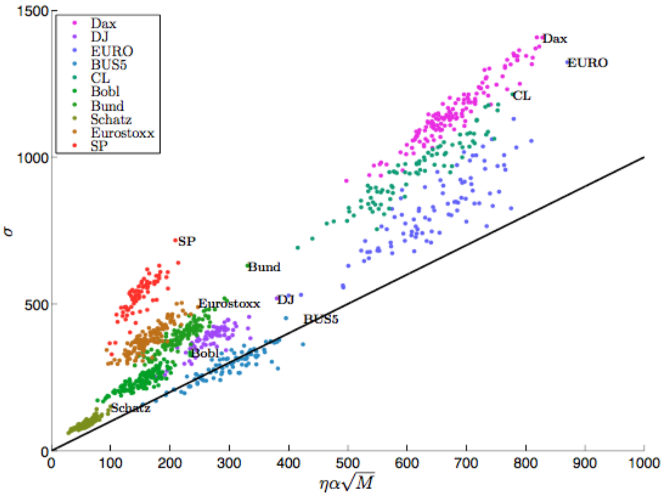

Figure 5: Cloud $(\eta\alpha\sqrt{M}, \sigma)$, for each day, for each asset

Note that ϕ includes operational costs and inventory control, then we write $\phi = kS$. And we test the daily regression

$$\sigma \quad = \quad p_1 \eta\alpha\sqrt{M} \ + \ p_2 S\sqrt{M} \ + \ p_3$$

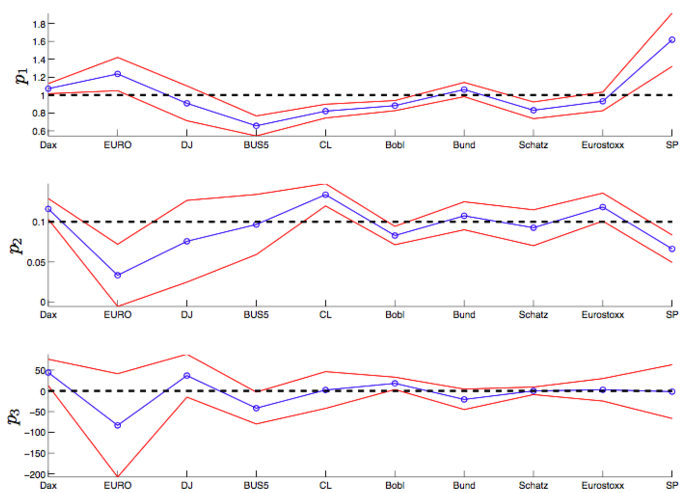

Figure 6: Results of the regression

We can find the result with a cost analysis. Indeed the average ex post cost of a market order is

$$\frac{\alpha}{2} \ - \ \eta\alpha.$$

On the other hand, the average P&L per trade of the market makers should be equal to the average cost of a market order, hence

$$\eta\alpha \quad = \quad c\frac{\sigma}{\sqrt{M}} \ + \ \phi.$$

7.4 Predicting consequences of tick value changes

We know that α too small encourages free-riding (directional HFT) and traditional market makers cannot fix their quotes. And α too large implies price sloppiness. Moreover it favors speed (race to the top of book).

What happens to η if one changes the tick value ? And how to obtain the following optimal situation:

- $S \sim 1$.
- η close to $\frac{1}{2}$.
- Cost of market orders = cost of limit orders = 0.

We assume that when changing the tick value:

- σ constant then

$$\eta_0\alpha_0\sqrt{M_0} \ + \ 0.1\alpha_0\sqrt{M_0} \quad = \quad \eta\alpha\sqrt{M} \ + \ 0.1\alpha\sqrt{M}$$

- Constant volume.
- Linear shape for the cumulative latent liquidity.

Then

$$\eta \quad \sim \quad (\eta_0 + 0.1)\left(\frac{\alpha_0}{\alpha}\right)^{\frac{1}{2}} \ - \ 0.1$$

$$\alpha^* \quad \sim \quad \left(\frac{\eta_0 + 0.1}{0.6}\right)^2 \alpha_0.$$