20

Applications to Finance

20.1 Portfolio Theory

One of the arch-problems in quantitative finance is portfolio construction. For example, one may consider an investment universe made of N stocks (or more generally N risky assets) that one should bundle up in a portfolio to achieve optimal performance, according to some quality measure that we will discuss below.

20.1.1 Returns and Risk Free Rate

We call $p_{i,t}$ the price of stock i at time t and define the returns over some elementary time scale (say one day) as

$$r_{i,t} := \frac{p_{i,t} - p_{i,t-1}}{p_{i,t-1}}. \tag{20.1}$$

The portfolio weight π_i is the dollar amount invested on asset i, which can be positive (corresponding to buys) or negative (corresponding to short sales). The total capital to be invested is C. Naively, one should have $C = \sum_i \pi_i$. But one can borrow cash, so that $\sum_i \pi_i > C$ and pay the risk free rate r_0 on the borrowed amount, or conversely under-invest in stocks ($\sum_i \pi_i < C$) and invest the remaining capital at the risk free rate, assumed to be the same r_0.[1] Then the total return of the portfolio (in dollar terms) is

$$R_t = \sum_i \pi_i r_{i,t} + (C - \sum_i \pi_i) r_0, \tag{20.2}$$

so that the *excess return* (over the risk free rate) is

$$R_t - Cr_0 := \sum_i \pi_i (r_{i,t} - r_0), \tag{20.3}$$

where $r_{i,t} - r_0$ is the excess return of asset i. From now on, we will denote $r_{i,t} - r_0$ by $r_{i,t}$. We will assume that these excess returns are characterized by some expected gains g_i and a covariance matrix $\mathbf{C}$, with

[1] In general, the risk free rate to borrow is different from the one to lend, but we will neglect the difference here.

https://doi.org/10.1017/9781108768900.024 Published online by Cambridge University Press

$$\mathbf{C}_{ij} = \mathrm{Cov}[r_i r_j]. \tag{20.4}$$

The problem, of course, is that both the vector of expected gains $\mathbf{g}$ and the covariance matrix $\mathbf{C}$ are unknown to the investor, who must come up with his/her best guess for these quantities. Forming expectations of future returns is the job of the investor, based on his/her information, anticipations, and hunch. We will not attempt to model the sophisticated process at work in the mind of investors, and simply assume that $\mathbf{g}$ is known. In the simplest case, the investor has no preferences and $\mathbf{g} = g\mathbf{1}$, corresponding to the same expected return for all assets. Another possibility is to assume that $\mathbf{g}$ is, for all practical purposes, a random vector in $\mathbb{R}^N$.

As far as $\mathbf{C}$ is concerned, the most natural choice is to use the sample covariance matrix $\mathbf{E}$, determined using a series of past returns of length T. However, as we already know from Chapter 17, the eigenvalues of $\mathbf{E}$ can be quite far from those of $\mathbf{C}$ when $q = N/T$ is not very small. On the other hand, T cannot be as large as one could wish, the most important reason being that the (financial) world is non-stationary. For a start, many large firms that exist in 2019 did not exist 25 years ago. More generally, it is far from clear that the parameters of the underlying statistical process (if such a thing exists) can be considered as constant in time, so mixing different epochs is in general not warranted. On the other hand, due to experimental constraints, the limitation of data points can be a problem even in a stationary world.

20.1.2 Portfolio Risk

The risk of a portfolio is traditionally measured as the variance of its returns, namely

$$\mathcal{R}^2 := \mathbb{V}[R] = \sum_{i,j} \pi_i \pi_j \mathrm{Cov}[r_i r_j] = \boldsymbol{\pi}^T \mathbf{C} \boldsymbol{\pi}. \tag{20.5}$$

Other measures of risk can however be considered, such as the *expected shortfall* S_p (or conditional value at risk), defined as

$$S_p = -\frac{1}{p} \int_{-\infty}^{R_p} \mathrm{d}R \, R \, P(R), \tag{20.6}$$

where R_p is the p-quantile, with for example $p = 0.01$ for the 1% negative tail events:

$$p = \int_{-\infty}^{R_p} \mathrm{d}R \, P(R). \tag{20.7}$$

If $P(R)$ is Gaussian, then all risk measures are equivalent and subsumed in $\mathbb{V}[R]$.

20.1.3 Markowitz Optimal Portfolio Theory

For the reader not familiar with Markowitz's optimal portfolio theory, we recall in this section some of the most important results. Suppose that an investor wants to invest in a portfolio containing N different assets, with optimal weights $\boldsymbol{\pi}$ to be determined. An intuitive strategy is the so-called mean-variance optimization: the investor seeks an allocation such that the variance of the portfolio is minimized given an expected return target. It is not hard to see that this mean-variance optimization can be translated into a simple quadratic optimization program with a linear constraint. Markowitz's optimal portfolio amounts to solving the following quadratic optimization problem:

$$\begin{cases} \min_{\boldsymbol{\pi} \in \mathbb{R}^N} \frac{1}{2} \boldsymbol{\pi}^T \mathbf{C} \boldsymbol{\pi} \\ \text{subject to } \boldsymbol{\pi}^T \mathbf{g} \geq \mathcal{G} \end{cases} \tag{20.8}$$

where $\mathcal{G}$ is the desired (or should we say hoped for) gain. Without further constraints – such as the positivity of all weights necessary if short positions are not allowed – this problem can be easily solved by introducing a Lagrangian multiplier γ and writing

$$\min_{\boldsymbol{\pi} \in \mathbb{R}^N} \frac{1}{2} \boldsymbol{\pi}^T \mathbf{C} \boldsymbol{\pi} - \gamma \boldsymbol{\pi}^T \mathbf{g}. \tag{20.9}$$

Assuming that $\mathbf{C}$ is invertible, it is not hard to find the optimal solution and the value of γ such that overall expected return is exactly $\mathcal{G}$. It is given by

$$\boldsymbol{\pi}_{\mathbf{C}} = \mathcal{G} \frac{\mathbf{C}^{-1} \mathbf{g}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}, \tag{20.10}$$

which, as noted above, requires the knowledge of both $\mathbf{C}$ and $\mathbf{g}$, which are *a priori* unknown. Note that even if the predictions g_i of our investor are completely wrong, it still makes sense to look for the minimum risk portfolio consistent with his/her expectations. But we are left with the problem of estimating $\mathbf{C}$, or maybe $\mathbf{C}^{-1}$ before applying Markowitz's formula, Eq. (20.10). We will see below why one should actually find the best estimator of $\mathbf{C}$ itself before inverting it and determining the weights.

What is the risk associated with this optimal allocation strategy, measured as the variance of the returns of the portfolio? If one knew the population correlation matrix $\mathbf{C}$, the *true* optimal risk associated with $\boldsymbol{\pi}_{\mathbf{C}}$ would be given by

$$\mathcal{R}^2_{\text{true}} := \boldsymbol{\pi}_{\mathbf{C}}^T \mathbf{C} \boldsymbol{\pi}_{\mathbf{C}} = \frac{\mathcal{G}^2}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}. \tag{20.11}$$

However, the optimal strategy (20.10) is not attainable in practice as the matrix $\mathbf{C}$ is unknown. What can one do then, and how poorly is the realized risk of the portfolio estimated?

20.1.4 Predicted and Realized Risk

One obvious – but far too naive – way to use the Markowitz optimal portfolio is to apply (20.10) using the SCM $\mathbf{E}$ as is, instead of $\mathbf{C}$. Recalling the results of Chapter 17, it is not hard to see that this strategy will suffer from strong biases whenever T is not sufficiently large compared to N.

https://doi.org/10.1017/9781108768900.024 Published online by Cambridge University Press

Notwithstanding, the optimal investment weights using the SCM $\mathbf{E}$ read

$$\boldsymbol{\pi}_{\mathbf{E}} = \mathcal{G}\frac{\mathbf{E}^{-1}\mathbf{g}}{\mathbf{g}^T\mathbf{E}^{-1}\mathbf{g}}, \tag{20.12}$$

and the minimum risk associated with this portfolio is thus given by

$$\mathcal{R}_{\text{in}}^2 = \boldsymbol{\pi}_{\mathbf{E}}^T\mathbf{E}\,\boldsymbol{\pi}_{\mathbf{E}} = \frac{\mathcal{G}^2}{\mathbf{g}^T\mathbf{E}^{-1}\mathbf{g}}, \tag{20.13}$$

which is known as the "in-sample" risk, or the *predicted* risk. It is "in-sample" because it is entirely constructed using the available data. The realized risk in the next period, with fresh data, is correspondingly called *out-of-sample*.

Using the convexity with respect to $\mathbf{E}$ of $\mathbf{g}^T\mathbf{E}^{-1}\mathbf{g}$, we find from the Jensen inequality that, for fixed predicted gains $\mathbf{g}$,

$$\mathbb{E}[\mathbf{g}^T\mathbf{E}^{-1}\mathbf{g}] \geq \mathbf{g}^T\mathbb{E}\big[\mathbf{E}\big]^{-1}\mathbf{g} = \mathbf{g}^T\mathbf{C}^{-1}\mathbf{g}, \tag{20.14}$$

where the last equality holds because $\mathbf{E}$ is an unbiased estimator of $\mathbf{C}$. Hence, we conclude that the in-sample risk is lower than the "true" risk and therefore the optimal portfolio $\boldsymbol{\pi}_{\mathbf{E}}$ suffers from an in-sample bias: its predicted risk underestimates the true optimal risk $\mathcal{R}_{\text{true}}$. Intuitively this comes from the fact that $\boldsymbol{\pi}_{\mathbf{E}}$ attempts to exploit all the idiosyncracies that happened during the in-sample period, and therefore manages to reduce the risk below the true optimal risk. But the situation is even worse, because the future out-of-sample or *realized* risk, turns out to be larger than the true risk. Indeed, let us denote by $\mathbf{E}'$ the SCM of this out-of-sample period; the *out-of-sample* risk is then naturally defined by

$$\mathcal{R}_{\text{out}}^2 = \boldsymbol{\pi}_{\mathbf{E}}^T\mathbf{E}'\,\boldsymbol{\pi}_{\mathbf{E}} = \frac{\mathcal{G}^2\mathbf{g}^T\mathbf{E}^{-1}\mathbf{E}'\mathbf{E}^{-1}\mathbf{g}}{(\mathbf{g}^T\mathbf{E}^{-1}\mathbf{g})^2}. \tag{20.15}$$

For large matrices, we expect the result to be self-averaging and given by its expectation value (over the measurement noise). But if the measurement noise in the in-sample period (contained in $\boldsymbol{\pi}_{\mathbf{E}}$) can be assumed to be independent from that of the out-of-sample period, then $\boldsymbol{\pi}_{\mathbf{E}}$ and $\mathbf{E}'$ are uncorrelated and we get, for $N \to \infty$,

$$\boldsymbol{\pi}_{\mathbf{E}}^T\mathbf{E}'\boldsymbol{\pi}_{\mathbf{E}} = \boldsymbol{\pi}_{\mathbf{E}}^T\mathbf{C}\boldsymbol{\pi}_{\mathbf{E}}. \tag{20.16}$$

Now, from the optimality of $\boldsymbol{\pi}_{\mathbf{C}}$, we also know that

$$\boldsymbol{\pi}_{\mathbf{C}}^T\mathbf{C}\boldsymbol{\pi}_{\mathbf{C}} \leq \boldsymbol{\pi}_{\mathbf{E}}^T\mathbf{C}\boldsymbol{\pi}_{\mathbf{E}}, \tag{20.17}$$

so we readily obtain the following general inequalities:

$$\mathcal{R}_{\text{in}}^2 \leq \mathcal{R}_{\text{true}}^2 \leq \mathcal{R}_{\text{out}}^2. \tag{20.18}$$

We plot in Figure 20.1 an illustration of these inequalities. One can see how using $\boldsymbol{\pi}_{\mathbf{E}}$ is clearly overoptimistic and can potentially lead to disastrous results in practice.

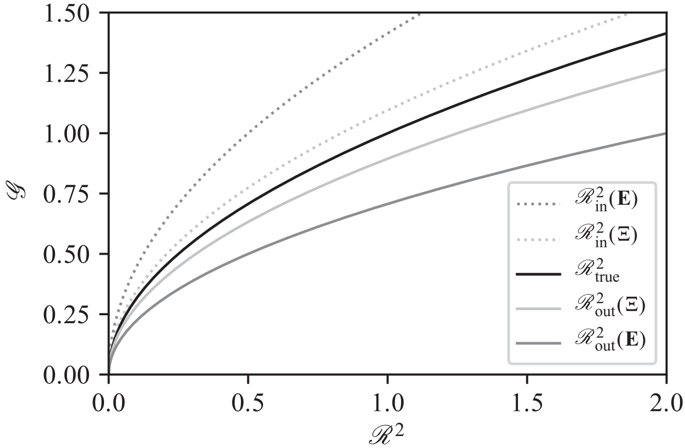

Figure 20.1 Efficient frontier associated with the mean-variance optimal portfolio (20.10) for $\mathbf{g} = \mathbf{1}$ and $\mathbf{C}$ an inverse-Wishart matrix with $p = 0.5$, for $q = 0.5$. The black line depicts the expected gain as a function of the *true* optimal risk (20.11). The gray lines correspond to the realized (out-of-sample) risk using either the SCM $\mathbf{E}$ or its RIE version Ξ. Both estimates are above the true risk, but less so for RIE. Finally, the dashed lines represent the predicted (in-sample) risk, again using either the SCM $\mathbf{E}$ or its RIE version Ξ. $\mathcal{R}$ and $\mathcal{G}$ in arbitrary units, such that $\mathcal{R}_{\text{true}} = 1$ for $\mathcal{G} = 1$.

This conclusion in fact holds for different risk measures, such as the expected shortfall measure mentioned in Section 20.1.2.

20.2 The High-Dimensional Limit

20.2.1 In-Sample vs Out-of-Sample Risk: Exact Results

In the limit of large matrices and with some assumptions on the structure $\mathbf{g}$, we can make the general inequalities Eq. (20.18) more precise using the random matrix theory tools from the previous chapters. Let us suppose that the vector of predictors $\mathbf{g}$ points in a random direction, in the sense that the covariance matrix $\mathbf{C}$ and $\mathbf{g}\mathbf{g}^T$ are mutually free. This is not necessarily a natural assumption. For example, the simplest "agnostic" prediction $\mathbf{g} = \mathbf{1}$ is often nearly collinear with the top eigenvector of $\mathbf{C}$ (see Section 20.4). So we rather think here of market neutral, sector neutral predictors that attempt to capture very idiosyncratic characteristics of firms.

Now, if $\mathbf{M}$ is a positive definite matrix that is free from $\mathbf{g}\mathbf{g}^T$, then in the large N limit:

$$\frac{\mathbf{g}^T\mathbf{M}\mathbf{g}}{N} = \frac{1}{N}\,\mathrm{Tr}[\mathbf{g}\mathbf{g}^T\mathbf{M}] \underset{\text{freeness}}{=} \frac{\mathbf{g}^2}{N}\tau(\mathbf{M}), \tag{20.19}$$

https://doi.org/10.1017/9781108768900.024 Published online by Cambridge University Press

where we recall that τ is the normalized trace operator. We can always normalize the prediction vector such that $\mathbf{g}^2/N = 1$, so setting $\mathbf{M} = \{\mathbf{E}^{-1}, \mathbf{C}^{-1}\}$, we can directly estimate Eqs. (20.13), (20.11) and (20.15) and find

$$\begin{aligned} \mathcal{R}^2_{\text{in}} &\to \frac{\mathcal{G}^2}{N\tau(\mathbf{E}^{-1})}, \\ \mathcal{R}^2_{\text{true}} &\to \frac{\mathcal{G}^2}{N\tau(\mathbf{C}^{-1})}, \\ \mathcal{R}^2_{\text{out}} &\to \frac{\mathcal{G}^2\tau(\mathbf{E}^{-1}\mathbf{C}\mathbf{E}^{-1})}{N\tau^2(\mathbf{E}^{-1})}. \end{aligned} \tag{20.20}$$

Let us focus on the first two terms above. For $q < 1$, we know from Eq. (17.14) that, in the high-dimensional limit, $\tau(\mathbf{C}^{-1}) = (1-q)\tau(\mathbf{E}^{-1})$. As a result, we have, for $N \to \infty$,

$$\mathcal{R}^2_{\text{in}} = (1-q)\mathcal{R}^2_{\text{true}}. \tag{20.21}$$

Hence, for any $q \in (0,1)$, we see that the in-sample risk associated with $\boldsymbol{\pi}_{\mathbf{E}}$ always provides an overoptimistic estimator. Even better, we are able to quantify precisely the risk underestimation factor thanks to Eq. (20.21).

Next we would like to find the same type of relation for the "out-of-sample" risk. In order to do so, we write $\mathbf{E} = \mathbf{C}^{\frac{1}{2}}\mathbf{W}_q\mathbf{C}^{\frac{1}{2}}$ where $\mathbf{W}_q$ is a white Wishart matrix of parameter q, independent from $\mathbf{C}$. Plugging this representation into Eq. (20.15), we find that the out-of-sample risk can be expressed as

$$\mathcal{R}^2_{\text{out}} = \frac{\mathcal{G}^2\tau(\mathbf{C}^{-1}\mathbf{W}_q^{-2})}{N\tau^2(\mathbf{E}^{-1})}$$

when $N \to \infty$. Now, since $\mathbf{W}_q$ and $\mathbf{C}$ are asymptotically free, we also have

$$\tau(\mathbf{C}^{-1}\mathbf{W}_q^{-2}) = \tau(\mathbf{C}^{-1})\,\tau(\mathbf{W}_q^{-2}). \tag{20.22}$$

Hence, using again Eq. (17.14) yields

$$\mathcal{R}^2_{\text{out}} = \mathcal{G}^2(1-q)^2\frac{\tau(\mathbf{W}_q^{-2})}{N\tau(\mathbf{C}^{-1})}. \tag{20.23}$$

Finally, we know from Eq. (15.22) that $\tau(\mathbf{W}_q^{-2}) = (1-q)^{-3}$ for $q < 1$. Hence, we finally get

$$\mathcal{R}^2_{\text{out}} = \frac{\mathcal{R}^2_{\text{true}}}{1-q}. \tag{20.24}$$

All in all, we have obtained the following asymptotic relation:

$$\frac{\mathcal{R}^2_{\text{in}}}{1-q} = \mathcal{R}^2_{\text{true}} = (1-q)\mathcal{R}^2_{\text{out}}, \tag{20.25}$$

which holds for a completely general $\mathbf{C}$.

Hence, if one invests with the "naive" weights $\boldsymbol{\pi}_{\mathbf{E}}$, it turns out that the predicted risk $\mathcal{R}_{\text{in}}$ underestimates the realized risk $\mathcal{R}_{\text{out}}$ by a factor $(1-q)$, and in the extreme case $N = T$ or $q = 1$, the in-sample risk is equal to zero while the out-of-sample risk diverges (for $N \to \infty$). We thus conclude that, as announced, the use of the SCM $\mathbf{E}$ for the Markowitz optimization problem can lead to disastrous results. This suggests that we should use a more reliable estimator of $\mathbf{C}$ in order to control the out-of-sample risk.

20.2.2 Out-of-Sample Risk Minimization

We insisted throughout the last section that the right quantity to control in portfolio management is the realized, out-of-sample risk. It is also clear from Eq. (20.25) that using the sample estimate $\mathbf{E}$ is a very bad idea, and hence it is natural to wonder which estimator of $\mathbf{C}$ one should use to minimize this out-of-sample risk? The Markowitz formula (20.10) naively suggests that one should look for a faithful estimator of the so-called precision matrix $\mathbf{C}^{-1}$. But in fact, since the expected out-of-sample risk involves the matrix $\mathbf{C}$ linearly, it is that matrix that should be estimated.

Let us show this using another route, in the context of rotationally invariant estimators, which we considered in Chapter 19. Let us define our RIE as

$$\Xi = \sum_{i=1}^{N} \xi(\lambda_i)\mathbf{v}_i\mathbf{v}_i^T, \tag{20.26}$$

where we recall that $\mathbf{v}_i$ are the sample eigenvectors of $\mathbf{E}$ and $\xi(\cdot)$ is a function that has to be determined using some optimality criterion.

Suppose that we construct a Markowitz optimal portfolio $\boldsymbol{\pi}_{\Xi}$ using this RIE. Again, we assume that the vector $\mathbf{g}$ is random, and independent from Ξ. Consequently, the estimate (20.19) is still valid, such that the realized risk associated with the portfolio $\boldsymbol{\pi}_{\Xi}$ reads, for $N \to \infty$,

$$\mathcal{R}^2_{\text{out}}(\Xi) = \mathcal{G}^2 \frac{\text{Tr}\left(\Xi^{-1}\mathbf{C}\Xi^{-1}\right)}{\left(\text{Tr}\,\Xi^{-1}\right)^2}. \tag{20.27}$$

Using the decomposition (20.26) of Ξ, we can rewrite the numerator as

$$\text{Tr}\left(\Xi^{-1}\mathbf{C}\Xi^{-1}\right) = \sum_{i=1}^{N} \frac{{\mathbf{v}_i}^T\mathbf{C}\mathbf{v}_i}{\xi^2(\lambda_i)}, \tag{20.28}$$

while the denominator of Eq. (20.27) is

$$\left(\text{Tr}\,\Xi^{-1}\right)^2 = \left(\sum_{i=1}^{N} \frac{1}{\xi(\lambda_i)}\right)^2. \tag{20.29}$$

Regrouping these last two equations allows us to express Eq. (20.27) as

https://doi.org/10.1017/9781108768900.024 Published online by Cambridge University Press

$$\mathcal{R}^2_{\text{out}}(\Xi) = \mathcal{G}^2 \sum_{i=1}^{N} \frac{\mathbf{v}_i{}^T \mathbf{C} \mathbf{v}_i}{\xi^2(\lambda_i)} \left(\sum_{i=1}^{N} \frac{1}{\xi(\lambda_i)} \right)^{-2}. \tag{20.30}$$

Our aim is to find the optimal shrinkage function $\xi(\lambda_j)$ associated with the sample eigenvalues $[\lambda_j]_{j=1}^N$, such that the out-of-sample risk is minimized. This can be done by solving, for a given j, the following first order condition:

$$\frac{\partial \mathcal{R}^2_{\text{out}}(\Xi)}{\partial \xi(\lambda_j)} = 0, \quad \forall j = 1, \dots, N. \tag{20.31}$$

By performing the derivative with respect to $\xi(\lambda_j)$ in (20.30), one obtains

$$-2 \frac{\mathbf{v}_j{}^T \mathbf{C} \mathbf{v}_j}{\xi^3(\lambda_j)} \left(\sum_{i=1}^{N} \frac{1}{\xi(\lambda_i)} \right)^{-2} + \frac{2}{\xi^2(\lambda_j)} \left(\sum_{i=1}^{N} \frac{\mathbf{v}_i{}^T \mathbf{C} \mathbf{v}_i}{\xi^2(\lambda_i)} \right) \left(\sum_{i=1}^{N} \frac{1}{\xi(\lambda_i)} \right)^{-3} = 0. \tag{20.32}$$

The solution to this equation is given by

$$\xi(\lambda_j) = A \mathbf{v}_j{}^T \mathbf{C} \mathbf{v}_j, \tag{20.33}$$

where A is an arbitrary constant at this stage. But since the trace of the RIE must match that of $\mathbf{C}$, this constant A must be equal to 1. Hence we recover precisely the oracle estimator that we have studied in Chapter 19.

As a conclusion, the optimal RIE (19.26) actually minimizes the out-of-sample risk within the class of rotationally invariant estimators. Moreover, the corresponding "optimal" realized risk is given by

$$\mathcal{R}^2_{\text{out}}(\Xi) = \frac{\mathcal{G}^2}{\operatorname{Tr}\left[(\Xi)^{-1}\right]}, \tag{20.34}$$

where we used the notable property that, for any $n \in \mathbb{Z}$,

$$\operatorname{Tr}[(\Xi)^n \mathbf{C}] = \sum_{i=1}^{N} \xi(\lambda_i)^n \operatorname{Tr}[\mathbf{v}_i \mathbf{v}_i^T \mathbf{C}] = \sum_{i=1}^{N} \xi(\lambda_i)^n \mathbf{v}_i^T \mathbf{C} \mathbf{v}_i \equiv \operatorname{Tr}[(\Xi)^{n+1}]. \tag{20.35}$$

20.2.3 The Inverse-Wishart Model: Explicit Results

In this section, we specialize the result (20.34) to the case when $\mathbf{C}$ is an inverse-Wishart matrix with parameter $p > 0$, corresponding to the simple linear shrinkage optimal estimator. First, we read from Eq. (15.30) that

$$\tau\left(\mathbf{C}^{-1}\right) = -\mathfrak{g}_{\mathbf{C}}(0) = 1 + p, \tag{20.36}$$

so that we get from Eq. (20.20) that, in the large N limit,

$$\mathcal{R}^2_{\text{true}} = \frac{\mathcal{G}^2}{N} \frac{1}{1+p}. \tag{20.37}$$

Next, we see from Eq. (20.34) that the optimal out-of-sample risk requires the computation of $\tau((\Xi)^{-1})$. In general, the computation of this quantity is highly non-trivial but some simplifications appear when **C** is an inverse-Wishart matrix. In the large-dimension limit, the final result reads

$$\tau((\Xi)^{-1}) = -\left(1 + \frac{q}{p}\right) \mathfrak{g}_{\mathbf{E}}\left(-\frac{q}{p}\right) = 1 + \frac{p^2}{p+q+pq}, \tag{20.38}$$

and therefore we have from Eq. (20.34)

$$\mathcal{R}^2_{\text{out}}(\Xi) = \frac{\mathcal{G}^2}{N} \frac{p+q+pq}{(p+q)(1+p)}, \tag{20.39}$$

and so it is clear from Eqs. (20.39) and (20.37) that, for any $p > 0$,

$$\frac{\mathcal{R}^2_{\text{out}}(\Xi)}{\mathcal{R}^2_{\text{true}}} = 1 + q\frac{pq}{(p+q)(1+p)} \geq 1, \tag{20.40}$$

where the last inequality becomes an equality only when $q = 0$, as it should.

It is also interesting to evaluate the in-sample risk associated with the optimal RIE. It is defined by

$$\mathcal{R}^2_{\text{in}}(\Xi) = \mathcal{G}^2 \frac{\mathrm{Tr}\big[(\Xi)^{-1}\mathbf{E}(\Xi)^{-1}\big]}{N\tau^2((\Xi)^{-1})}, \tag{20.41}$$

where the most challenging term is the numerator. Using the fact that the eigenvalues of Ξ are given by the linear shrinkage formula (19.49), one can once again find a closed formula.[2] The final result is written

$$\mathcal{R}^2_{\text{in}}(\Xi) = \frac{\mathcal{G}^2}{N} \frac{p+q}{(1+p)(p+q(p+1))}, \tag{20.42}$$

and we therefore deduce with Eq. (20.37) that, for any $p > 0$,

$$\frac{\mathcal{R}^2_{\text{in}}(\Xi)}{\mathcal{R}^2_{\text{true}}} = 1 - \frac{pq}{p+q(1+p)} \leq 1, \tag{20.43}$$

where the inequality becomes an equality for $q = 0$ as above.

Finally, one may easily check from Eqs. (20.25), (20.40) and (20.43) that

$$\mathcal{R}^2_{\text{in}}(\Xi) - \mathcal{R}^2_{\text{in}}(\mathbf{E}) \geq 0, \qquad \mathcal{R}^2_{\text{out}}(\Xi) - \mathcal{R}^2_{\text{out}}(\mathbf{E}) \leq 0\,, \tag{20.44}$$

showing explicitly that we indeed reduce the overfitting by using the oracle estimator instead of the SCM in the high-dimensional framework: both the in-sample and out-of-sample risks computed using Ξ are closer to the true risk than when computed with the raw empirical matrix **E**. The results shown in Figure 20.1 correspond to the inverse-Wishart case with $p = q = \frac{1}{2}$.

Exercise 20.2.1 Optimal portfolio when the true covariance matrix is Wishart

In this exercise we continue the analysis of Exercise 19.2.2 assuming that we measure an SCM from data with a true covariance given by a Wishart with parameter q_0.

[2] Details of this computation can be found in Bun et al. [2017].

https://doi.org/10.1017/9781108768900.024 Published online by Cambridge University Press

(a) The minimum risk portfolio with expected gain G is given by

$$\boldsymbol{\pi} = G\frac{\mathbf{C}^{-1}\mathbf{g}}{\mathbf{g}^T\mathbf{C}^{-1}\mathbf{g}}, \tag{20.45}$$

where $\mathbf{C}$ is the covariance matrix (or an estimator of it) and $\mathbf{g}$ is the vector of expected gains. Compute the matrix Ξ by taking the matrix $\mathbf{E}_1$ and replacing its eigenvalues λ_k by $\xi(\lambda_k)$ and keeping the same eigenvectors. Use the result of Exercise 19.2.2(c) for $\xi(\lambda)$, if some λ_k are below 0.05594 or above 3.746 replace them by 0.05594 and 3.746 respectively. This is so that you do not have to worry about finding the correct solution $\mathfrak{t}_{\mathbf{E}}(z)$ for z outside of the bulk.

(b) Build the three portfolios $\boldsymbol{\pi}_{\mathbf{C}}$, $\boldsymbol{\pi}_{\mathbf{E}}$ and $\boldsymbol{\pi}_{\Xi}$ by computing Eq. (20.45) for the three matrices $\mathbf{C}$, $\mathbf{E}_1$ and Ξ using $G = 1$ and $\mathbf{g} = \mathbf{e}_1$, the vector with 1 in the first component and 0 everywhere else. These three portfolios correspond to the true optimal, the naive optimal and the cleaned optimal. The true optimal is in general unobtainable. For these three portfolios compute the in-sample risk $\mathcal{R}_{\text{in}} := \boldsymbol{\pi}^T\mathbf{E}_1\boldsymbol{\pi}$, the true risk $\mathcal{R}_{\text{true}} := \boldsymbol{\pi}^T\mathbf{C}\boldsymbol{\pi}$ and the out-of-sample risk $\mathcal{R}_{\text{out}} := \boldsymbol{\pi}^T\mathbf{E}_2\boldsymbol{\pi}$.

(c) Comment on these nine values. For $\boldsymbol{\pi}_{\mathbf{C}}$ and $\boldsymbol{\pi}_{\mathbf{E}}$ you should find exact theoretical values. The out-of-sample risk for $\boldsymbol{\pi}_{\Xi}$ should better than for $\boldsymbol{\pi}_{\mathbf{E}}$ but worse than for $\boldsymbol{\pi}_{\mathbf{C}}$.

20.3 The Statistics of Price Changes: A Short Overview

20.3.1 Bachelier's First Law

The simplest property of financial prices, dating back to Bachelier's thesis, states that typical price variations grow like the square-root of time. More formally, under the assumption that price returns have zero mean (which is usually a good approximation on short time scales), then the *price variogram*

$$\mathcal{V}(\tau) := \mathbb{E}[(\log p_{t+\tau} - \log p_t)^2] \tag{20.46}$$

grows linearly with time lag τ, such that $\mathcal{V}(\tau) = \sigma^2\tau$.

20.3.2 Signature Plots

Assume now that a price series is described by

$$\log p_t = \log p_0 + \sum_{t'=1}^{t} r_{t'}, \tag{20.47}$$

where the return series r_t is covariance-stationary with zero mean and covariance

$$\operatorname{Cov}(r_{t'}, r_{t''}) = \sigma^2 C_r(|t' - t''|). \tag{20.48}$$

https://doi.org/10.1017/9781108768900.024 Published online by Cambridge University Press

The case of a random walk with uncorrelated price returns corresponds to $C_r(u) = \delta_{u,0}$, where $\delta_{u,0}$ is the Kronecker delta function. A trending random walk has $C_r(u) > 0$ and a mean-reverting random walk has $C_r(u) < 0$. How does this affect Bachelier's first law?

One important implication is that the volatility observed by sampling price series on a given time scale τ is itself dependent on that time scale. More precisely, the volatility at scale τ is given by

$$\sigma^2(\tau) := \frac{\mathcal{V}(\tau)}{\tau} = \sigma^2(1)\left[1 + 2\sum_{u=1}^{\tau}\left(1 - \frac{u}{\tau}\right)C_r(u)\right]. \tag{20.49}$$

A plot of $\sigma(\tau)$ versus τ is called a *volatility signature plot*. The case of an uncorrelated random walk leads to a flat signature plot. Positive correlations (which correspond to trends) lead to an increase in $\sigma(\tau)$ with increasing τ. Negative correlations (which correspond to mean reversion) lead to a decrease in $\sigma(\tau)$ with increasing τ.

20.3.3 Volatility Signature Plots for Real Price Series

Quite remarkably, the volatility signature plots of most liquid assets (stocks, futures, FX, ...) are nowadays almost flat for values of τ ranging from a few seconds to a few months (beyond which it becomes dubious whether the statistical assumption of stationarity still holds). For example, for the S&P500 E-mini futures contract, which is one of the most liquid contracts in the world, $\sigma(\tau)$ only decreases by about 20% from short time scales (seconds) to long time scales (weeks). For single stocks, however, some interesting deviations from a flat horizontal line can be detected, see Figure 20.2. The exact form of a volatility signature plot depends on the microstructural details of the underlying asset, but most liquid contracts in this market have a similar volatility signature plot.

20.3.4 Heavy Tails

An overwhelming body of empirical evidence from a vast array of financial instruments (including stocks, currencies, interest rates, commodities, and even implied volatility) shows that unconditional distributions of returns have *fat tails*, which decay as a power law for large arguments and are much heavier than the tails of a Gaussian distribution.

On short time scales (between about a minute and a few hours), the empirical density function of returns r can be fit reasonably well by a Student's t-distribution, see Figure 20.3. Student's t-distributions read

$$P(r) = \frac{1}{a\sqrt{\pi\mu}}\frac{\Gamma\left(\frac{1+\mu}{2}\right)}{\Gamma\left(\frac{\mu}{2}\right)}\left(1 + \frac{r^2}{a^2\mu}\right)^{-\frac{1+\mu}{2}}, \tag{20.50}$$

where a is a parameter fixing the scale of r. Student's t is such that $P(r)$ decays for large r as $|r|^{-1-\mu}$, where μ is the *tail exponent*. Empirically, the tail parameter μ is consistently found to be around 3 for a wide variety of different markets (see Fig. 20.3), which suggests

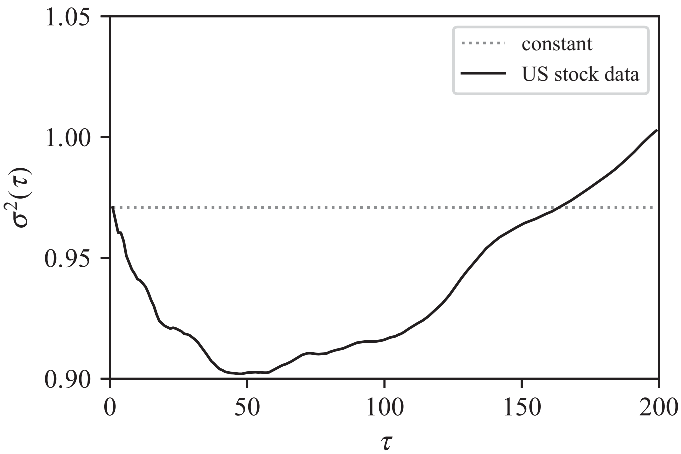

Figure 20.2 Average signature plot for the normalized returns of US stocks, where the x-axis is in days. The data consists of the returns of 1725 US companies over the period 2012–2019 (2000 business days), returns are normalized by a one-year exponential estimate of their past volatility. To a first approximation $\sigma^2(\tau)$ is independent of τ. The signature plot allows us to see deviations from this pure random walk behavior. One can see that stocks tend to mean-revert slightly at short times ($\tau < 50$ days) and trend at longer times. The effect is stronger on the many low liquidity stocks included in this dataset.

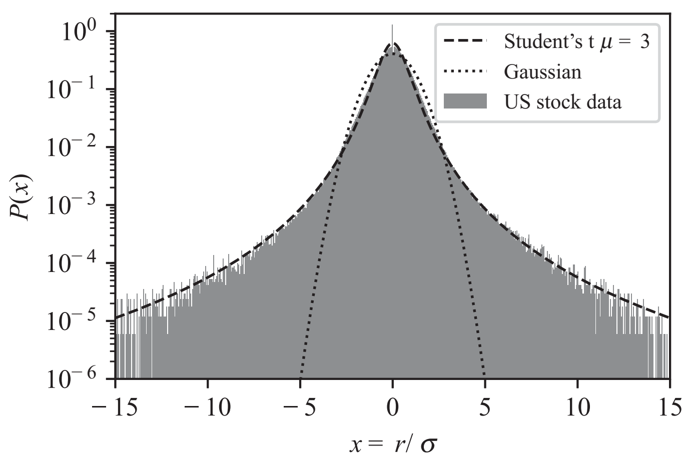

Figure 20.3 Empirical distribution of normalized daily stock returns compared with a Gaussian and a Student's t-distribution with $\mu = 3$ and the same variance. Same data as in Figure 20.2.

some kind of universality in the mechanism leading to extreme returns. This universality hints at the fact that fundamental factors are probably unimportant in determining the amplitude of most large price jumps. Interestingly, many studies indeed suggest that large price moves are often not associated with an identifiable piece of news that would rationally explain wild valuation swings.

https://doi.org/10.1017/9781108768900.024 Published online by Cambridge University Press

20.3.5 Volatility Clustering

Although considering the unconditional distribution of returns is informative, it is also somewhat misleading. Returns are in fact very far from being IID random variables – although they are indeed nearly uncorrelated, as their (almost) flat signature plots demonstrate. Therefore, returns are not simply independent random variables drawn from the Student's t-distribution. Such an IID model would predict that upon time aggregation the distribution of returns would quickly converge to a Gaussian distribution on longer time scales. Empirical data indicates that this is not the case, and that returns remain substantially non-Gaussian on time scales up to weeks or even months.

The dynamics of financial markets is in fact highly intermittent, with periods of intense activity intertwined with periods of relative calm. In intuitive terms, the volatility of financial returns is itself a dynamic variable that changes over time with a broad distribution of characteristic frequencies, a phenomenon called heteroskedasticity. In more formal terms, returns can be represented by the product of a time dependent volatility component σ_t and an IID directional component ε_t,

$$r_t := \sigma_t \varepsilon_t. \tag{20.51}$$

In this representation, ε_t are IID (but not necessarily Gaussian) random variables of unit variance and σ_t are positive random variables with long memory. This is illustrated in Figure 20.4 where we show the autocorrelation of the squared returns, which gives access to $\mathbb{E}[\sigma_t^2 \sigma_{t+\tau}^2]$.

It is worth pointing out that volatilities σ and scaled returns ε are in fact not independent random variables. It is well documented that positive past returns tend to decrease

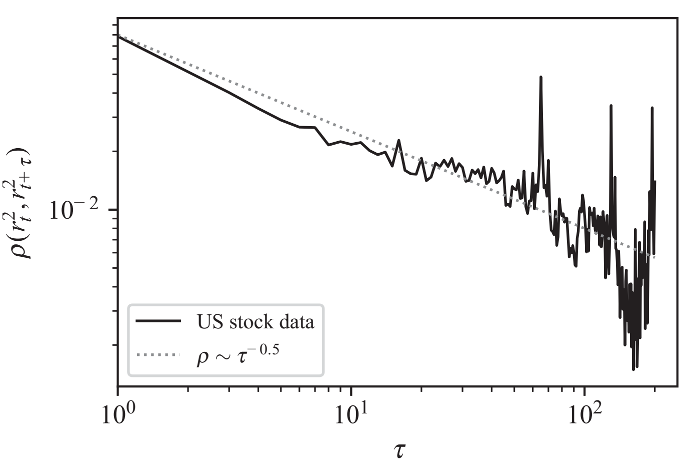

Figure 20.4 Average autocorrelation function of squared daily returns for the US stock data described in Figure 20.3. The autocorrelation decays very slowly with the time difference τ. A power law $\tau^{-\gamma}$ with $\gamma = 0.5$ is plotted to guide the eye. Note the three peaks at $\tau = 65$, 130 and 195 business days correspond to the periodicity of highly volatile earning announcements.

https://doi.org/10.1017/9781108768900.024 Published online by Cambridge University Press

future volatilities and that negative past returns tend to increase future volatilities (i.e. $\mathbb{E}[\varepsilon_t\sigma_{t+\tau}] < 0$ for $\tau > 0$). This is called the *leverage effect*. Importantly, however, past volatilities do not give much information on the sign of future returns (i.e. $\mathbb{E}[\varepsilon_t\sigma_{t+\tau}] \approx 0$ for $\tau < 0$).

20.4 Empirical Covariance Matrices

20.4.1 Empirical Eigenvalue Spectrum

We are now in position to investigate the empirical covariance matrix $\mathbf{E}$ of a collection of stocks. For definiteness, we choose $q = 0.25$ by selecting $N = 500$ stocks observed at the daily time scale, with time series of length $T = 2000$ days. The distribution of eigenvalues of $\mathbf{E}$ is shown in Figure 20.5. We observe a rather broad distribution centered around 1, but with a slowly decaying tail and a top eigenvalue λ_1 found to be $\sim$100 times larger than the mean. The top eigenvector $\mathbf{v}$ corresponds to the dominant risk factor. It is closely aligned with the uniform mode $[\mathbf{e}]_i = 1/\sqrt{N}$, i.e. all stocks moving in sync – hence the name "market mode" to describe the top eigenvector. Numerically, one finds $|\mathbf{v}^T\mathbf{e}| \approx 0.95$.

20.4.2 A One-Factor Model

The simplest model aimed at describing the co-movement of different stocks is to assume that the return $r_{i,t}$ of stock i at time t can be decomposed into a common factor f_t and IID idiosyncratic components $\varepsilon_{i,t}$, to wit,

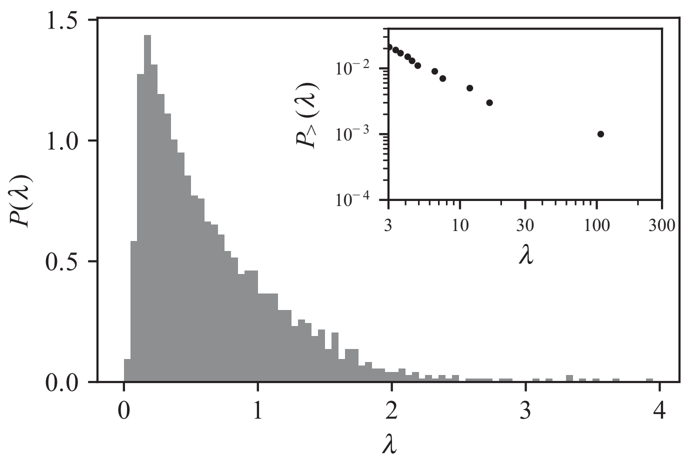

Figure 20.5 Eigenvalue distribution of the SCM, averaged over for three random sets of 500 US stocks, each measured on 2000 business days. Returns are normalized as in Figure 20.3, corresponding to $\bar{\lambda} = 0.97$. The inset shows the complementary cumulative distribution for the largest eigenvalues indicating a power-law behavior for large λ, as $P_>(\lambda) \approx \lambda^{-4/3}$. Note the largest eigenvalue $\lambda_1 \approx 0.2N$, which corresponds to the "market mode", i.e. the risk factor where all stocks move in the same direction.

https://doi.org/10.1017/9781108768900.024 Published online by Cambridge University Press

$$r_{i,t} = \beta_i f_t + \varepsilon_{i,t}, \tag{20.52}$$

where f_t is often thought of as the "market factor". Assuming further that $f_t, \varepsilon_{i,t}$ are uncorrelated random variables of mean zero and variance, respectively, σ_f^2 and σ_ε^2, the covariance matrix $\mathbf{C}_{ij}$ is simply given by

$$\mathbf{C}_{ij} = \beta_i \beta_j \sigma_f^2 + \delta_{ij} \sigma_\varepsilon^2. \tag{20.53}$$

Hence, the matrix $\mathbf{C}$ is equal to $\sigma_\varepsilon^2 \mathbf{1}$ plus a rank-1 perturbation $\sigma_f^2 \boldsymbol{\beta}\boldsymbol{\beta}^T$. The eigenvalues are all equal to σ_ε^2, save the largest one, equal to $\sigma_\varepsilon^2 + \sigma_f^2 |\boldsymbol{\beta}|^2$. The corresponding top eigenvector $\mathbf{u}_f$ is parallel to $\boldsymbol{\beta}$. When the β_i's are not too far from one another, this top eigenvector is aligned with the uniform vector $\mathbf{e}$, as found empirically.

From the analysis conducted in Chapter 14, we know that the eigenvalues of the empirical matrix corresponding to such a model are composed of a Marčenko–Pastur "sea" between $\lambda_- = \sigma_\varepsilon^2 (1 - \sqrt{q})^2$ and $\lambda_+ = \sigma_\varepsilon^2 (1 + \sqrt{q})^2$, and an outlier located at

$$\lambda_1 = \sigma_\varepsilon^2 (1 + a)(1 + \frac{q}{a}), \tag{20.54}$$

with $a = \sigma_f^2 |\boldsymbol{\beta}|^2 / \sigma_\varepsilon^2$, and provided $a > \sqrt{q}$ (see Section 14.4). Since $|\boldsymbol{\beta}|^2 = O(N)$, the last condition is easily satisfied for large portfolios. When $a \gg 1$, one thus finds

$$\lambda_1 \approx \sigma_f^2 |\boldsymbol{\beta}|^2. \tag{20.55}$$

Since empirically $\lambda_1 \approx 0.2N$ for the correlation matrix for which $\operatorname{Tr} \mathbf{C} = N$, we deduce that for that normalization $\sigma_\varepsilon^2 \approx 0.8$. The Marčenko–Pastur sea for the value of $q = 1/4$ used in Figure 20.5 should thus extend between $\lambda_- \approx 0.2$ and $\lambda_+ \approx 1.8$. Figure 20.5 however reveals that $\sim$20 eigenvalues lie beyond λ_+, a clear sign that more factors are needed to describe the co-movement of stocks. This is expected: the industrial sector (energy, financial, technology, etc.) to which a given stock belongs is bound to have some influence on its returns as well.

20.4.3 The Rotationally Invariant Estimator for Stocks

We now determine the optimal RIE corresponding to the empirical spectrum shown in Figure 20.5. As explained in Chapter 19, there are two possible ways to do this. One is to use Eq. (19.44) with an appropriately regularized empirical Stieltjes transform – for example by adding a small imaginary part to λ equal to $N^{-1/2}$. The second is to use a cross-validation method, see Eq. (19.88), which is theoretically equivalent as we have shown in Section 19.6. The two methods are compared in Figure 20.6, and agree quite well provided one chooses a slightly higher, effective value q^*, so as to mimic the effect of temporal correlations and fluctuating variance that lead to an effective reduction of the size of the sample (see the discussion in Section 17.2.3).

The shape of the non-linear function $\xi(\lambda)$ is interesting. It is broadly in line with the inverse-Wishart toy model shown in Figure 19.2: $\xi(\lambda)$ is concave for small λ, becomes approximately linear within the Marčenko–Pastur region, and becomes convex for larger λ.

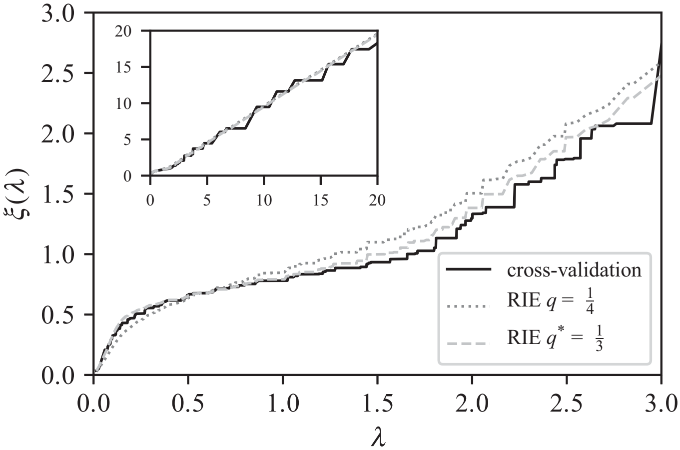

Figure 20.6 Non-linear shrinkage function $\xi(\lambda)$ computed using cross-validation and RIE averaged over three datasets. Each dataset consists of 500 US stocks measured over 2000 business days. Cross-validation is computed by removing a block of 100 days (20 times) to compute the out-of-sample variance of each eigenvector (see Eq. (19.88)). RIE is computed using the sample Stieltjes transform evaluated with an imaginary part $\eta = N^{-1/2}$. Results are shown for $q = N/T = 1/4$ and also for $q^* = 1/3$, chosen to mimic the effects of temporal correlations and fluctuating variance that lead to an effective reduction of the size of the sample (cf. Section 17.2.3). All three curves have been regularized through an isotonic fit, i.e. a fit that respects the monotonicity of the function.

For very large λ, however, $\xi(\lambda)$ becomes linear again (not shown), in a way compatible with the general formula Eq. (19.57). The use of RIE for portfolio construction in real applications is discussed in the references below.

Bibliographical Notes

- For a general introduction on Markowitz portfolio construction, see
  - E. J. Elton and M. J. Gruber. Modern portfolio theory, 1950 to date. *Journal of Banking & Finance*, 21(11):1743–1759, 1997,
  - M. Rubinstein. Markowitz's portfolio selection: A fifty-year retrospective. *The Journal of Finance*, 57(3):1041–1045, 2002,
  - P. N. Kolm, R. Ttnc, and F. J. Fabozzi. 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research*, 234(2):356–371, 2014.
- For an application of the RIE technique to portfolio construction, see
  - J. Bun, J.-P. Bouchaud, and M. Potters. Cleaning correlation matrices. *Risk magazine*, 2016,
  - J. Bun, J.-P. Bouchaud, and M. Potters. Cleaning large correlation matrices: Tools from random matrix theory. *Physics Reports*, 666:1–109, 2017,

https://doi.org/10.1017/9781108768900.024 Published online by Cambridge University Press

– P.-A. Reigneron, V. Nguyen, S. Ciliberti, P. Seager, and J.-P. Bouchaud. Agnostic allocation portfolios: A sweet spot in the risk-based jungle? *The Journal of Portfolio Management*, 46 (4), 22–38, 2020,

and references therein.

- For reviews on the main stylized facts of asset returns, see, e.g.
  – R. Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001,
  – J.-P. Bouchaud and M. Potters. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press, Cambridge, 2nd edition, 2003,
  – A. Chakraborti, I. M. Toke, M. Patriarca, and F. Abergel. Econophysics review: I. Empirical facts. *Quantitative Finance*, 11(7):991–1012, 2011,
  – J.-P. Bouchaud, J. Bonart, J. Donier, and M. Gould. *Trades, Quotes and Prices*. Cambridge University Press, Cambridge, 2nd edition, 2018.
- For early work on the comparison between the Marčenko–Pastur spectrum and the eigenvalues of financial covariance matrices, see
  – L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters. Noise dressing of financial correlation matrices. *Physical Review Letters*, 83:1467–1470, Aug 1999,
  – V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley. Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6):066126, 2002.
- For a study of the dependence of the instantaneous covariance matrix on some dynamical indicators, and the use of RMT in this case, see
  – P.-A. Reigneron, R. Allez, and J.-P. Bouchaud. Principal regression analysis and the index leverage effect. *Physica A: Statistical Mechanics and its Applications*, 390(17):3026–3035, 2011,
  – A. Karami, R. Benichou, M. Benzaquen, and J.-P. Bouchaud. Conditional correlations and principal regression analysis for futures. *preprint arXiv:1912.12354*, 2019.