Modern Portfolio Theory Revisited: from Real Traders to New Methods

THÈSE N^{O} 5255 (2012)

PRÉSENTÉE LE 20 JANVIER 2012

À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE DE BIOPHYSIQUE STATISTIQUE
PROGRAMME DOCTORAL EN PHYSIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

David MORTON DE LACHAPELLE

acceptée sur proposition du jury:

Prof. N. Grandjean, président du jury
Prof. P. De Los Rios, Dr M. Marsili, directeurs de thèse
Prof. J.-Ph. Bouchaud, rapporteur
Dr F. Hashemi, rapporteur
Prof. R. Mantegna, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2012

Acknowledgement

I am very much grateful to Swissquote Bank for the opportunity to carry out my thesis within the Quantitative Asset Management Dept. A warm thank you to the wonderful QAM team members for your constant support and all our fruitful discussions. I am truly indebted to my thesis Director Paolo de Los Rios and co-Director Matteo Marsili for their trust in me and my work, and to the three reviewers of my dissertation: Jean-Philippe Bouchaud, Rosario Mantegna, and Fariba Hashemi, for their involvement in reviewing the manuscript.

I dedicate this thesis to my family and friends who all came along for the ride and without whom none of this would have been possible. Special thanks go to my Dad, my Bro, Mum, Isa, SQ Noyau Dur, NT4R, and Aitana & EPFL friends.

Last but not least, very special thanks to you my dear and sweet Eva for your unflagging love and support.

Résumé

Le comportement des traders en-ligne est tout d'abord analysé puis modélisé. On montre que l'investisseur moyen se comporte comme un minimiseur de *mean-variance* en finance. Sous cette description, on montre également que les frais de transaction des courtiers jouent un rôle primordial dans l'explication des motifs d'investissement révélés par l'analyse, et en particulier dans celle d'une relation semi-empirique reliant l'investissement moyen à la valeur en portefeuille. Puisque les investisseurs en-ligne tiennent compte des coûts de transaction dans leur stratégie d'investissement, il sont également sensibles aux coût élevés de rotation des positions de leur portefeuille. Des solutions pour éviter de trop importantes fluctuations du portefeuille sont étudiées : tout d'abord dans le cas unidimensionnel, où l'on montre que les estimateurs de dispersion minimale améliorent la précision ainsi que la variance des estimations de volatilité et de valeur-à-risque, puis dans un environnement *multi-asset* où la matrice de covariance idéale devrait avoir de bonnes propriétés de conditionnement afin de maintenir le turnover du portefeuille à un niveau raisonnable. Plusieurs résultats théoriques sont obtenus en utilisant la théorie des matrices aléatoires, comme par exemple une équation pour la transformée de Stieltjes de la densité des valeurs propres de la matrices de corrélation de variables i.i.d., construite à l'aide d'un profile de poids décroisant dans le temps. Les résultats trouvés dans le cas unidimensionnel sont généralisés, menant à l'introduction des matrices de covariance à longue mémoire. Finalement, un actuel "fléau de la dimension" en finance est abordé par la généralisation de la méthode du *Spectral Coarse Graining*, une technique de réduction des réseaux complexes qui est étendue ici à la simplification du problème de l'optimisation de la *mean-variance*.

Mots-clés : allocation de portefeuille ; optimisation de la *mean-variance* ; matrices aléatoires ; valeur-à-risque ; volatilité ; stabilité des estimateurs ; conditionnement des matrices ; *spectral coarse-graining*

Summary

In the first place the behavior of (online) traders on markets is analyzed and modeled, and it is shown that the average investor behaves as a *mean-variance* optimizer in finance. Within this description, transaction costs play a key role in explaining observed investment patterns and in particular an important uncovered relation between average investment and portfolio value. As online investors take into account transaction costs in their investment strategy, they are also sensitive to high portfolio rebalancing costs. Solutions to avoid high portfolio turnovers are investigated: first in the one-dimensional case, where it is shown that estimators with minimal-dispersion improve both the accuracy and the variance of volatility and Value-at-Risk forecasts; second, in a multi-asset environment where the ideal covariance matrix must have good conditioning properties to maintain reasonable portfolio turnover. Theoretical results are derived using Random Matrix Theory, as for instance an equation for the Stieltjes transform of the eigenvalue density of i.i.d. correlation matrices with general time-decreasing weight profiles. Results found in the one-dimensional case are generalized leading to long-memory covariance matrices. Finally, a "curse of dimensionality" in portfolio allocation is tackled by generalizing the Spectral Coarse Graining, a method of reduction for complex networks, that is extended to the simplification of the mean-variance optimization problem.

Keywords : portfolio allocation ; mean-variance optimization ; random matrices ; value-at-risk ; volatilty ; estimator stability ; matrix conditioning ; spectral coarse-graining

Contents

1

Motivations

1.1 A time-proof theory

More than 50 years have elapsed since economist and Nobel Laureate Harry Markovitz put forward his theory of portfolio selection (1952). Fifty years during which the theory has been disputed, sometimes challenged, refined and improved to finally become one of the most significant breakthroughs of quantitative methods into the world of finance. Along with younger Black and Scholes pricing scheme for derivatives (1973) and Engle's autoregressive description of asset returns (1982), Modern Portfolio Theory (MPT) is today a major pillar of the financial investment worldwide and as such an essential ingredient of our market economy. But how did this theory gain such acclaim so that, time passing by, it is still at the very root of investment strategies deployed by many fund managers and financial institutions worldwide? What makes this theory so important, and what makes it last when many others have not survived the test of time? What does Markovitz's framework teach us about markets and its constituting element, the investor? What are tomorrows' challenges to be faced by portfolio allocation and how can the theory be improved to tackle this task? These are some of the questions that will be touched upon in this dissertation.

1.2 Agents building their portfolios: when costs and size matter

Empirical facts at the agent level

At the very ground of the financial system lies the investor. The price of an asset fluctuates, rises up, or suddenly collapses as a result of actions taken by investors—at least most of the times[1]. It seems therefore natural to start investigating financial markets by giving a close look at its constituting and determining elements. Two serious problems make this tempting approach difficult. Real investors greatly differ in means and strategies, which requires a theory capable of capturing heterogeneity at the agent level while still offering understandable and tractable outputs. This framework should be developed by taking into account empirical facts about real investors, which raises the issue of the availability of datasets with rich enough information on the market participants.

This is where online brokers can help. Swissquote Bank (the largest Swiss online broker) owns a dataset of comprehensive details about million of orders sent to the market by more than 180'000 online investors over the past 10 years. The level of details is impressive, as is the complexity of the data.

The dissertation opens with a paper, recently published in New Journal of Physics, that is the fruit of an intense collaboration with Damien Challet. In this study, we only touch the tip of the iceberg representing this huge amount of data, but the preliminary observations are shown to be extremely encouraging. Two of them have particularly caught my attention: (1) be they individuals, companies, or asset managers, online traders behave collectively as mean-variance optimizers and (2) a significant number of online investors take into account transaction costs in their investment strategy, and therefore do care about high portfolio turnovers. The first observation places the mean-variance optimization (MVO), cornerstone of Markowitz's theory, at a leading position of the first principles driving financial markets. The second one has lead me to wonder about the causes and remedies for high turnovers in the mean-variance framework.

Stable estimators: the issue of portfolio turnovers and transaction costs

When I started to think about this issue, I realized that volatility was given considerably more attention than portfolio turnover in MVO. This was surprising at first sight because the high

[1] Technical failures of markets may happen sometimes; when they occur market places have the means to restore their system, e.g. by canceling and reseting at a safer moment all the pending orders.

costs induced by important portfolio turnovers are one of the main causes of losses within the population of active traders. Why is it so, then? Perhaps because the major market players (e.g. financial institutions or hedge funds) seem to be rather insensitive to transaction costs (they usually negotiate flat-fee rates), or maybe because, at least in the mean-variance allocation framework, costly portfolio rebalancing can be easily avoided by adding "brute-force" constraints into the minimization process.

In my opinion if these were true reasons they would suffer from major drawbacks. On the one hand, as shown in this dissertation, a significant part of the market participants do care about high rebalancing costs of their positions in portfolio. They may not be the most influential ones, but their collective role has to do with fluctuations, liquidity, transmission of "information", and is probably as important as the one played by the biggest players (e.g. unlike institutions they were recently found to be consistently contrariant hence providing liquidity to the market).

On the other hand, adding constraints to the mean-variance optimization in order to limit transaction costs inevitably increases the portfolio *in-sample* volatility, which in turn induces higher *out-of-sample* risk leading to the opposite of the desired outcome. My goal is to shown that low portfolio turnovers, much like low volatilities, can and should result from an efficient and stable estimation of risk instead of from some ad hoc "fixing method".

In Chapter 3 I analyze the weighted-estimators of volatility that produce estimates with minimum dispersion and minimum mean-squared error. In particular, I show that these estimators exhibit long-memory, much like the volatility autocorrelation upon which they strongly depend.

The analysis is extended in Chapter 4 to handle multi-asset portfolios by introducing and analyzing a class of weighted estimators, the long-memory covariance matrices, which are shown to improve the accuracy of covariance forecasts while offering better control on portfolio turnovers than exponentially-weighted estimators. In this chapter, I make extensive use of Random Matrix Theory, a subject that was introduced to me by Olivier Lévèque who became my main collaborator on this project.

The last part of the dissertation was also motivated by practical considerations that showed up when evolving in the Swissquote environment. It is related to diversification and a so-called "curse of dimensionality" in the world of (online) brokers.

The curse of dimensionality: diversification and correlations in the stock market

Diversification is an essential ingredient of modern portfolio allocation. An investor who is able to diversify his investments across, say many stocks (at reasonable costs), reduces substantially the risk of losing an important part of his capital after e.g. the bankruptcy of any of them. With the *specific risk* diversified out, risk management mainly follows statistical and technical considerations; in short there is no need to scrutinize the results of a company to decide whether it is worth buying shares.

But is diversification always possible? As depicted in Fig. 1.1 cross-correlations in the stock market have been skyrocketing since year 2000. One therefore expects that investing in N stocks today does not have the same diversification power as the one that same investment would have had in the eighties or in the nineties. In other words to maintain a constant level of *effective* diversification in their portfolios asset managers should invest in more and more stocks as time goes by.

The latter constraint does pose a technical challenge. Large portfolios of stocks need constant rebalancing to stick to the target of their investment strategy (e.g. the mean-variance portfolio). This in turn is only feasible if one is able to recompute the target portfolio on a frequent basis and in an efficient way. In the mean-variance allocation this possibility is by no means guaranteed. Indeed, with thousands of stocks in portfolio even the more efficient algorithms would struggle recomputing at such a high frequency (e.g. intra-day) the optimal mean-variance portfolio.

Furthermore, the last few years have seen the democratization by online brokers of e-banking solutions based on quantitative allocation methods. When thousands of accounts are concerned with a similar but not identical investment strategy (i.e. the global strategy may be the same but parameter values are usually specific to the investor, e.g. risk tolerance), it leverages the computing power required to maintain the strategy.

Along with Paolo De Los Rios I have developed a spectral-based method of *coarse graining* to reduce to its *effective size* the risk estimator used in the mean-variance optimization (MVO), namely the covariance matrix of returns. The method was recently introduced by Paolo and David Gfeller as a mean to simplify complex networks while preserving important properties. To fit in the financial context it has been first thoroughly formalized (Appendix B), and then extended to the simplification of MVO in finance (Chapter 5).

The last research project of this thesis is directed to the biggest market players who are not

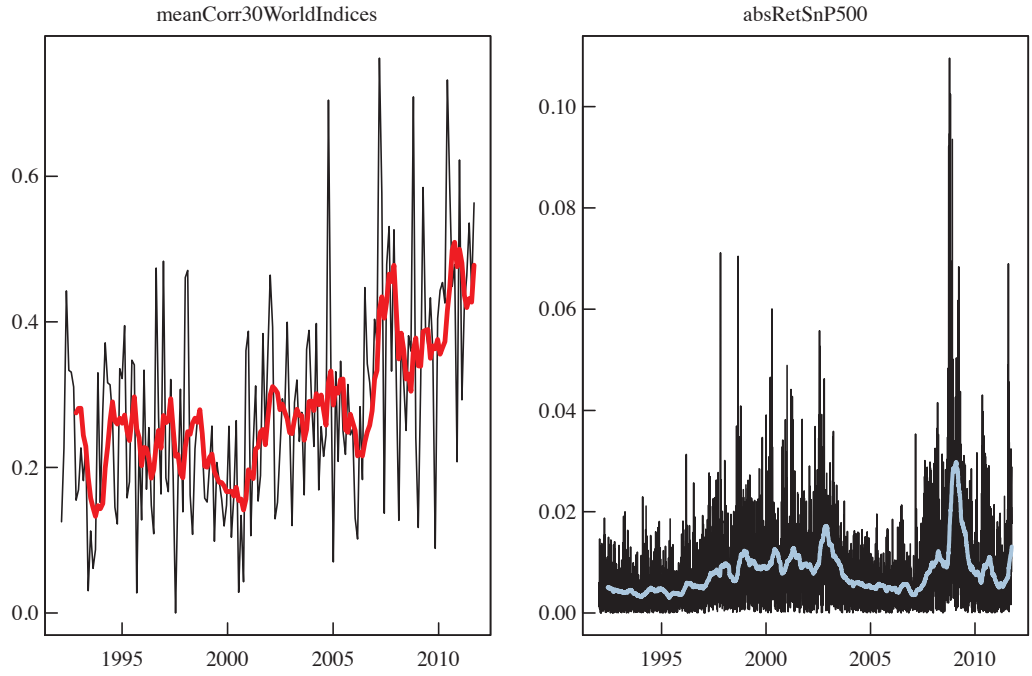


Figure 1.1: *Left plot: mean (Pearson) correlation between the daily returns of 30 major world stock indices. Right plot: the somewhat unrelated daily absolute returns of the S&P500 (i.e. a proxy for daily volatility on the stock market). Six-month moving averages are shown for better readability.*

sensitive to high portfolio turnovers and who do care about maintaining a high level of effective diversification in an increasingly correlated market.

2

Portfolio allocation: the perspective of online investors

2.1 Motivations

The availability of large data sets on financial markets is one of the main reasons behind the number and variety of works devoted to their analysis in various fields, and especially so in Econophysics since physicists much prefer to deal with very large data sets. At the macroscopic level, the analysis of millions of tick-by-tick data points uncovered striking regularities of price, volume, volatility, and order book dynamics (see [61, 15, 30, 13] for reviews). Since these phenomena are caused by the behavior of individual traders, news, and the interplay between the two, finding a microscopic mechanism that allows agent-based models to reproduce some of these stylized facts is an important endeavor meant to give us insight on the causes for large fluctuations, be it herding [29], competition for predictability [23], portfolio optimization leading to market instability [71], or chaotic transitions [19].

Market phenomenology appears as a typical example of collective phenomena to the eyes of statistical physicists. Thus, the temptation to regard the numerous power-laws found in empirical works as signatures of criticality is intense. But if the former are really due to a phase transition, one wishes at least to know what the phases are, which is hard to guess from the data alone. According to early herding theoretical models [29], the phase transition may lie in the density of social communication and imitation, and is of percolation type, thereby linking

power-law distributed price and volume, criticality and agent-behavior. The standard Minority Game [25] has also a single phase transition point where market predictability is entirely removed by the agents, without any specular effect on price and volume; on the other hand, grand-canonical MGs [75, 53, 21, 22] that allow the agents not to play have a semi-line of critical points that do produce stylized facts of price, volume and volatility dynamics; in the framework of statistical physics, the phase transition is due to symmetry breaking, i.e., it is a transition between predictable and perfectly efficient markets; this also suggests that the emergence of large fluctuations is due to market efficiency.

There are of course many other possible origins of power-laws in financial markets that have nothing to do with a second order phase transition. The simplest mechanism is to consider multiplicative random walks with a reflecting boundary [64]. Long-range memory of volatility is well-reproduced in agent-based models whose agents act or do nothing depending on a criterion based on a random walk [14]. Assuming pre-existing power-law distributed wealth, an effective theory of market phenomenology links the distributions of price returns, volume, and trader wealth [37]. On the other hand, markets are able to produce power-law distributed price returns by simple mechanisms of limit order placement and removal without the need for wealth inequality [24, 36]. However, in turn, one needs to explain why limit orders are placed in such manner; the heterogeneity of time scales may provide an explanation of order placement far away from best prices if power-law distributed [58], but additional work is needed in order to explain order placement near best prices, which causes these large price moves. Finally, a recent simple model of investment with leverage is able to reproduce some stylized facts [78].

But mechanisms alone may not be sufficient to replicate the full complexity of financial markets, as some part of it may lie instead in the heterogeneity of the agents themselves. While the need for heterogeneous agents in this context is intuitive (see e.g. [2]), there is no easily available data against which to test or to validate microscopically an agent-based model. Even if it is relatively easy to design agent-based models that reproduce some of the stylized facts of financial markets (see e.g [60, 20, 19, 23, 1]), one never knows if this is achieved for good reasons, except for volatility clustering [14]: it is to be expected that real traders behave sometimes at odds with one's intuition. Thus, without data about the traders themselves, one is left with the often frustrating and time-consuming task of reverse-engineering the market in order to determine the good ingredients indirectly. Some progresses have been made recently with the analysis of transactions in Spanish stock market aggregated by brokers [81], hence with mesoscale resolu-

tion.

Data on trader behavior is found in the files of brokers, usually shrouded in secrecy. But this lack of data accessibility is not entirely to blame for the current ignorance of real-trader dynamics: researchers, even when given access to broker data, have focused on trading gains and behavioral biases, often with factor-based analyses (see e.g. [6, 7, 32]).

We aim at providing a coherent picture of how various types of traders behave and interact, making it possible for agent-based models to rest on a much more solid basis. This work is the first step towards establishing stylized facts about trader characteristics and behavior. One of the most important aspects will be to characterize the heterogeneity of the traders in all respects (account value, turnover, trading frequency, behavioral biases, etc.) and the relationships between these quantities in probability distribution, not with factors. This chapter is first devoted to the description of the large data set that we use; it then focuses on the relationship between trader account value, turnover per transaction and transaction costs, both empirically and theoretically. We will show that while the traders have a spontaneous tendency to build equally-weighted portfolios, the number of stocks in a portfolio increases non-linearly with their account value, which we link to portfolio optimization and broker transaction fee structure.

2.2 Description of the data

Our data are extracted from the database of the largest Swiss on-line broker, Swissquote Bank SA (further referred to as Swissquote). The sample contains comprehensive details about all the 19 million electronic orders sent by 120'000 professional and non-professional on-line traders from January 2003 to March 2009. Of these orders, 65% have been canceled or have expired and 30% have been filled; the remaining 5% percent were still valid as of the 31st of March 2009. Since this study focuses on turnover as a function of account value, we chose to exclude orders for products that allow traders to invest more than their account value, also called leveraging, i.e., orders to margin-calls markets such as the foreign exchange market (FOREX) and the derivative exchange EUREX. The resulting sample contains 50% of orders for derivatives, 40% for stocks, and 4% for bonds and funds. Finally, 70% of these orders were sent to the Swiss market, 20% to the German market and about 10% to the US market.

Swissquote clients consist of three main groups: *individuals*, *companies*, and *asset managers*. In-

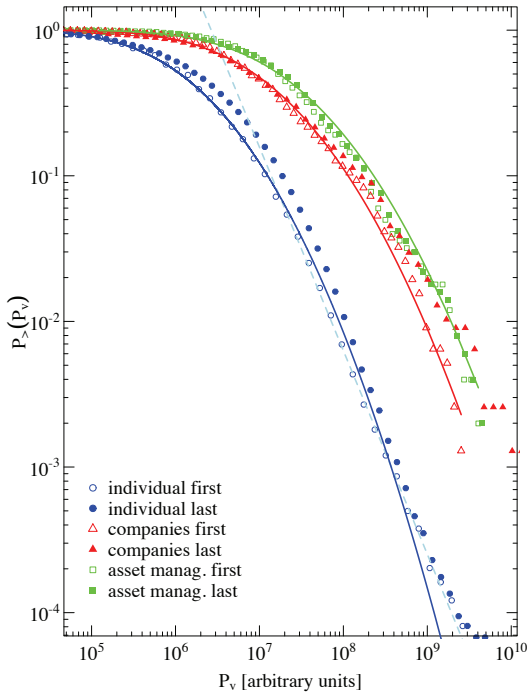


Figure 2.1: *Reciprocal cumulative distribution function of the portfolio value P_v for the three categories of clients at the time of their first (empty symbols) and last (filled symbols) transactions. Several models have been fitted to the data by Maximum Likelihood Estimation (MLE): the Student distribution (Pareto with plateau), the Weibull (stretched exponential), and the log-normal distribution. The best candidate, determined graphically and via bootstrapping the Kolmogorov Smirnov test [26] was found to be the log-normal distribution, which is the only one shown here for the sake of clarity. The dashed line in light blue results from a MLE fit to the tail of the individual traders with the Pareto distribution $p(x) \sim (x/x_{\min})^{-\gamma}$ (see section 2.3.1).*

dividual traders, also referred to as *retail clients*, are mainly non-professional traders acting for their own account. The accounts of companies are usually managed by individuals trading on behalf of a company and, as we shall see, behave very much like retail clients, albeit with a larger typical account value. Finally, asset managers manage accounts of individuals and/or companies, some of them dealing with more than a thousand clients; their behavior differ markedly from that of the other two categories of clients.

2.3 Results

2.3.1 Account values

Numerous studies have been devoted to the analysis and modeling of wealth dynamics and distribution among a population (see [84] and references therein). The general picture is that in a population, a very large majority lies in the exponential part of the reciprocal cumulative

Table 2.1: *Results of the fits of Pareto law $(x/x_{min})^{\gamma}$ to the account value P_v of individuals.*

individuals	γ	x_{min}
first transaction	$2.33 \in [2.29, 2.37]_{95}$	$2.30 \cdot 10^6 \in [1.99 \cdot 10^6, 2.59 \cdot 10^6]$
last transaction	$2.39 \in [2.33, 2.44]_{95}$	$3.73 \cdot 10^6 \in [3.15 \cdot 10^6, 4.29 \cdot 10^6]$

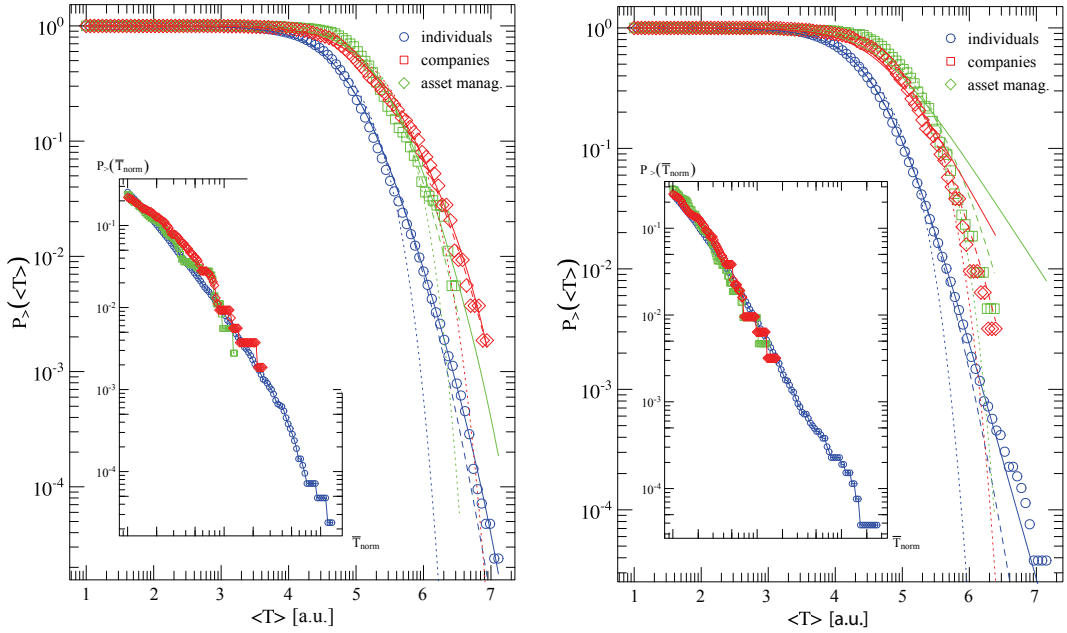
Table 2.2: *Parameter values and 95% confidence intervals for the MLE fit of the account values to the log-normal distribution $\ln N(\mu, \sigma^2)$. For each category of investors, the first and second row correspond to the account value at the time of the first, respectively the last transaction (see text). Note that portfolio values have been multiplied by an arbitrary number for confidentiality reasons. This only affects the value of μ.*

	μ	σ
individuals	13.94 ± 0.02	2.87 ± 0.01
	14.25 ± 0.02	2.01 ± 0.01
companies	16.0 ± 0.2	2.0 ± 0.1
	15.9 ± 0.2	2.4 ± 0.1
asset managers	16.7 ± 0.2	1.8 ± 0.1
	16.7 ± 0.2	2.0 ± 0.1

distribution function, while the wealth of the richest people is Pareto-distributed, i.e., according to a power-law.

The account value of Swissquote traders is by definition the sum of all their assets (cash, stock, bonds, derivatives, funds, deposits), and denoted by P_v. In order to simplify our analysis, we compute P_v once per day after US markets close and take this value as a proxy for the next day's account value. Figure 2.1 displays this distribution computed at the time of the first and last transactions of the clients. Results are shown for the three main categories of clients. Maximum likelihood fits to the tail of the individual traders to the Pareto model $p(x) \sim (x/x_{\min})^{-\gamma}$ were performed using the BC_a bootstrap method of [34] and determining the parameter $x_{\min}$ by minimizing the Kolmogorov-Smirnov statistics as in [26]. Results are reported in table 2.1.

The values of γ are in line with the wealth distribution of all major capitalistic countries (see [76] for a possible origin of Pareto exponents between 2.3 and 2.5). Thus the retail clients are most probably representative of the Swiss population. The account value distributions of companies and asset managers have no clear power-law tails, in agreement with the results of a recent model that suggests a log-normal distribution of mutual fund asset sizes [72]. Consequently, figure 2.1 also reports a fit of the data to log-normal distributions $\ln N(\mu, \sigma^2)$, which approximate more faithfully $P_>(P_v)$ than the Student and the Weibull distributions for the three categories of clients, except its extreme tail in the case of retail clients.



(a) *Stocks* **(b)** *Derivatives*

Figure 2.2: *Reverse cumulative distribution function of the mean turnover per transaction for the three categories of clients, and for both stock and derivative transactions. In the insets, the tail part of the RCDF of* $\langle T_{norm} \rangle = \langle T \rangle / mean(\langle T \rangle)$. *The solid curves are maximum likelihood fits to (2.1) for stocks and (2.2) for derivatives. The dotted lines are fits to the Weibull distribution and the dashed lines to the log-normal distribution.*

2.3.2 Mean turnover

The turnover of a single transaction i, denoted by T_i is defined as the price paid times the volume of the transaction and does not include transaction fees. We have excluded the traders that have leveraged positions on stocks, hence $T_i \leq P_v$; more generally one wishes to determine how the average turnover of a given trader relates to his portfolio value. In passing, since $P(P_v)$ has fat tails, the only way the distribution of T can avoid having fat tails is if the typical turnover is proportional to $\log(P_v)$. We denote by $\langle T \rangle$ the mean turnover per transaction for a given client over the history of his activities.

Figure 2.2 reports its reciprocal cumulative distributions functions (RCDF) for stocks and derivatives for the three categories of clients; all RCDFs have a first plateau and then a fat tail. For stocks, the tails are not a pure power laws, but they are for derivatives. Indeed, fitting the RCDFs with Weibull, log-normal and Zipf-Mandelbrot distribution with an exponential cut-off, defined as

$$F_{>}^{(1)}(x) = \frac{c^{\gamma} e^{-\beta x}}{(c+x)^{\gamma}}, \tag{2.1}$$

clearly shows that the latter is the only one that does not systematically underestimate the tail

Table 2.3: *Results of the maximum likelihood fit of $P_>(\langle T \rangle)$ with (2.1) and (2.2) for the three categories of clients. The 95% confidence intervals reported in smaller character are computed by the biased-corrected accelerated (BC_a) bootstrap method of [34].*

	Stocks (2.1)		Derivatives (2.2)
	γ	$\beta \cdot 10^{-6}$	γ
individuals	1.97 [1.83,2.10]	0.98 [0.46,1.5]	1.98 [1.91,2.15]
companies	1.29 [1.52,1.89]	1.66 [0.44,2.3]	-
asset managers	1.93 [1.47,2.93]	0.91 [-7.8,4.5]	-

of the RCDF for stocks; estimated values of β and γ given in table 2.3.

The RCDFs related to the turnover of transactions on derivative products have clearer power-law tails for retail clients, which we fitted with a standard Zipf-Mandelbrot function, defined as

$$F_>^{(2)}(x) = \frac{c^\gamma}{(c+x)^\gamma}. \tag{2.2}$$

The parameters estimated are to be found in table 2.3; because of the power-law nature of this tail, fits with Weibull and log-normal distributions are not very good in the tails. While the decision process that allocates a budget to each type of product may be essentially the same, the buying power is larger for derivative products, which may explain the absence of a cut-off. Fits for companies and asset managers is very difficult and mostly non-conclusive because of unsufficient sample size; the good quality of the tail collapse (see inset) tends to indicate that the three distributions are identical, but we could not fit the RCDF of companies and asset managers with (2.2); as reported in figure 2.2b, log-normal distributions are adequate choices in these cases; since the quality of the fits are poor, we do not report the resulting parameters.

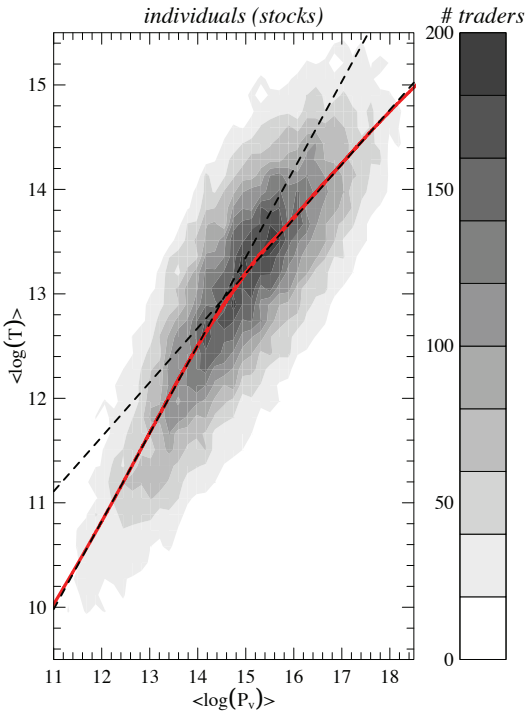


Figure 2.3: *Density plot of the average* $\log T$ *vs the average* $\log P_v$*, robust non-parametric fit (red line), and linear fits (dashed lines)*

2.3.3 Mean turnover vs account value

The relationship between $\langle T \rangle$ vs $\langle P_v \rangle$ is important as it dictates what fraction of their investable wealth the traders exchange in markets. We first produce a scatter plot of $\langle \log T \rangle$ vs $\langle \log P_v \rangle$ (figure 2.4). In a log-log scale plot, it shows a cloud of points that is roughly increasing. A density plot is however clearer for retail clients as there are many more points (figure 2.3).

These plots make it clear that there are simple relationships between $\log T$ and $\log P_v$. A robust non-parametric regression method [27] reveals a double linear relationship between $\langle \log T \rangle$ and $\langle \log P_v \rangle$ for all three categories of investors (see figures 2.4 and 2.3):

$$\langle \log T \rangle = \beta_x \langle \log P_v \rangle + a_x \tag{2.3}$$

where $x = 1$ when $\langle \log P_v \rangle < \Theta_1$ and $x = 2$ when $\langle \log P_v \rangle > \Theta_2$. Fitted values with confidence intervals are reported in table 2.5.

This result is remarkable in two respects: (i) the double linear relation, not obvious to the naked eye, separates investors into two groups (ii) the ranges of values where the transition occurs is very similar across the three categories of traders.

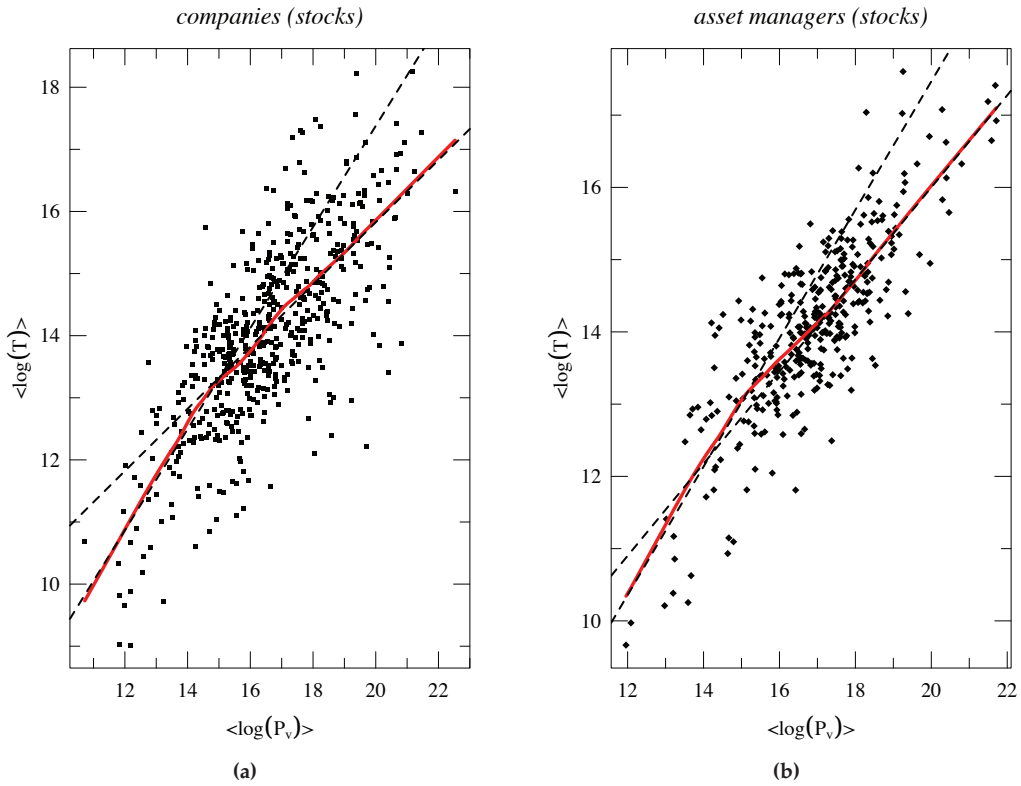


Figure 2.4: *Density plot of the average* $\log T$ *vs the average* $\log P_v$*, robust non-parametric fit (red line), and linear fits (dashed lines)*

Table 2.5: *Parameter values and 95% confidence intervals for the double linear model (2.4). For each category of investors, the first and second row correspond respectively to* $\langle \log P_v \rangle \leq \Theta_1$ *an* $\langle \log P_v \rangle \geq \Theta_2$*. For confidentiality reasons, we have multiplied* P_v *and* T *by a random number. This only affects the true values of* a_x *and* Θ *in the table.*

	β_x	a_x	ξ	Θ	R^2
individuals	0.84 ± 0.02	0.73 ± 1.25	0.71	14	0.52
	0.54 ± 0.01	5.07 ± 0.15	0.77	14.5	0.40
companies	0.81 ± 0.13	1.12 ± 8.17	0.88	15.5	0.47
	0.50 ± 0.07	5.82 ± 1.65	1.00	15.6	0.33
asset managers	0.89 ± 0.20	-0.31 ± 0.76	0.62	15.5	0.52
	0.63 ± 0.08	3.28 ± 5.78	0.62	16.5	0.46

The relationships above only applies to averages over all the agents. This means that there are some intrinsic quantities that make all the agents deviate from this average line. Detailed examination of the regression residuals show that the latter are for the most part (i.e. more than 95%) normally distributed with constant standard deviations ξ_x and that the residuals deviating from the normal distributions are not fat-tailed. This directly suggests the simple relation for individual traders

$$T^i = e^{a_x + \delta^i a_x} (P_v^i)^{\beta_x} \leq e^{\Theta_x} \tag{2.4}$$

where T^i and P_v^i are respectively the turnover and portfolio value of investor i, and $\delta^i a_x$ are i.i.d. $N(0, \xi_x^2)$ idiosyncratic variations independent from P_v that mirror the heterogeneity of the agents. As we shall see, portfolio optimization with heterogeneous parameters yields this precise relationship.

2.3.4 Turnover rescaled by account value

Let us now measure the typical fraction of wealth exchanged in a single transaction, defined as $Q = \left\langle \frac{T}{P_v} \right\rangle$. Since the inverse of this ratio is an indirect (and imperfect) proxy of the number N of assets that a trader owns, it also indicates how well diversified his investments are, hence, it can be viewed a simple proxy of the risk profiles of the agents.

2.3.4.1 data

Figure 2.5 shows that the distributions look exponential to a naked eye for about 90% of the individuals and nearly 80% of the companies, while that of the asset managers is rapidly more complex that a simple exponential. We derive exact relationships for this quantity in subsection 2.3.4.2 that show that these distributions are in fact not exponential but log-normal.

The resulting picture is that only a small fraction of customers trade a large fraction of their wealth on average. Interestingly, these figures show a clear difference between the three categories of clients. As discussed above, figure 2.5 roughly reflects the risk profile of the different types of customers: less than 10% of asset managers trade on average more than 20% of their clients' capital in a single transaction; this rises to 30% for companies, and 45% for retail clients. Note however that despite the fact that the account values of companies and asset managers are comparable, companies tend to have a Q closer to that of the individuals; this suggests either

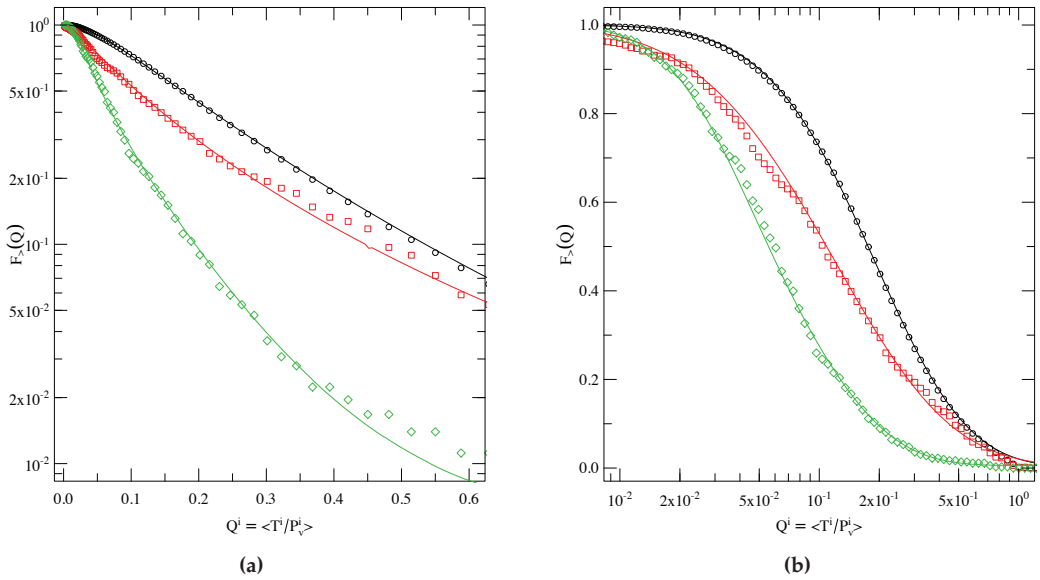


Figure 2.5: *Reverse cumulative distribution function of* $Q = \left\langle \frac{T}{P_v} \right\rangle$*, the mean ratio of the turnover over the portfolio value for individual traders (black), companies (red) and asset managers (green). Left plot is in lin-log scale and right plot is in log-lin scale. Solid lines come from theoretical predictions of section 2.3.4.2.*

that companies hold a smaller N than asset managers for the same account value, or that asset managers tend to make smaller adjustments to the quantities of assets.

2.3.4.2 theory

Since we know the distributions of T, P_v and their relationship, we are in a position to derive analytical expressions for $Q^i = \left\langle \frac{T(t)}{P_v(t)} \right\rangle$ of investor i. The distribution of Q across the population of on-line investors can be easily found using (2.4) and the distribution of P_v. Let $P_{T,P_v}(t, p_v)$ denote the joint distribution of T and P_v:

$$P_{Q=\frac{T}{P_v}}(q) = \int_0^\infty p_v P_{T,P_v}(qp_v, p_v)\, dp_v = \int_0^\infty p_v P_{T|P_v}(qp_v|p_v) P_{P_v}(p_v)\, dp_v. \tag{2.5}$$

Let us now assume for the sake of clarity that $T = e^{a+\delta a} P_v^\beta$. Given P_v, the turnover T follows a log-normal distribution with mean $\log p_v + a$ and variance ξ^2. Substituting $P_{T|P_v}(t|p_v) = \ln N\left(\log p_v + a, \xi^2\right)$ in (2.5) leads after some simplifications to

$$P_Q(q) = \int_0^\infty \frac{1}{\sqrt{2\pi\xi^2} q} \exp\left(-\frac{\left(\log(qp_v^{1-\beta}) - a\right)^2}{2\xi^2} \right) P_{P_v}(p_v)\, dp_v, \tag{2.6}$$

and

$$F_Q(q) = \int_0^q P_Q(x)\,dx = \int_0^\infty \frac{1}{2}\mathrm{erfc}\left(\frac{a - \log(qp_v^{1-\beta})}{\sqrt{2}\xi}\right) P_{P_v}(p_v)\,dp_v, \tag{2.7}$$

where $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-y^2}\,dy$ is the complementary error function. As expected, when $\beta = 0$ (i.e. T and P_v are independent), we recover the product of the two marginal distributions. On the other hand, when $\beta = 1$, i.e., when T is proportional to P_v, $P_Q(q) = \ln N\left(a, \xi^2\right)$, which is the distribution of the factor $e^{a+\delta a}$. For other values of β the functions P_Q and F_Q cannot be determined analytically unless P_{P_v} takes a particular form as shown below. However, the moments of $P_Q(q)$ can be arranged in a simpler form:

$$E(q^n) = \int_0^\infty q^n P_Q(q)dq = e^{na+\frac{1}{2}n^2\xi^2}\int_0^\infty \frac{1}{p_v^{n(1-\beta)}} P_{P_v}(p_v)\,dp_v, \tag{2.8}$$

that is, the (log-normal) moments of T/P_v times an integral term smaller or equal to 1 (because in practice $P_{P_v}(p_v) > 1$)[1]. Hence, the relation $E(q^n) \leq e^{na+\frac{1}{2}n^2\xi^2}$ with equality when $\beta = 1$ holds for any distribution of the account value P_v.

In section 2.3.1, we have shown that the distribution of P_v is well-approximated by a log-normal distribution. This particular choice of distribution makes the previous integrals analytically tractable. Indeed, with $P_{P_v} = \ln N(\mu, \sigma^2)$ straight integration of (2.6) leads to $P_Q = \ln N(M, S^2)$, where $M = a - (1-\beta)\mu$ and $S^2 = \xi^2 + (1-\beta)^2\sigma^2$. This simple result has some practical interest: given the distribution parameters and the coupling factor β, one can draw realistic q factors for agent-based modeling as $Q = e^{M+SX}$, where X is $\mathcal{N}(0,1)$ distributed. Furthermore, in the next section, we show how the value of β may be inferred from the transaction cost structure, which decreases the number of parameters to four.

Figure 2.5 confirms the validity of the above theoretical results, once expanded to the case of a bi-linear relation between T and P_v. It is noteworthy that the continuous lines are no fits on empirical q factors, but use instead the results of the separate fits on the turnover and account distributions.

[1] Mathematically, all the moments of Q always exist since $\beta \leq 1$ and $P_v(p_v)$ must decay faster than p_v^{-1} to be a valid distribution.

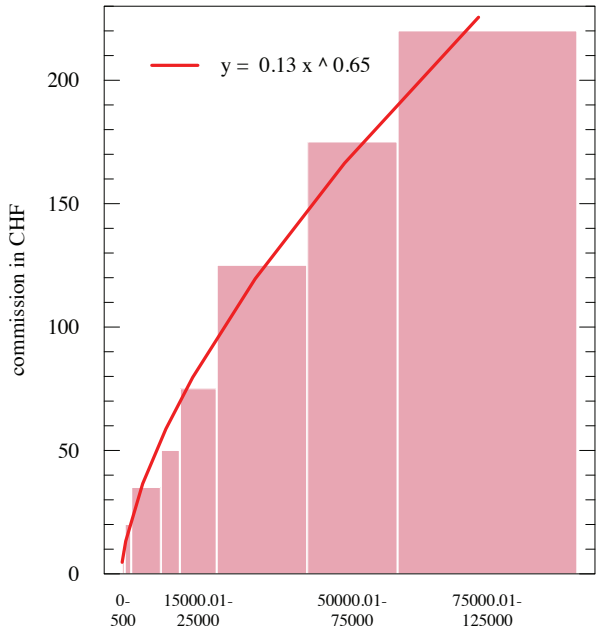


Figure 2.6: *Swissquote fee curve for the Swiss stock market. Commissions based on a sliding scale of costs are common practice in the world of on-line finance. The red line results from a non-linear fit to equation 2.10. Parameter values are* $C = 0.13 \in [0.05, 0.5]_{95}$ *and* $\delta = 0.63 \in [0.5, 0.74]_{95}$*, where the 95% confidence intervals are obtained from the* BC_a *bootstrap method of [34].*

2.4 The influence of transaction costs on trading behavior: optimal mean-variance portfolios

Apart from risk profiles, education, and typical wealth, the differences in the turnover as a function of wealth observed above between the three populations of traders may also lie in the difference of their actual transaction cost structure. Swissquote current standard structure for the Swiss market (its shape is very similar for European and US markets) is shown in figure 2.6; it is a piece-wise constant, non-linear looking function. Fitting all segments to equation 2.10 gives $\delta = 0.63 \in [0.5, 0.74]_{95}$. The fee structure of most brokers is not set in stone and can be negotiated. A frequent request is to have a flat fee, i.e. a fixed cost per transaction corresponding to a constant function. Since quite clearly the negotiation power of large clients or of clients that carry out many transactions is more important, asset managers are more likely to obtain a more favorable fee structure than basic retail clients.

Since buying some shares of an asset is the result of unconscious or calculated portfolio construction process, one first needs a theoretical reference point with which to compare the population

characteristics as measured in the previous subsection. In other words, we shall use results from portfolio optimization theory with non-linear transaction cost functions to understand the results of the previous subsection.

Quite curiously, all analytical papers in the literature on optimal portfolios either neglect transaction costs or assume constant or linear transaction cost structures; non-linear structures are tackled numerically; thus, we incorporate the specific non-linear transaction cost structure faced by the traders under investigation in the classic one-shot portfolio optimization problem studied by Brennan [18], who restricted its discussion to fees proportional to the number of securities, in other words, a flat fee per transaction.

Building optimal mean-variance stock portfolios consists for a given agent in selecting which stock to invest in and in what proportion by maximizing the expected portfolio growth, usually called return, while trying to reduce the resulting a priori risk. One cost function that corresponds to such requirements is

$$L_\lambda(R) = \lambda E(R)\text{-Var}(R), \tag{2.9}$$

where R is the stochastic return of the portfolio over the investment horizon (e.g., one month, one year) and λ tunes the trade-off between risk and return; as such, it can be interpreted as a measure of an investor's attitude towards risk: the larger λ , the more risk-adverse the investor.

The return of the portfolio can be decomposed into contributions from risky assets (stocks, derivatives, etc.), the interests of the amount kept in cash, and the total relative cost of broker commission, which we denote as $R = R^{risky} + R^{cash} - R^{cost}$. Mathematically,

- $R^{risky} = \Sigma_{i=1}^{N} x_i R_i$, where R_i is the return of stock i over this horizon, x_i is the fraction of the total wealth invested in this stock, and N is the total number of investable assets; we shall denote the total fraction of wealth invested in risky assets by $x = \sum_{i=1}^{N} x_i$;
- $R^{cash} = (1 - x)r$, where r is the interest rate;
- $R^{cost} = \frac{\sum_{i=1}^{N} F(x_i P_v)}{P_v}(1 + r)$, where $F(x)$ is the amount charged by a broker to exchange an amount x of cash into shares or vice-versa.

The focus of this section is to derive explicit relationships between F, the number of assets to hold in a portfolio, and the account value P_v. Whereas previous works only considered special

cases for F that are not compatible with the fees structure of Swissquote, we need to introduce a cost function that can accommodate all the standard broker commission schemes. The two extreme cases are i) flat-fee per transaction, i.e., a fixed cost that does not depend on the amount exchanged ii) a proportional scheme, possibly with a maximum fee. Swissquote's standard scheme stands in between and is well approximated by a power-law with a maximum fee F_{max}. We hence choose

$$F(x_i P_v) = \min\left(C(x_i P_v)^{\delta}, F_{max}\right), \tag{2.10}$$

where δ interpolates between a flat-fee ($\delta = 0$), as in [18], and a proportional scheme ($\delta = 1$) via a power-law, and C is a constant.

Following the well-known one-factor model of Sharpe [73], we assume that the return of asset i follows the global market's return R_M with an idiosyncratic proportionality factor β_i. More specifically,

$$R_i = \beta_i(R_M - r) + r + \varepsilon_i, \tag{2.11}$$

where ε_i is an uncorrelated white noise $E(\varepsilon_i) = E(\varepsilon_i\varepsilon_j) = E(R_M\varepsilon_i) = 0$. This equation means that the systematic idiosyncratic part of R_i only applies to the return above the risk-free interest rate, also called market risk premium.

This completely specifies the functional L_λ. Returning to (2.9), one first computes the expectation and variance of the portfolio return:

$$\begin{aligned} E(R) &= \sum_{i=1}^{N} x_i E(R_i) + (1-x)r - \frac{\sum_{i=1}^{N} F(x_i P_v)}{P_v}(1+r), \\ &= (E(R_M) - r)\sum_{i=1}^{N} x_i\beta_i + r - \frac{(1+r)C}{P_v^{1-\delta}}\sum_{i=1}^{N} x_i^{\delta}, \end{aligned} \tag{2.12}$$

and

$$\begin{aligned} \mathrm{Var}(R) &= \mathrm{Var}(R^{risky}) \\ &= \mathrm{Var}(R_M)\sum_{i=1}^{N}(x_i\beta_i)^2 + \sum_{i=1}^{N} x_i^2 \mathrm{Var}(\varepsilon_i). \end{aligned} \tag{2.13}$$

Note that, since here the risk-free rate is non-random, the portfolio variance is independent of both the risk-free investment and broker commission; this does not hold for the expected return.

In principle, the functional L depends on N, the number of assets in the portfolio, λ the risk parameter, and x_i the fraction of account value to invest in risky product i. Assuming that x_i is

constant for all i (i.e. equally-weighted allocation), we are left with only three parameters since $x_i = x/N$. Thus, from the optimization of the resulting functional one can obtain a relationship between any two of these parameters. We are mostly interested in N as a function of x.

2.4.1 Non-linear relationship between account value and number of assets

We will first assume that agents seek the optimal fraction of their account value x^* to invest in N securities—*N being known*—given the risk free rate r and broker commission $F(x_i W)$. The optimal solution is simply obtained by setting $x_i = x/N$ in (2.12) and (2.13), and by equating to zero the derivative of (2.9) with respect to x. This leads to the following transcendental equation for x^*:

$$x^* = \frac{\lambda}{2} \frac{\bar{\beta}(E(R_M) - r) - \delta(1+r)C(\frac{N}{x^* P_v})^{1-\delta}}{\bar{\beta}^2 \mathrm{Var}(R_M) + \frac{1}{N}\overline{\mathrm{Var}}(\varepsilon)}, \tag{2.14}$$

where $\bar{\beta} = \frac{1}{N}\sum_{i=1}^{N} \beta_i$ and $\overline{\mathrm{Var}}(\varepsilon) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{Var}(\varepsilon_i)$ is the mean idiosyncratic volatility. Provided the investor risk tolerance λ has been reliably estimated, which is usually a complex task [83], and that Sharpe model is adequate, (2.14) can be used directly in a real-world portfolio optimization problem. The β_i and ε_i are then obtained by regressing the returns of all the stocks with (2.11); the optimal solution is expected to be reliable in the absence of significant residual correlations between ε_i and ε_j. In the more common situation where λ is unknown, one can derive a second equation for the optimal number of securities under the assumption that portfolios are sufficiently homogeneous, or that the investment horizon is long enough so as to have $\bar{\beta}$ and $\overline{\mathrm{Var}}(\varepsilon)$ independent from N. As shown in figure 2.7, $\bar{\beta}$ on the US stock market is persistently close to one for various time horizons and values of N, consistently with the homogeneous assumption. Taking a few technical precautions into account ([18]), the differentiation of the Lagrangian (2.9) with respect to N leads to

$$\lambda = \frac{\overline{\mathrm{Var}}(\varepsilon) P_v^{1-\delta}}{(1-\delta)C(1+r)\left(\frac{N^*}{x}\right)^{2-\delta}}, \tag{2.15}$$

where it is assumed that $\delta < 1$ since for $\delta = 1$ the optimum investment does not depend on N through the cost function. According to (2.15), the agent risk tolerance increases with their account value P_v, in agreement with various survey studies on the risk tolerance of actual

investors (see the literature review of [82]). Using (2.14) and (2.15) to get rid of λ, we obtain

$$N^{2-\delta}\left(1+\frac{\delta}{1-\delta}\frac{K}{N}\right)=K\frac{\bar{\beta}(E(R_M)-r)}{(1-\delta)C(1+r)}\left(xP_v\right)^{1-\delta}, \tag{2.16}$$

where K is the ratio of residual risk to market risk defined as

$$K=2\left(\frac{\bar{\beta}^2\mathrm{Var}(R_M)}{\overline{\mathrm{Var}}(\varepsilon)}+\frac{1}{N}\right)^{-1}\underset{N\gg 1}{\approx}2\frac{\overline{\mathrm{Var}}(\varepsilon)}{\bar{\beta}^2\mathrm{Var}(R_M)}. \tag{2.17}$$

Given the desired level of systematic risk x, (2.16) can be solved for N numerically in an actual portfolio optimization. Further insight is gained by considering the high diversification limit $N\gg 1$, which yields $1+\frac{\delta}{1-\delta}\frac{K}{N}\approx 1$ in (2.16) and thus

$$N=\left(K\frac{\bar{\beta}(E(R_M)-r)}{(1-\delta)C(1+r)}\right)^{\frac{1}{2-\delta}}\left(xP_v\right)^{\frac{1-\delta}{2-\delta}}, \tag{2.18}$$

where K is given by the right-hand side of (2.17). The latter equation generalizes [18] to the case of a varying cost impact represented here by the parameter δ (i.e. the result of [18] is recovered by setting $\delta=0$ and $\beta_i=1$ in (2.18)). These results can be further generalized to non-equally weighted portfolios by differentiating (2.9) with respect to x_i and assuming again an homogeneous condition for the β_is.

In essence, (2.18) says that the number of securities held in an equally-weighted mean-variance portfolio with Sharpe-like returns is related to the amount invested as

$$\log(N)=\frac{1-\delta}{2-\delta}\log(xP_v)+\kappa \tag{2.19}$$

in the high diversification limit, where κ is the pre-factor of $\left(xP_v\right)^{\frac{1-\delta}{2-\delta}}$ in (2.18). The last equation gives N as a function of P_v for a predefined x in the optimal portfolio. The heterogeneity of the traders, beyond their account value, is not apparent yet, but may occur both in x and κ: first each trader may have his own preference regarding the fraction of this account to invest in risky assets, x; therefore one should replace x by x^i; next, κ includes both a term related to transaction costs, which does vary from trader to trader, and some measures and expectation of market returns and variance; each trader may have his own perception or way of measuring them, hence κ should also be replaced by κ^i. Finally, both terms can be merged in the same constant term $\zeta^i=\frac{1-\delta}{2-\delta}\log(x^i)+\kappa^i$. This explains how the heterogeneity of the traders is the

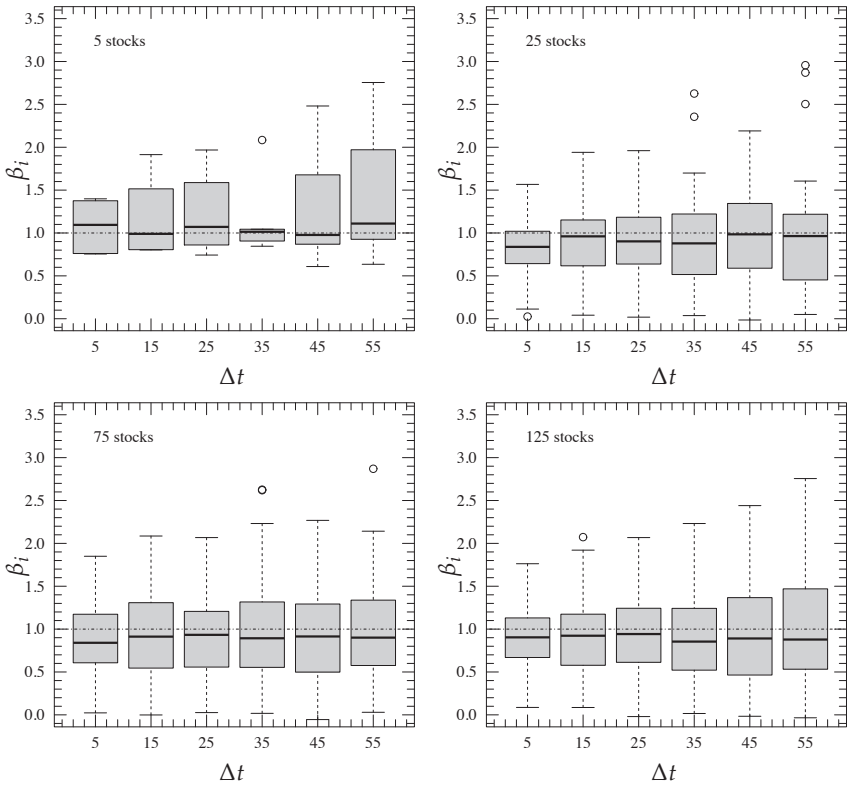


Figure 2.7: *Box-plot of empirical βs obtained from the regression of several US stocks on the S&P500. The observation period covers 2001 to 2008 and returns are computed on various time horizons Δt (in days). Results show that $\bar{\beta} = \frac{1}{N}\sum_{i=1}^{N}\beta_i \approx 1$ for all values of Δt and (even small) N, consistently with the homogeneous assumption of section 2.4.1.*

cause of fluctuations in the kind of relationships we are interested in.

2.5 Turnover, number of assets and account value

The result above only links N with P_v, but one also wishes to obtain relationships that involve the turnover per transaction, T. Whereas in section 2.3, we have characterized the turnover of any transaction, the results of section 2.4 rest on the assumption that the agents build their portfolio by selecting a group of assets and stick to them over a period of time. This, obviously, does not include the possibility of speculating by a series of buy and sell trades on even a single asset, nor portfolio rebalancing which consists in adjusting the relative proportions of some assets. We thus have to find a way to differentiate between portfolio building, rebalancing and speculation. Here, we shall focus on portfolio building in order to test and link the results of section 2.4 to those of section 2.3.

We have found a simple effective method that can separate portfolio-building transactions from

the other ones: we assume that the transactions of trader i that correspond to the building of his portfolio are restricted the first transaction of assets not traded previously; sell orders are ignored, since Swissquote clients cannot short sell easily. In other words, if trader i owns some shares of assets A, B, and C and then buys some shares of asset D, the corresponding transaction is deemed to contribute to his portfolio building process; the set of such transactions is denoted by Φ_i, while the full set of transactions is denoted by Ω_i. Any subsequent transaction of shares of assets A, B, C, or D are left out of Φ_i. The number of different assets that trader i owns is supposed to be $N_i \simeq |\Phi_i|$ where $|X|$ is the cardinal of set X; this approach assumes that a trader always owns shares in all the assets ever traded; surprisingly, this is by large the most common case. We shall drop the index i from now on.

Let us now focus on $T_\Phi = \sum_{k\in\Phi} T_k$, the total turnover that helped building his portfolio (i.e. the turnover on the transactions that allowed him to build his portfolio). We should first check how it is related to the total portfolio value P_v. Let us define $\langle P_v \rangle_\Phi$, the account value of a trader averaged at the times at which he trades a new asset.Plotting $\log \langle P_v \rangle_\Phi$ against $\log T_\Phi$ gives a cloudy relationship, as usual, but the fitting it with $\log \langle P_v \rangle_\Phi = \chi \log T_\Phi$ gives $\chi = 1.03 \pm 0.02$ for individuals, $\chi = 0.99 \pm 0.02$ for asset managers and $\chi = 1.00 \pm 0.01$ for companies with an adjusted $R^2 = 0.99$ in all cases. This relationship trivially holds for the traders who buy all their assets at once, as assumed in the portfolio model. The traders who do not lie on this line either hold positions in cash (in which case this line is a lower bound), or do not build their portfolio in a single day: they pile up positions in derivative products or stocks whose price fluctuations are the origin of the devations from the line. But the fact that the slope is close to 1 means that the average fluctuation is zero, hence, that on average trades do not make money from the positions taken on new stocks. The consequence of this is that $\log P_v$ can be replaced by $\log T_\Phi$ in (2.19), thus, setting $x = 1$,

$$\log N = \frac{1-\delta}{2-\delta} \log T_\Phi + \kappa \tag{2.20}$$

The $x = 1$ assumption is in fact quite reasonable: most Swissquote traders do not use their trading account as savings accounts and are fully invested; we do not know what amount they keep on their other bank accounts.

A robust non-parametric fit does reveal a linear relationship between $\log N$ and $\log T_\Phi$ in a given

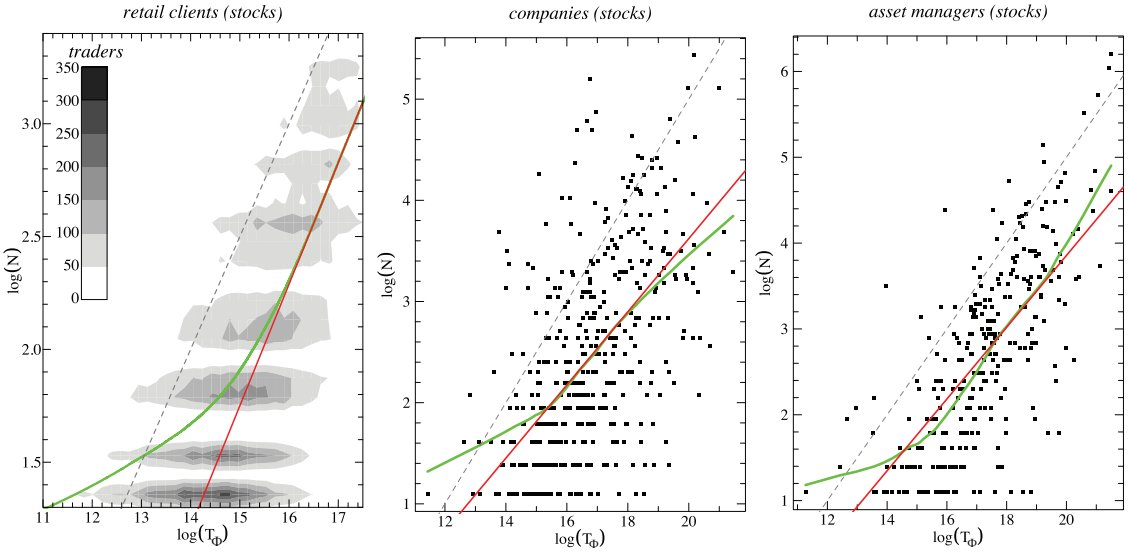


Figure 2.8: *Turnover of transactions contributing to the building of a portfolio T_Φ versus the number N of assets held by a given trader at the time of the transaction. Green lines: non-parametric fit; red lines: fits of the linear part of the non-parametric fit. From left to right: companies, asset managers, and individuals.*

Table 2.6: *Slope α linking $\log T_\Phi$ and $\log N$ for the three trader categories.*

	individuals	companies	asset managers
α	0.52 ± 0.02	0.36 ± 0.14	0.44 ± 0.13
$\log T_\Phi \in$	$[16, 19]$	$[17, 19.8]$	$[15.8, 18]$

Table 2.7: *Results of the double linear regression of $\log \langle T \rangle_\Phi$ versus $\log \langle P_v \rangle_\Phi$. For each category of investors, the first and second row correspond respectively to $\log \langle P_v \rangle_\Phi \leq \Theta_1$ an $\log \langle P_v \rangle_\Phi \geq \Theta_2$, where $\Theta_{1,2}$ have been determined graphically using the non-parametric method of [27] as in section 2.3.3. Parameters are as in the double linear model (2.4). For confidentiality reasons, we have multiplied P_v and T by a random number, which only affects the true values of $\Theta_{1,2}$ and of the ordinate a_x.*

	β_x	a_x	ξ	Θ	R^2
individuals	0.85 ± 0.02	0.71 ± 0.16	0.65	14.5	0.59
	0.51 ± 0.01	5.62 ± 0.17	0.76	15	0.31
companies	0.83 ± 0.17	1.03 ± 2.47	0.86	15.5	0.42
	0.62 ± 0.14	3.99 ± 2.55	0.93	17	0.32
asset managers	0.84 ± 0.25	0.45 ± 3.77	0.79	15.95	0.50
	0.73 ± 0.17	1.72 ± 3.23	0.72	18	0.41

region $(N, T_\Phi) \in \Gamma$ (figure 2.8). In this region, we have

$$\log N = \alpha \log T_\Phi + \beta, \tag{2.21}$$

which gives

$$\alpha = \frac{1-\delta}{2-\delta}. \tag{2.22}$$

We still need to link $\langle T \rangle_\Phi$ and $\langle P_v \rangle_\Phi^\beta$. While section 3 showed that the unconditional averages lead to $\langle T \rangle \sim \langle P_v \rangle^\beta$, one also finds that $\langle T \rangle_\Phi \sim \langle P_v \rangle_\Phi^\beta$. Therefore, one can write

$$\log \langle T \rangle_\Phi = \beta \log \langle P_v \rangle_\Phi + \text{cst}. \tag{2.23}$$

Thus, one is finally rewarded with the missing link

$$\beta = \frac{1}{2-\delta}, \tag{2.24}$$

which directly involves the transaction cost structure in the relationship between turnover and portfolio value, as argued in section 3[2]. This relationship allows us to close the loop as we are now able to relate directly the exponents linking T, N, and P_v. Going back to section 2.3, one understands that the existence of a bi-linear relationship between log-turnover and log-account value, i.e., of two values of β for each of the three categories of clients, is linked to two values of δ: a flat flee structure or the disregard for transaction costs leads to $\beta = \frac{1}{2}$, while proportional fees ($\delta = 1$) give $\beta = 1$.

Let us finally discuss the empirical values of α, β, and δ against their theoretical counterparts, which is summarized in table 2.8.

1. *Small values of* T_Φ: it was impossible to measure α in that case since the non-parametric fit shows a non-linear relationship in the log-log plot for retail clients, which we trust more since they have many many more points than the graphs for the two other categories of clients. But it may not make sense to expect a linear relationship since such a relationship is only expected for N large enough ($N \geq 10$ in practice) and a small T_Φ is related to a small N. Thus we can only test $\beta = 1/(2-\delta)$. The reported value of β is consistent accross all

[2]Note that this relationship can be obtained directly by assuming that all the transactions happen at the same time, hence that $T = (xP_v)/N$, which leads straightforwardly to (2.24).

the clients. Retail clients have a larger $\delta_{eff} = 2 - \frac{1}{\beta}$ that the estimated δ_{SQ}. Since the shape of the fee structure is discontinuous, the values of these exponents can hardly be expected to match. However, fitting the whole curve structure may be problematic in this context: indeed, the traders with a typical small value of T_Φ see a more linear relationship in the region of small transaction value that when considering the whole curve; for instance, removing the two largest segments from the fee structure yields $\delta'_{SQ} = 0.74 \in [0.43, 0.79]$, which is not far of δ_{eff}.

2. *Large values of* T_Φ: the relationships between all the exponents are verified for the three categories of clients. While not very impressive for companies and asset managers, this result is much stronger in the case of retail clients since the relative uncertainties associated with each measured exponent are small (1-2%). The value of β_{retail} is of particular interest as it corresponds $\delta_{eff} = 0$, or equivalently, to a flat fee structure. Going back to the fees structure of Swissquote, one finds that that the transition happens when the relative transaction cost falls below some threshold (we cannot give its precise value for confidentiality reasons; it is smaller than 1%). A possible explanation is that either some traders with a high enough average turnover have a flat-fee agreement with Swissquote and that the rest of them simply act as if they were not able to take correctly into account transaction costs. Since not all traders have a flat-fee aggrement, one must conclude that some traders have indeed some problems estimating small relative fees and simply disregard them. The reported value of β for companies and asset managers is larger that β_{retail}, but it is more likely than not that the small sample size is responsible for this discrepancy, since these two categories of clients have a greater propensity to negociate a flat-fee structure.

3. *Transition between the two regimes:* the transitions between the standard Swissquote and an effective flat-fee structure happens occur *at the same average value* of T for the three categories of traders (idem for T_Φ). Since there is no automatic switching between fee structures at Swissquote for any predefined value of transaction value, one is lead to conclude that this transition has behavioural origins, which is also responsible for the value at which the transition takes place which, in passing, corresponds to the end of the plateau of the RCDF of P_v in the case of retail clients ($e^{15} \simeq 3.27 \cdot 10^6$). As a consequence, it is likely that the traders tend to either neglect or consider as constant transaction fees smaller than some threshold when they build their portfolio.

Table 2.8: *Table summarising the empirical and theoretical relationships between α, β, and δ.*

small T_ϕ	individuals	companies	asset managers
β	0.85 ± 0.02	0.83 ± 0.17	0.84 ± 0.25
$\log T_\Phi <$	14.5	17	18
$\delta_{eff} = 2 - \frac{1}{\beta}$	0.82 ± 0.02	0.80 ± 0.20	0.81 ± 0.30
δ_{SQ}	$0.63 \in [0.50, 0.74]$	$0.63 \in [0.50, 0.74]$	$0.63 \in [0.50, 0.74]$
δ'_{SQ}	$0.74 \in [0.43, 0.79]$	$0.74 \in [0.43, 0.79]$	$0.74 \in [0.43, 0.79]$
$\tilde{\beta} = \frac{1}{2-\delta_{SQ}}$	$0.73 \in [0.66, 0.74]$	$0.73 \in [0.66, 0.74]$	$0.73 \in [0.66, 0.74]$

large T_ϕ	individuals	companies	asset managers
β	0.51 ± 0.01	0.62 ± 0.14	0.73 ± 0.17
$\log T_\Phi >$	15	17	18
$\delta_{eff} = 2 - \frac{1}{\beta}$	0.04 ± 0.02	0.39 ± 0.23	0.63 ± 0.23
$\alpha_{eff} = \frac{1-\delta_{eff}}{2-\delta_{eff}}$	0.49 ± 0.01	0.38 ± 0.09	0.27 ± 0.08
α	0.52 ± 0.02	0.36 ± 0.14	0.44 ± 0.13
$\log T_\Phi \in$	$[16, 19]$	$[17, 19.8]$	$[15.8, 18]$

2.6 Discussion and outlook

We have been able to determine empirically a bilinear relationship between the average log-turnover and the average log-account value and have argued that it comes from the transaction fee structure of the broker and its perception by the agents. A theoretical derivation of optimal simple one-shot mean-variance portfolios with non-linear transaction costs predicted relationships between turnover, number of different asset in the portfolio and log-account values that could be verified empirically. This means that the populations of traders do take correctly *on average, i.e. collectively,* the transaction costs into account and act *collectively* as mean-variance equally-weighted portfolio optimizers. This is not to say that each trader is a mean-variance optimizer, but that the population taken as a whole behaves as such—with differences across populations, as discussed in the previous section. This to be related to findings of Kirman's famous work on demand and offer average curves in Marseille's fish market [47] and more generally as what has become known as the wisdom of the crowds (see [77] for an easy-to-read account).

The fact that the turnover depends in a non-linear way on the account value implies that linking the exponents of the distributions of transaction volume, buying power of large players in financial markets, and price return is more complex that previously thought [37]. It has also implications for agent-based models, which from now on must take into account the fact that

the real traders do invest into a number of assets that depends non-linearly on their wealth.

Future research will address the relationship between account value and trading frequency, which is of utmost importance to understand if the many small trades of small investors have a comparable influence on financial market than those of institutional investors. This will give an understanding of whom provides liquidity and what all the non-linear relationships found above mean in this respect. This is also crucial in agent-based models, in which one often imposes such relationship by hand, arbitrarily; reversely, one will be able to validate evolutionary mechanisms of agent-based model according to the relationship between trading frequency, turnover, number of assets and account value they achieve in their steady state.

3

Volatility long-memory and min-dispersion profiles

3.1 Weighting the past: why and how

Weighted averages are extensively used in a number of fields for smoothing and forecasting time series data. In a nutshell, when a series is believed to be stationary and to exhibit some patterns, one can use regularities that showed up in the past for predicting future values. In many instances though, motives are fairly simple and principally consist in clusters of values with comparable amplitudes (e.g. high and low volatility periods). Forecasting is then done by giving more weight to recent past values, which can be achieved by a time-decreasing weight sequence.

Mainly for convenience, the most used and studied weighted estimators are the simple (SMA) and exponentially-weighted (EMA) moving averages. The properties of the latter, which exploits geometrically decreasing weights, have been thoroughly studied in econometrics, and also by practitioners of finance [46]. In the financial context simple averages are poor forecasters as they treat equally long and short past market events, whereas EMA, while showing better accuracy, exhibit a high intrinsic variance which can create instabilities and high portfolio turnovers when used in portfolio allocation.

The high intrinsic dispersion of EMA also raises the question of its efficiency: since the variance of $\bar{x}$ is a measure of the dispersion of the estimator around its expected value, it gives

no information on its accuracy and efficiency, which from a statistical viewpoint are assessed respectively via the bias $B(\bar{x}) = E(\bar{x} - \mu)$ and the mean squared error $\text{MSE}(\bar{x}) = E((\bar{x} - \mu)^2)$ (assuming the true value μ exists). Nevertheless, an unbiased estimator that minimizes $V(\bar{x})$ as well minimizes the MSE for $\text{MSE}(\bar{x}) = V(\bar{x}) + B(\bar{x})^2$. Hence if one is able to reduce the intrinsic variance of a weighted estimator while not simultaneously harming its accuracy, one ends up with a more efficient solution, that is an estimator with lower MSE. More precisely, when restricting to normalized weighted averages, i.e. $\bar{x} = \sum_t w_t x_t$ with $\sum_t w_t = 1$, we readily see that the estimator with minimum MSE is the one with minimum dispersion as the bias $B(\bar{x}) = E(\bar{x} - \mu) = E(x_t) - \mu$ is independent of the particular weight sequence w_t. Importantly, since the MSE impacts the stability of the prediction, low MSE estimators are useful in many places in Finance, and particularly in the mean-variance portfolio allocation that is the main object of this dissertation.

Constructing efficient estimators is a task that traditionally belongs to prediction theory. To do so one typically resorts to least-squares or generalized least-squares estimators because they have minimum MSE when the residual errors are not too much correlated.[1] Unfortunately, linear residual errors on financial series can still be highly correlated and one has to turn to more sophisticated non-linear models (e.g. of the GARCH family) to remove the residual serial correlations at a satisfactory level. Such models do not allow one to build efficient estimators easily unless ad hoc assumptions on the dynamics of the true value are made (i.e. a model). On a different level, one may have doubts about the existence of unbiased estimators with practical interest in finance: the owner of a constantly unbiased estimator would benefit from an advantage that should quickly be arbitraged out by other market participants (according to the efficient market hypothesis).

Estimators with minimal variance, as we will see, do not suffer from the same drawback. They can be computed under the sole weak stationarity of the underlying process, and as they do not depend on the value to be estimated, they hold under no assumptions on the residual errors. Their interest is related to the stability of risk predictions and is therefore directly connected to portfolio turnovers, which was shown in the preceding chapter to be a major concern to the majority of online investors.

[1] This follows from the Gauss-Markov theorem, which states that the best unbiased linear estimator is the one with minimum MSE provided that the residual errors are uncorrelated and homoscedastic (i.e. have equal variance). Generalized least square estimation extends this result to the case of heteroscedastic data with some degree of *known* correlations.

3.2 Variance of weighted averages and autocorrelation

Consider the *real-valued weakly stationary* process $\{x_t\}_{t\in T}$, $T = \{1, \ldots, t_{\max}\}$ and its associated weighted average defined over $\{1, \ldots, M\} \subset T$ as

$$\bar{x} = \sum_{t=1}^{M} w_t x_t. \tag{3.1}$$

Notice that in a model view one commonly writes $x_t = \sum_{s=1}^{M} w_s x_{t-s}$, whereas here we use the previous definition and keep in mind that x_1 corresponds to the most recent value. Also note that we will later restrict to more specific weight sequences such as decreasing, positive or normalized ones, but this is not required at this stage.

Recall that the variance of $\bar{x}$ takes the form

$$\begin{aligned} V(\bar{x}) &= E\left((\bar{x} - E(\bar{x}))^2\right) \\ &= \sum_{s,t=1}^{M} E\left((w_s x_s - E(\bar{x}))\,(w_t x_t - E(\bar{x}))\right) \\ &= \sum_{s,t=1}^{M} \operatorname{cov}(w_s x_s, w_t x_t) \\ &= 2\sum_{s\le t}^{M} w_s w_t \operatorname{cov}(x_s, x_t) - \sum_{t=1}^{M} w_t^2 \sigma_t^2, \end{aligned} \tag{3.2}$$

where we have introduced the covariance function of x_t and defined $\sigma_t^2 = \operatorname{cov}(x_t, x_t)$. Now since x_t is weakly stationary its autocorrelation function ρ is well-defined and we can write $\sigma_t^2 = \sigma^2$ and $\rho(h) = \operatorname{cov}(x_s, x_t)/\sigma^2$, where $h = |s - t|$. With these simplified notations, Eq. (3.2) becomes

$$V(\bar{x}) = \sigma^2 \left[2\sum_{h=1}^{M-1} \rho(h) \left(\sum_{t=1}^{M-h} w_t w_{t+h} \right) + \sum_{t=1}^{M} w_t^2 \right]. \tag{3.3}$$

This equation says that the sample autocorrelation function (ACF) of the process has a great influence on the dispersion of the estimator $\bar{x}$. For instance for positively correlated processes the higher the autocorrelation the higher the variance of the estimation as prescribed by the central limit theorem. It also suggests that to any particular ρ should correspond a weight sequence minimizing $V(\bar{x})$, which is the question addressed in Sec. 3.6.

Because of its key-role for understanding and controlling the variance of weighted estimators in finance the sample ACF of volatility deserves a particular attention.

3.2.1 Describing the whole volatility autocorrelation function

By volatility we mean here some positive power of the absolute returns $|r|^{\alpha}$, usually taken as 1 or 2. On the big picture the sample autocorrelation function (ACF) of absolute returns decays at an hyperbolic rate $\rho(h) \sim h^{-\delta}$ where $\delta < 1$ is reported to lie between 0.2 and 0.5 for most financial assets [28]. If this description is true then volatility is said to exhibit long memory in the sense that $\sum_h |\rho(h)|$ diverges. Since analyzes leading to these conclusions are done on series with limited length they should benefit from occasional updates, which is the first purpose of what follows.

We have analyzed the daily return sample ACF of 16 European stock indices, 200 European stocks, 1'250 US stocks, 100 Swiss stocks, and 300 stocks of the world market over the period 1990-2011, and have reached the following conclusions:

1. In many instances the sample ACF is equally well described by (at least) four different functions which despite presenting similar short-term decay have very different convergence properties as $h \to \infty$.

2. The very short-term behavior of $\rho(h)$, despite its obvious influence on the dispersion of weighted estimators (more on that later), is usually not given much attention in the literature. For instance a number of statistical models have been designed to capture and reproduce the "observed" long-memory of volatility, but as far as we know, none of them tries to accommodate the short-time behavior of ρ.
   Here we take the viewpoint that an accurate description of the whole sample ACF of the volatility is key to building stable estimators with minimal dispersion.

The general shape of the volatility autocorrelation is illustrated in Fig. 3.1 using the sample ACF of the absolute returns of the S&P500 since 1950. For several assets ρ starts to decrease gently over the first few lags leading to a "plateau" on a lin-log scale (this feature is particularly pronounced for stock indices). Then the ACF decreases faster but still slowly enough to look either like a power-law on a log-log scale or like a logarithm on a lin-log scale (and perhaps like many other things...). Here comes the limitation of the "fit-by-eye" technique often seen in the literature to differentiate e.g. a power-law or a logarithmic tail from anything else. Indeed, this technique is valid as long as the tail of the function spans several decades. Unfortunately this is not true here since the stable part of the sample ACF is comprised in $2/\sqrt{t_{\max}} < \rho_{stable} < 1$

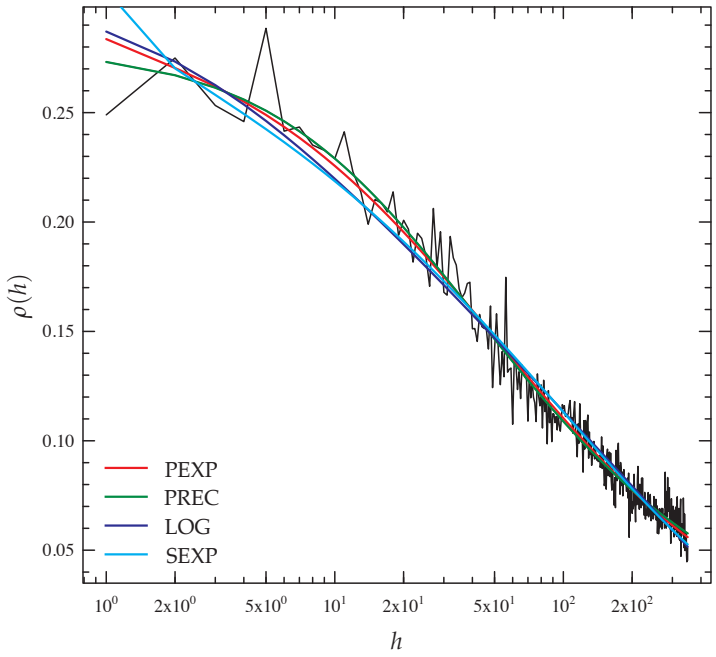


Figure 3.1: *Sample autocorrelation function of the daily absolute returns of the S&P500 over the period 1950-2011. Lines are non-linear fits to the heavy-tailed functions defined in this paragraph. In general PREC and PEXP are better candidates than LOG and SEXP to describe the stable part of the sample ACF, in particular on stock indices, but the latter do almost as good as the former on individual stocks.*

that is in $\sim 10^{-2} < \rho_{stable} <\sim 10^{-1}$ for daily returns. In conclusion, on a logarithmic scale and over a single decade many heavy-tailed functions look the same to the naked eye, and therefore one should resort to more accurate fitting techniques to discriminate between several candidate functions which one is the best.

Based on our observations and on previous ones reported in the literature, we have introduced four different candidates to describe the sample ACF of absolute returns: (1) over the whole range where it is stable (not only its "tail"), and (2) using as few parameters as possible.

- PREC has power long-term decay. It tends to the step function as the memory parameter δ goes to infinity:

$$\rho_{prec}(h) = \frac{\rho_0}{1 + (\frac{h-1}{b})^{\delta}} \quad \delta > 0,\ b > 0. \tag{3.4}$$

- PEXP has power long-term decay. It tends to the exponential as the memory parameter δ goes to infinity:

$$\rho_{pexp}(h) = \frac{\rho_0}{(1 + \frac{h-1}{b})^{\delta}} \quad \delta > 0,\ b > 0. \tag{3.5}$$

- LOG has logarithmic long-term decay:

$$\rho_{log}(h) = \rho_0 \left(1 - \delta \log\left(1 + \frac{h-1}{b}\right)\right) \quad 1 > \delta > 0,\, b > 0. \tag{3.6}$$

- SEXP has stretched-exponential long-term decay and thus can accommodate both constant and exponential autocorrelations:

$$\rho_{sexp}(h) = \rho_0 \exp\left(-\left(\frac{h-1}{b}\right)^{\delta}\right) \quad 1 > \delta > 0,\, b > 0. \tag{3.7}$$

PREC and PEXP are natural extensions of a pure power-decreasing tail to describe the short-term behavior of the ACF. SEXP roughly interpolates between the two previous ones and LOG has been introduced here as it was reported to fit well the volatility ACF in [87]. Notice also that even though all these functions are heavy-tailed the series $\sum_h |\rho(h)|$ diverges for LOG and PREC,PEXP when $\delta \leq 1$, whereas it converges for SEXP and PREC, PEXP with $\delta > 1$, which emphasizes the clear difference between their asymptotic behavior.

For each function we have computed the non-linear least-squares estimates of ρ_0, b, and δ. Since the number of fitting parameters is the same for the three models a fair comparison is possible by looking at the residual sum-of-squares. The latter are shown in Fig. 3.2 for European stocks and indices, and for stocks of the Swiss and US markets. Note that similar results are obtained for securities of the world stock market. The results reveal a notable difference between the volatility ACF of aggregated returns (i.e. indices) and that of individual stocks. While the latter are well described by the four candidate functions, indices on the other hand are significantly and systematically better described by PREC and PEXP which outperform LOG and SEXP. Looking into detail at the numbers we find that the best model on average over all five data sets (including securities from the world stock market) is PREC, closely followed by PEXP.

Fit results are summarized below and in Fig. 3.3 and Fig. 3.4. They are not discussed at length nor reported with confidence intervals because of high statistical errors. We note however that all parameters are significant except for b in LOG which is regularly found to be not significant on individual stocks. Estimates of the memory parameter δ in PREC are in line with known results on the volatility ACF tail (see Fig. 3.3), whereas the same parameter in PEXP is often significantly higher (as is the "plateau" b). Estimates of δ in PREC and PEXP for 75% of the

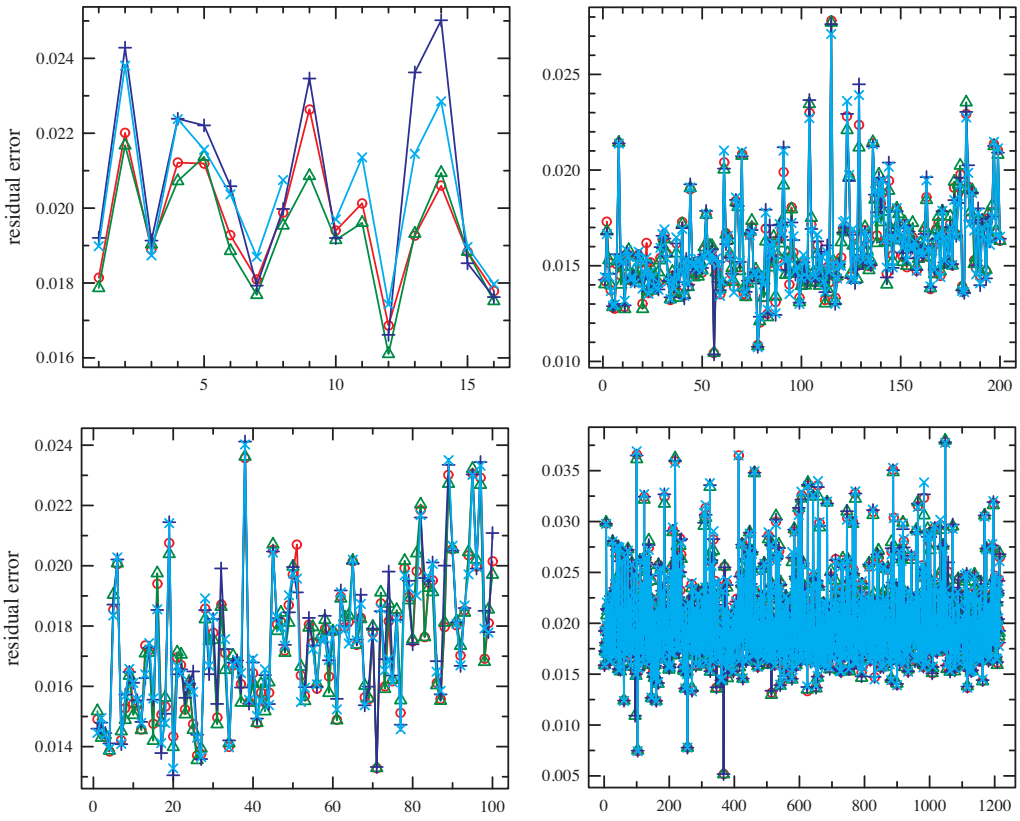


Figure 3.2: *Residual sum-of-squares of the ACF non-linear fits to the functions of Sec. 3.2.1. Top left to right: European indices and stocks. Bottom left to right: Swiss stocks and US stocks. Green triangles: PREC, red circles: PEXP, dark blue crosses: LOG, and light-blue stars: SEXP. On stock indices PREC and PEXP systematically outperform LOG and SEXP. On individual stocks all four functions give similar results, even though PREC gives on average better results.*

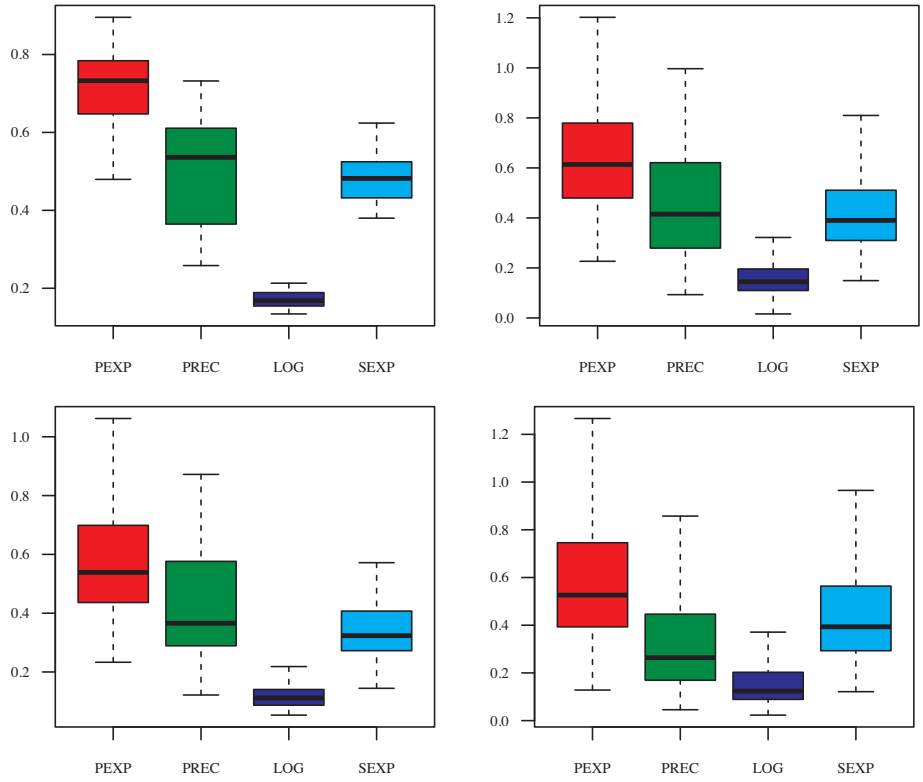


Figure 3.3: *Box plots of the ACF fit results for the "memory" parameter δ. Top left to right: European indices and stocks. Bottom left to right: Swiss stocks and US stocks.*

securities belong to:

	EU Ind.	EU St.	CH	US
PREC	$(0.38, 0.6)$	$(0.23, 0.61)$	$(0.29, 0.59)$	$(0.17, 0.43)$
PEXP	$(0.65, 0.78)$	$(0.48, 0.78)$	$(0.42, 0.71)$	$(0.40, 0.77)$

In conclusion, our results favor the two power-decreasing candidates over LOG and SEXP for describing the stable part of the volatility ACF. Another argument in the same direction is the following: if we admit that the ACF of $|r|^\alpha$ should be described by the same function for any α, we can repeat the above analysis and compare the different models for α varying, say from 1 to 2. Doing so clearly reveals an increasing discrepancy in favor of the power-law functions.[2] Nevertheless a logarithmic description of the autocorrelation on individual stocks gives more stable results and sometimes can be achieved with only 2 parameters, which may be useful in some applications involving individual stocks.

[2]Notice that by increasing α one pushes up the noise level and thus makes the fitting results less accurate.

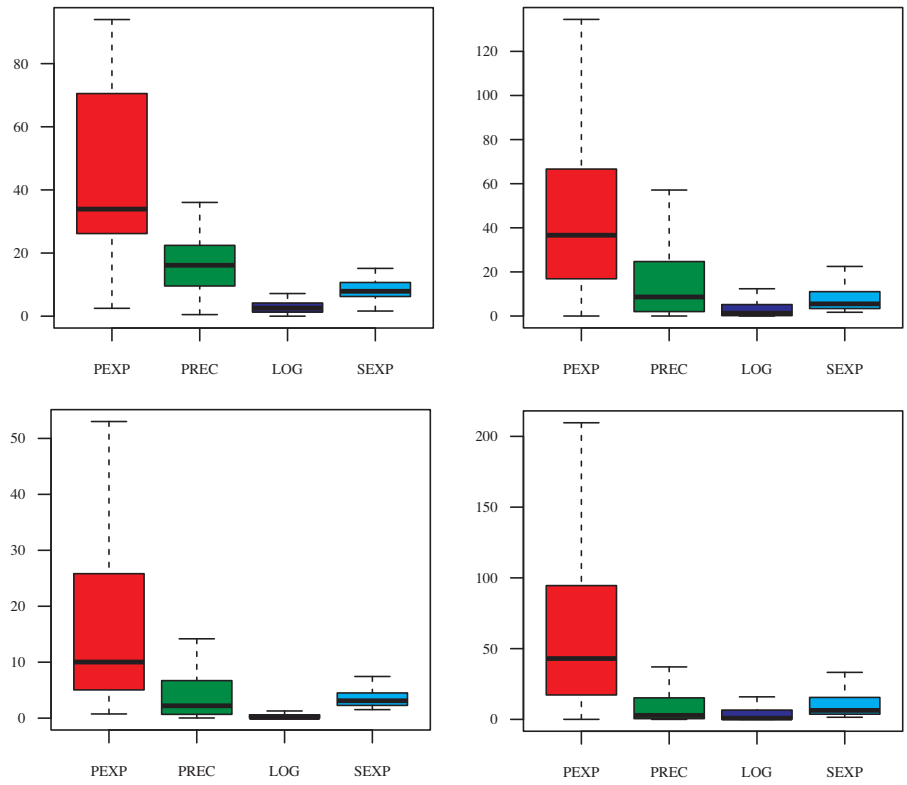


Figure 3.4: *Box plots of the ACF fit results for the "plateau" parameter b. Top left to right: European indices and stocks. Bottom left to right: Swiss stocks and US stocks.*

3.3 Asymptotic behavior of the variance

It is interesting to examine the variance of the most used moving average estimators (SMA and EMA) on the two important exponential and power decays of the autocorrelation function. Exponential autocorrelations show up in Markov processes, autoregressive conditional heteroscedastic models (ARCH) [35], and in many empirical time series. On the other hand power-decaying autocorrelations are characteristic of fractionally integrated generalized ARCH (FIGARCH [4]), and most notably of stock volatility as seen above.

Let us come back to Eq. (3.3). Setting first $w_t = 1/M$ we obtain

$$V(\bar{x}) = \frac{\sigma^2}{M}\left(2\sum_{h=1}^{M-1}\rho(h)\left(1-\frac{h}{M}\right)+1\right). \tag{3.8}$$

Assuming now $\rho(h) \sim \exp(-h/\tau)$ with $\tau \geq 1$ leads to

$$V(\bar{x}) \sim \frac{\sigma^2}{M}\left(\frac{e^{\frac{1}{\tau}}+1}{e^{\frac{1}{\tau}}-1}+O\left(\frac{1}{M}\right)\right), \tag{3.9}$$

which means that the variance of simple averages on exponentially autocorrelated processes goes to zero as fast as the uncorrelated variance $V(\bar{x}) = \frac{\sigma^2}{M}$. It is shown in Sec. 3.5 that simple moving averages are asymptotically of minimal dispersion on these processes in the sense that they minimize $V(\bar{x})$ when $M \to \infty$.

Let us now consider the hyperbolic decay $\rho(h) \sim h^{-\delta}$, $0 < \delta < 1$. Then x_t exhibits long-memory and the variance of $\bar{x}$ takes the form

$$V(\bar{x}) \sim \frac{\sigma^2}{M^\delta} \left(\frac{2}{(1-\delta)(2-\delta)} + O\left(\frac{1}{M^{1-\delta}} \right) \right), \tag{3.10}$$

where the asymptotic expansion of the generalized harmonic number $H_{M,\delta} = \sum_{h=1}^{M} h^{-\delta}$ has been used in the derivation. This result shows that the dispersion of simple averages on long-memory processes decays at an hyperbolic rate governed by δ.

A common practice in finance is to use exponential moving averages to forecast volatility, sometimes regardless of the process autocorrelation structure [46]. Let us look at the variance of the one-step forecaster $\hat{x}_1 \equiv \bar{x}$ computed with geometric weights on both exponential and hyperbolic autocorrelated processes. For the normalized geometric weights $w_t \sim \alpha^{t-1}$ Eq. (3.3) becomes

$$V(\bar{x}) \approx \sigma^2 \left(\frac{1-\alpha}{1+\alpha} \right) \left(\sum_{h=1}^{M-1} \rho(h) \alpha^h \left(1 - \alpha^{2(M-h)} \right) + 1 \right), \tag{3.11}$$

where we have neglected the terms α^M. Now in the limit $M \to \infty$ we find the following non-zero limits for $\rho(h) \sim \exp(-h/\tau)$

$$V(\bar{x}) \sim \sigma^2 \left(\frac{1-\alpha}{1+\alpha} \right) \left(\frac{e^{\frac{1}{\tau}} + \alpha}{e^{\frac{1}{\tau}} - \alpha} \right), \tag{3.12}$$

and for $\rho(h) \sim h^{-\delta}$

$$V(\bar{x}) \sim \sigma^2 \left(\frac{1-\alpha}{1+\alpha} \right) \left(\frac{2\Gamma(1-\delta)}{\log(\frac{1}{\alpha})^{1-\delta}} - 1 \right), \tag{3.13}$$

where $\Gamma(x)$ stands for the gamma function.[3] Hence unlike simple averages exponential moving averages have *non-zero* asymptotic variance on both exponential and hyperbolic autocorrelations, which means that there exists some threshold value $M_{\max} < M$ above which the dispersion of

[3]The first formula is exact as $M \to \infty$, but the second one is only accurate for small values of δ as it was obtained by replacing sums with integrals in Eq. (3.11). Fortunately, as can be seen in Fig. 3.5, this is true of typical values observed in practice ($\delta \approx 0.3$).

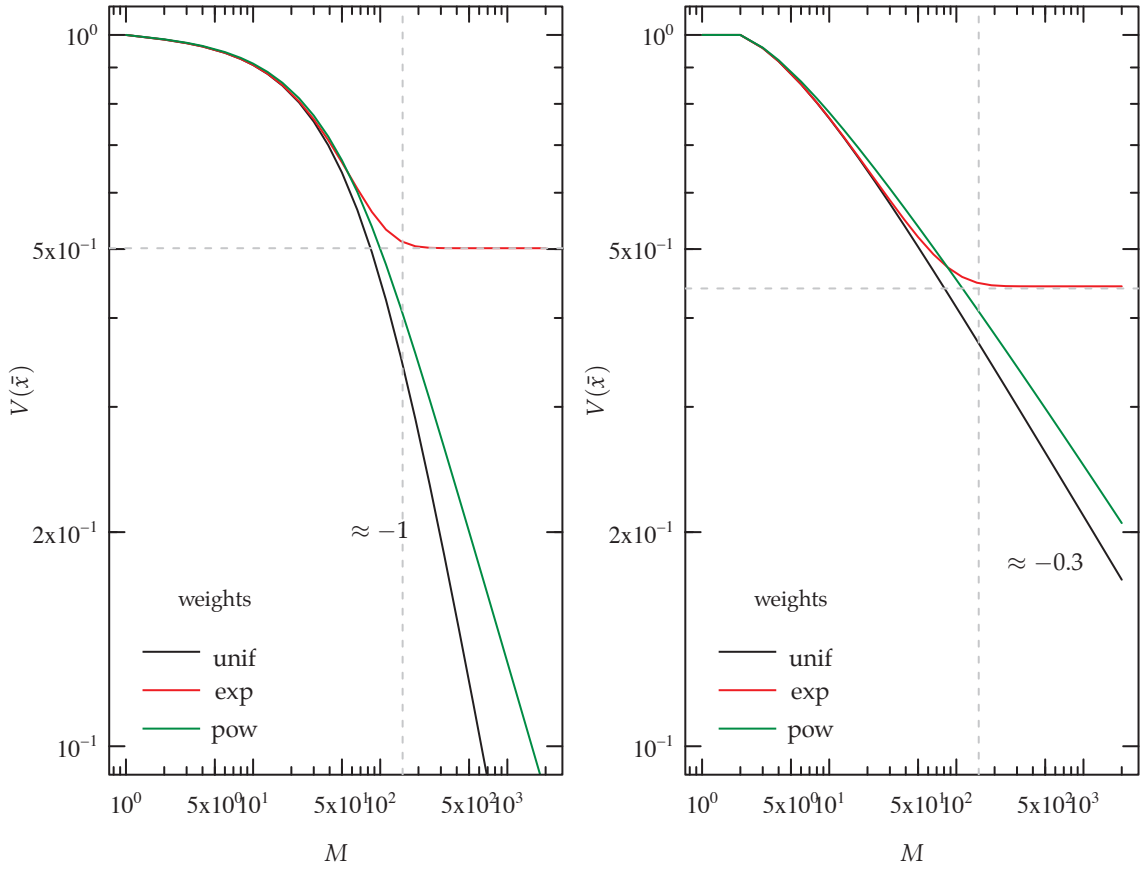


Figure 3.5: *Asymptotic variance of moving averages as a function of the sample size M. The left plot shows the effects of uniform, exponential, and power-decreasing weights on exponential autocorrelation. The horizontal dashed line is the limit Eq. (3.12) and the vertical one is placed at $M = 150$. The slope of the variance for uniform (and power-decaying weights) is equal to -1 in agreement with the theory (Eq. (3.9)). The right plot shows the same effects on power-decreasing autocorrelation like that of volatility. The limiting line is Eq. (3.13) and the parameters in this case are $\delta = 0.3$ and $\alpha = 0.97$ which are typical values seen in practice [46]. The slope of the variance in the power and uniform cases is ≈ -0.3 in agreement with the theory (Eq. (3.10)). The limitation of EMA (in red) is that no variance reduction can be achieved beyond 150 observations.*

$\bar{x}$ cannot be significantly reduced by adding more historical data. As shown in Fig. 3.5 (the vertical dashed lines) this threshold is remarkably low for the typical values of α used in volatility forecasting, for instance by RiskMetrics [46]. More precisely we see that the dispersion of EMA depends on the last 150 observations and cannot be reduced further by adding observations beyond this threshold; whatever the form of the autocorrelation. The rest of the chapter is devoted to building estimators with better efficiency by taking into account the peculiar structure of the volatility autocorrelation.

3.4 Weight profiles inducing minimal dispersion

Our approach is to compute the weight sequences that minimize $V(\bar{x})$ under constraints that are relevant to the financial application and particularly to the forecast of volatility and Value-at-Risk. These constraints are chosen so as to preserve the accuracy of the estimation.

In the first place, it is suitable to rewrite Eq. (3.3) more compactly as

$$V(\bar{x}) = \sigma^2 \mathbf{w}^t R \mathbf{w}, \tag{3.14}$$

where R is the $M \times M$ Toeplitz matrix of autocorrelations of x_t, that is $R_{st} = R_{|s-t|} = \rho(h)$, $h = |s - t|$. Since $V(\bar{x})$ is non-negative for any weight sequences, and in fact strictly positive for non-trivial ones, R is symmetric positive-definite which implies that $V(\bar{x})$ has a unique global minimum (i.e. the minimizing problem is strictly convex). In order to avoid the trivial solution given by the null sequence it is natural to restrict to normalized weights. Note that although normalization to unity seems natural in this context it is by no means mandatory and indeed taking $\sum_t w_t = b < 1$ may sometimes be desirable (in a model-based view, this allows one to define min-dispersion ARCH processes with finite variance). Imposing $b = 1$ leads to a result well-known in prediction theory that is similar to Markovitz's minimum variance portfolio [70]:

$$\mathbf{w}^* = \frac{R^{-1}\mathbf{1}}{\mathbf{1}^t R^{-1} \mathbf{1}}, \tag{3.15}$$

where $\mathbf{1}^t = (1, \ldots, 1)$. Introducing the optimal weights Eq. (3.15) in Eq. (3.14) we obtain the corresponding minimal variance

$$\min_{\mathbf{w} \in \mathbb{R}^M,\, \mathbf{w}^t \mathbf{1} = 1} V(\bar{x}) = \frac{\sigma^2}{\mathbf{1}^t R^{-1} \mathbf{1}}. \tag{3.16}$$

Despite their similarity, the cross-correlation problem of Markovitz and the time-correlation problems addressed here have two interesting differences due to the weak-stationarity assumption on x_t. First, one can factor σ^2 out of the covariance in Eq. (3.2) and thus express the solution in terms of the autocorrelation function alone. This is useful because R can be computed in a much faster way than Σ as it is fully determined by only M numbers. Furthermore, in many instances a great deal is known about ρ and therefore about R. For example explicit autocorrelations have been obtained by econometricians for the most important representatives of the

GARCH family (ARCH, GARCH, IGARCH, FIGARCH, etc) [9, 12, 67, 4]. In the next section we use these results to gain some insight into the solution and derive an explicit solution for processes with exponentially decreasing autocorrelation.

3.4.1 Adding up relevant constraints

The other major difference between the variance minimization of asset portfolios and that of weighted averages comes from the peculiar structure of R. By the weak-stationarity of x_t the Toeplitz matrix R is symmetric about its two diagonals. This property is conserved by matrix inversion and therefore yields $w_t^* = w_{M-t+1}^* \ \forall t$ in Eq. (3.15). This means that the M numbers required to generate R define in reality a sequence of $M/2$ "independent" optimal weights, the second half of $\mathbf{w}^*$ being just a mirror of the first one. This situation lacks financial intuition and should be remedied in one of the following two ways in order two preserve the accuracy of the estimator: either by dropping the second half of $\mathbf{w}^*$ and renormalizing the resulting sequence, or alternatively by imposing decreasing weights in the minimization at the cost of harming the minimal variance. Additionally, the latter option has the benefit of smoothing out the weight profile which may lead to improved stability in applications.

Another natural constraint to consider at this stage is the non-negativity of w_t. Indeed, all the significant weights on autocorrelated processes are positive. Hence restricting to positive sequences only slightly harms the minimal dispersion for the benefit of turning $\mathbf{w}$ into a well-defined discrete density function of time. Such a density is useful as it is required for the forecast of volatility or Value-at-Risk (VaR) with the historical method used in Sec. 3.7.

In summary the original minimum dispersion problem quite naturally extends to

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^M} \sigma^2 \mathbf{w}^t R \mathbf{w} & \\ \mathbf{w}^t \mathbf{1} &= 1 \\ w_t &\geq 0 \ \forall t \\ w_t &\geq w_{t+1} \ \forall t. \end{aligned} \tag{3.17}$$

It is worth mentioning that Eq. (3.17) can be further extended to handle optimal weights conditional on a given range of values taken by the estimator. For instance, one may investigate whether the shape of the optimal profile depends on the magnitude of the estimation by adding

the constraint $\sum_t w_t x_t \geq S$ and varying S (e.g. how does vary the shape of optimal profiles between high and low volatility periods?). By analogy with Markovitz's problem [63], the curve $(S, V(S))$ then defines an "efficient frontier" in which a point is mapped to the set of decreasing densities over the time index.

3.4.2 Solving the min-dispersion problem

Computing the optimal weights with Eq. (3.15) requires the inversion of a symmetric positive-definite Toeplitz matrix, which can be achieved efficiently by using ultra-fast algorithms (i.e. algorithms that require $O(n \log^2 n)$ flops). As the autocorrelation function can be computed in $O(n \log n)$ time by means of Fast Fourier Transform (FFT), the whole process is indeed very fast. In the case of Eq. (3.17) the minimization problem admits no explicit solutions because of the inequality constraints. However, several methods minimizing a quadratic function subject to linear constraints make use of R^{-1} or of some matrix decomposition of R (e.g. Cholesky's). The latter can be easily and efficiently worked out by algorithms specially designed to handle symmetric Toeplitz matrices. In summary, solving the minimum dispersion problem, whether its solution is given explicitly or not, is always much faster and easier than solving Markovitz's mean-variance problem. This is useful to practitioners as it allows them to compute optimal estimators virtually on any time scale (possibly also on intra-day horizons) and by using as much historical data as available.

3.5 Exponentially autocorrelated processes: explicit results

Interestingly, the estimator with minimal dispersion can be worked out in closed form for processes with strictly exponential autocorrelation. This follows from the advantageous tridiagonal structure taken by R^{-1} in this case. Indeed, if $\rho(h)$ can be written $\rho(h) = \alpha^h$ for all $h \geq 0$, then it is straightforward to verify that

$$R^{-1} = \frac{1}{1-\alpha^2} \begin{pmatrix} 1 & -\alpha & 0 & \dots & 0 \\ -\alpha & 1+\alpha^2 & -\alpha & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & \dots & -\alpha & 1+\alpha^2 & -\alpha \\ 0 & \dots & 0 & -\alpha & 1 \end{pmatrix}, \tag{3.18}$$

a result that was first reported in [66]. Using Eq. (3.18) in Eq. (3.15) and Eq. (3.16), one immediately finds

$$w_t^* = \frac{1}{M(1-\alpha)+2\alpha} \begin{cases} 1 & t = 1, M \\ 1-\alpha & t \in \{2, \ldots, M-1\} \end{cases} \tag{3.19}$$

and

$$V(\bar{x}; \mathbf{w}^*) = \sigma^2 \frac{1+\alpha}{M(1-\alpha)+2\alpha} \approx \frac{\sigma^2}{M} \frac{1+\alpha}{1-\alpha}. \tag{3.20}$$

As expected, in the absence of autocorrelation $\alpha = 0$ one is left with $w_s^* = 1/M$ and $V(\bar{x}) = \sigma^2/M$, whereas for perfectly correlated series all weights are zero except for $w_1^* = w_M^* = 1/2$ which yields $V(\bar{x}) = \sigma^2$. We notice also that for $M \to \infty$ the optimal weights tend to $1/M$ and $V(\bar{x}, \frac{1}{M})$ is as in Eq. (3.9), which shows that uniform averages are asymptotically of minimal variance on exponentially autocorrelated processes.

We consider now a correlated process and its optimal estimator with length M, whose variance is given by Eq. (3.16). If the same process were uncorrelated its variance would be σ^2/M_{eff} with $M_{eff} < M$, so that by comparing the two expressions one can define an effective number of "uncorrelated observations" in the correlated process as

$$M_{eff} = \sum_{i,j=1}^{M} R_{ij}^{-1}. \tag{3.21}$$

On processes with exponential ACF this yields

$$\frac{M_{eff}}{M} = \frac{1-\alpha+\frac{2\alpha}{M}}{1+\alpha} \approx \frac{1-\alpha}{1+\alpha},$$

which is the factor previously found in Eq. (3.11) (and incidentally the largest eigenvalue of R^{-1} in the limit $M \to \infty$).

A number of statistical processes seen in finance and natural sciences have exponentially decreasing autocorrelation. The popular Ornstein-Uhlenbeck process, used to model mean-reversion for instance in statistical arbitrage, features a "true" exponential decay of its autocorrelation $\rho(h) = \exp(-h/\tau)$ for $h \geq 0$. The weights defined by Eq. (3.19) are therefore of minimal dispersion on this process. In contrast, other useful exponentially correlated processes as the GARCH(1,1) have autocorrelation defined for $h \geq 1$ as $\rho(h) = \rho_0 \alpha^{h-1}$ [67] which in general can-

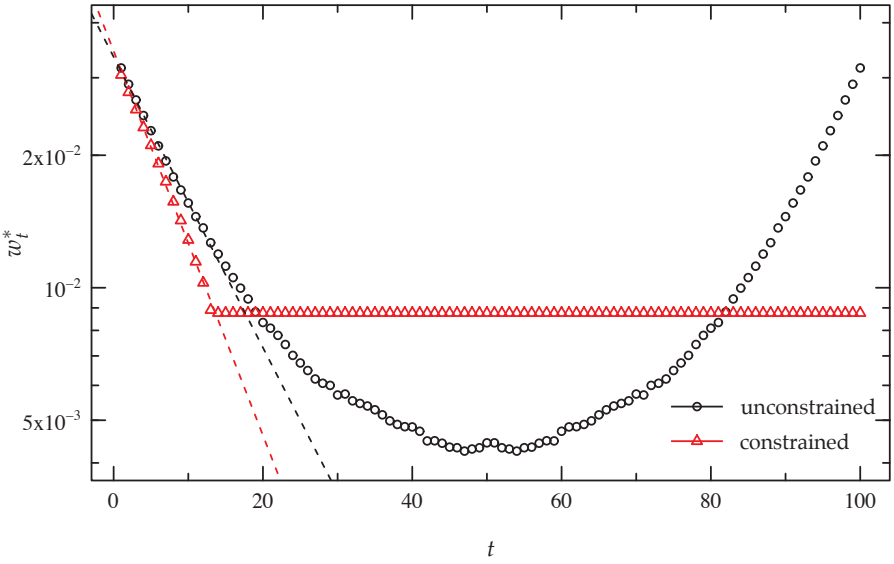


Figure 3.6: *Weights inducing minimum dispersion on moving average estimates on a GARCH(1,1). Black circles show the unconstrained solution (except for normalization) given by Eq. (3.15) and red triangles correspond to the solution of the constrained problem Eq. (3.17). In both cases the short-term decay is exponential as expected from Eq. (3.22).*

not be written in the adequate form to invoke Eq. (3.18). For such processes the inverse of the autocorrelation matrix reads $R_{pert}^{-1} = ((1-\rho_0)I + \rho_0 R)^{-1}$, which can sometimes be expanded as

$$R_{pert}^{-1} = \theta \left(R^{-1} + \sum_{k\geq 0} (-1)^{k+1} \left((1-\theta) R^{-1} \right)^{k+2} \right) \tag{3.22}$$

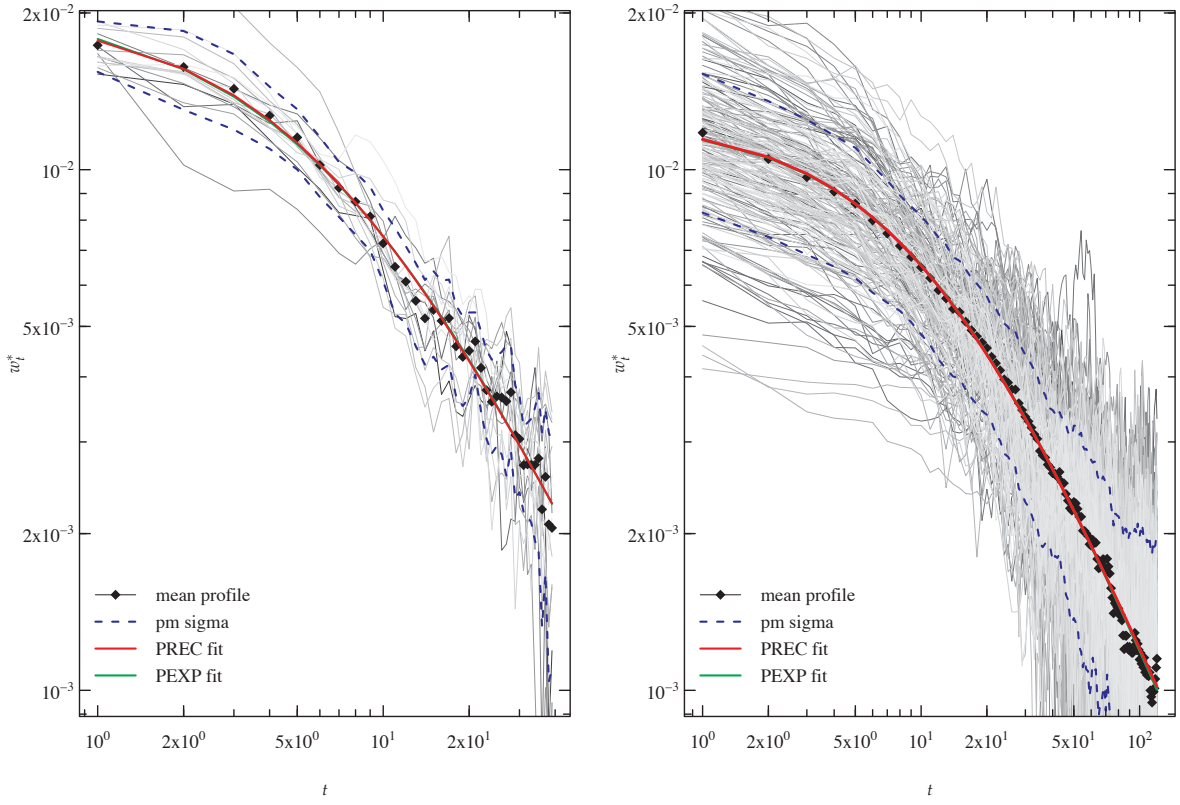
where $\theta = 1/\rho_0$. Although the expansion Eq. (3.22) is only valid under certain circumstances[4], it suggests a sum of exponentially decreasing corrections to R^{-1} which in turn induce short-term exponential decreasing components in $\mathbf{w}^*$ (R^{-1} is tridiagonal implies $(R^{-1})^2$ is four-band, $(R^{-1})^3$ is five-band, etc.). These observations are recapped in Fig. 3.6.

3.6 Profiles of volatility with minimum dispersion

We have computed the weight profiles inducing minimal variance (i.e. Eq. (3.15)) in the estimator

$$\overline{|r|} = \sum_{t=1}^{M} w_t |r_t|,$$

[4]The expansion follows from a direct application of the identities $(A+B)^{-1} = A^{-1} - A^{-1}(A^{-1}+B^{-1})^{-1}A^{-1}$ and $(I+A)^{-1} = \sum_k A^k$. For the latter to converge one needs here $(1-\theta)\lambda_{\max}(R^{-1}) \leq 1$, where it can be shown that $\lambda_{\max}(R^{-1}) \to (1-\alpha)/(1+\alpha)$ as $M \to \infty$).



(a) *European stock indices* **(b)** *US stocks*

Figure 3.7: *Volatility profiles with min-dispersion computed with Eq. (3.15) on 16 European Indices and 1'250 US stocks. Fits of PREC and PEXP on the mean profiles are also shown. Unlike that on individual stocks fits on the average curve are robust and give very close results for both PREC and PEXP.*

where $|r_t|$ stands for the absolute log-return of the securities in our five datasets. Unsurprisingly, and because of the tight connexion between the process min-dispersion profile and its autocorrelation, the results presented in Fig. 3.7 describe a functional form that is close to the ones proposed in Sec. 3.2.1 to describe the volatility ACF. The main difference here is the clear discrepancy of the fits between the four functions used to describe the optimal weighting. Indeed, PREC clearly and almost systematically outperforms PEXP, SEXP and especially LOG in this task. This can be seen in Fig. 3.8 and Table 3.1 where the residual sum-of-squares and the cumulated errors are compared. At this stage we can therefore safely discard LOG as a good candidate to describe the optimal volatility profiles.

Analysis of the fit results reveals once again an important dispersion of the fitted values and a relatively high statistical error. This error is considerably reduced if one instead tries to fit the mean optimal profile, which is just the average curve over all the optimal weights in the dataset. The mean profile is interesting in that it confirms PREC and PEXP as the best candidates

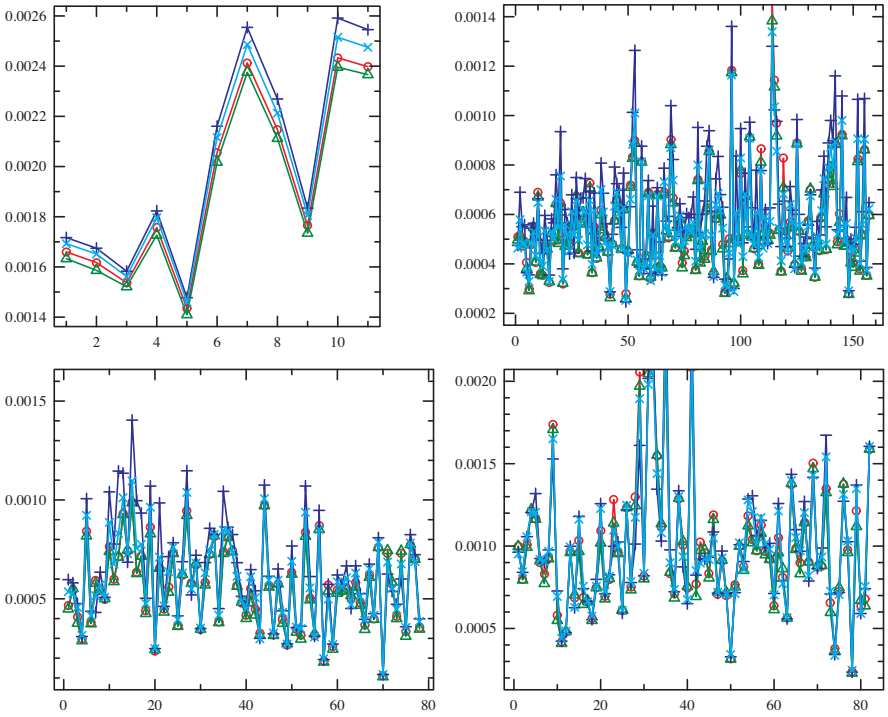


Figure 3.8: *Residual sum-of-squares of the optimal profiles non-linear fits to the functions of Sec 3.2.1. Top left to right: European indices and stocks. Bottom left to right: Swiss and US stocks. Green triangles: PREC, red circles: PEXP, dark blue crosses: LOG, light-blue stars: SEXP. For all datasets PREC almost systematically outperforms the other candidates and LOG is systematically behind the three others.*

Table 3.1: *Cumulative residual sum-of-squared exceeding the best fitting profile. PREC gives systematically the minimal cumulative error. It is set to zero and the other errors are expressed relatively to PREC.*

	PEXP	LOG	SEXP
US	3%	20%	8%
CH	4%	13%	6%
EUR	2%	16%	3%
W	3%	18%	4%

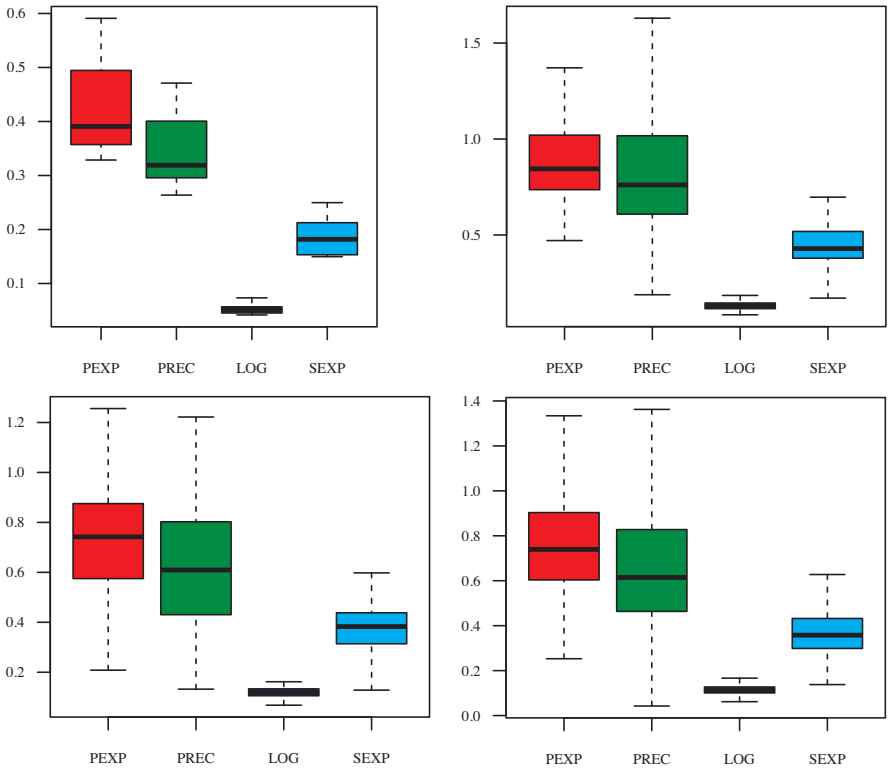


Figure 3.9: *Box plots of the fit results of the weight profiles for the "memory" parameter δ. Top left to right: European indices and stocks. Bottom left to right: Swiss and US stocks.*

for describing the min-dispersion weighting of volatility. Nevertheless, one should treat with caution the fitted values on the mean profile because of the cross-correlations between stocks of a same dataset. This difficulty can be overcome by applying for instance the random-phase method used in [85], which creates surrogate uncorrelated series from originally cross-correlated ones while preserving asymptotically their autocorrelation. Alternatively, one can use the results of Fig. 3.9 and Fig. 3.10 but given the effectiveness and ease of implementation of the computation methods required, we recommend instead to follow the procedure in Sec. 3.4.2.

3.7 Forecasting Value-at-Risk with min-dispersion profiles

Value-at-Risk (VaR) is a widely used technique to estimate the probability of losses of an asset portfolio. Since 2004 it is the recommended approach to assess market risk in the Basel II Accords on banking laws and regulations (despite regular criticisms made by experts about severe potential consequences when misused in risk management). Basically, a VaR estimate computed on the returns of a portfolio answers the question: What is the most I can expect

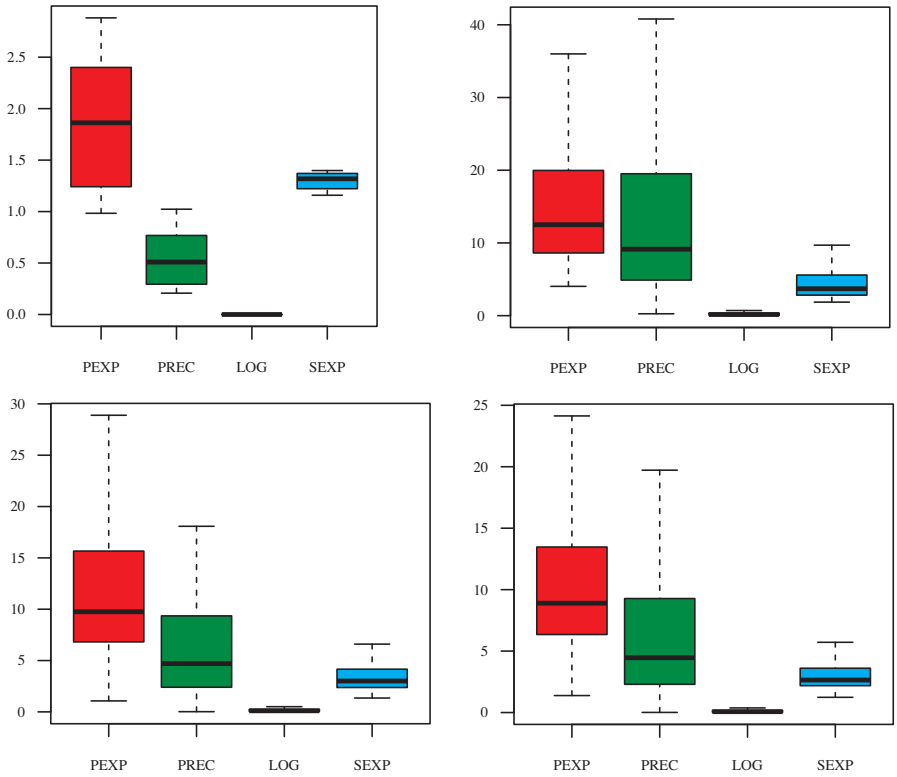


Figure 3.10: *Box plots of the fit results to the weight profiles for the "plateau" parameter b. Top left to right: European indices and stocks. Bottom left to right: Swiss and US stocks.*

to lose—with 95%/99%/.... confidence—in percentage of my portfolio value over the next day/week/months/...? One sees that to function VaR must be fed with two essential ingredients: a *confidence level* (usually, 95% or 99%) and a *forecast horizon* (usually one day or one week). Here we have set the confidence level at 95% and the forecast horizon at one day to take advantage of more historical data in the estimation process.

3.7.1 Historical estimation of VaR with non-uniform time-weighting

VaR can be estimated in a number of ways as for instance with dynamical models of volatility, Monte-Carlo simulations, extreme event theory, or with the historical method used here. Ultimately the choice of one method or another is based on practical considerations and in this regard the historical method is shown to be stable, easy to implement (it is indeed the easiest one), and it gives the fastest results as it only requires to reorder historical data (see Fig. 3.11 for an explanation of the method). These features make the historical method suitable to brokers managing a very large number of portfolios and willing to communicate VaR estimates to their clients on a regular basis. This is the case of Swissquote Bank which provides weekly estimates

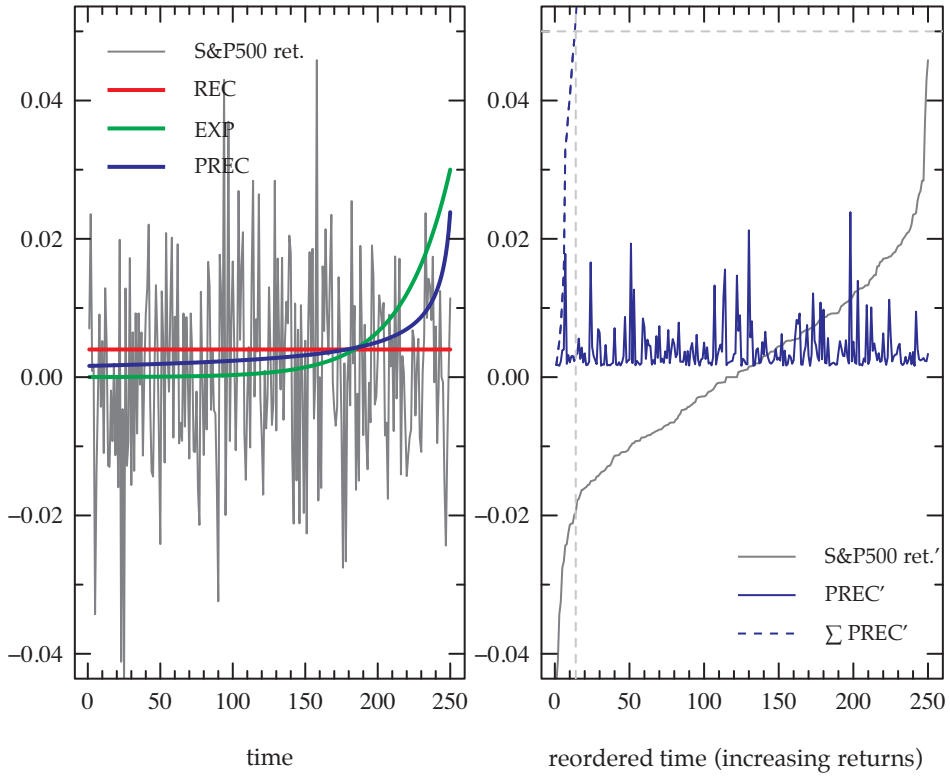


Figure 3.11: *Historical method of VaR forecasting with non-uniform weight densities. The VaR confidence level is fixed at 95% and the forecast horizon at one day. Left plot displays 250 daily returns of the S&P500 along with the weight profiles compared in this paragraph, namely the uniform (REC), exponential (EXP), and parametric power-decay PREC defined in Sec. 3.2.1 and fitted in Sec. 3.4. The parameters of PREC correspond to the median values measured on the US stock market $\delta = 0.77$ and $b = 9$ days Fig. 3.9 and Fig. 3.10. All the weight profiles are normalized to one. At the end of the period $t = 250$, VaR is estimated in the following way: (1) each return r_t is attached a weight w_t (i.e. a probability) according to the moment of its appearance in the time-window (2) returns are reordered from worst to best $r_t \to r'_k$ (right plot) (3) the cumulative sum of the reorganized weights w'_k is computed (dashed blue line) (4) the reordered return r'_p attached to the first weight such that $\sum_{k=1}^{p} w'_k \geq$ conf. level is the VaR estimate (strictly speaking, $r'_p = VaR$ if $w'_p =$ conf. level, otherwise an interpolation is used, e.g. $VaR = (r'_p + r'_{p-1})/2$). In this case $p = 13$ and $VaR(95\%) = -0.016$. Traditionally the historical method makes use of uniform weights. It has been extended here to more general weight densities.*

of VaR to its 180'000 clients using a improved version of the historical method.

3.7.2 Backtesting results: comparing the weight profiles

We have restricted our VaR computations to single assets in order to focus on dynamical properties and avoid potential difficulties due to cross-correlations (see e.g. [79] for methods dealing with VaR and cross-correlation estimation and in multi-asset portfolios). Fig. 3.11 illustrates the estimation of VaR on the S&P500 daily returns with the historical method extended to non-uniform weight densities. When the VaR estimation is repeated daily over a long period of time by means of a rolling window, one obtains the plot shown in Fig. 3.12. This plot gives much insight into the differences between the VaR estimates obtained with the simple (REC),

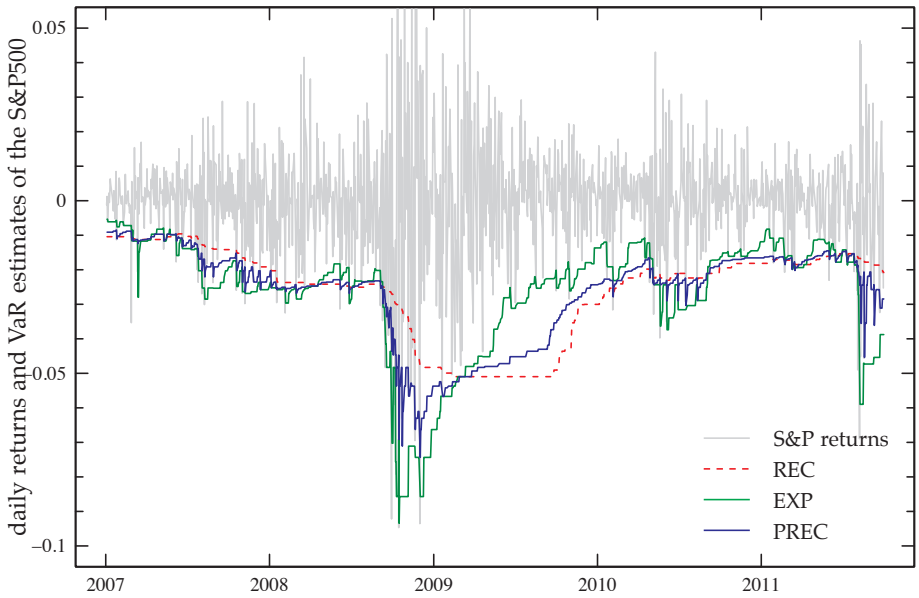


Figure 3.12: *Daily returns of the S&P500 along with VaR(5%) forecasts computed with the extended historical method (Fig. 3.11) over a moving window of 250 days. The three forecasts shown here have been obtained with uniform, exponential (EMA) and power-decaying (PREC) weight profiles (see Sec. 3.2.1 and Sec. 3.4). For EMA the typical value $\alpha = 0.97$ [46] has been used. The parameters of PREC are $b = 9$ days and $\delta = 0.77$ which are the median values of the fit results of Eq. (3.4) on the optimal weights of the US stock market. For a fair comparison between EMA and PREC, these values do not come from the minimization of the dispersion on the S&P500 absolute returns.*

exponential (EXP) and PREC moving averages. By definition of the Value-at-Risk a *VaR violation* occurs every time the S&P500 returns cross the VaR estimate curve. We readily see that most of the VaR violations of REC happened in the second halves of 2007 and 2008 whereas in contrast VaR violations of EXP and PREC tend to be less clustered in time—except in 2009 where PREC adopted a pretty conservative posture all year long due to the volatility boom caused by the end of 2008 recession. In practice we would like our VaR estimator to be accurate (i.e. VaR violations should occur on 5% of the days), and reactive enough so as to produce time-independent VaR violations (as is seemingly true of EXP). Therefore testing accuracy alone is not sufficient as it may happen that estimators have the right number of violations but that the latter are clustered in time (see VaR results for REC below).

In order to systematically assess and compare the quality of VaR forecasts produced by the different weight profiles, we have conducted two statistical tests constituting our *VaR backtesting* strategy. The first one is an independence test: given a binary sequence of 0 and 1, where 1 stands for a VaR violation, what is the probability of rejecting independence while the violations indeed occurred independently. The second one is a binomial test: given an *independent* sequence of VaR violations, what is the probability of wrongly rejecting the hypothesis that the number of

violations agrees with the VaR confidence (i.e. here 5 violations out of 100 days)? In both cases the probability is called the p-value of the test and is returned by any software implementing these tests (e.g. in R the functions `binom.test` and `runs.test`).

The p-values computed from the violation binary sequences of 1'250 US stocks are presented in Fig. 3.13. Results show that most of the VaR estimations performed with REC have failed the independence test (62%), as expected from Fig. 3.12, in the sense that they show a p-value smaller that the confidence threshold for the test (here set to 0.05). Of the total number of stocks 5% have failed the binomial test. This number remains pretty much the same when restricting to stocks that passed the independence which seems to indicate that there are no clear correlations between failing at one and the other test. The table below gives the percentage of failures at both tests for the three profiles. Results indicate a slightly more important dependence in the binary series of VaR violations of PREC as compared to EXP. On the other hand PREC shows a low rate of failures at the binomial test (slightly less than 1%) where EXP failure rate is above 3%.

	independence	binomial
REC	62%	5%
EXP	14%	3%
PREC	18%	1%

In a reassuring way the percentage of simultaneous failures at both tests is 0.4% for PREC and 0.5% for EXP, whereas it lies above 3.6% for REC.

Another observation arising from the examination of Fig. 3.12 is the lower variance of REC and PREC VaR estimates with respect to that of EXP. This is not surprising given the results of Sec. 3.3 and the fact that PREC has been designed so as to minimize the dispersion of volatility forecasts—volatility and VaR are closely related since they both assess market risk. The variance of the VaR results for the three profiles is compared in Fig. 3.14. As can be seen PREC does not give the minimum dispersion which most of the time is achieved by the simple moving average. The main reason is that PREC parameter values are the same for all stocks (we did so for the sake of a fair comparison with EXP and REC). Given the high dispersion observed on the fit results of Sec. 3.4 this choice is by far not optimal on many stocks.

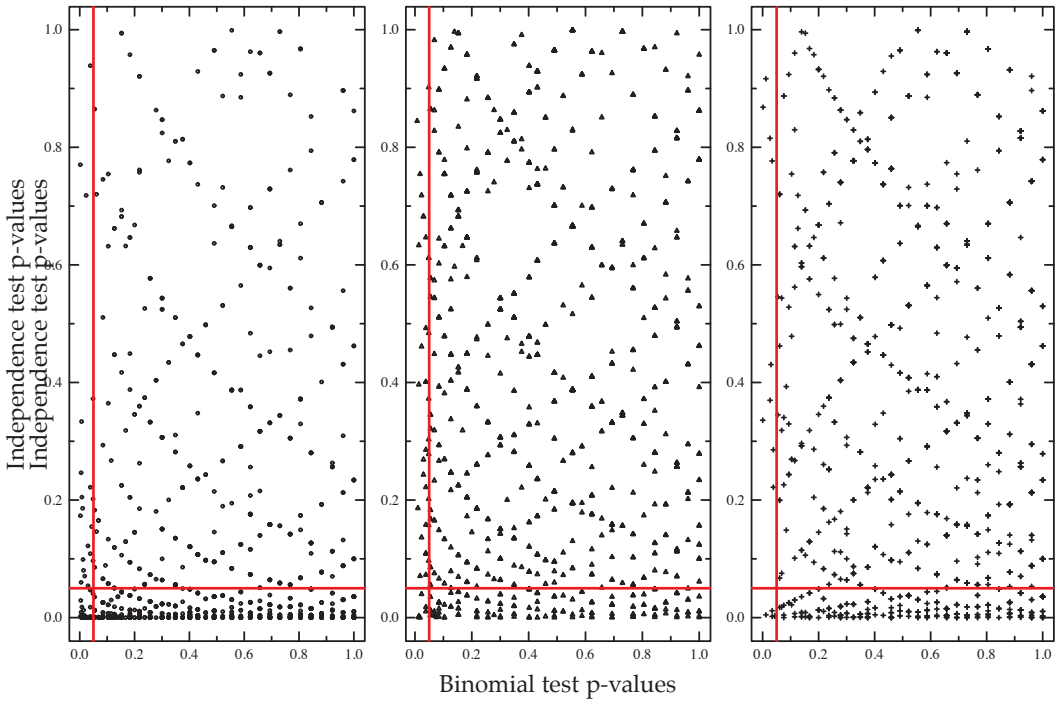


Figure 3.13: *p-values of the independence test with respect to those of the VaR violation test used to asses the quality of VaR forecasts. From left to right results correspond to the REC, EXP, and PREC weight profiles. Solid red lines show the 5% significance level of the test. Series with p-values below the significant level are said to fail the test with a risk of error below 5%. For instance, series with p-values belonging to the bottom-left rectangle fail both tests. The quality of their VaR forecasts is therefore highly questionable.*

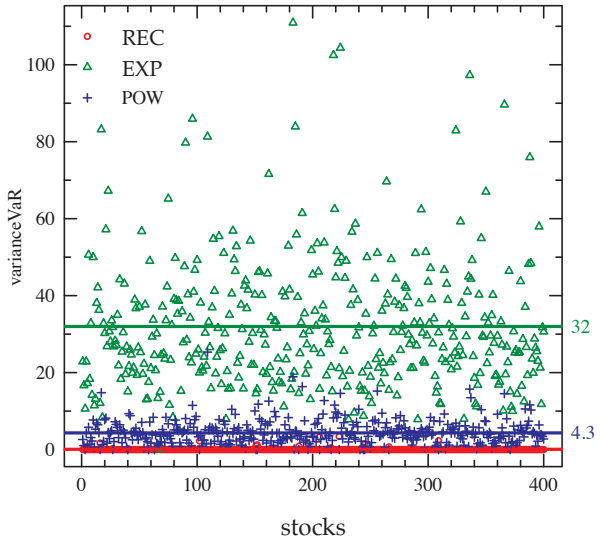


Figure 3.14: *Excess variance of VaR(95%) estimates on 400 US stocks for the three weight profiles REC, EXP and PREC. The smallest variance (generally obtained with simple moving averages) is taken as benchmark and the two others are expressed relatively to this value in percentage. Solid lines represent the mean over the 1'250 US stocks. On average EMA estimates are 32% more volatile than REC's and 28% more volatile than the ones computed with the PREC profile.*

3.8 Concluding remarks

We have seen that even a simplified description of the min-dispersion weights taking into account the autocorrelation structure of the volatility was able to reduce in a highly significant way the variance of VaR estimates, as compared with exponentially moving average, while inducing much less dependence in the sequence of VaR violations than simple moving averages. Furthermore, long-memory VaR estimators are more accurate than EWA and SMA in the sense that the number of their VaR violations, measured out-of-sample on 1'250 US stocks, was shown to be significantly closer to the prescribed confidence level. Although our backtesting procedure should be extended to other forecast horizons and confidence levels, these promising results so far suggest that long-memory estimators are more efficient than the commonly used alternatives with flat or exponential memory. This is not a surprise: to comparable or better accuracy, estimators with smaller dispersion are expected to have better efficiency. All these effects should be enhanced with the non-parametric form of the profile that one can obtain from direct minimization of a quadratic Toeplitz form under suitable constraints so as to preserve the accuracy and stability of the estimator (i.e. imposing non-decreasing and positive weights). As discussed in Sec. 3.4.2, the minimization process can be performed in a stable and efficient way by using appropriate algorithms implemented in open-source (e.g. R, Scilab) or commercial software solutions (e.g. Matlab or Mathematica). Their efficiency and ease of implementation should make long-memory estimators attractive to risk managers supervising a large number of portfolios.

4

Long-memory covariance matrices

4.1 Time-weighted covariance matrices: the issue of conditioning

We have seen in the previous chapter that one-dimensional weighted risk estimators can produce estimates with lower dispersion and higher accuracy when computed with a suitable power-decaying weight profile. It is natural at this stage to wonder about the extension of these results to the multidimensional case. In other words: what if one uses power-decaying weights to build a *long-memory covariance matrix of returns*? At first sight, this estimator should have less intrinsic dispersion, give more accurate predictions than the usual sample covariance matrix, and therefore should reduce portfolio turnover in mean-variance allocation while achieving lower *ex-post* volatility.

Things are not so simple though. In fact the construction of weighted covariance matrices raises important questions that should be addressed prior to feeding one's optimizer with them. The case of Exponential Moving Averages (EMA) is instructive in this regard. As expected, EMA covariance estimators are generally more accurate than simple moving average ones (in the sense that optimal portfolios solving the mean-variance optimization problem can reach lower out-of-sample volatility when computed with EMA covariance matrices as shown in this chapter), but it may produce so high turnovers that the resulting transaction costs in a real-world allocation may ruin the whole investment strategy (this will be demonstrated at the end of the chapter). Why is it so? Causes have to do with the (bad) conditioning of the weighted estimator.

To see this consider an investor who buys at $t - \Delta t$ the mean-variance portfolio $\mathbf{x}(t - \Delta t)$ computed with the return covariance matrix forecast $\hat{\Sigma}(t)$. Then at t the investor rebalances his positions to match the optimal portfolio at this date. The turnover of this operation depends partly on the intrinsic error of the estimator, partly on price movements between $t - \Delta t$ and t. Assume that Δt can be reduced sufficiently so that price fluctuations become negligible compared to the estimator intrinsic errors (e.g. to some sufficiently small intra-day horizon) and denote by $E = \hat{\Sigma} - \Sigma$ the forecast error matrix. Then it follows from standard matrix algebra that the absolute turnover is bounded by

$$|\Delta \mathbf{x}(t)| \lesssim \frac{\kappa(\Sigma)}{1 - \kappa(\Sigma)\frac{|E|}{|\Sigma|}} \frac{|E|}{|\Sigma|}, \tag{4.1}$$

where $\kappa(\Sigma)$ is the condition number of Σ defined as $\kappa(\Sigma) = |\Sigma||\Sigma^{-1}|$ if Σ is invertible and ∞ otherwise. Relation (4.1), which holds for any matrix norm and compatible vector norm, suggests that the conditioning of $\kappa(\Sigma)$ plays a key role in determining the sensitivity of the turnover to forecast errors (strictly speaking the sensitivity of the bound). In particular, for small $|E|$ the bound is of the order of $\kappa(\Sigma)(|E|/|\Sigma|)$,[1] that is, of the order of the relative forecast errors provided that $\kappa(\Sigma)$ remains small, which is the "ideal" situation we were expecting. On the other hand, for ill-conditioned matrices $\kappa(\Sigma)$ can take very large values, which implies that even small relative forecast errors may give rise to high portfolio turnovers and thus may cause high rebalancing costs. This is precisely the drawback of time exponentially-weighted covariance estimators.

Detailed examination reveals that the spectral shortcoming of EMA, much like its dynamical limitations discussed in Sec. 3.3, can be explained by the "steepness" of the exponential decay. Indeed, it is shown in [70] that the smallest eigenvalue of EMA matrices tends exponentially fast to zero as the "effective" number of observations captured by the exponential profile decreases. This inevitably leads to ill-conditioned estimator. The situation gets worse when the number of assets increases because in order to preserve the estimator accuracy weight profiles should not scale with the number of assets.

In this chapter we first derive a general formula and a simple algorithm for the computation of the eigenvalue density of general weighted covariance estimators. We then focus on power-

[1] This relation becomes $\kappa(\Sigma)(|E|/|\Sigma| + |e|/|r|)$ when forecasts on the returns and corresponding errors $e = \hat{r} - r$ are included.

decaying profiles, in particular on PREC which was shown to provide the best parametric description of the volatility min-dispersion weights among four heavy-tailed profiles, and derive exact results for its eigenvalue density and edge spectrum in two asymptotic limits. We finally present results of backtested strategies on real data with the newly introduced long-memory estimators.

4.2 Theory

We first give a formal definition of the weighted covariance estimator used throughout the text, and then proceed to its spectral analysis in the case of independent asset returns with zero mean and unit variance (i.e. a correlation matrix). Since we deal here with finite-length series the sample correlation matrix of independent returns is not the identity, and therefore its spectrum is not trivially 1. Our spectral analysis has two goals: (1) building a null model for the eigenvalue density of weighted correlation matrices of independent returns, and (2) assessing the effects of specific weight profiles on the estimator conditioning, with particular focus on power-decaying profiles.

4.2.1 Weighted covariance estimator

Let H be the $N \times T$ matrix of returns. Our aim in this section is to study the dependence of the spectrum of weighted correlation matrices on the chosen weight profile. Towards this goal, we will consider first the i.i.d. model, that is, empirical correlation matrices of the form

$$C = \frac{1}{N} H \operatorname{diag}(\mathbf{w}_N)\, H^t, \tag{4.2}$$

where H is an $N \times T$ matrix with i.i.d. entries—having zero mean and unit variance—and $\mathbf{w}_N = (w_N(1), \ldots, w_N(T))$ is a given profile, exhibiting some limiting behavior as N and T become large. Restricting ourselves to matrices with entries having all unit variance allows us to focus on *correlation effects*; the impact of the profile on the variance of the individual observed processes has been discussed at length in the previous chapter.

4.2.2 Random Matrix Theory in a nutshell

Random Matrix Theory (RMT) deals with matrices whose entries are random variables. The theory aims at computing the (joint) distribution of the matrix eigenvalues given the distribution of its entries. In some cases, RMT allows one to find the distribution of other quantities of interest as extreme eigenvalues and eigenvector components. In this chapter, we focus on the eigenvalue distribution and the edge spectrum as these are of primary interest in portfolio allocation (see later).

Given an $N \times N$ symmetric and non-negative definite matrix X with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_N \geq 0$, we define its empirical eigenvalue distribution as[2]

$$F_N(x) = \frac{1}{N} |\{k \in \{1, \ldots, N\} : \lambda_k \leq x\}|, \quad x \in \mathbb{R},$$

where $|A|$ denotes the cardinality of the set A. Let also

$$g_N(z) = \frac{1}{N} \operatorname{Tr}\left((X - zI)^{-1}\right) = \int_{\mathbb{R}} \frac{1}{x - z} \, dF_N(x), \quad z \in \mathbb{C}\backslash\mathbb{R},$$

denote the Stieltjes' transform of F_N. We say that F_N converges weakly a.s. (i.e. almost surely) to a (deterministic) limiting distribution function F if $F_N(x)$ converges to $F(x)$ a.s., for every $x \in \mathbb{R}$ continuity point of F. This implies in particular that

$$\lim_{N \to \infty} g_N(z) = g(z) = \int_{\mathbb{R}} \frac{1}{x - z} \, dF(x), \ a.s. \quad \forall z \in \mathbb{C}\backslash\mathbb{R}. \tag{4.3}$$

Assuming that $F(x)$ admits the density $p(x)$, one can further invoke the Frobenius-Perron inversion formula to obtain

$$p(x) = \frac{dF}{dx}(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \operatorname{Im} g(x + i\varepsilon), \quad x \in \mathbb{R}. \tag{4.4}$$

Furthermore, a recent result [74] states that for correlation matrices as in Eq. (4.2), Eq. (4.4) reduces to

$$p(x) = \frac{1}{\pi} \operatorname{Im} g(x), \quad x \in \mathbb{R}\backslash\{0\}, \tag{4.5}$$

which will be useful for the actual computation of the eigenvalue density function.

[2] The theory is valid for Hermitian matrices with complex entries. We restrict here to real non-negative definite matrices for the sake of simplicity.

4.2.3 Main result

In the following, we derive a general equation for the Stieltjes transform of the asymptotic eigenvalue distribution of weighted correlation matrices, that is valid under mild assumptions on the profile and that is found to be easily solvable in some limit for various families of profiles. The theorem below can be directly deduced from Marčenko and Pastur's main result [62].

Theorem 1. • *Let $c_0 > 0$ be fixed and T, N be integers tending to infinity with a fixed ratio, that is:*

$$\lim_{N,T\to\infty} \frac{T}{N} = c_0.$$

- *Let H be an $N \times T$ matrix with i.i.d. real entries h_{jk} having zero mean and unit variance.*

- *Let $(w_N(k))_{k\in\{1,..,T\}}$ be positive real numbers such that the sequence of empirical distribution functions*

$$F_W^{(N)}(x) = \frac{1}{T}\,|\{k \in \{1,\dots,T\} \,:\, w_N(k) \le x\}|, \quad x \ge 0,$$

converges weakly towards a given distribution function $F_W(x)$ as $N, T \to \infty$.

- *Let C be the $N \times N$ matrix defined as $C = \frac{1}{N} H\, diag(w_N)\, H^t$, that is,*

$$C_{ij} = \frac{1}{N}\sum_{k=1}^{T} w_N(k)\, h_{ik}h_{jk}, \quad i,j \in \{1,\dots,N\}.$$

Then the empirical eigenvalue distribution of C converges weakly almost surely as $N, T \to \infty$ towards a deterministic distribution, whose Stieltjes' transform $g(z)$ is solution of the equation

$$g(z) = -\left(z - c_0 \int_0^\infty \frac{x}{1 + x\, g(z)}\, dF_W(x)\right)^{-1}, \quad z \in \mathbb{C}\backslash\mathbb{R}. \tag{4.6}$$

Let us now introduce a continuous and decreasing function $w : [0, c_0] \to \mathbb{R}_+$ such that

$$\lim_{N\to\infty} w_N(\lfloor Nt \rfloor) = w(t), \quad \forall t \in [0, c_0]. \tag{4.7}$$

We will see below examples of sequences w_N and functions w satisfying this assumption. Let us first see how Eq. (4.7) allows us to rewrite Eq. (4.6) in a simpler form. Using this assumption as

well as the one made in Theorem (1), we obtain that for any $x \geq 0$ continuity point of F_W,

$$\begin{aligned} F_W(x) &= \lim_{N,T\to\infty} \frac{1}{T} \left|\{k \in \{1,\ldots,T\} \,:\, w_N(k) \leq x\}\right| \\ &= \frac{1}{c_0} \left|\{t \in [0,c_0] \,:\, w(t) \leq x\}\right| = \frac{1}{c_0}\, w^{-1}(x), \end{aligned}$$

where w^{-1} denotes the formal inverse of w (which exists as w is decreasing). Therefore, $c_0\, F_W(x) = w^{-1}(x)$ for all continuity points of F_W. Making thus the change of variable $t = w^{-1}(x)$ in Eq. (4.6) gives

$$g(z) = -\left(z - \int_0^{c_0} \frac{w(t)}{1 + w(t)\, g(z)}\, dt\right)^{-1}, \quad z \in \mathbb{C}\backslash\mathbb{R}. \tag{4.8}$$

For a given constant c_0 and a given profile w, Eq. (4.8) does not have in general a closed-form solution as the integral on the right-hand side can be some special function. Even though $g(z)$ may be computed in a number of cases (see next section), it is often helpful for comprehension's sake and numerical efficiency to obtain analytical results. To this aim, a possible simplification in the analysis of Eq. (4.8) can be made by letting the factor c_0 tend to infinity. But this requires to assess under which condition on w Eq (4.8) admits a unique solution in the limit $c_0 \to \infty$. This study has been rigorously conducted in Appendix A.1.

It is shown that if $w : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous and decreasing function such that $w \in L^2(\mathbb{R}_+)$, then the limiting equation

$$g(z) = -\left(z - \int_0^{\infty} \frac{w(t)}{1 + w(t)\, g(z)}\, dt\right)^{-1}, \quad z \in \mathbb{C}\backslash\mathbb{R}. \tag{4.9}$$

admits a unique solution which is the Stieltjes transform of a distribution.

Alternatively, one can obtain the same result via the more recent R-transform method by generalizing the calculation in [70] for uniform covariance matrices, to estimators with arbitrary weight profiles. We present this alternate and less rigorous derivation below as a simple way of obtaining the main formula without relying on Marčenko and Pastur's heavy machinery.

Derivation of Eq. (4.9) by the R-Transform method For simplicity we will assume that the entries of the matrix H are i.i.d. Gaussian with zero mean and unit variance (the Gaussian assumption is crucial for the R-transform). Let $g_N(z)$ be the Stieltjes transform of C^N (i.e. C with $N < \infty$) and $g(z) := \lim_{N\to\infty} g_N(z)$. We define the R-transform of C as $R(z) := g^{-1}(-z) - \frac{1}{z}$,

where $z \in \mathbb{C}$; by definition, $g(g^{-1}(-z)) = -z$.

We consider the rank-one matrices $\delta C_j^N := \frac{1}{N} w_N(j)\, h_j\, h_j^t$. Notice that because the entries of H are Gaussian and independent, these matrices are both independent and rotationally invariant (i.e. they are independent Wishart matrices). The idea is to use the additivity of the R-transform for such matrices [**?**, 10], which is expressed as $R(z) = \sum_{j=1}^{\infty} R_{\delta C_j}(z)$, and then to compute $g(z)$ and finally $p(x)$ by Eq. (4.5).

The multiplication of h_j by δC_j^N leads to $\delta C_j^N h_j = w_N(j)\, \frac{\|h_j\|^2}{N}\, h_j$, so that δC_j^N has one eigenvalue equal to $w_N(j)\, \frac{\|h_j\|^2}{N}$ and $N-1$ zero eigenvalues. By the law of large numbers, the non-zero eigenvalue is approximately equal to $w_N(j)$ for large N. This allows us to calculate the Stieltjes transform of δC_j^N in this asymptotic limit:

$$g_{\delta C_j}(z) \equiv \frac{1}{N} \mathrm{Tr}((\delta C_j^N - zI)^{-1}) \simeq \frac{1}{N} \left(\frac{1}{w_N(j) - z} - \frac{N-1}{z} \right).$$

Inverting $g_{\delta C_j}(z)$ to find $R_{\delta C_j}(z)$, we first obtain the following quadratic equation in z:

$$xz^2 + (1 + wx)z - (1 - \frac{1}{N})w \simeq 0, \tag{4.10}$$

where we have used the abbreviations $x := g_{\delta C_j}(z)$ and $w := w_N(j)$. The solutions of Eq. (4.10) read

$$\begin{aligned} z_\pm &\simeq \frac{1}{2x} \left(wx - 1 \pm (1 + wx) \sqrt{1 - \frac{4wx}{N(1+wx)^2}} \right) \\ &\simeq \frac{1}{2x} \left(wx - 1 \pm \left[(1 + wx) - \frac{2wx}{N(1+wx)} \right] + O\left(N^{-2}\right) \right). \end{aligned} \tag{4.11}$$

We consider the solutions to first order in $1/N$. Inverting Eq. (4.11) to find $x(z_+)$ (with $O(N^{-2})$ dropped), we realize that $\lim_{\Im z \downarrow 0} \Im x(\Re z + i \Im z) = 0$, so that $x(z_+)$ cannot be the Stieltjes transform of a distribution. On the other hand, the solution z_- leads to a valid Stieltjes transform, which yields

$$g_{\delta C_j}^{-1}(x) \simeq -\frac{1}{x} + \frac{w_N(j)}{N(1 + w_N(j)x)}.$$

Then the R-transform of W can be easily calculated as

$$R_{\delta C_j}(z) = g_{\delta C_j}^{-1}(-z) - \frac{1}{z} \simeq \frac{w_N(j)}{N(1 - w_N(j)z)} \Rightarrow R(z) \simeq \frac{1}{N} \sum_{j=1}^{\infty} \frac{w_N(j)}{1 - w_N(j)z},$$

and thus

$$g^{-1}(z) = R(-z) - \frac{1}{z} \simeq \frac{1}{N}\sum_{j=1}^{\infty} \frac{w_N(j)}{1+w_N(j)z} - \frac{1}{z} \underset{N\to\infty}{\longrightarrow} \int_0^{\infty} \frac{w(t)}{1+w(t)\,z} dt - \frac{1}{z}.$$

Finally, replacing z by $g(z)$ in the above equation and rearranging terms yields the final result established before:

$$g(z) = -\left(z - \int_0^{\infty} \frac{w(t)}{1+w(t)\,g(z)}\, ds\right)^{-1}.$$

In the above model, T accounts for the amount of available historical data, which is assumed here to be of the order of the number of assets N. As the value of N is typically large in our setting, this assumption may sound unpractical at first sight. In our model, the fact that the profile w decays over time compensates for this unrealistic assumption, as fewer and fewer weight is put on old data as time goes by. Comparing Eq. (4.8) and Eq. (4.9), we see that as $c_0 \to \infty$, the former equation converges to the latter. Provided that the profile w decays sufficiently fast to zero, the solution of the resulting equation gives therefore a reasonable approximation to the solution of Eq. (4.8).

One could argue that the right model is to consider a perhaps large but fixed amount of available historical data T, independent of the number of assets N. Our viewpoint on this problem is on the contrary that considering the present asymptotic regime leads to explicit answers regarding the dependency of the spectrum on the shape of the profile. These are important in practical applications and might never have been obtained by setting T to be constant.

Next, we consider particular examples of profiles w satisfying the above assumptions. Interpreting w as a weighting profile, a natural assumption to add is the profile normalization, i.e.

$$\|w\|_{c_0} = \int_0^{c_0} w(t; c_0)\, dt = 1, \tag{4.12}$$

which, as $c_0 \to \infty$, reads

$$\|w\|_1 = \lim_{c_0\to\infty} \|w\|_{c_0} = \int_0^{\infty} w(t)\, dt = 1. \tag{4.13}$$

In Eq. (4.13), it is understood that $w(t) \equiv \lim_{c_0\to\infty} w(t; c_0)$. Notice that Eq. (4.13) together with the assumption that w is continuous and decreasing implies that $w \in L^2(\mathbb{R}_+)$. Indeed,

$$\|w\|_2^2 = \int_0^{\infty} w(t)^2\, dt \le w(0) \int_0^{\infty} w(t)\, dt < \infty. \tag{4.14}$$

Algorithme 4.1 Compute the asymptotic spectral density of weighted correlation matrices

Require: λ, $k_{\max}$, tol
1: $g_0 \leftarrow$ random starter in $\mathbb{C} \setminus \mathbb{R}$
2: **for** $k = 1$ to $k_{\max}$ **do**
3: $\quad g_k \leftarrow g_{k-1} - G_\lambda(g_{k-1}) / G'_\lambda(g_{k-1})$
4: $\quad$ **if** $|g_k - g_{k-1}| \leq$ tol **then**
5: $\quad\quad g^* \leftarrow g_k$
6: $\quad\quad$ **exit loop**
7: $\quad$ **end if**
8: **end for**
9: $p(\lambda) \leftarrow \frac{1}{\pi} \mathrm{Im}\, g^*$

4.2.4 Computing the spectral density of i.i.d. weighted correlation matrices

Consider the following complex function of $g \in \mathbb{C} \setminus \mathbb{R}$:

$$G_z(g) = g \int_0^{c_0} \frac{w(t; c_0)}{1 + w(t; c_0)\, g}\, dt - g\, z - 1.$$

If g is as in Theorem 1, then the equation $G_z(g) = 0$ has a unique solution for any $c_0 \in]1, \infty]$ and positive decreasing profile w (and integrable in the case $c_0 = \infty$). It is easy to show that $G_z(g)$ is holomorphic with respect to g and has derivative

$$G_z(g)' = \int_0^{c_0} \frac{w(t; c_0)}{1 + w(t; c_0)\, g} dt - g \int_0^{c_0} \left(\frac{w(t; c_0)}{1 + w(t; c_0)\, g} \right)^2 dt - z.$$

The well-known Newton-Raphson method is particularly suitable for finding the zero of a holomorphic function with known derivative. Algorithm 4.1 provides a straightforward version of this recursive scheme adapted to the computation of $p(\lambda)$. The algorithm converges very quickly to the solution provided that the initial value g_0 is in the solution basin of attraction, which can be (almost surely) ensured by randomly choosing $g_0 \in \mathbb{C} \setminus \mathbb{R}$. Note that the algorithm crucially relies on the fact that $\lim_{z \to \lambda} G_z(g(z)) = G_\lambda(g(\lambda))$, which follows from Eq. (4.5).

4.3 Long-memory covariance matrices

4.3.1 Old and new weight profiles

As discussed in Chapter 3, the most popular weight sequences used in volatility forecasting are the uniform and exponential weightings. They are now defined formally and so as to satisfy the assumptions of Theorem 1.

The uniform weight sequence (REC):

$$w_N(k) = \frac{1}{c_0} 1_{\{1 \le k \le c_0 N\}} \Rightarrow w(t; c_0) = \frac{1}{c_0} 1_{[0,c_0]}(t). \tag{4.15}$$

The Exponential Moving Average (EMA): EXP Up to an appropriate normalizing factor, the exponential weight sequence and profile are given by

$$w_N(k) \sim \left(1 - \frac{1}{cN}\right)^k \Rightarrow w(t; c_0) = \frac{e^{-\frac{t}{c}}}{c\left(1 - e^{-\frac{c_0}{c}}\right)} \Rightarrow \lim_{c_0 \to \infty} w(t; c_0) = \frac{1}{c} e^{-\frac{t}{c}}. \tag{4.16}$$

We define similarly PREC and PEXP, which have been shown in Sec. 3.4 to give best fit results to the min-dispersion weights. As will be demonstrated theoretically and by simulation in this chapter, these heavy-tailed profiles do spare the conditioning of Σ even for large portfolio size. EXP, on the other hand, is not suitable for portfolio allocation as it induces very high turnovers for the optimal values of the parameter c (Sec. 4.4).

PREC produces an "almost" flat short-time decay:

$$w_N(k) \sim \frac{1}{1 + \left(\frac{k}{cN}\right)^\gamma} \Rightarrow w(t; c_0) = \frac{K_0}{1 + (\frac{t}{c})^\gamma}, \tag{4.17}$$

where $K_0 = \left(c_0 \, {}_2F_1\left(1, \frac{1}{\gamma}; 1 + \frac{1}{\gamma}; -\left(\frac{c_0}{c}\right)^\gamma\right)\right)^{-1}$ is such that $\|w(t; c_0)\|_{c_0} = 1$ (Eq. (4.12)) and ${}_2F_1(a, b; c; z)$ is the hypergeometric function [44]. For $c_0 = \infty$, PREC becomes

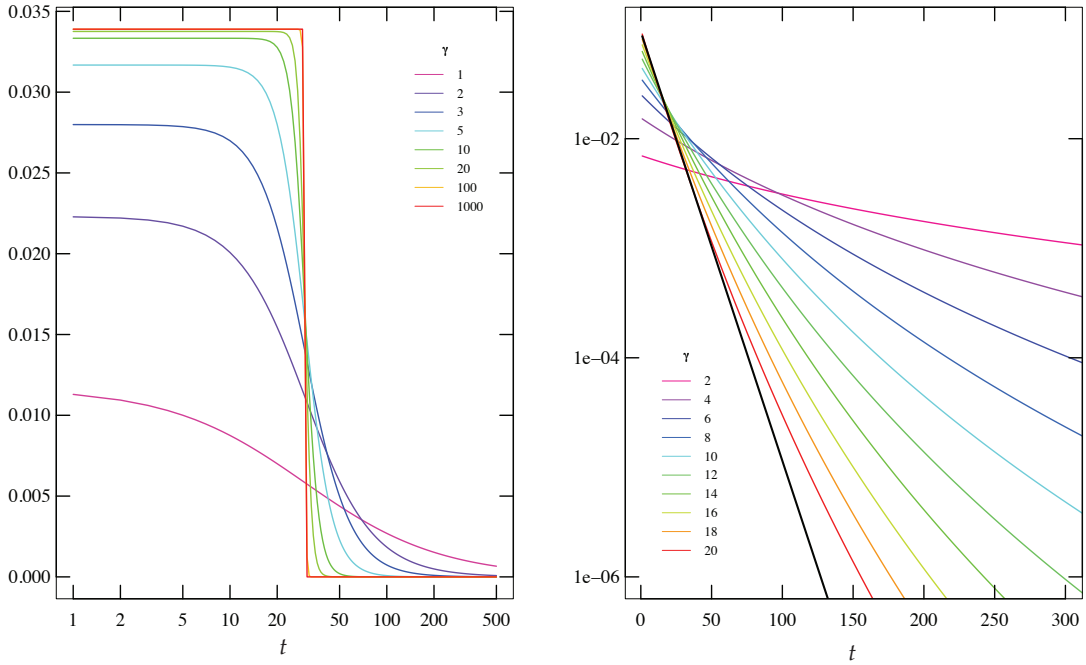
$$w(t) = \frac{\gamma}{\pi c} \frac{\sin(\frac{\pi}{\gamma})}{1 + (\frac{t}{c})^\gamma}. \tag{4.18}$$

PEXP provides an alternate version of power-decaying profile with exponential short-time behavior:

$$w_N(k) \sim \frac{1}{\left(1 + \frac{k}{cN}\right)^\gamma} \Rightarrow w(t; c_0) = \frac{Q_0}{(1 + \frac{t}{c})^\gamma}, \tag{4.19}$$

where $Q_0 = \frac{(\gamma - 1)}{c - c^\gamma (c + c_0)^{1-\gamma}}$. In the asymptotic case $c_0 = \infty$ PEXP reads

$$w(t) = \frac{\gamma - 1}{c\,(1 + \frac{t}{c})^\gamma}. \tag{4.20}$$



(a) *PREC tending to REC for high values of γ*

(b) *PEXP decaying as EXP (black line) for high values of γ*

Figure 4.1: *As γ is increased PREC tends to the uniform weighting (left plot) and PEXP resembles EXP over a larger range of values (right plot). This will be used in Sec. 4.4 to analyze the effects of long-memory weights in a real-world portfolio allocation.*

The rate of decay in both sequences is tuned up by the memory parameter γ: for high values of γ PREC and PEXP resemble respectively REC and EXP (see Fig. 4.1).

	$w_N(k)$	$w(t)$	$\bar{t}$	d	$\delta_{0.5}$
	$\frac{1}{N}\sum_{k=1}^{\infty} w_N(k) = 1$	$\int_0^{\infty} w(t)\,dt = 1$	$\int_0^{\infty} t\,w(t)\,dt$	$\int_0^{d} w(t)\,dt = 1-\varepsilon$	$\frac{w(\delta_{0.5})}{w(0)} = 0.5$
REC	$\frac{1}{c_0}\,1_{\{1\le k\le c_0 N\}}$	$\frac{1}{c_0}\,1_{[0,c_0]}(t)$	$\frac{c_0}{2}$	$\begin{cases}(1-\varepsilon)c_0 & d\le c_0\\ 0 & d>c_0\end{cases}$	/
EXP	$\frac{1}{c}\left(1-\frac{1}{cN}\right)^k$	$\frac{1}{c}\,e^{-\frac{t}{c}}$	c	$c\,\log(\frac{1}{\varepsilon})$	$c\,\log 2$
PREC ($\gamma>1$)	$\frac{K_N}{1+(\frac{k}{cN})^{\gamma}}$	$\frac{\gamma}{\pi c}\,\frac{\sin(\frac{\pi}{\gamma})}{1+(\frac{t}{c})^{\gamma}}$	$c\,(2\cos\frac{\pi}{\gamma})^{-1}$ ($\gamma>2$)	$\sim c\,(\frac{1}{\varepsilon})^{\frac{1}{\gamma}-1}$	c
PEXP ($\gamma>1$)	$\frac{Q_N}{(1+\frac{k}{cN})^{\gamma}}$	$\frac{\gamma-1}{c(1+\frac{t}{c})^{\gamma}}$	$c\,(\gamma-2)^{-1}$ ($\gamma>2$)	$c\left((\frac{1}{\varepsilon})^{\frac{1}{\gamma-1}}-1\right)$	$c\left(2^{\frac{1}{\gamma}}-1\right)$

Table 4.1: *The weight profiles of Sec. (4.3.1) for $c_0=\infty$ and their main attributes: the mean number of observations $\bar{t}$, the effective sample size at a given tolerance level $d(\varepsilon)$, and the median lag $\delta_{0.5}$. Attributes are useful for comparison purpose and to better understand the role of the parameters: e.g. c is the mean number of observations in EXP and the return half-life time in PREC. As can be seen c is proportional to all the measures of the effective number of observations (i.e. information) captured by the profile.*

Remark 2. Obviously, one can define other heavy tailed functions exhibiting some sort of long-memory. For instance, in Sec. 3.6 we have fitted a logarithmic profile and the stretched exponential SEXP: $w(t) \sim \exp(-(x/c)^\gamma)$, $\gamma \in [0,1]$, to the min-dispersion weight profiles on the one-dimensional volatility estimator[3]. Although giving fairly good results, SEXP and LOG were generally outperformed by the two power-decaying profiles PREC and PEXP defined above.

4.3.2 Eigenvalue density and edge spectrum of PREC estimators

We work out the spectral density and the edge spectrum of the i.i.d. PREC estimator in the two asymptotic cases $c_0 < \infty$ and $c_0 = \infty$. This profile was shown to provide the best parametric description of min-dispersion volatility estimators in Sec. 3.6. PEXP does not lead in general to closed-form calculations (except for a few integer values of γ), nevertheless, we show numerically that the spectrum of PREC and PEXP are much alike because of their similar long-term decay.

Spectral density

Finite T/N ratio For $T/N = c_0 < \infty$, the definite integral in Eq. (4.8) can be expressed in terms of the hypergeometric function [44]:

$$\int_0^{c_0} \frac{w(t)}{1+w(t)g} dt = \frac{c_0 K_0}{1+K_0 g} \, {}_2F_1\left(1, \frac{1}{\gamma}; 1+\frac{1}{\gamma}; -\frac{\left(\frac{c_0}{c}\right)^\gamma}{1+K_0 g}\right), \tag{4.21}$$

where the value of the profile normalizing factor K_0 is given in Section 4.3.1. Although this is sufficient for computing the asymptotic spectrum via the method of Section 4.2.4, we can gain significant insight and computing power in some cases of interest by deriving a closed-form equation for g. Indeed, a general expression for $\gamma = 1/p$, $p \in \mathbb{N}$ is obtained in Appendix A.1; in the most simple case, $\gamma = 1$, one finds

$$1 + zg = c\, K_0 g \log\left(1 + \frac{c_0}{c\,(1+K_0 g)}\right).$$

Consistently, in the limit of infinite information $c \to \infty$ one recovers the Marčenko and Pastur equation for g [62]. Other closed-form expressions for general γ require to consider the limit $T/N = \infty$.

[3] For $\gamma = 1$ SEXP is equal to EXP, whereas it decreases slower and slower for values of γ approaching zero. SEXP can therefore be seen as a *tradeoff* between REC and EXP.

Infinite T/N ratio Letting $c_0 = \infty$ in Eq. (A.2) and assuming $\gamma > 1$ leads to

$$\int_0^\infty \frac{w(t)}{1+w(t)g}dt = \left(\frac{1}{1+Kg}\right)^{\frac{\gamma-1}{\gamma}},$$

where from Sec. 4.3.1 $K = \frac{\gamma}{\pi c}\sin(\frac{\pi}{\gamma})$. Using this result in Eq. (4.9) leads to the following general equation for Stieltjes transform of PREC

$$1 + zg = g(1+Kg)^{\frac{1}{\gamma}-1}. \tag{4.22}$$

Interestingly, for $\gamma = q/p$ with $q > p \geq 1$ two integers, the valid transform of p is a root of a complex polynomial of degree $2q - p$ given by

$$(1+Kg)^{q-p}(1+zg)^q - g^q. \tag{4.23}$$

This shows that p admits a closed form only for $\gamma = 2$ and $\gamma = 3/2$ for which its transform is solution of a polynomial of degree 3, respectively of degree 4. In all other cases, Eq. (4.22) has to be solved numerically, for instance with Algorithm 4.1. The PREC density for several values of γ and c is compared with an empirical histogram in Fig. 4.2.

Edge spectrum

The edge spectrum of weighted estimators is important in portfolio allocation, for it is related to the conditioning of the estimator via the condition number $\kappa(C) = \lambda_{\max}/\lambda_{\min}$. The edge spectrum of C is also useful as it gives the limits of the i.i.d. return model; any eigenvalue outside these limits can be regarded as "informative" from a statistical (and thus a financial) point of view. The idea has been successfully used in portfolio allocation to reduce the "out-of-sample" volatility of the optimal portfolio (see [17] for a review of the eigenvalue filtering and conditioning techniques).

The approach of Marčenko and Pastur [62] to obtain the edge spectrum of matrices verifying Theorem 1 readily carries over to the case $T/N = \infty$. It is applied below to PREC estimators.

Denoting by $B \equiv g^{-1}$ the inverse function of the Stieltjes transform, the bounds of the spectrum

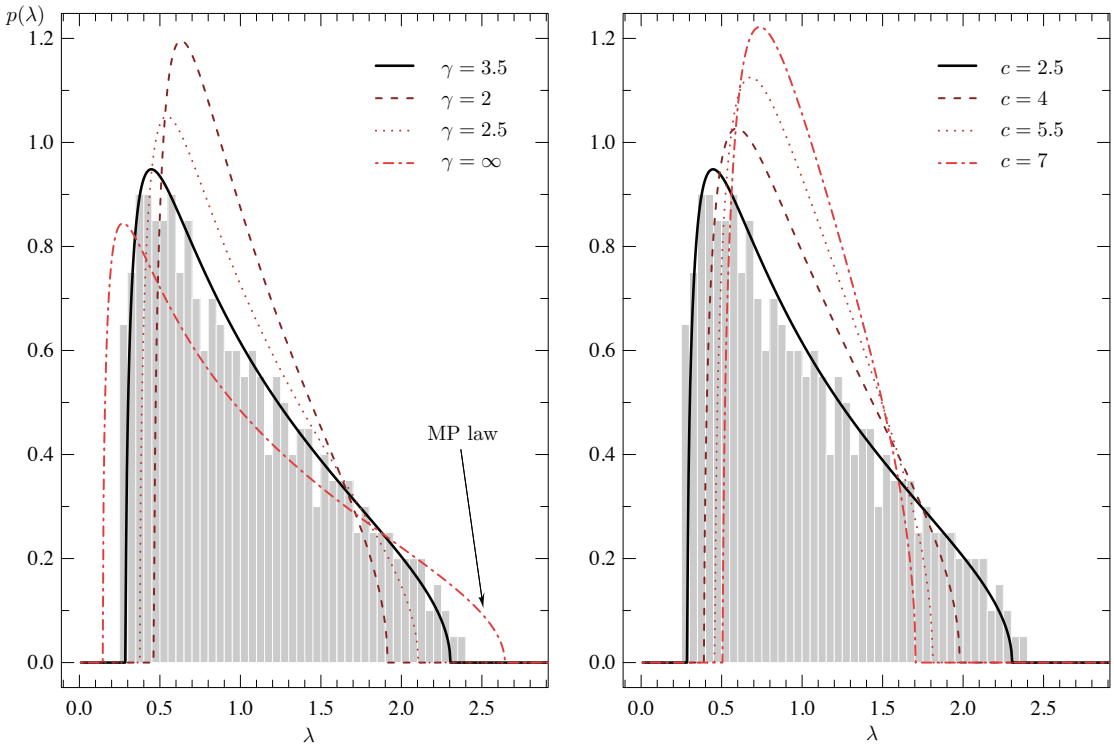


Figure 4.2: *The asymptotic spectral density of PREC in the limit $c_0 = \infty$ (Eq. 4.22) compared with the spectral histogram of a 400×2000 matrix of i.i.d Student returns. The setting is $c = 2.5$ and $\gamma = 3.5$. Dashed lines show the density upon varying the memory parameter γ (left plot) and median lag c (right plot). For $\gamma \gg 1$ PREC tends to the Marčenko and Pastur law. As the information c captured by PREC increases the density peaks around one (i.e. the correlation "noise dressing" vanishes), and in the limit $c \to \infty$ the spectrum is exactly equal to one, that is the spectrum of the identity matrix in this limit.*

are given by $B(y_\pm)$, where $y_\pm$ are such that $B'(y_\pm) = 0$. From Eq. (4.22), $B(y)$ verifies

$$B(y) = (1 + Ky)^{\frac{1}{\gamma}-1} - \frac{1}{y}, \tag{4.24}$$

whose derivative can be written

$$B'(y) = K\left(\frac{1}{\gamma} - 1\right)(1 + Ky)^{-1}\left(\frac{1}{y} + B\right) + \frac{1}{y^2}. \tag{4.25}$$

Setting $B'(y) = 0$ and rearranging terms gives a second-degree equation for y:

$$(\gamma - 1)By^2 - y - \frac{\gamma}{K} = 0. \tag{4.26}$$

When introduced in Eq. (4.24), the solutions of Eq. (4.26) lead to the following equations for the upper, resp. the lower, bound of the spectrum:

$$\lambda_{1,N} = \left(\frac{K \mp \sqrt{K\left[K + 4\gamma(\gamma-1)\lambda_{1,N}\right]}}{2(\gamma-1)\lambda_{1,N}} + 1\right)^{\frac{1}{\gamma}-1} \pm \frac{2(\gamma-1)\lambda_{1,N}}{\sqrt{\frac{4\gamma(\gamma-1)\lambda_{1,N}}{K} + 1} \mp 1}. \tag{4.27}$$

Hence for given values of γ and c the limiting eigenvalues can easily be computed by solving numerically Eq. (4.27). Interestingly, a much simpler expression is found by expanding the right-hand side of this equation to first order in λ_N. Solving the resulting linear equation leads to

$$\lambda_N \underset{\lambda_N \to 0^+}{\approx} \frac{\pi^{\frac{(\gamma-1)^3}{\gamma^2}}}{\gamma}\left(\frac{c(\gamma-1)}{\gamma^2 \sin\frac{\pi}{\gamma}}\right)^{\gamma-1} \sim c^{\gamma-1}. \tag{4.28}$$

Hence in the critical regime $\lambda_N \approx 0$ the smallest eigenvalue of PREC matrices decays as a power of c at a rate that is governed by the memory parameter γ. Even though we haven't found such a simple expression for λ_1, we have observed that $\lambda_1 \sim c^{-\alpha}$ with $\alpha \approx 2$ for small c and different values of γ (see Fig. 4.3). From this observation, we therefore conclude that the condition number of PREC i.i.d. correlation matrices decreases with c roughly as $\kappa(C) \sim 1/c^{\gamma+1}$ for $\gamma > 1$. By the similarity between both power-decaying profiles the same conclusion holds for PEXP matrices (see the insets in Fig. 4.3).

4.3.3 Simulation results

As shown in Fig. 4.3, Eq. (4.28) agrees well with simulations based on randomly generated returns, even when the latter are drawn from independent but *non-identical* Student distributions[4]. The simulation goes as follows: we draw a matrix $H \in \mathbb{R}^{400\times500}$ of 400 independent Student returns with mean zero and degree of freedom chosen randomly in $[2, 5]$. Then the return cross-correlation matrices are computed for values of c between 1 and 200 days. The dashed line results from a linear fit to the tail of the curves. For values of $\gamma > 1.5$, we find a good agreement with the theoretical bound $\lambda_N \sim c^{\gamma-1}$ obtained by considering $\lambda_N \ll 1$ (Eq. (4.28)). On the right plot, we examine the behavior of the maximum eigenvalue. Fits show that $\lambda_1 \sim c^{\alpha}$ with $\alpha \approx 2$ for all $\gamma \in \{1.25, 1.5, 2, 2.5, 3\}$ and $c \in [1, 10]$. A more intricate relation between λ_1, c and γ is observed for higher values of c.

It was shown in [70] that the smallest eigenvalue of EMA matrices goes exponentially fast to zero as the information c captured by the profile decreases:

$$\lambda_N \sim \exp(-1/c).$$

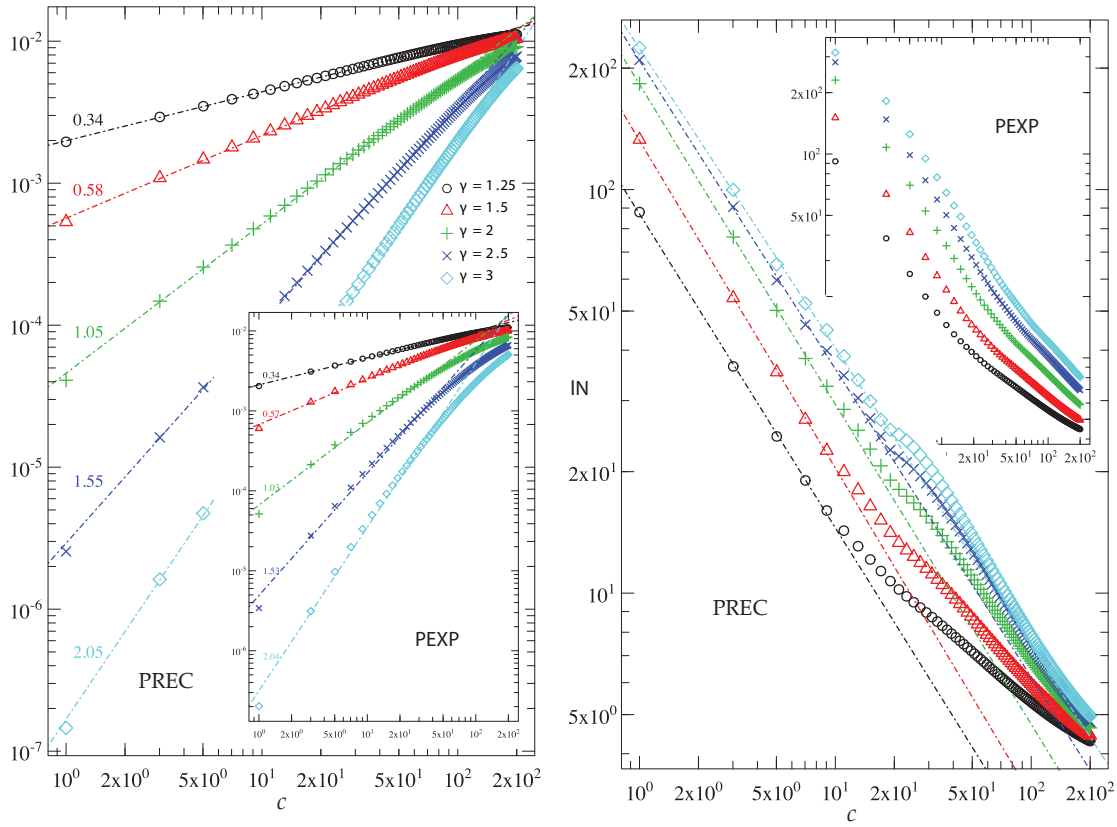
As can be seen in Fig. 4.4, the smallest eigenvalue of PREC and PEXP matrices is much larger than that of EXP when optimal values (obtained in the one-dimensional case) are plugged in. Furthermore, power-decaying profiles produce a null model with narrower spectral bandwidth, which is a useful feature for the eigenvalue-based filtering techniques [17].

4.4 Turnover and long-memory profiles in portfolio allocation

Simulations The goal of this application is twofold (1) assessing the consequences of ill-conditioned estimators on both realized volatility and portfolio turnover, and (2) showing how long-memory added to the exponential profile can decrease the turnover in MVO while reducing (to some extent) the realized volatility.

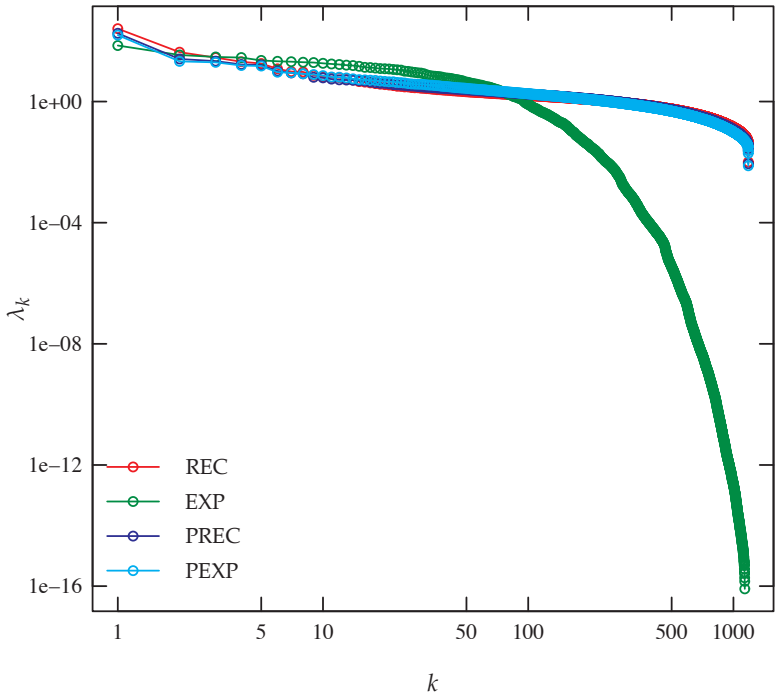
We have run many long-only mean-variance allocations using REC (uniform), EXP, and PEXP covariance matrices (with and without the expected return constraint), and averaged over subsamples to improve the readability of the results. The simulation parameters are given below (without the expected return constraint).

[4] This choice is closer to empirical observations than returns drawn from the Gaussian distribution [15].

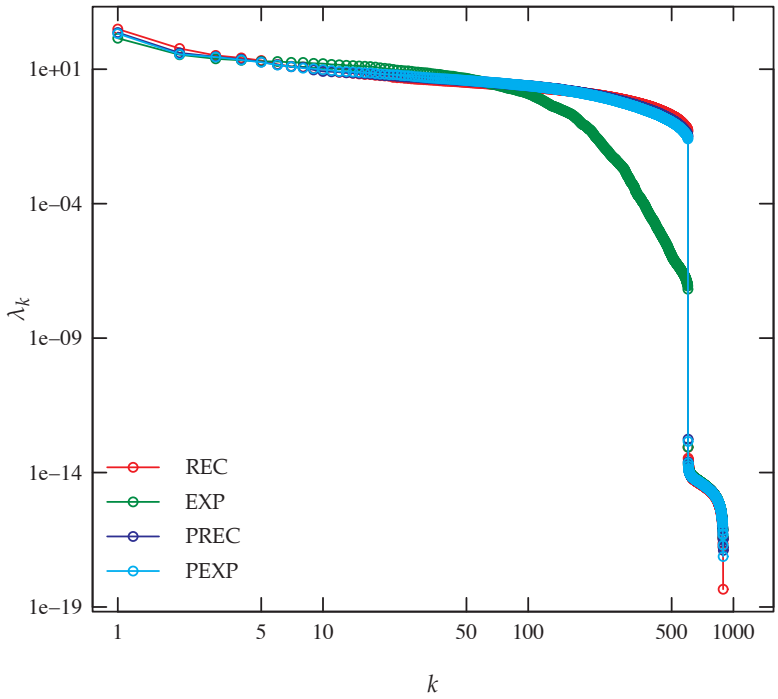


(a) *smallest eigenvalue* λ_N **(b)** *Largest eigenvalue* λ_1

Figure 4.3: *Left panel: main plot: smallest eigenvalue* λ_N *as a function of the median lag c for PREC-weighted covariance matrices (log-log plot). For each* $\gamma \in \{1.25, 1.5, 2, 2.5, 3\}$*, we draw a matrix* $H \in \mathbb{R}^{400\times 500}$ *of 400 independent Student returns with mean zero and degree of freedom chosen at random in* $[2, 5]$*. Then the PREC estimator is computed for values of c between 1 and 200 days. The dashed lines result from a linear fit to the tail of the curves. Inset: Same plot for PEXP matrices. Right panel: same plots for the largest eigenvalue* λ_1*. In this case the fits show that* $\lambda_1 \sim c^{\alpha}$ *with* $\alpha \approx 2$ *for all* γ *and* $c \in [1, 10]$*.*



(a) *Daily returns*



(b) *Weekly returns: the matrix is rank-deficient (zero eigenvalues to numerical precision are shown).*

Figure 4.4: *Spectrum of weighted correlation matrices. Data are the daily returns of 1'200 stocks of the NYSE over the period 1996-2008. The profile parameters are chosen according to their optimal forecast values obtained in the one-dimensional case, that is: $c_{exp} = 33$ days corresponds to the typical RiskMetrics value $\lambda = 0.97$ [46]. For power-decaying profiles the values $(c_{prec}, \gamma_{prec}) = (4, 0.6)$ and $(c_{pexp}, \gamma_{pexp}) = (8, 0.75)$ are the medians of the fit results obtained on the US stock markets for these two profiles (see Sec. 3.4). Plots show the better conditioning of power-decaying estimators over EXP. Similar conclusions hold for rank-deficient correlation matrices (i.e. the number of assets exceeds the length of time series), that is, the magnitude of the first non-zero eigenvalue is substantially larger with power-estimators than with EXP.*

Simulation settings

Market	*NYSE*
Calibration period	*1998-2006*
Observation period	*2006-2010*
Return horizon	*weekly*
Portfolio rebalancing	*weekly*
Allocation method	*long-only MVO*
Subsampling	*300 stocks / 1'200*
Nb. subsamples	*50*

The realized volatility σ was defined as follows: the optimal portfolio was allocated every Monday and the realized daily volatility was computed as the square-root of the mean square returns over the week. Then the daily volatility was annualized multiplying by $\sqrt{250}$.

To better illustrate the relation between the realized volatility σ and the portfolio turnover Δ (annualized), we have plotted $\sigma(\Delta)$ for different values of the profile parameters.

Plot results The result for the uniform weighting (REC) is the red diamond in the top-left corner in Fig. 4.5. As expected, REC shows low turnover and high realized volatility as compared to other (decreasing) profiles. The red squares are the results for EMA matrices. They were obtained by letting c_{exp} take the following values (from left to right on the plot)

c_{exp}:	400	200	100	66	50	40	**33**	20	15

where the RiskMetrics value ($\lambda = 0.97$) corresponds to $c_{exp} = 33$ weeks. We see that on average EMA achieves a significant, although not striking, reduction of σ with respect to REC's. The minimal realized volatility is obtained for $c_{exp} = 100$ ($\lambda = 0.99$), which is in agreement with the minimal value obtained by [68] with similar EMA matrices but on daily returns. We notice that the RiskMetrics value $c_{exp} = 33$, considered optimal in the one-dimensional case (in the sense of forecasting), is not optimal when used in a multi-asset environment.

Unlike σ, the annualized turnover is very sensitive to the variations of c_{exp}, which can be explained by the high sensitivity of the smallest eigenvalue of EMA matrices to the variations of

this parameter. For instance, the annualized turnover observed for $c_{exp} = 100$ is half the value observed for $c_{exp} = 33$, that is, the latter strategy may be on average twice as costly as the optimal one.

To see the theoretical results of the previous section at work—in particular our considerations on the edge spectrum and the conditioning of weighted covariance matrices—we have imagined a way to integrate long-memory in EMA while preserving the shape of the short-time exponential decay. This can be done in a simple way: set γ to a very large value so that PEXP is as close as possible to the exponential profile. Then for t not too large the two profiles must coincide, that is, one should have

$$\text{PEXP} \approx \text{EXP} \Rightarrow c_{pexp} = \gamma c_{exp}, \tag{4.29}$$

which follows from equating the first order terms in the two Taylor expansions. As can be seen in Fig. 4.6, this method allows one to preserve in PEXP the short-time exponential decay of EXP while progressively adding long-memory by decreasing γ.

For each value of c_{exp}, we have defined 10 PEXP profiles using the method above and the following values for γ

γ:	1.2	2	3	4	10	15	30	80	150	500

.

For each γ we have run the same simulations as before and averaged the annualized out-of-sample volatility and turnover over 50 subsamples. The results on the plot Fig. 4.5 show that adding long-memory efficiently and substantially decreases the turnover, at least for $c_{exp} \geq 100$. As suggested by the theoretical results in the previous chapter, this illustrates the high sensitivity of portfolio turnovers to the tail of the weight profiles (since in this case the short-term profile remains exponential).

4.5 Conclusion

The dynamical and spectral analysis of weighted correlation matrices has lead us to introduce a class of weight profiles with long-memory, which, while maintaining a small number of fitting parameters, are able to capture the persistence of volatility and to spare the conditioning of the estimator.

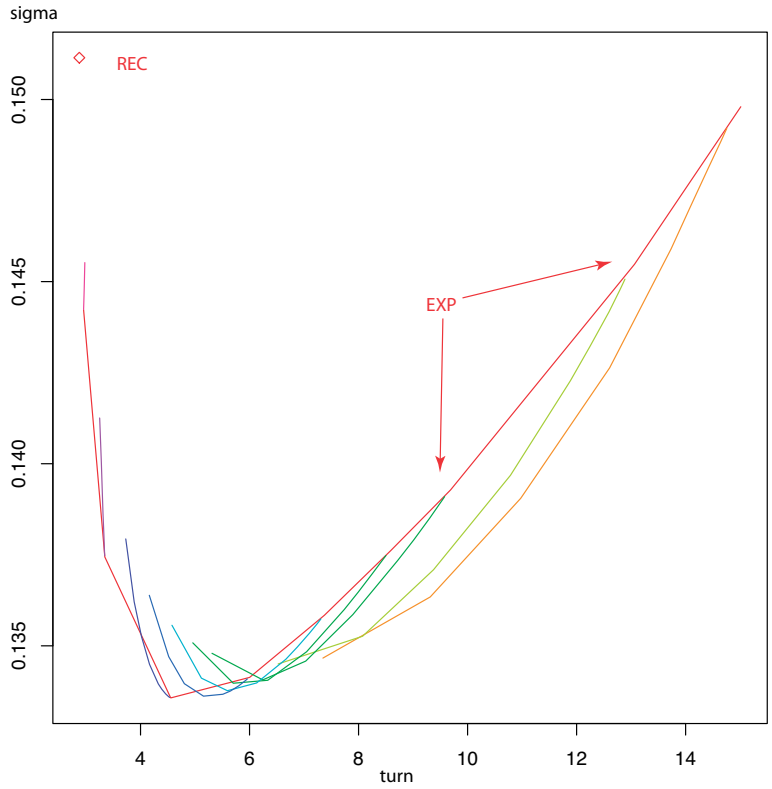


Figure 4.5: *Annualized out-of-sample volatility and turnover for the MVO problem in which the covariance matrix is estimated with uniform (REC), exponential (EXP), and power-decaying (PEXP) weight profiles (details in the text).*

We have derived a spectral null model for weighted correlation matrices with i.i.d. entries that is valid in the two limits $\lim_{N,T\to\infty} T/N = c_0 < \infty$ and $\lim_{N,T\to\infty} T/N = \infty$, where N is the number of assets and T the number of observations. Through an appropriate change of variable, these results readily extend to the case $0 < c_0 < 1$, that is, when the number of assets exceeds the number of observations. Finally, the introduction of "long-memory" in EMA covariance matrices was shown do decrease substantially the MVO portfolio turnover, which we believe provides another evidence that long-memory is key to improving the efficiency of one-dimensional and multi-dimensional estimators.

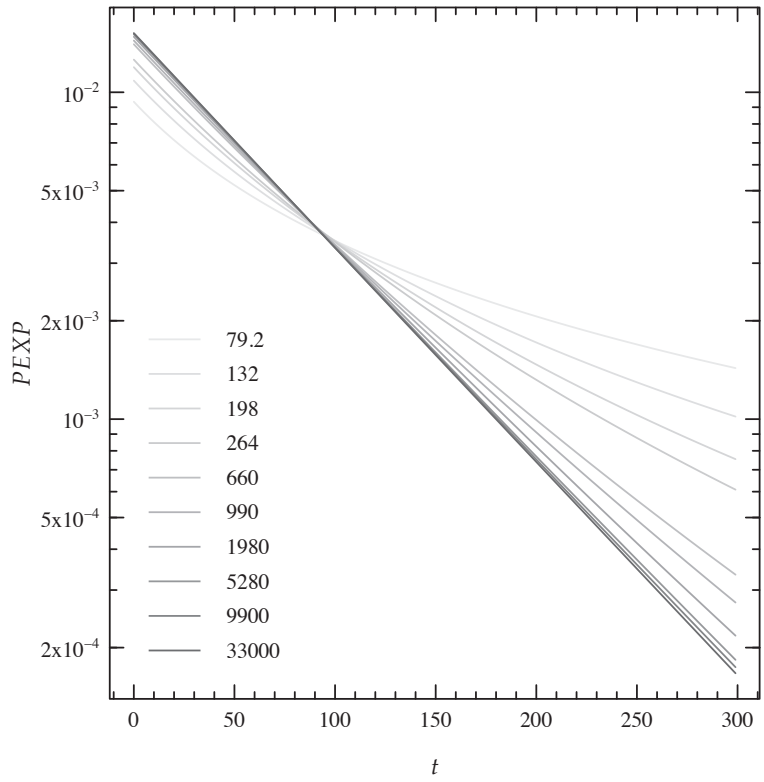


Figure 4.6: *PEXP can be used to add long-memory in EMA while preserving the shape of the exponential short-time decay. The mean number of observations in EXP is fixed, here at $c_{exp} = 66$, and the power-decaying profiles PEXP are defined by $c_{pexp} = \gamma c_{exp}$, where γ takes the values used in this section. The legend displays the values obtained for c_{pexp}.*

5

Simplifying the Mean-Variance Optimization

To address the topical "curse of dimensionality" (see Sec. 1.2) in portfolio allocation, we propose a spectral-based method that reduces the sample covariance matrix of asset returns and simplifies the complexity of the mean-variance optimization (MVO). In contrast with current trends (e.g. factor models), our method is fully algebraic and thus does not rely on any ad hoc assumptions on the statistical properties of the observed returns. It applies to the usual (weighted and non-weighted) covariance and correlation matrices as well as to other less common but useful robust estimators. The method produces a "coarse-grained" matrix that is always better conditioned than the input, in addition to featuring straightforward financial interpretation.

5.1 Exploiting stock correlations to reduce the covariance matrix of returns

Modern portfolio theory relies on the minimization of expected risk balanced with the maximization of expected returns. In this context, we have already seen that the sample covariance matrix plays a key role as it embeds the idiosyncratic risk associated with individual fluctuations of prices, as well as the systematic component rooted in the co-movement of correlated returns. Whether it is analyzed in its raw form (weighted or not), decomposed, or filtered out [16], the covariance matrix of returns is the topic of an ever-growing number of papers by researchers from various horizons. Very recently, the effective size of the mean-variance portfolio (a measure of its actual diversification) was found to be much smaller than its actual size [65].

This raises a question that is directly related to the so-called *curse of dimensionality* faced by today's asset managers: the proliferation of financial instruments to be handled by portfolio allocators poses complex practical challenges. We argue that the latter may be overcome, at least partially, if one is able to reduce the sample matrix of returns to its effective size and conduct the MVO over the reduced estimator. The reduction, however, should meet some basic requirements to fit in mean-variance optimizers: (1) it should be linear and (2) the reduced covariance estimators $\widetilde{\Sigma}$ should not be ill-conditioned since the optimal portfolio depends on $\widetilde{\Sigma}^{-1}$ [70]. Furthermore, to be efficient the reduction is expected to preserve in $\widetilde{\Sigma}$ most of the relevant financial information contained in the original covariance matrix Σ and thus to be the result of some function optimization over the original asset space.

We will see that this suggests the use of *semi-projectors* to achieve the reduction as in the Spectral Coarse Graining (SCG) of graphs introduced in [39], and formalized and expanded in [31]. We expose in this chapter the main elements of the theory and present preliminary out-of-sample results of real-world allocation backtests (our priority here is to set the theoretical frame).

5.2 Different types of covariance estimators

Let us start with N time series of observed log-returns arranged in the $N \times T$ matrix H. Mathematically, the elements of H are defined as $h_{it} = \log p_i(t) - \log p_i(t - \Delta t)$, where $p_i(t)$ is the price of asset i at time t and Δt is the return time-horizon, usually ranging from a few minutes to several months. The (unweighted) sample covariance matrix is defined as

$$\Sigma^{unb} = \frac{1}{T-1}\left(HH^t - \frac{1}{T}H\mathbf{1}H^t\right), \tag{5.1}$$

with $\mathbf{1}$ standing for the $T \times T$ matrix of ones. It is readily checked that Σ^{unb}_{ij} is the usual unbiased estimate of $\mathrm{cov}(H_i, H_j) = \mathrm{E}(H_iH_j) - \mathrm{E}(H_i)\mathrm{E}(H_j)$. In practice, the mean of H_i is often much smaller than its fluctuations, so it is customary to consider instead the (biased) sample covariance matrix of centered returns, already encountered in Chapter 4, and given by

$$\Sigma = \frac{1}{T}HH^t. \tag{5.2}$$

The estimator in Eq. (5.2) is the maximum likelihood estimator (MLE) of the covariance matrix for normally distributed returns. The normality of asset returns being highly challenged for decades (see e.g. [28] for a review of the most important stylized facts about markets) robust MLE accounting for the return excess kurtosis have been recently proposed as alternatives in portfolio allocation. Other popular estimators include the time-weighted covariance matrices analyzed in Chapter (4).

The reduction method presented here applies to the estimators of Eq. (5.1) and Eq. (5.2), to time-weighted covariance matrices, as well as to many robust estimators. This flexibility, which follows from the pure algebraic nature of the coarse-graining operation, makes the method suitable to applications in other fields of research as for instance to the analysis of gene expression data in bioinformatics.

5.3 Coarse-graining the Mean-Variance Optimization problem

Under its general form the Mean-Variance Optimization (MVO) problem consists in finding the N asset portfolio $\mathbf{x}^*$, solution of

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^N} \mathbf{x}^t\Sigma\mathbf{x} &- \alpha\,\mathbf{x}^t\mathbf{r} \\ A\mathbf{x} &\leq \mathbf{b}, \end{aligned} \tag{5.3}$$

where (without loss of generality) Σ is the covariance matrix Eq. (5.2), $\mathbf{r}$ is the asset vector of expected returns, and α is the risk tolerance factor describing the investor's attitude toward risk (i.e. a small α indicates risk-aversion whereas a large α a risk-prone attitude). All the linear constraints are expressed in a canonical form by $A\mathbf{x} \leq \mathbf{b}$; beside normalization these are typically restrictions to: long-only positions ($x_i \geq 0$), mean expected return above the threshold μ ($\sum_i x_i r_i \geq \mu$), minimum diversification ($x_i \leq d$), etc.[1]

The *coarse-grained* version of Eq. (5.3) is defined as

$$\begin{aligned} \min_{\mathbf{y}\in\mathbb{R}^{\tilde{N}}} \mathbf{y}^t\widetilde{\Sigma}\mathbf{y} &- \alpha\,\mathbf{y}^t\mathbf{q} \\ B\mathbf{y} &\leq \mathbf{c}, \end{aligned} \tag{5.4}$$

[1] Equality constraints are formed from two inequalities, e.g. $\sum_i x_i \geq 1$ and $\sum_i x_i \leq 1$

where $\widetilde{N} < N$ and the reduced covariance matrix results from a mapping $\widetilde{\Sigma} \equiv M(\Sigma)$ from the space of non-negative definite matrices in $\mathbb{R}^{N\times N}$ to the same space in $\mathbb{R}^{\widetilde{N}\times\widetilde{N}}$.

In order to account for the linearity of the constraints in MVO, it is convenient to consider a linear reduction of Σ, that, is to write

$$\widetilde{\Sigma} = R\Sigma R^t = \frac{1}{T}(RH)(RH)^t = \frac{1}{T}\widetilde{H}\widetilde{H}^t, \tag{5.5}$$

where the *coarse graining operator* R is a $\widetilde{N} \times N$ real matrix such that the rows of $\widetilde{H} = RH$ are linear combinations of the asset returns. The financial interpretation of $\widetilde{H}$ is then straightforward: $\widetilde{H}$ is the return matrix of $\widetilde{N}$ *sub-portfolios* built from the N assets according to the weights embedded in R (hence $\widetilde{\Sigma}$ is merely the covariance matrix of $\widetilde{H}$).[2]

For any portfolio $\mathbf{y} \in \mathbb{R}^{\widetilde{N}}$, the vector $\mathbf{x} = R^t\mathbf{y}$ defines a portfolio in $\mathbb{R}^N$ such that the coarse-grained MVO problem can be conveniently expressed in terms of R and of the original settings as

$$\begin{aligned} \min_{\mathbf{y}\in\mathbb{R}^{\widetilde{N}}} \ \mathbf{y}^t R\Sigma R^t\mathbf{y} \quad &- \quad \alpha\, \mathbf{y}^t R\mathbf{r} \\ AR^t\mathbf{y} \quad &\leq \quad \mathbf{b}. \end{aligned} \tag{5.6}$$

Transforming back the minimizer $\mathbf{y}^*$ of Eq. 5.6 with R^t defines in a natural way the *coarse-grained portfolio*

$$\mathbf{x}^*_{cg} = R^t\mathbf{y}^* \in \mathbb{R}^N.$$

In general $\mathbf{x}^*_{cg}$ is expected to be different from the optimal portfolio $\mathbf{x}^*$ that is solution of the original problem in $\mathbb{R}^N$. The key point then is to find $\widetilde{N}$ and R such that the *out-of-sample* behavior of $\mathbf{x}^*_{cg}$ is close enough to the one of $\mathbf{x}^*$ so that $\mathbf{x}^*_{cg}$ can be safely used as a proxy for $\mathbf{x}^*$ in an actual portfolio allocation. Succeeding in this task leads to an improvement of the performance of the optimizer by direct reduction of its workload. The expected gain of this operation crucially depends on the effective size $\widetilde{N}$, which decreases as the cross-correlation between assets increases as in the world stock market since 2000 (Fig. 1.1).

[2]Taking the same R to left-and right-multiply Σ in Eq. 5.5 ensures the nonnegative definiteness of $\widetilde{\Sigma}$: $\mathbf{y}'\widetilde{\Sigma}y = (R'\mathbf{y})'\Sigma(R'\mathbf{y}) \geq 0\,\forall\mathbf{y}$ since Σ is nonnegative definite.

5.4 Building the coarse graining operator

Obviously R should be built so as to preserve in $\widetilde{\Sigma}$ the relevant financial information contained in Σ—ideally, we would like to only remove noise by coarse-graining Σ. But that is not all. We have already discussed the dangers of badly conditioned estimators in MVO (Sec. 4.1). It is therefore essential for the coarse-graining operation not to spoil the conditioning of the original estimator; as argued below this is ensured if R takes the form of a semi-projector.

Semi-projectors and eigenvalue interlacing

A semi-projector R is a matrix whose cross-product defines a projector $R^tR = P = P^2$. From this definition, it can be seen that the range of P is spanned by the rows of R, and thus the idea will be to construct appropriately R so as to preserve in $R\Sigma R^t$ the most "informative" eigenpairs of Σ (which lie at the extremity of the spectrum [69]).

Given a semi-projector $R \in \mathbb{R}^{\widetilde{N}\times N}$, what can be said about the conditioning of $R\Sigma R^t$? Let $0 \leq \lambda_N \leq \lambda_{N-1} \leq \ldots \leq \lambda_1$ denote the eigenvalues of Σ and $\kappa^{-1}(\Sigma) = \lambda_N/\lambda_1$ its inverse condition number (in Euclidean norm); Σ is said to be well-conditioned if $\kappa^{-1}(\Sigma) \approx 1$ and ill-conditioned when $\kappa^{-1}(\Sigma) \approx 0$. By the Poincaré eigenvalue separation theorem [49], the eigenvalues of Σ and $\widetilde{\Sigma} = R\Sigma R^t$ interlace as

$$\lambda_{N-\widetilde{N}+i} \leq \tilde{\lambda}_i \leq \lambda_i, \tag{5.7}$$

which implies that $\widetilde{\Sigma}$ is never worse conditioned than Σ (in practice it can be significantly better conditioned). Indeed

$$\kappa^{-1}(\widetilde{\Sigma}) = \frac{\tilde{\lambda}_{\widetilde{N}}}{\tilde{\lambda}_1} \geq \frac{\tilde{\lambda}_{\widetilde{N}}}{\lambda_1} \geq \frac{\lambda_N}{\lambda_1} = \kappa^{-1}(\Sigma).$$

Hence the linear reduction of Σ through semi-projectors produces reduced estimators with a narrower spectral bandwidth.[3] Nevertheless, this result, which does not depend explicitly on the semi-projector R, is not sufficient for assessing the effects of the reduction on the individual eigenvalues of Σ. Since we want to preserve in $\widetilde{\Sigma}$ a few *prescribed* eigenpairs of Σ, it is crucial to better understand the connexion between R, the spectrum of Σ, and that of $\widetilde{\Sigma}$.

[3] A similar result holds for the eigenvectors of Σ and $\widetilde{\Sigma}$ [42]

Location of the coarse-grained eigenvalues

The connexion between the eigenvalues of $\widetilde{\Sigma}$ and those of Σ is uncovered below by following a perturbative approach. We start from the equality

$$\tilde{\Sigma}(R\mathbf{v}) = \lambda(R\mathbf{v}) + R\Sigma e_P(\mathbf{v}), \tag{5.8}$$

where λ is an eigenvalue of Σ for the vector $\mathbf{v}$ (i.e. $\Sigma\mathbf{v} = \lambda\mathbf{v}$), and $e_P(\mathbf{v}) = \mathbf{v} - P\mathbf{v}$ is such that $\|e_P(\mathbf{v})\|$ measures the distance between $\mathbf{v}$ and the projection $P\mathbf{v}$. If trivially $\mathbf{v} = P\mathbf{v}$ then $R\mathbf{v}$ is an eigenvector of the coarse-grained matrix $\tilde{\Sigma}$ for the eigenvalue λ.

In the following, we first show that when $\mathbf{v} \neq P\mathbf{v}$ the norm of $R\Sigma e_P(\mathbf{v})$ provides an estimate of the distance between λ and the closest eigenvalue of $\tilde{\Sigma}$. Ultimately, we derive a bound that is used to optimize the coarse-graining operation. The derivation is not complicated, but it requires some basic facts about matrices and norms that are briefly stated below.

The spectral norm of a matrix M is defined as $\|M\| \equiv \max\{\sqrt{\mu} : \mu \in \mathrm{sp}(M^tM)\}$, such that for P a projector and R a semi-projector $\|P\| = \|R\| = 1$ since $\{0,1\} = \mathrm{sp}(P) = \mathrm{sp}(R^tR) = \mathrm{sp}(P^tP)$.[4] The spectral norm is a vector space norm with the following additional properties [49]: (1) $\|M_1M_2\| \leq \|M_1\|\,\|M_2\|$ (i.e. sub-multiplicativity) and (2) $\|M\mathbf{x}\| \leq \|M\|\,\|\mathbf{x}\|$ $\forall\mathbf{x}$ (i.e. consistency with the vector Euclidean norm).

If λ is not an eigenvalue of $\tilde{\Sigma}$, then $\det(\tilde{\Sigma} - \lambda I_N) \neq 0$ and Eq. 5.8 can be rewritten as

$$R\mathbf{v} = (\tilde{\Sigma} - \lambda I_N)^{-1} R\Sigma e_P(\mathbf{v}). \tag{5.9}$$

Taking the spectral norm on both sides of Eq. 5.9, we obtain

$$\|R\mathbf{v}\| \leq \|(\tilde{\Sigma} - \lambda I_N)^{-1}\|\|R\Sigma e_P(\mathbf{v})\|.$$

Since $\tilde{\Sigma}$ is symmetric, there is a unitary matrix $\tilde{U}$ such that $\tilde{\Sigma} = \tilde{U}\tilde{D}\tilde{U}^t$, with $\tilde{D}$ a diagonal matrix

[4] In contrast with the usual Euclidean norm in the space of matrices: $\|R\|_F = \|P\|_F = \mathrm{rank}(P) \geq 1$.

containing the eigenvalues of $\tilde{\Sigma}$. Hence

$$\begin{aligned} \|(\tilde{\Sigma} - \lambda I_N)^{-1}\| &= \|\tilde{U}^t(\tilde{D} - \lambda I_N)^{-1}\tilde{U}\| \\ &\leq \|(\tilde{D} - \lambda I_N)^{-1}\| \\ &\leq \max_{\tilde{\lambda}} |\tilde{\lambda} - \lambda|^{-1} \\ &\leq (\min_{\tilde{\lambda}} |\tilde{\lambda} - \lambda|)^{-1}, \end{aligned} \tag{5.10}$$

where we have used the sub-multiplicativity of the spectral norm and the fact that $\|\tilde{U}\| = \|\tilde{U}^t\| = 1$. Finally, using Eq. 5.10 in Eq. 5.9 leads to

$$\min_{\tilde{\lambda}} |\lambda - \tilde{\lambda}| \leq \frac{\|R\Sigma e_P(\mathbf{v})\|}{\|R\mathbf{v}\|} \leq \|\Sigma\| \frac{\|e_P(\mathbf{v})\|}{\|P\mathbf{v}\|},$$

since $\|P\mathbf{v}\| \leq \|R^t\|\|R\mathbf{v}\| = \|R\mathbf{v}\|$. We can express this result in a more compact form by introducing θ_λ, the angle between $\mathbf{v}$ and $P\mathbf{v}$. Assuming as usual $\|\mathbf{v}\| = 1$, we finally have

$$\min_{\tilde{\lambda}} |\lambda - \tilde{\lambda}| \leq \lambda_1 \tan\theta_\lambda, \tag{5.11}$$

where we have made the substitution $\|\Sigma\| = \lambda_1$. Eq. 5.11 provides an upper bound on the *maximum* eigenvalue error caused by the coarse graining operation. It asserts that at least one eigenvalue of the coarse-grained matrix $R\Sigma R^t$ lies at a distance $\lambda_1 \tan\theta_\lambda$ of λ. As expected, when $\theta_\lambda = 0$ (e.g. the projection is the identity or its range is $\mathbf{v}$) the equality $\lambda = \tilde{\lambda}$ holds, that is λ is perfectly preserved in $\tilde{\Sigma}$.

We now focus on the edge spectrum. When substituting λ for the extreme eigenvalues λ_1 and λ_N, the left-hand side of Eq. 5.11 is known exactly by the interlacing inequalities (5.7). Two new bounds that depend explicitly on R follow:

$$\tilde{\lambda}_1 \geq \lambda_1(1 - \tan\theta_1) \text{ and} \tag{5.12}$$

$$\tilde{\lambda}_{\tilde{N}} \leq \lambda_N + \lambda_1 \tan\theta_N. \tag{5.13}$$

From Eq. 5.12 and Eq. 5.13 one can readily bound the *coarse-grained* reverse condition number $\tilde{\kappa}^{-1} = \tilde{\lambda}_N / \tilde{\lambda}_1$ with respect to κ^{-1} and R:

$$\tilde{\kappa}^{-1} \leq \frac{\kappa^{-1} + \tan\theta_N}{1 - \tan\theta_1}, \tag{5.14}$$

which is tight in the sense that $\theta_N = \theta_1 = 0$ imply $\tilde{\kappa}^{-1} = \kappa^{-1}$. Notice that, when it is smaller than one, the right-hand side of Eq. 5.14 is a limiting factor to the improvement of the conditioning by the coarse graining. In order to avoid this limitation one should take θ_1 and θ_N large enough so as to have

$$\tan\theta_1 + \tan\theta_N \geq 1 - \kappa^{-1}, \tag{5.15}$$

which in practice might not be desirable nor even always possible.

Minimizing the eigenvalue error to preserve relevant information in the coarse-grained matrix

Eq. (5.11) is useful because it provides an upper bound on the *maximum* eigenvalue error caused by the coarse graining operation. This suggests that the preservation of $(\lambda, \mathbf{v})$ in $\widetilde{\Sigma}$ can be achieved by minimizing $\tan\theta_\lambda$, for example with the methods presented in [31] and in Appendix B.

Notice that in this study we restrict to non-negative entries in R in order to avoid having to deal with short positions; short-selling is of course possible in MVO but it makes the whole allocation process trickier. Furthermore, in absence of additional information, it is fair to say that all stocks belonging to the same sub-portfolio are to be treated equally (i.e. following a max. entropy argument), or in other words that all sub-portfolios should be equally-weighted.

Under these assumptions, and given the size of the coarse-grained universe $\widetilde{N}$, the problem becomes combinatoric and amounts to finding the partition of assets that minimizes the objective function (see Sec. B.5.3)

$$\|\mathbf{v} - P\mathbf{v}\|^2 = \sum_{\alpha=1}^{\widetilde{N}} \sum_{i\in\alpha} [v(i) - (P\mathbf{v})(i)]^2 = \sum_{\alpha=1}^{\widetilde{N}} \sum_{i\in\alpha} \left(v(i) - \frac{1}{|\alpha|} \sum_{j\in\alpha} v(j) \right)^2, \tag{5.16}$$

where $\sum_{i\in\alpha}$ stands for the sum over the assets in sub-portfolio α of cardinality $|\alpha|$, and $\sum_{\alpha=1}^{\widetilde{N}}$ denotes the sum over all the sub-portfolios in the partition. We have designed a dynamic programming algorithm to solve exactly this problem (Algorithm B.5.3.1). This algorithm finds the partition with size $\widetilde{N}$ minimizing Eq. (5.16) (in fact all the minimizing partitions with size 2 to $\widetilde{N}$) in time $O(\widetilde{N}N^2)$ and memory load $O(N^2)$, which is suitable to the financial application considered here.

In general, one may want to rely not only on the first eigenpair of Σ to build the coarse-grained portfolio but rather on, say, the top q ones. In that case, the problem becomes that of minimizing

$\sum_{k=1}^{q} \|\mathbf{v}_k - P\mathbf{v}_k\|^2$ where each term in the sum is similar to Eq. (5.16). Unfortunately this problem is known to be NP hard for $q \geq 2$, which means that in practice it can only be tackled via approximate methods. In our simulations, we have used the efficient k-means algorithm [3] to carry out the minimization (see Sec. B.5 and following for the use of approximate methods in this context).

5.5 Simulations and preliminary results

We show here how to build a coarse-grained portfolio with similar ex-post returns as the ones of the Markovitz's portfolio, but with an effective size that is half the dimension of the asset universe.

In this application H is the 2-week return matrix of 200 stocks quoted on the New York Stock Exchange in the period 1996-2008. The dimension of the coarse-grained matrix $\widetilde{N} = 100$ was found by using the method suggested in [65], which is similar as the one used to derive Eq. (3.21) in Sec. 3.19 (i.e. time-correlations are to replaced be cross-correlations).

We proceed as follows: we estimate the sample (unweighted) covariance matrix of returns over the period 1996-2007 and compute the optimal portfolio $\mathbf{x}^*$, solution of Eq. (5.3) with $\alpha = 0$ and $x_i \geq 0 \, \forall i$; we also add the diversification constraints $x_i \leq 0.1 \, \forall i$ and $\sum_i x_i \hat{r}_i \geq \bar{r}$. The out-of-sample returns of the MVO portfolio are compared with the ones of the coarse-grained portfolios $\mathbf{x}^*_{cg}$ obtained from Eq. (5.6) according to two different grouping strategies:

1. using a random grouping (i.e. the null model),
2. by minimizing $\sum_{k=1}^{5} \|\mathbf{v}_k - P\mathbf{v}_k\|^2$ with the spectral-based method of Section. 5.4, where the eigenvectors were the five largest ones in the correlation matrix of the residuals of the returns regressed on the first eigenvector of Σ (i.e. this was shown to improve the stability of the groups, see next section).

The whole procedure was repeated 10 times for 200 stocks chosen at random among a total of 1'000 securities, and results averaged over the subsamples to improve readability.

As shown in Fig. 5.1, the out-of-sample returns of $\mathbf{x}^*_{cg}$ and $\mathbf{x}^*$ are in remarkable agreement during the whole observation period. In contrast, the random grouping strategy leads to significantly more volatile ex-post returns. The excess volatility observed for the null model comes from the

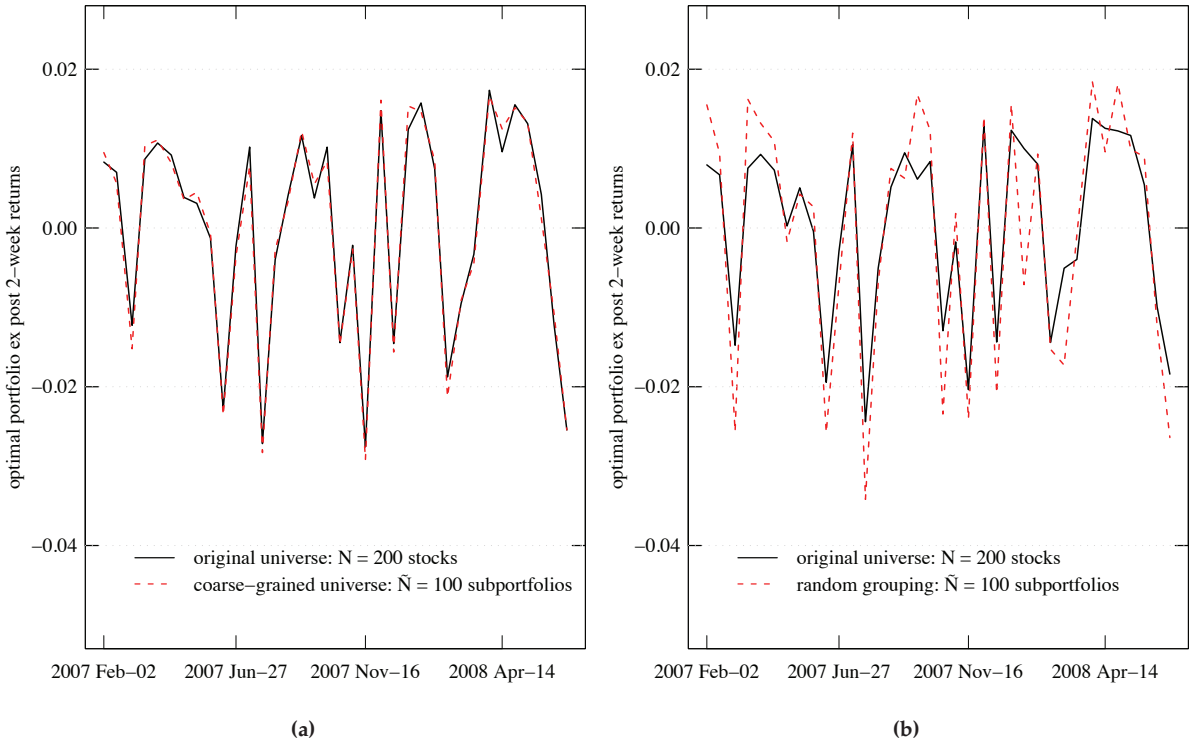


Figure 5.1: *Comparison of the ex-post two-week returns of the MVO portfolio with the ones of coarse-grained portfolios obtained from (a) the grouping of the stocks according to the spectral-based method of Section 5.4 with* $\widetilde{N} = 100$*, and (b) a random grouping of the stocks into 100 sub-portfolios. Results are averaged over 10 subsamples of 200 stocks drawn at random in a universe of 1'000 US stocks.*

loss of information induced by randomly grouping the stocks, which can be better quantified by comparing the annualized information ratios (IR) over the observation period. The annualized IR is defined as $\sqrt{24}\,\hat{\mathrm{E}}(r - r_M)/\hat{\sigma}(r - r_M)$, where r stands for the portfolio two-week returns and r_M denotes the returns of the Standard & Poor's 500 index. To fix ideas the threshold value reached by the top-quartile asset managers is 0.5 [45].

The information ratios averaged over the 10 subsamples are

$$\begin{array}{ccc} \mathbf{x}^* & \mathbf{x}^*_{cg} & \mathbf{x}^*_{rdm} \\ 0.57(0.2) & 0.58(0.2) & 0.41(0.2) \end{array}',$$

where the standard deviation is indicated in parentheses. The relevant information is preserved by a spectral-based clustering whereas it is dramatically harmed by blindly grouping the stocks which here leads to an average IR dropping below 0.5.

5.6 Discussion and conclusion

The close resemblance of the ex-post returns of $\mathbf{x}^*$ and $\mathbf{x}^*_{cg}$ is a direct consequence of the spectral-based method adopted to group the stocks: the most significant eigenvectors of the return correlation matrix contain important information about the correlation structure between the stock returns [69], which can be exploited to efficiently reduce the sample covariance matrix in the sense of the Mean-Variance Optimization. The underlying mechanism can be understood from the following simple heuristic: stocks with highly correlated returns are likely to behave *on average* similarly in the next period if the correlation pattern persists. Grouping those stocks in the same sub-portfolio is reasonable from a portfolio allocation perspective, since it is the grouping that yields minimum loss of diversification. The key issue then turns out to be the reliable identification of persistent correlation patterns among returns, which can be addressed via a number of techniques as the spectral one used here. Indeed, the largest eigenvectors of the return correlation matrix are known to give a rough picture of important and stable correlation structures, e.g. industrial sectors [5] [69]. This effect is enhanced if one considers instead the correlation matrix C_{res} of the residuals $\varepsilon_i(t)$ of the regression

$$r_i(t) = a_i + b_i r_M(t) + \varepsilon_i(t),$$

where r_M is the return on the largest eigenvector of the correlation matrix (regarded as a portfolio) [71]. Beside the eigenvector-based clustering, a number of methods have been developed to find "natural" groups of stocks from the return cross-correlation matrix, e.g. correlation screening filters out less significant correlations ultimately leading to trees [61] or planar graphs [80] of correlations; other approaches include the Super-Paramagnetic Clustering of [11] and more recently the Maximum Likelihood Approach of [41]. Since the spectral coarse graining accepts virtually any grouping, further tests should tell us which of these clusterings performs best when applied to the coarse graining of MVO.

The main benefit of reducing the covariance matrix of returns in portfolio allocation is to improve the performance of the optimizer by reducing its workload. This may be particularly useful for asset managers dealing with a very large number of portfolios or managing very large portfolios (several thousands of assets or more). In this case, it might be desirable to use the same grouping

[5] The time-stability of the main eigenspace of covariance estimators has been seriously questioned in [86]. Nevertheless, it seems clear that important correlation patterns embedded in this space show enough persistence to be efficiently exploited, even if they may appear "deformed" in an intricate way over time.

over a certain period of time in order to improve the reallocation process and avoid too large portfolio turnovers.

A

Proofs of Chapter 4

A.1 Proof of Eq. 4.9

This proof was done by Olivier Lévèque.

We show that under the assumptions made on w, namely that it is continuous, decreasing and such that $w \in L^2(\mathbb{R}_+)$, Eq. (4.9) admits a unique solution which is the Stieltjes transform of a distribution. First, recall the definition of g:

$$g(z) = \int_{\mathbb{R}} \frac{1}{x - z}\, d\mu(x), \ a.s. \quad \forall z \in \mathbb{C}\backslash\mathbb{R}.$$

as we know from the start that μ is supported on the positive axis (W being positive semi-definite). This implies that if $\operatorname{Re} z \leq 0$, then

$$\operatorname{Re} g(z) = \int_0^\infty \frac{x - \operatorname{Re} z}{(x - \operatorname{Re} z)^2 + (\operatorname{Im} z)^2}\, d\mu(x) \geq 0.$$

On the other hand, we know that if $\operatorname{Im} z \geq 0$, then $\operatorname{Im} g(z) \geq 0$ also. So it is sufficient to prove that if $z \in \mathbb{C}$ is such that $\operatorname{Re} z \leq 0$ and $\operatorname{Im} z > A = \|w\|_2 + \varepsilon$ (by assumption $A < \infty$), then the map $F : \mathbb{C} \to \mathbb{C}$ defined as

$$F(g) = -\left(z - \int_0^\infty \frac{w(t)}{1 + w(t)\, g}\, dt\right)^{-1}, \quad g \in \mathbb{C},$$

is a contraction on $\mathbb{C}_{++} = \{g \in \mathbb{C} : \text{Re } g \geq 0, \text{Im } g \geq 0\}$, that is, there exists $K < 1$ such that

$$|F(g_2) - F(g_1)| \leq K\,|g_2 - g_1|, \quad \forall g_1, g_2 \in \mathbb{C}_{++}. \tag{A.1}$$

It is easily checked that given the assumptions made on z, F indeed maps $\mathbb{C}_{++}$ into $\mathbb{C}_{++}$. Let us then check inequality (A.1):

$$|F(g_2) - F(g_1)| = \left|\frac{1}{G(g_1)} - \frac{1}{G(g_2)}\right| = \frac{|G(g_2) - G(g_1)|}{|G(g_1)|\,|G(g_2)|},$$

where

$$G(g) = z - \int_0^\infty \frac{w(t)}{1 + w(t)\,g}\,dt.$$

As Im $g \geq 0$ by assumption, $|\text{Im } G(g)| \geq |\text{Im } z| \geq A$. We also have

$$\begin{aligned} |G(g_2) - G(g_1)| &\leq \int_0^\infty w(t) \left|\frac{1}{1 + w(t)\,g_2} - \frac{1}{1 + w(t)\,g_1}\right| dt \\ &\leq \int_0^\infty w(t)^2 \frac{|g_2 - g_1|}{|1 + w(t)\,g_1|\,|1 + w(t)\,g_2|}\,dt \leq \|w\|_2^2\,|g_2 - g_1|, \end{aligned}$$

as Re g_1, Re $g_2 \geq 0$. So we obtain finally that

$$|F(g_2) - F(g_1)| \leq \frac{\|w\|_2^2}{A^2}\,|g_2 - g_1|,$$

proving the above claim, since $\frac{\|w\|_2^2}{A^2} < 1$. This implies that Eq. (4.9) admits a unique solution $g(z)$ when $z \in \mathbb{C}$ is such that Re $z \leq 0$ and Im $z \geq A$. As we also know that $g(z)$ is the Stieltjes transform of a distribution, it is analytic on $\mathbb{C}\backslash\mathbb{R}$, so its value on the upper half plane is uniquely determined by its value on $\{z \in \mathbb{C} : \text{Re } z \leq 0, \text{Im } z \geq A\}$ (and a similar argument applies to the lower half plane). This finally proves that Eq. (4.9) admits a unique solution for all $z \in \mathbb{C}\backslash\mathbb{R}$.

A.2 Spectrum of PREC i.i.d. correlation matrices for $\lim_{N,T} \frac{T}{N} = c_0 < \infty$

When $c_0 < \infty$, the definite integral in Eq. (4.8) can be expressed in terms of the hypergeometric function [44]:

$$\int_0^{c_0} \frac{w(t)}{1 + w(t)g}dt = \frac{c_0 K_0}{1 + K_0 g}\,{}_2F_1\left(1, \frac{1}{\gamma}; 1 + \frac{1}{\gamma}; -\frac{\left(\frac{c_0}{c}\right)^\gamma}{1 + K_0 g}\right), \tag{A.2}$$

where the normalizing constant $K_0 = \left(c_0\, {}_2F_1\left(1, \frac{1}{\gamma}; 1+\frac{1}{\gamma}; -\left(\frac{c_0}{c}\right)^{\gamma}\right)\right)^{-1}$. Although this is sufficient for computing the spectrum with Algorithm 4.1, one can gain some insight into the solution by expanding the hypergeometric function, and even obtain a closed form in some cases of interest.

In series representation ${}_2F_1(a,b;c;z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}$, where $(x)_k$ stands for $(x)_k = \Gamma(x+k)/\Gamma(x) = x(x+1)\ldots(x+k-1)$. Using this expansion in Eq. (A.2) and simplifying leads to

$$\begin{aligned} {}_2F_1\left(1, \frac{1}{\gamma}; 1+\frac{1}{\gamma}; -\frac{\left(\frac{c_0}{c}\right)^{\gamma}}{1+K_0 g}\right) &= \sum_{k=0}^{\infty} \frac{(1)_k \left(\frac{1}{\gamma}\right)_k}{\left(1+\frac{1}{\gamma}\right)_k} \frac{(-1)^k}{k!} \left(\frac{\left(\frac{c_0}{c}\right)^{\gamma}}{1+K_0 g}\right)^k \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{1+\gamma k} \left(\frac{\left(\frac{c_0}{c}\right)^{\gamma}}{1+K_0 g}\right)^k, \end{aligned} \tag{A.3}$$

which is convergent since Eq. (A.2) admits a unique solution by Theorem 1. Consider the case $\gamma = 1/p$, $p \in \mathbb{N}$; Eq. (A.3) can be written

$$\begin{aligned} {}_2F_1\left(1, p; 1+p; -\frac{\left(\frac{c_0}{c}\right)^{1/p}}{1+K_0 g}\right) &= p \sum_{k=0}^{\infty} \frac{(-1)^k}{p+k} \left(\frac{\left(\frac{c_0}{c}\right)^{1/p}}{1+K_0 g}\right)^k \\ &= p \sum_{m=p}^{\infty} \frac{(-1)^{m-p}}{m} \left(\frac{\left(\frac{c_0}{c}\right)^{1/p}}{1+K_0 g}\right)^{m-p} \\ &= p(-1)^{p+1} \frac{c}{c_0} (1+K_0 g)^p \sum_{m=p}^{\infty} \frac{(-1)^{m+1}}{m} \left(\frac{\left(\frac{c_0}{c}\right)^{1/p}}{1+K_0 g}\right)^m, \end{aligned}$$

where we have made the substitution $m = k+p$. Then, using $\log(1+\omega) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \omega^k$, $|\omega| < 1$, we finally obtain

$${}_2F_1\left(1, p; 1+p; -Z_p\right) = \begin{cases} \frac{1}{Z_1} \log\left(1+Z_1\right) & p = 1 \\ \\ \frac{p\,(-1)^{p+1}}{Z_p^p} \left[\sum_{j=1}^{p-1} \frac{(-1)^j}{j} Z_p^j + \log\left(1+Z_p\right)\right] & p > 1 \end{cases} \tag{A.4}$$

where $Z_p = \left(\frac{c_0}{c}\right)^{1/p} / \left(1+K_0 g\right)$. The simplest form is obtained for $\gamma = p = 1$, which, from Eq. (A.2) and Eq. (4.8), leads to the following equation for the Stieltjes transform of the asymp-

totic spectral distribution:

$$1 + zg = cK_0 g \log\left(1 + \frac{c_0}{c\,(1 + K_0 g)}\right).$$

The next expression corresponds to $\gamma = 0.5$ and is obtained by letting $p = 2$ in Eq. (A.4):

$$1 + zg = 2cK_0 g\left(\sqrt{c_0/c} - (1 + K_0 g)\log\left(1 + \frac{\sqrt{c_0/c}}{1 + K_0 g}\right)\right).$$

B

The Spectral Coarse Graining of matrices: theory and examples from complex network theory

B.1 Introduction

In order to provide a size-reduction scheme that is not only effective at simplifying the system's complexity, but that also preserves its behavior, we have introduced a spectral-based method that we have named Spectral Coarse Graining (SCG) [39, 40]. SCG allows to go beyond the classical clustering techniques, by not only identifying groups, but also assembling them in a coarse-grained arrangement while protecting some targeted features of the original system; typically, these features can be readily expressed in terms of the spectral properties of the system's interaction matrix.

In this work our first aim is to frame SCG on robust mathematical foundations by showing that it can be cast as a projection, which, when duly chosen, causes the least possible perturbation on some prescribed eigenpairs of the interaction matrix.

Our second goal is to present different algorithms to carry out the SCG of a matrix. Each specific implementation has both strengths and weaknesses that are going to be addressed. Finally, some examples drawn from graph and network theory will be provided to allow for a better assessment of the techniques in practical applications.

B.1.1 Overview

Our text is structured as follows. We first draw up a formal frame for the spectral coarse graining of matrices in §B.2 and §B.3. The effects of spectral coarse graining on the eigenpairs of a matrix are analyzed in §B.4 borrowing techniques from matrix perturbation theory. Then, in a slightly less formal style, we describe our methods and algorithms in §B.5, followed by some applications of the spectral coarse graining to graph theory in §B.6. Finally, we present our conclusions and sketch some possible developments of Spectral Coarse Graining in §B.7.

B.1.2 Notations and Style

The conjugate, transpose, and conjugate transpose of a matrix M are denoted $\overline{M}$, M^t, respectively M^*. The abbreviation $\mathrm{sp}(M)$ stands for the spectrum of $M \in \mathbf{C}^{n\times n}$. The spectral norm, or 2-norm, of M is denoted by $\|M\|$. The couple (λ, v), where $Mv = \lambda v$, is called (right) eigenpair of M; it is called *zero (right) eigenpair* if $\lambda = 0$. Eigenvalues are assumed sorted as $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ and eigenvectors are normalized except where noted. We state our results for right eigenvectors only since the translation to left eigenvectors is straightforward by transposing M; therefore, we omit the qualifier "right" to designate a right eigenpair/eigenvector.

B.2 A Note on Projectors in $\mathbf{C}^{n\times n}$

We introduce this section by recalling some fundamentals about projectors in $\mathbf{C}^{n\times n}$.

By definition, a matrix $P \in \mathbf{C}^{n\times n}$ is a projection matrix or simply a projector if $P^2 = P$. Besides, if $P = P^*$ the projector is said to be orthogonal (with respect to the canonical Hermitian scalar product).

The *range* and the *null space* of P are the sets of y such that $y = Pz$ for $z \in \mathbf{C}^n$, respectively the sets of $z \in \mathbf{C}^n$ such that $Pz = 0$. By complementarity, we have the decomposition $z = Pz + (I_n - P)z$ for any $z \in \mathbf{C}^n$, where Pz is in the range and $(I_n - P)z$ is in the null space of P.

From $P(P - I_n) = 0$, the minimal polynomial of P factors into distinct roots and thus a projector is always diagonalizable with eigenvalues 0 and 1. There is equality between the number of one eigenvalues, the dimension of the range and the rank of P.

The next result provides a useful decomposition of a projector that will be used throughout the text.

Theorem 3. *A matrix $P \in \mathbf{C}^{n\times n}$ is a projector of rank k if and only if there exists two matrices $L, R \in \mathbf{C}^{k\times n}$ such that $P = R^*L$ and $LR^* = I_k$. Furthermore, the rows of R span the range of the projector and the rows of L span the orthogonal complement of its null space.*

Proof. The first statement is clearly necessary since $P^2 = R^*(LR^*)L = P$, and $\text{rank}P = \text{rank}(R^*L) = \text{rank}R = k$ by the properties of the rank and the fact that L is row-full rank. To show it is sufficient recall that, up to a row-column permutation in V, the eigen-decomposition of P reads

$$P = V \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} V^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} E & F \\ G & H \end{pmatrix} = \underbrace{\begin{pmatrix} A \\ C \end{pmatrix}}_{R^*} \underbrace{\begin{pmatrix} E & F \end{pmatrix}}_{L},$$

with $LR^* = EA + FC = I_k$ by definition of R and L. For the second statement, P acting on R^* gives $PR^* = R^*LR^* = R^*$ so that the rows of $\overline{R}$ belong to the range of P. As they are k independent vectors they actually span the latter. Finally, for all $z \in \mathbf{C}^n$ we have $Pz = R^*Lz = 0 \Leftrightarrow Lz = 0$ (by left-multiplication with L), which shows that the null space of P is orthogonal to the rows of $\overline{L}$ (with respect to the canonical Hermitian scalar product). □

When P is Hermitian, the spectral theorem yields $V^{-1} = V^*$ in the proof of Theorem 3. This shows the following result.

Corollary 4. *A matrix $P \in \mathbf{C}^{n\times n}$ is an orthogonal projector of rank k if and only if there exists a matrix $R \in \mathbf{C}^{k\times n}$ such that $P = R^*R$ and $RR^* = I_k$. The rows of R span the range of the projection.*

Even though there is an infinite number of RL-decompositions one can associate to a projector, when defining the matrices R and L through an eigen matrix of P, as in the proof of Theorem 3, there is no need to know beforehand the range and the null space of the projection—whose bases are provided by the decomposition[1].

The next and last result of this section is useful to assess the *non-orthogonality* of a projector.

Theorem 5. *For any projector $P \in \mathbf{C}^{n\times n}$ of rank $k > 0$, $\|P\| \geq 1$ with equality if and only if P is orthogonal.*

[1]Such a decomposition provides particular bases for the range and the null space of P that are orthogonal between each other (i.e. by the condition $LR^* = I_k$).

Proof. The proof can be found in different places, e.g. [51]. □

Remark 6. For any sub-multiplicative matrix norm $\|\cdot\|$ and non-zero projector $P \in \mathbf{C}^{n\times n}$, $\|P\| = \|P^2\| \leq \|P\|^2 \Rightarrow \|P\| \geq 1$. Hence the spectral norm in Theorem 5 is to some extent an optimum choice. For example, it is not difficult to show that $\|P\|_F \geq \operatorname{rank} P$, where $\|\cdot\|_F$ is the Frobenius norm.

B.3 Exact Coarse Graining

The matrices R and L of a projection can be used to shrink a matrix M down to a new matrix LMR^*. The purpose of this section is to introduce the formalism used throughout the text to deal with this *coarse graining transformation*, as well as to link it with the projection PMP.

Definition 7. The linear map $L \cdot R^* : \mathbf{C}^{n\times n} \to \mathbf{C}^{\tilde{n}\times\tilde{n}}$ is a coarse graining transformation if (1) $\tilde{n} \leq n$ and (2) $L, R \in \mathbf{C}^{\tilde{n}\times n}$ are such that $LR^* = I_{\tilde{n}}$. For $M \in \mathbf{C}^{n\times n}$ the matrix $\widetilde{M} \equiv LMR^*$ is a *coarse graining* (CG) of M.

The matrices L and R are referred to as the *semi-projectors* and $P = R^*L$ as the projector induced by the coarse graining $L \cdot R^*$. If the semi-projectors are equal, we call $L = R$ the *semi-orthogonal projector* of the coarse graining. If $P = P^*$ the coarse-graining is said to be orthogonal. Besides, if M is Hermitian (symmetric) and $L = R$, the coarse-graining is called Hermitian (symmetric), since $\widetilde{M}$ is also Hermitian (symmetric). The matrix $\widetilde{M}$ is sometimes called the *coarse-grained matrix* and its eigenvalues the *coarse-grained eigenvalues*.

As shown by the following proposition, there is a simple one-to-one mapping between the non-zero eigenpairs of LMR^* and PMP.

Proposition 8. *Let $M \in \mathbf{C}^{n\times n}$ and let $L, R \in \mathbf{C}^{\tilde{n}\times n}$ be such that $\tilde{n} \leq n$, $P = R^*L$ and $LR^* = I_{\tilde{n}}$. For every eigenpair $(\tilde{\lambda}, \tilde{v})$ of LMR^*, $(\tilde{\lambda}, R^*\tilde{v})$ is an eigenpair of PMP. Furthermore, for every non-zero eigenpair (μ, w) of PMP (μ, Lw) is an eigenpair of LMR^*.*

Proof. By definition of R and L, we have $LMR^*\tilde{v} = \tilde{\lambda}\tilde{v} \Leftrightarrow LMP(R^*\tilde{v}) = \tilde{\lambda}\tilde{v} \Leftrightarrow PMP(R^*\tilde{v}) = \tilde{\lambda}(R^*\tilde{v})$. Since $R^*\tilde{v} \neq 0$ ($R^*\tilde{v} = 0 \Rightarrow \tilde{v} = 0$, which is impossible), $(\tilde{\lambda}, R^*\tilde{v})$ is an eigenpair of PMP. On the other hand, suppose there exists (μ, w), an eigenpair of PMP, such that $\mu \neq 0$. Left-multiplying $PMPw = \mu w$ by L yields $LMR^*(Lw) = \mu(Lw)$. We note that $Lw \neq 0$ otherwise

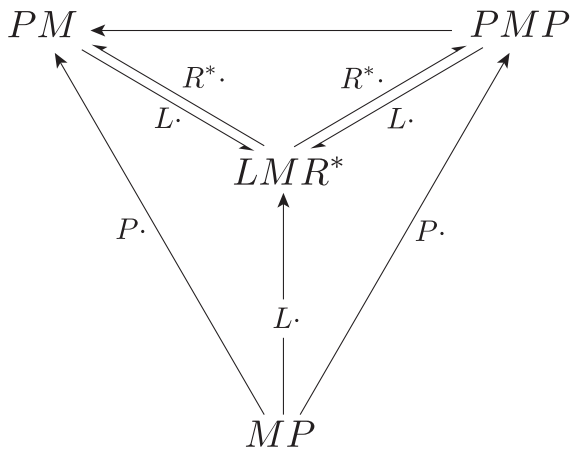


Figure B.1: *In this diagram $M, P \in \mathbf{C}^{n\times n}$ and $P = R^*L$ is a projector such that $LR^* = I_{\tilde{n}}$, with $\tilde{n} \leq n$. It is easy to show that $sp(PM) = sp(PMP) = sp(MP) = sp(LMR^*) \cup \{0\}$. A directed edge goes from A to B if there is a mapping from the non-zero eigenpairs of A to the non-zero eigenpairs of B. Furthermore, left-multiplying an eigenvector of A that is associated with a non-zero eigenvalue by the label of an edge gives the corresponding eigenvector of B. For instance, if (λ, v) is an eigenpair of MP and $\lambda \neq 0$ then (λ, Pv) is an eigenpair of both PM and PMP. Proposition 8 establishes the one-to-one mapping between the non-zero eigenpairs of LMR^* and PMP.*

we would have $Pw = 0 \Rightarrow \mu w = 0 \Rightarrow w = 0$, which is impossible. Thus (μ, Lw) is an eigenpair of LMR^*. □

Remark 9. As depicted in Figure B.1, similar mappings exist between the eigenpairs of LMR^* and those of some combinations of products between M and P. We stress that for a given P all the relations hold for any choice of L and R as long as $P = R^*L$ and $LR^* = I_{\tilde{n}}$.

Consider the following eigenvalue equations for M and $\widetilde{M}$ serving to introduce the next definition:

$$\begin{aligned} Mv_i = \lambda_i v_i \quad &\Rightarrow \quad (LM)v_i = \lambda_i L v_i \text{ and} \\ \widetilde{M}\tilde{v}_\alpha = \tilde{\lambda}_\alpha \tilde{v}_\alpha \quad &\Leftrightarrow \quad (LM)R^*\tilde{v}_\alpha = \tilde{\lambda}_\alpha \tilde{v}_\alpha. \end{aligned}$$

Definition 10. Let $\widetilde{M} \in \mathbf{C}^{\tilde{n}\times\tilde{n}}$ be a coarse graining of M with projector P and semi-projectors L, R. Let also (λ_i, v_i) and $(\tilde{\lambda}_\alpha, \tilde{v}_\alpha)$ be eigenpairs of M, respectively of $\widetilde{M}$. $|\lambda_i - \tilde{\lambda}_\alpha|$ is the absolute eigenvalue shift between both eigenpairs. We also define the two eigenvector shifts $e_R(v_i, \tilde{v}_\alpha) \equiv v_i - R^*\tilde{v}_\alpha$, $e_L(v_i, \tilde{v}_\alpha) \equiv Lv_i - \tilde{v}_\alpha$, and the vector $e_P(v_i) \equiv v_i - Pv_i$.

Together, $e_R(v_i, \tilde{v}_\alpha)$, $e_L(v_i, \tilde{v}_\alpha)$, and $|\lambda_i - \tilde{\lambda}_\alpha|$ form the *eigenpair shifts*. These are used to estimate the accuracy of a coarse graining with respect to the eigenpair (λ_i, v_i)—the smaller the eigenpair

shifts the more *accurate* the coarse graining for (λ_i, v_i). An important particular case, called *exact coarse graining*, is when the three eigenpair shifts are zero for some α.

Definition 11. The matrix $\widetilde{M}$ is an exact coarse graining of M for the eigenpair (λ, v) if there is an eigenpair $(\tilde{\lambda}, \tilde{v})$ of $\widetilde{M}$ such that $\tilde{\lambda} = \lambda$, $e_R(v, \tilde{v}) = 0$ and $e_L(v, \tilde{v}) = 0$.[2]

Remark 12. The trivial exact coarse graining of M for (λ, v) is obtained by setting $R = L = v^*$; this yields $LR^* = 1$ and $LMR^* = \lambda$.

An exact coarse graining for (λ, v) means that there is a one-to-one mapping between (λ, v) and an eigenpair of $\widetilde{M}$. As a consequence, no information about (λ, v) is lost in the transformation, and we say that (λ, v) is *exactly preserved* in $\widetilde{M}$.

It is sometimes convenient to define the exact coarse graining in terms of v and P instead of the unknown eigenpair $(\tilde{\lambda}, \tilde{v})$.

Proposition 13. *$\widetilde{M}$ is an exact coarse graining of M for the eigenpair (λ, v) if and only if v* is in the range of P (i.e. $e_P(v) = 0$).

Proof. Substituting $\tilde{v} = Lv$ into $R^*\tilde{v} = v$ shows that $Pv = v$. Conversely, left-multiplying by LM the equation $e_P(v) = v - Pv$ and rearranging yields

$$LMPv = LMv - LMe_P(v) \Leftrightarrow \widetilde{M}Lv = \lambda Lv - LMe_P(v).$$

If $Pv = v$ then $Lv \neq 0$ and there is an eigenpair $(\tilde{\lambda}, \tilde{v})$ of $\widetilde{M}$ such that $\tilde{\lambda} = \lambda$ and $\tilde{v} = Lv$; this in turn implies $R^*\tilde{v} = v$. □

B.4 Approximate Coarse Graining

Whenever $\|e_P(v)\| > 0$ the coarse graining for (λ, v) is said to be approximate. In this case, it is interesting to have an estimate of the distance between the eigenpairs of M and $\widetilde{M}$. We address this question here by means of techniques developed in matrix perturbation theory [50, 49, 43].

The idea is first to bound the smallest distance between λ and an eigenvalue of the matrix PMP in terms of the difference between Pv and an actual eigenvector of PMP. Then, the same bound

[2] Notice that for $\lambda \neq 0$, $\tilde{\lambda} = \lambda$ and $e_R(v, \tilde{v}) = 0$ readily imply $e_L(v, \tilde{v}) = 0$

holds for the smallest distance between λ and an eigenvalue of LMR^* through the one-to-one correspondence between the non-zero eigenpairs of PMP and LMR^* (Proposition 8).

Theorem 14. *Let $M, P \in \mathbf{C}^{n\times n}$ with P a ñ-rank projector. Let also $Q^*(PMP)Q = D + N$ be a Schur decomposition of PMP, where D is diagonal and N strictly upper triangular. For any eigenpair (λ, v) of M such that $Pv \neq 0$,*

$$\min_{\tilde{\lambda}\in sp(PMP)} |\lambda - \tilde{\lambda}| \leq \max(\beta, \beta^{1/p}),$$

where p is the smallest integer such that $N^p = 0$, and β is given by

$$\beta = \frac{\|PMe_P(v)\|}{\|Pv\|} \sum_{k=0}^{p-1} \|N\|^k.$$

Proof. This theorem is the *a posteriori* version of Theorem 7.2.3 in [43] so its demonstration follows similar lines, except for the first part. The theorem clearly holds if $\lambda \in \mathrm{sp}(PMP)$. If $\lambda \notin \mathrm{sp}(PMP)$, consider the identity

$$PMP(Pv) = \lambda(Pv) - PMe_P(v). \tag{B.1}$$

As $Pv \neq 0$, Equation B.1 allows us to write $1 \leq \|(\lambda I_n - PMP)^{-1}\| \|PMe_P(v)\| / \|Pv\|$. Using the Schur decomposition of PMP and the invariance under unitary transformations of the spectral norm, we have $\|(\lambda I_n - PMP)^{-1}\| = \|(\lambda I_n - D - N)^{-1}\|$. The proof is completed by the same technique as in the proof of Theorem 7.2.3 of [43, p.321].

□

The relevant factors influencing the accuracy of a coarse graining can easily be uncovered by breaking down the upper bound of Theorem 14. Indeed, let β be defined as in Theorem 14, then

$$\beta \leq \|M\| \Lambda(PMP) \|P\| \frac{\|e_P(v)\|}{\|Pv\|}, \tag{B.2}$$

where $\Lambda(PMP) = \sum_{k=0}^{p-1} \|N\|^k \geq 1$ with equality if and only if PMP is a normal matrix (by the spectral theorem), and by Theorem 5 $\|P\| \geq 1$ with equality if and only if P is orthogonal. The quantity $\Lambda(PMP)$ is an estimate of PMP's departure from normality and $\|P\|$ of P's departure from "orthogonality".

In the important coarse graining of Hermitian matrices with orthogonal projectors, Theorem 14

and Equation B.2 become notably simpler.

Corollary 15. *Let $M \in \mathbf{C}^{n\times n}$ be a Hermitian matrix and $P \in \mathbf{C}^{n\times n}$ be an* orthogonal *$\tilde{n}$-rank projector. Then*

$$\min_{\tilde{\lambda}\in sp(PMP)} |\lambda - \tilde{\lambda}| \leq \max_{\mu\in sp(M)} |\mu| \frac{\|e_P(v)\|}{\|Pv\|} = \max_{\mu\in sp(M)} |\mu| \|e_P(v)\| \left(1 + O(\|e_P(v)\|^2)\right).$$

Furthermore, if M is non-singular,

$$\min_{\tilde{\lambda}\in sp(PMP)} \frac{|\lambda - \tilde{\lambda}|}{|\lambda|} \leq \kappa(M) \frac{\|e_P(v)\|}{\|Pv\|} = \kappa(M) \|e_P(v)\| \left(1 + O(\|e_P(v)\|^2)\right),$$

where $\kappa(M) \equiv \|M\|\|M^{-1}\| \geq 1$ is the condition number of M expressed in the spectral norm.

Proof. Assuming M and P Hermitian implies $\max(\beta, \beta^{1/p}) = \beta$ in Theorem 14 (since $p = 1$), and $1 = \Lambda(PMP) = \|P\|$ in Equation B.2. We also notice that the decomposition $v = Pv + e_P(v)$ yields $1 = \|v\|^2 = \|Pv\|^2 + \|e_P(v)\|^2$ so that $\|Pv\|^{-1} = (1 - \|e_P(v)\|^2)^{-1/2} = 1 + O(\|e_P(v)\|^2)$. Finally, the relative bound follows since for M a non-singular normal matrix $\kappa(M) = \max_{\mu\in \mathrm{sp}(M)} |\mu| (\min_{\mu\in \mathrm{sp}(M)} |\mu|)^{-1}$ when the spectral norm is used [49]. □

The following result provides a lower bound on the eigenvector shifts in terms of $\|e_P(v)\|$.

Proposition 16. *Let $\widetilde{M} \in \mathbf{C}^{\tilde{n}\times\tilde{n}}$ be a coarse graining of $M \in \mathbf{C}^{n\times n}$ with projector P and semi-projectors L, R. Let also (λ_i, v_i) and $(\tilde{\lambda}_\alpha, \tilde{v}_\alpha)$ be eigenpairs of M and $\widetilde{M}$ respectively. The following relations hold for any $i \in \{1, \ldots, n\}$ and any $\alpha \in \{1, \ldots, \tilde{n}\}$: (1) $e_L(v_i, \tilde{v}_\alpha) = L e_R(v_i, \tilde{v}_\alpha)$ and (2) $e_P(v_i) = (I_n - P) e_R(v_i, \tilde{v}_\alpha)$. As a consequence, when P is orthogonal,*

$$\|e_P(v_i)\| \leq \|e_R(v_i, \tilde{v}_\alpha)\| \leq 2.$$

Proof. Relation (1) is immediate. Relation (2) can be derived as follows:

$$\begin{aligned} e_P(v_i) &= v_i - R^* L v_i + R^* \tilde{v}_\alpha - R^* \tilde{v}_\alpha \\ &= (v_i - R^* \tilde{v}_\alpha) - R^* (L v_i - \tilde{v}_\alpha) \\ &= e_R(v_i, \tilde{v}_\alpha) - R^* e_L(v_i, \tilde{v}_\alpha) \\ &= (I_n - P) e_R(v_i, \tilde{v}_\alpha). \end{aligned}$$

The bound is a straightforward consequence of the definitions and of Theorem 5. □

Proposition 16 shows that $\|e_R(v, \tilde{v})\|$ cannot be decreased faster than $\|e_P(v)\|$ for any $\tilde{v}$; this is a useful fact to have in mind when trying to minimize the eigenpairs shifts (§B.5). We add that there is no simple relation between $\|e_P(v_i)\|$ and $\|e_L(v_i, \tilde{v}_\alpha)\|$ as it is easy to find examples for either $\|e_P(v_i)\| \lesseqgtr \|e_L(v_i, \tilde{v}_\alpha)\|$.

We conclude this section by mentioning an important theorem that can help localize the coarse-grained eigenvalues about the original spectrum. This result is cited as the *Poincaré separation theorem* in [49].

Theorem 17. *Let $\widetilde{M} = LMR^* \in \mathbf{C}^{\tilde{n}\times\tilde{n}}$ be a coarse graining of M such that (1) M is Hermitian and (2) $P = R^*L$ is orthogonal. Then, the eigenvalues of $\widetilde{M}$ interlace the eigenvalues of M in the following way*

$$\lambda_{n-\tilde{n}+i} \leq \tilde{\lambda}_i \leq \lambda_i.$$

Proof. The proof given in [49, p.190] holds for $L = R$. The generalization to an arbitrary RL-decomposition of P is straightforward since $\mathrm{sp}(LMR^*) = \mathrm{sp}(\widehat{L}M\widehat{R}^*)$ as long as $P = \widehat{R}^*\widehat{L}$ and $\widehat{L}\widehat{R}^* = I_{\tilde{n}}$ (Proposition 8 and Remark 9). □

In [39, Table 1] one can observe the left-positioning of the eigenvalues of the coarse-grained matrix about the original spectrum. Notice that when $\tilde{n} = n - 1$ the rank of the eigenvalues is automatically preserved since in that event

$$\lambda_n \leq \tilde{\lambda}_{\tilde{n}} \leq \lambda_{n-1} \leq \cdots \leq \lambda_2 \leq \tilde{\lambda}_1 \leq \lambda_1.$$

The presence of $\|P\|$ in our upper bounds suggests that orthogonal projectors are less harmful to eigenvalues in coarse graining. Other relevant factors with potential effects on the accuracy of a CG are the departure from normality of PMP (closely related with that of M) and the conditioning of M. In particular, when P is orthogonal and M is Hermitian, the presence of $\kappa(M)$ in Theorem 15 suggests that matrices amenable to numerical computation (i.e. with low $\kappa(M)$) should be amenable to coarse graining as well.

As the upper bound of Theorem 15 suggests the accuracy of a coarse graining can be independent of the magnitude of the original eigenvalue (since the upper bound on the minimum relative eigenvalue shift can cancel independently of $|\lambda|$); numerous simulations validate this

observation (see also the discussion of Figure B.3). Finally, in the coarse graining of Hermitian matrices with orthogonal projector the coarse-grained eigenvalues interlace the original spectrum as described by the Poincaré separation theorem.

B.5 Optimizing Coarse Graining under Constraints

When coarse graining real-world systems—e.g. oscillator networks while retaining their synchronization properties [40], or large graphs while preserving the mean first passage time of random walks on them [39]—one aims at preserving an arbitrary number of M's eigenpairs given some problem-specific constraints on the transformation. In this section, we define a constrained minimization problem and solve it taking into account the material previously introduced.

We have seen in Proposition 13 that an eigenpair (λ, v) of M is *exactly* preserved in $\widetilde{M} = LMR^*$ if and only if $\|e_P(v)\| = 0$. This yields the following definition of a generic *coarse graining problem*.

Problem 18. Given $M \in \mathbf{C}^{n\times n}$ and (λ, v) an eigenpair of M to be preserved by the coarse graining, the problem is to find a projector $\widehat{P}$ that solves

$$\min_{P\in\mathfrak{C}} \|e_P(v)\|$$

where $\mathfrak{C}$ is a set of projectors in $\mathbf{C}^{n\times n}$ described by some *ad hoc* constraints $c_1, \dots c_r$ (e.g. $c_1 : P \in \mathbf{R}^{n\times n}$, $c_2 : P = P^t$, $c_3 : P_{ij} \geq 0$, etc).

Assuming a solution exists, $\widehat{P}$ is called a *minimizer* of Problem 18. Once a minimizer has been found, one can compute an RL-decomposition of $\widehat{P}$, for instance as in the proof of Theorem 3, and finally the coarse graining LMR^*.

Remark 19. In the absence of constraints, the minimization of $\|e_P(v)\|$ with respect to P leads to $P = vv^*$. Then the natural RL-decomposition is provided by the trivial exact coarse graining (Remark 12).

Remark 20. The term *Spectral Coarse Graining* (SCG) is used to stress that the projector of a coarse graining is chosen so as to preserve one or more (see §B.5.4) eigenpairs of the original matrix.

In the next two sections we introduce two important constraints, namely the *grouping by partitioning* and the *homogeneous mixing*.

B.5.1 Partitioning

In most applications, the entries of a $n \times n$ matrix quantify the pairwise interactions between *n objects*, or *entity*, labelled $\{1, \ldots, n\}$. Conceptually, as in data clustering [52], the grouping of the objects can be done either by partitioning $\{1, \ldots, n\}$, or by allowing every object to have a non-zero degree of membership to all the groups (the latter is called "soft-grouping"). We focus on the more common partitioning and introduce below the formalism to deal with it in coarse graining.

Definition 21. Let γ be the mapping from $\{1, \ldots, n\}$ to one of its partitions Γ, such that $\gamma(i)$ indexes the group (block) of i in Γ. We use Greek letters α, β to label the elements of Γ and Roman letters i, j for the elements of $\{1, \ldots, n\}$. To simplify notation α can represent either the group or its index, so that the reader will encounter the notations $i \in \alpha$, $\sum_{\alpha=1}^{\tilde{n}} \cdot$, and $|\alpha|$ to designate the number of elements in group α. The so-called *partitioning constraint* is the requirement for the product $(LMR^*)_{\alpha\alpha}$ not to mix up entries of M whose indexes don't belong to α. Under partitioning, the semi-projectors L and R are defined as

$$L_{\alpha j} = \ell_{\alpha j}\delta_{\alpha\gamma(j)} \quad \text{and} \quad R_{\alpha j} = r_{\alpha j}\delta_{\alpha\gamma(j)}, \tag{B.3}$$

where $\delta_{\alpha\gamma(j)}$ is equal to one if j belongs to α and zero otherwise, and $\ell_{\alpha j}, r_{\alpha j} \in \mathbf{C}$ for all $\alpha \in \{1, \ldots, \tilde{n}\}$, $i \in \{1, \ldots, n\}$.

The next example should clarify the purpose of Equation B.3.

Example 22. When partitioning is imposed, the condition $LR^* = I_{\tilde{n}}$ (Definition 7) reads

$$(LR^*)_{\alpha\beta} = \sum_{j=1}^{n} \ell_{\alpha j}\bar{r}_{\beta j}\delta_{\alpha\gamma(j)}\delta_{\beta\gamma(j)} = \delta_{\alpha\beta} \Leftrightarrow \sum_{j\in\alpha} \ell_{\alpha j}\bar{r}_{\alpha j} = 1 \ \forall\alpha, \tag{B.4}$$

and the entries of P become

$$P_{ij} = \sum_{\alpha=1}^{\tilde{n}} R^*_{i\alpha}L_{\alpha j} = \ell_{\gamma(j)j}\bar{r}_{\gamma(i)i}\delta_{\gamma(i)\gamma(j)}. \tag{B.5}$$

Hence, P can be put in block form where each block corresponds to a unique group in Γ. Finally,

the coarse graining of M reads

$$\widetilde{M}_{\alpha\beta} = (LMR^*)_{\alpha\beta} = \sum_{\substack{i\in\alpha \\ j\in\beta}} \ell_{\alpha i}\bar{r}_{\beta j}M_{ij}, \tag{B.6}$$

such that, as expected, $\widetilde{M}_{\alpha\alpha}$ is a linear combination of M's entries whose indexes belong exclusively to α.

Remark 23. When partitioning is the only constraint the minimizing projector is given by

$$\widehat{P}_{ij} = \frac{v(i)\overline{v(j)}}{\sum_{k\in\gamma(i)}|v(k)|^2}\delta_{\gamma(i)\gamma(j)}, \tag{B.7}$$

where $v(i)$ denotes component i of v. It is straightforward to check that $\widehat{P}$ in Equation B.7 is indeed an (orthogonal) projector and that $e_{\widehat{P}}(v) = 0$ for any partition of $\{1,\ldots,n\}$.

B.5.2 Homogeneous Mixing

The homogeneous mixing constraint is imposed to ensure that objects belonging to the same group are identical in the projected system (i.e. that they are indistinguishable with respect to the interactions described by PMP), and therefore that they can be merged into a single entity in the coarse-grained system described by LMR^*.

Definition 24. Let Γ be a partition of $\{1,\ldots,n\}$. The constraint of homogeneous mixing is the requirement that for any $x \in \mathbf{C}^n$ and any $\alpha \in \Gamma$, component i of Px is the same for all $i \in \alpha$ (i.e. $(Px)(i) = \text{constant } \forall i \in \alpha$).

The Homogeneous mixing constraint, which has roots in physics, was implicitly assumed in [39, 40].

B.5.3 Minimizing $\|e_P(v)\|$: Methods and Analysis

In this section, we show how to minimize $\|e_P(v)\|$ under the partitioning and the homogeneous mixing constraints. For simplicity, we impose $P \in \mathbf{R}^{n\times n}$ and $v \in \mathbf{R}^n$; the case $v \in \mathbf{C}^n$ is treated in §B.5.4.

For $v \in \mathbf{R}^n$, the general form of $\|e_P(v)\|^2$ is

$$\|v - Pv\|^2 = \sum_{\alpha=1}^{\tilde{n}} \sum_{i \in \alpha} [v(i) - (Pv)(i)]^2. \tag{B.8}$$

The homogeneous mixing allows us to write $\|e_P(v)\|^2 = \sum_{\alpha=1}^{\tilde{n}} \sum_{i \in \alpha} [v(i) - v_{avg}(\alpha)]^2$ for some $v_{avg} \in \mathbf{R}^{\tilde{n}}$. We see that $\|e_P(v)\|^2$ is minimum if $\sum_{i \in \alpha} [v(i) - v_{avg}(\alpha)]^2$ is minimum for each α, that is for $v_{avg}(\alpha) = \frac{1}{|\alpha|} \sum_{i \in \alpha} v(i)$ as it can be readily verified, for example, by deriving $\|e_P(v)\|^2$ with respect to $v_{avg}(\alpha)$. This gives us the optimal form of $(Pv)(i)$, that is $(Pv)(i) = \frac{1}{|\gamma(i)|} \sum_{j \in \gamma(i)} v(j)$. It is not difficult to see that for such P the partitioning constraint is satisfied, and thus the problem reduces to finding the partition of $\{1, \ldots, n\}$ minimizing

$$\|e_P(v)\|^2 = \sum_{\alpha=1}^{\tilde{n}} \sum_{i \in \alpha} \left(v(i) - \frac{1}{|\alpha|} \sum_{j \in \alpha} v(j) \right)^2. \tag{B.9}$$

We present below three methods to tackle this problem. The first method finds a *true* minimizing partition of $\|e_P(v)\|^2$ in polynomial time and memory load, whereas the other two find an *approximate* solution in less time and memory. We stress that the use of approximate methods, which may be appropriate if one deals with very large systems, turns out to be indispensable when the SCG aims at preserving several eigenpairs (see §B.5.4).

B.5.3.1 Optimal Minimization of $\|e_P(v)\|$

The naive approach consists in generating all the partitions of $\{1, \ldots, n\}$ and extract a minimizer of $\|e_P(v)\|$. This method, however, turns out to be infeasible in practice (even for moderately small $\tilde{n}$)[3], and we must follow different lines to obtain a solution in reasonable time. We propose an algorithm in the spirit of Dynamic Programming [8], which relies on both the sub-optimality of the problem and the reusability of previously computed values to boost computation. A similar approach was independently taken in [55] in the context of image analysis.

Step 1 Sort the components of v in increasing order: $v(1) \leq v(2) \leq \cdots \leq v(n)$.

[3]The number of different partitions of $\{1, \ldots, n\}$, known as the Bell number, grows exponentially with n. E.g. for $n = 20$ it already exceeds 10^{13}!

Step 2 For all $i \leq j$, compute

$$c_v(i,j) \equiv \sum_{i\leq k\leq j}\left(v(i) - \frac{1}{j-i+1}\sum_{i\leq k\leq j} v(j)\right)^2 = \sum_{i\leq k\leq j} v(k)^2 - \frac{1}{j-i+1}\left(\sum_{i\leq k\leq j} v(k)\right)^2. \tag{B.10}$$

Step 3 Starting with $F_v(1,j) = c_v(1,j)$, compute $F_v(\tilde{n},n)$ recursively by the following formula

$$F_v(\alpha,j) = \min_{\alpha-1\leq q<j}\left(F_v(\alpha-1,q) + c_v(q+1,j)\right), \tag{B.11}$$

and store at each step the minimizer of Equation B.11:

$$Q_{\alpha j} \equiv \arg\min_{\alpha-1\leq q<j}\left(F_v(\alpha-1,q) + c_v(q+1,j)\right).$$

According to Equations B.10 and B.11, $F_v(\tilde{n},n)$ is the minimum of $\|e_P(v)\|$ over all the partitions with $\tilde{n}$ groups.

Step 4 Starting from $Q_{\tilde{n}n}$ work out the minimizing partition corresponding to $F_v(\tilde{n},n)$ by backtracking through the matrix Q.[4]

Provided c_v is computed in time $O(n^2)$—which can always be achieved computing the right-hand side of Equation B.10—it can be seen that Algorithm B.5.3.1 finds the minimizing partitions with size 1 to $\tilde{n}$ in time $O(\tilde{n}n^2)$ and memory load $O(n^2)$.

B.5.3.2 Approximate Minimization of $\|e_P(v)\|$: Fixed-Size Intervals Method

For very large systems (i.e. $n \gtrsim 10^4$), one has to rely on approximate methods to minimize $\|e_P(v)\|$. We discuss below in some detail the partitioning of v into fixed-sized intervals, as employed in [39, 40], and give arguments to explain the very accurate coarse grainings obtained by this simple method.

Recall that the eigenvector v is assumed normalized. Cut v into $m \geq \tilde{n}$ intervals $I_1,\ldots,I_m$ of respective length $\varepsilon_1,\ldots,\varepsilon_m$ and denote by $|I_k|$ the number of components falling into I_k; by definition $\tilde{n}$ is the number of non-empty intervals. We have seen above that $\sum_{\alpha=1}^{\tilde{n}}\sum_{i\in\alpha}(v(i) -$

[4] The same procedure can be applied to retrieve from Q all the minimizing partitions with $n' \leq \tilde{n}$ groups.

$v_{avg}(\alpha))^2$ is minimum for $v_{avg}(\alpha) = \frac{1}{|\alpha|}\sum_{i\in\alpha} v_\alpha$. Hence,

$$\|e_P(v)\|^2 = \sum_{\alpha=1}^{\tilde{n}} \sum_{i\in\alpha} \left(v(i) - \frac{1}{|\alpha|} \sum_{i\in\alpha} v_\alpha \right)^2 \leq \sum_{k=1}^{m} \sum_{i\in I_k} \left(\frac{\varepsilon_k}{2}\right)^2 = \frac{1}{4}\sum_{k=1}^{m} |I_k|\varepsilon_k^2. \tag{B.12}$$

Ideally, the I_k should be chosen so as to minimize the right-hand side of Equation B.12. For simplicity though, we consider here fixed-size intervals such that $\varepsilon_k = \varepsilon\,\forall k$, which yields immediately $\|e_P(v)\| \leq \frac{\varepsilon}{2}\sqrt{n}$. Let $\delta(v)$ be defined as $\delta(v) = \max_i v(i) - \min_i v(i) \leq 2$; then $\varepsilon = \delta(v)/m$ and we have

$$\|e_P(v)\| \leq \frac{\delta(v)\sqrt{n}}{2m} \leq \frac{\sqrt{n}}{\tilde{n}}. \tag{B.13}$$

Hence, when P is orthogonal, as in Theorem 15,

$$\frac{\|e_P(v)\|}{\|Pv\|} = \|e_P(v)\|(1 + O(\|e_P(v)\|^2) \leq \frac{\delta(v)\sqrt{n}}{2m}\left(1 + O\left(\frac{n}{\tilde{n}^2}\right)\right). \tag{B.14}$$

Remark 25. Equations B.13 and B.14 provide informative bounds on the eigenvalue shifts. Indeed, provided fixed-size intervals or a better partitioning is used (e.g. the optimal partitioning of §B.5.3.1):

- The minimum eigenvalue shifts go to zero at least as $\tilde{n}^{-1}$.
  - If the accuracy is the same, the SCG of large matrices may achieve better dimension reduction than the coarse graining of small matrices. E.g. $\|e_P(v)\| \leq 0.1$ is obtained either for $n = 10^4$ and $\tilde{n} = 10^3$, or for $n = 10^6$ and $\tilde{n} = 10^4$, which improves by a factor 10 the ratio $n/\tilde{n}$.
  - The factor $\delta(v)$ in Equation B.13 can affect substantially the accuracy of a coarse graining. Indeed, it is common to observe values of $\delta(v)$ much smaller than 2, especially for particular eigenvectors of some random matrices (see Figure B.3).

B.5.3.3 Approximate Minimization of $\|e_P(v)\|$: Fixed-Size Intervals+k-means Method

One can usually improve the result of the fixed-size intervals method by running the so-called "k-means" algorithm [59, 48] on the obtained partition. Starting from a partition of $\{1,\ldots,n\}$ with k groups, k-means finds at each step a new partition of same cardinality such that $\|e_P(v)\|$ (of Equation B.9) is smaller for the new partition than for the former. The algorithm keeps running until it gets stuck—usually—in a *local* minimum of $\|e_P(v)\|$.

Hence, even though the final partition driven by k-means is seldom the "absolute" minimizer of $\|e_P(v)\|$, it is certainly a better one than the partition obtained by the fixed-size intervals method when the latter is used to initialize k-means.

B.5.4 Several Eigenpairs and Complex Case

B.5.4.1 Several Eigenpairs: Exact SCG

In the absence of constraints on the coarse graining, it is sometimes possible to preserve exactly an arbitrary subspace of M's eigenspace. Suppose $M \in \mathbf{C}^{n\times n}$ is diagonalizable. Compute a projector P onto the subspace of $\mathbf{C}^n$ spanned by the eigenvectors of M to be preserved, and along the subspace spanned by the remaining eigenvectors; since M is diagonalizable both subspaces are complementary (but not necessarily orthogonal). Compute an RL-decomposition of P and finally the coarse-grained matrix $\widetilde{M} = LMR^*$.

This approach is popular in statistics to reduce multidimensional data sets, wherein it is known as *Principal Component Analysis* (PCA) [54]. In this context, M is the data covariance matrix and its eigenvectors are the (independent) directions along which the variance of the data is maximum; as for the variance along each eigenvector it is given by the corresponding eigenvalue of M. PCA usually preserves the largest eigenvalues as they explain the most variance.

When partitioning is imposed, the projector of Equation B.7 is generally useless to preserve exactly several eigenpairs. However, if the groups are made up of objects with equal components in the eigenvectors to be preserved, the projector of Equation B.7 takes the form $\widehat{P}_{ij} = |\gamma(i)|^{-1}\delta_{\gamma(i)\gamma(j)}$, and $\|e_P(v)\| = 0$ for all these eigenvectors. This technique can be used to eliminate zero eigenpairs resulting from possible row duplication in M (columns if left eigenvectors are considered).

B.5.4.2 Several Eigenpairs: Approximate SCG

Method 1 Let $(\lambda_1, v_1), \ldots, (\lambda_s, v_s)$ be s eigenpairs of M, *not necessarily ordered*, to be preserved by the coarse graining. We assume the conditions of §B.5.3 hold, that is P and the v_k are real, and we impose the homogeneous mixing constraint. Suppose one of the three methods of §B.5.3 has been applied on each v_k and has given the s partitions $\Gamma_1 = \{\alpha_{11}, \ldots, \alpha_{1\tilde{n}_1}\}, \ldots, \Gamma_s = \{\alpha_{s1}, \ldots, \alpha_{s\tilde{n}_s}\}$, where α_{ki} denotes group i of partition k. Let $\gamma(i)$ stand for the final group of i.

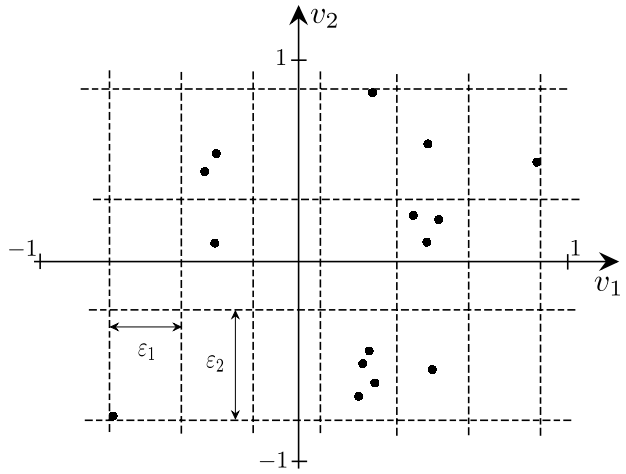


Figure B.2: *This figure depicts the construction of the final partition* Γ*, when fixed-size intervals of size* ε_1 *and* ε_2 *are used to bin the components of* v_1*, respectively of* v_2 *(Method 1 of §B.5.4.2). All objects whose components end up in the same box are grouped together in* Γ*.*

Two objects i and j are grouped together in the final partition Γ if they are grouped together within each Γ_k. Formally, Γ is defined implicitly as follows. For all $i, j \in \{1, \ldots, n\}$,

$$\gamma(i) = \gamma(j) \overset{d}{\Leftrightarrow} \exists k_1, \ldots, k_s \text{ such that } i, j \in \alpha_{1k_1} \cap \cdots \cap \alpha_{sk_s}. \tag{B.15}$$

Even though each Γ_k minimizes $\|e_P(v_k)\|$, Γ generally does not minimize any of the $\|e_P(v_k)\|$. The case $s = 2$ of this construction is illustrated in Figure B.2.

Method 2 Another way to preserve $(\lambda_1, v_1), \ldots, (\lambda_s, v_s)$ is by trying to minimize the overall sum

$$\sum_{k=1}^{s} \|e_P(v_k)\|^2 \equiv \sum_{k=1}^{s} \sum_{\alpha=1}^{\tilde{n}} \sum_{i \in \alpha} \left(v_k(i) - \frac{1}{|\alpha|} \sum_{i \in \alpha} v_k(i) \right)^2 = \sum_{\alpha=1}^{\tilde{n}} \sum_{i \in \alpha} \|w_i - \frac{1}{|\alpha|} \sum_{i \in \alpha} w_i\|^2, \tag{B.16}$$

where $w_i = (v_1(i), \ldots, v_s(i))^t$ and $\frac{1}{|\alpha|} \sum_{i \in \alpha} w_i$ is the barycentre of the w_i belonging to group α. This amounts to clustering n points (of a s-dimensional space) into $\tilde{n}$ groups by minimizing the overall intra-group variance. This approach was proposed in [56] for the coarse graining of stochastic matrices, but the problem was previously shown NP-hard for $s > 1$ and $\tilde{n} \geq 2$ in [33]. As a consequence, k-means is commonly employed to drive an approximate minimizer of Equation B.16 in reasonable time.

The main drawback of k-means here is that for large $\tilde{n}$ the k-means solution can be much poorer than the true minimizer of Equation B.16, due to the multiplication of "trapping" local minima in

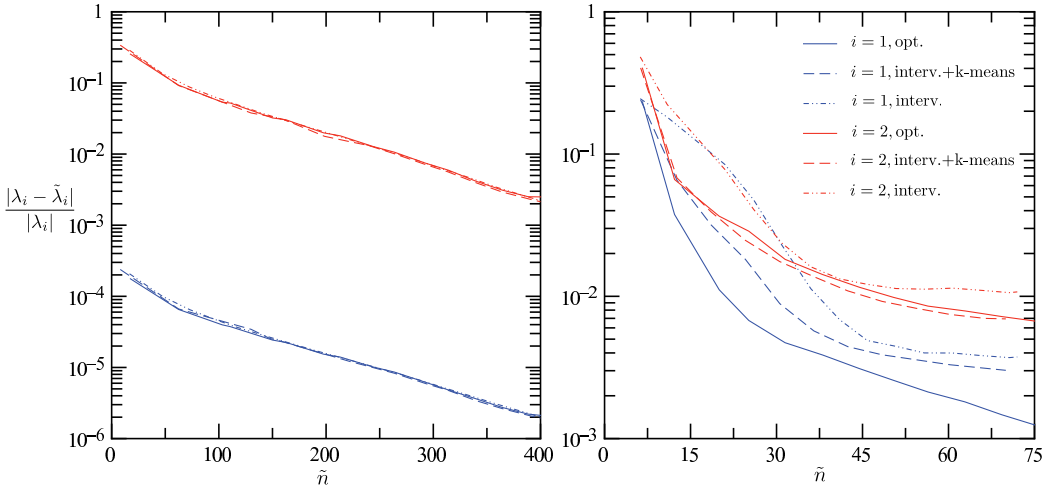


Figure B.3: *These figures depict the relative eigenvalue shifts when coarse graining for λ_1 and λ_2, by Method 1 of §B.5.4.2, two different classes of matrices in $\mathbf{R}^{500\times500}$. (1) On the left-hand side, the results for symmetric matrices with positive i.i.d. entries. In this case, the $v(i)$ are densely distributed with no particular pattern. We observe that the three methods, optimal ("——"), intervals+kmeans ("– – –"), and intervals ("– · ·–"), give equally accurate coarse grainings. The important gap between both shifts follows from the fact that $\delta(v_1) = \max_i v_1(i) - \min_i v_1(i) \approx 10^{-3}$ whereas $\delta(v_i) \approx 1\,\forall i \in \{2,\ldots,n\}$ (Remark 25). We stress that the relative eigenvalue shifts for $i > 2$, not shown here, are all identical to the case $i = 2$ (i.e. the relative accuracy is—generally—independent of the magnitude of the eigenvalues; see also Summary B.4). (2) On the right-hand side, the results for adjacency matrices of Barabasi–Albert random graphs [5]—the latter have correlated entries drawn from $\{0,1\}$. For these matrices, the $v(i)$ are sparsely distributed and a stripe pattern can be observed. As a consequence, the groups are better identified by the optimal algorithm, which outperforms the approximate methods. Results have been averaged over a thousand realizations of both matrices.*

$\sum_{k=1}^{s} \|e_P(v_k)\|^2$. More importantly, minimizing Equation B.16 does not allow precise individual control of the eigenpair shifts, which may be problematic in applications wherein the eigenpairs to be preserved are not all equally relevant to the system.

In other words, because of its greater flexibility, Method 1 of this paragraph is generally a better choice than Method 2 to preserve several eigenpairs in SCG. In particular, Algorithm B.5.3.1 of §B.5.3.1 can be used to find the true minimizing partitions of the $\|e_P(v_k)\|$, which makes further consideration about the (physical) meaning of the groups more pertinent than with approximate methods. On the other hand, Method 2 allows one to fix $\tilde{n}$ at the cost of unpredictable eigenpair shifts. Therefore, the latter might be considered if the size of the coarse grained matrix is imposed beforehand, and one is not concerned with precise control of individual eigenpair shifts.

B.5.4.3 Complex Eigenpairs

For v a non-real eigenvector, one can carry out the minimization of $\|e_P(v)\|$ as the minimization of both $\|e_P(\Re v)\|$ and $\|e_P(\Im v)\|$ by either of the methods above. Indeed, provided $v_{avg}(\alpha) =$

$\frac{1}{|\alpha|}\sum_{i\in\alpha} v(i)$,

$$\begin{aligned}
\|e_P(v)\|^2 &= \sum_{\alpha=1}^{\tilde{n}}\sum_{i\in\alpha} |v(i) - v_{avg}(\alpha)|^2 \\
&= \sum_{\alpha=1}^{\tilde{n}}\sum_{i\in\alpha} \left(\Re\left[v(i) - v_{avg}(\alpha)\right]\right)^2 + \left(\Im\left[v(i) - v_{avg}(\alpha)\right]\right)^2 \\
&= \sum_{\alpha=1}^{\tilde{n}}\sum_{i\in\alpha} \left(\Re[v](i) - \Re[v]_{avg}(\alpha)\right)^2 + \left(\Im[v](i) - \Im[v]_{avg}(\alpha)\right)^2 \\
&= \|e_P(\Re v)\|^2 + \|e_P(\Im v)\|^2.
\end{aligned}$$

The need to minimize two eigenvector shifts in the complex case is not surprising since, for M a real matrix, the preservation of an eigenpair automatically implies the preservation of its conjugate. In this case, Method 1 of §B.5.4.2 allows the differentiation between the preservation of $\Re v$ and $\Im v$.

The accuracy of the methods presented in this section is compared in Figure B.3.

B.6 Application to Graph Theory

The spectral coarse graining (SCG) of graphs under constraint has been initially introduced through the SCG of stochastic matrices in [38, 56, 39] and through the SCG of Laplacian matrices in [40]. The main goal in these works was to reduce large graphs while preserving their spectral-related features (i.e. features of the system related to the spectral properties of the associated interaction matrix). In the following we recast these SCG in the framework of this article, and introduce the SCG of a graph adjacency matrix.

Definition 26. Let $G(V, E)$ be a weighted strongly connected graph (i.e. there is a path between any pair of vertices in G). V is the set of vertices, labeled from 1 to n, and E is the set of edges; to every edge is associated a weight $e_{ij} \geq 0$ and, with no mention of the contrary, G is supposed directed, that is e_{ij} is not necessarily equal to e_{ji}.

1. The adjacency matrix A is defined as $A_{ij} = e_{ij}$, with $A_{ij} = 0$ indicating that there is no edge from vertex i to vertex j. By definition G is undirected if A is symmetric.

   (a) We define the graph Laplacian matrix $\mathcal{L}$ as $\mathcal{L}_{ij} = \delta_{ij}\sum_{k=1}^{n} A_{ik} - A_{ij}$.

(b) The row-stochastic matrix W is defined as $W_{ij} = \dfrac{A_{ij}}{\sum_{k=1}^{n} A_{ik}}$.

Definition 27. Let $G(V, E)$ be a weighted graph with edge weights $\{e_{ij}\}$. Let Γ be a partition of V. A coarse-grained graph $\widetilde{G}(\widetilde{V}, \widetilde{E})$ with respect to Γ is a graph where (1) to each vertex in $\widetilde{V}$ corresponds one and only one group in Γ, and (2) every edge weight $\tilde{e}_{\alpha\beta}$ is a linear combination of the elements in $\{e_{ij} | i \in \alpha, j \in \beta \text{ and } \alpha, \beta \in \widetilde{V}\}$.

Remark 28. A crucial condition to the coarse graining of A, $\mathcal{L}$ and W is that the transformation preserves the structure of the original matrix, that is $\widetilde{A} \equiv LAR^t$, resp. $\widetilde{L} \equiv L\mathcal{L}R^t$ and $\widetilde{W} \equiv LWR^t$, must be an adjacency, respectively a Laplacian and a stochastic matrix, of a graph; this is called the *structural constraint*.

B.6.1 Adjacency Matrices

$G(V, E)$ is a weighted graph with n vertices and A its associated adjacency matrix.

If A is symmetric, and the coarse-grained adjacency matrix LAR^t is to be symmetric as well, the natural choice is to coarse-grain A with $L = R$—which implies that the projector $P = R^t R$ is orthogonal. In addition, we impose the homogeneous mixing constraint of Section B.5.2 (i.e. $(Px)(i)$ is constant within each group and $\forall x \in \mathbf{R}^n$). Recall that $\|e_P(v)\|^2 = \sum_{i\in\alpha}[v(i) - (Pv)(i)]^2$ is minimum if $(Pv)(i) = \frac{1}{|\gamma(i)|}\sum_{j\in\gamma(i)} v(j)$, and the partitioning constraint is automatically satisfied by P (§B.5). A simple way to satisfy homogeneous mixing, while still ensuring $RR^t = I_{\tilde{n}}$, is to define R as

$$R_{\alpha j} = \frac{1}{\sqrt{|\alpha|}}\delta_{\alpha\gamma(j)}.$$

The coarse-grained matrix $\widetilde{A}$ is the adjacency matrix of the coarse-grained graph $\widetilde{G}(\widetilde{V}, \widetilde{E})$, with edge weights given by

$$\tilde{e}_{\alpha\beta} = \widetilde{A}_{\alpha\beta} = \frac{1}{\sqrt{|\alpha||\beta|}} \sum_{\substack{i\in\alpha \\ j\in\beta}} A_{ij}.$$

Obviously, taking $L = R$ is also possible if LAR^t is not to be symmetric (e.g. $G(V, E)$ is directed), as long as this choice does not violate other *ad hoc* constraints.

B.6.2 Laplacian Matrices

In this example, $G(V, E)$ is a weighted undirected graph with n vertices and $\mathcal{L}$ is its associated Laplacian matrix. Under these conditions, it can be shown that $\mathcal{L}$ is symmetric semi-positive definite.

By definition the rows of $\mathcal{L}$ sum up to zero, which in this case is equivalent to saying that $\mathcal{L}$ has a unique zero eigenvalue with corresponding right eigenvector $v_n = (1, \dots, 1)^t \in \mathbf{R}^n$—the zero eigenvalue has multiplicity one since G is connected.

The structural constraint is satisfied if $\widetilde{\mathcal{L}} = L\mathcal{L}R^t$ also has a unique zero eigenvalue with right eigenvector $\tilde{v}_{\tilde{n}} = (1, \dots, 1)^t \in \mathbf{R}^{\tilde{n}}$. A natural choice is then to choose R such that $R^t\tilde{v}_{\tilde{n}} = v_n$, which implies $r_{\alpha j} = \delta_{\alpha\gamma(j)}$ under partitioning (Equation B.3). Since $LR^t = I_{\tilde{n}}$, it follows that the rows of L verify $\sum_{j=1}^{n} \ell_{\alpha j}\delta_{\alpha\gamma(j)} = \sum_{j\in\alpha} \ell_{\alpha j} = 1$, which can be achieved by taking $\ell_{\alpha j} = |\alpha|^{-1}$. This results in the following definitions:

$$L_{\alpha j} = \frac{1}{|\alpha|}\delta_{\alpha\gamma(j)} \quad \text{and} \quad R_{\alpha j} = \delta_{\alpha\gamma(j)}. \tag{B.17}$$

As with adjacency matrices, this choice of L and R is optimum in the sense that $(Pv)(i) = \frac{1}{|\gamma(i)|}\sum_{j\in\gamma(i)} v(j)$ for all i (§B.5). Hence, the homogeneous mixing is satisfied and the minimization of $\|e_P(v)\|$ to find the optimal partition can be carried out as described in §B.5.

Let us examine the main properties of this coarse graining.

P1 A straightforward calculation shows that $\widetilde{\mathcal{L}}\tilde{v}_{\tilde{n}} = 0$.

P2 $(Px)(i)$ is constant over each group $\forall x \in \mathbf{R}^n$.

P3 Denote by H the $\tilde{n} \times \tilde{n}$ matrix $H_{\alpha\beta} = \sqrt{|\alpha|}\delta_{\alpha\beta}$ and consider the change of basis $H\widetilde{\mathcal{L}}H^{-1} = (HL)\mathcal{L}(R^tH^{-1})$. It is easy to see that $HL = (R^tH^{-1})^t$, such that $\widetilde{\mathcal{L}}$ is similar to a symmetric semi-positive definite matrix (though it is not symmetric since $L \neq R$).

P4 Since the projector $P = R^tL$ is orthogonal, the coarse-grained eigenvalues obey the interlacing of the Poincaré separation theorem (Theorem 17). As a consequence, the zero eigenvalue of $\widetilde{\mathcal{L}}$ is unique.

Properties P1 and P4 ensure that the matrix $\widetilde{\mathcal{L}}$ is the Laplacian matrix of a connected weighted graph $\widetilde{G}(\widetilde{V}, \widetilde{E})$ in the sense of Definition 26. To find the edge weights $\tilde{e}_{\alpha\beta}$, we first notice that

$L\mathcal{L}R^t = LDR^t - LAR^t = \widetilde{D} - \widetilde{A}$, where $\widetilde{A}_{\alpha\beta} = \frac{1}{|\alpha|}\sum_{\substack{i\in\alpha\\ j\in\beta}} A_{ij}$ and $\widetilde{D}_{\alpha\beta} = \delta_{\alpha\beta}\sum_{\omega=1}^{\tilde{n}} \widetilde{A}_{\beta\omega}$. Hence, $\widetilde{A}$ is the adjacency matrix of $\widetilde{G}$ and the edge weights are given by

$$\tilde{e}_{\alpha\beta} = \widetilde{A}_{\alpha\beta} = \frac{1}{|\alpha|}\sum_{\substack{i\in\alpha\\ j\in\beta}} A_{ij}.$$

As argued in [40] the coarse graining presented in this paragraph can be used to reduce efficiently large graphs of coupled oscillators while preserving their synchronization properties (the latter are related to the ratio $\lambda_1(\mathcal{L})/\lambda_{n-1}(\mathcal{L})$).Stochastic Matrices

For this last application, $G(V, E)$ is a strongly connected weighted graph with n vertices and row-stochastic matrix W.

The matrix W gives the transition probability distribution of a Markov chain on G. Since G is strongly connected W has a unique eigenvalue $\lambda = 1$ associated to a right eigenvector $v_1 = (1, \ldots, 1)^t \in \mathbf{R}^n$, and to a left eigenvector $p_1 \in \mathbf{R}^n$ with $p_1(i) > 0\,\forall i$. The components of p_1 give, up to a scalar multiplication, the stationary distribution of the Markov chain. Importantly, for undirected graphs p_1 (unnormalized) is given by $p_1(i) = \sum_{j=1}^{n} A_{ij}$.

In order to satisfy the structural constraint, we require that the coarse-grained matrix $\widetilde{W}$ is row-stochastic, that is $\lambda = 1$ must be an eigenvalue of $\widetilde{W}$ with corresponding right eigenvector $\tilde{v}_1 = (1, \ldots, 1)^t \in \mathbf{R}^{\tilde{n}}$.

In addition, we demand to preserve exactly the stationary state p_1 in $\widetilde{W}$, that is, we look for semi-projectors L and R such that $p_1^t R^t$ is a left eigenvector of LWR^t with eigenvalue equal to one.

To this aim the following choice for R and L was independently proposed in [38, 56, 39] (although column-stochastic matrices were considered there):

$$L_{\alpha j} = \frac{p_1(j)}{\sum_{k\in\gamma(j)} p_1(k)}\delta_{\alpha\gamma(j)} \quad \text{and} \quad R_{\alpha j} = \delta_{\alpha\gamma(j)}. \tag{B.18}$$

We verify immediately the following properties:

P1 $P = R^t L$ is indeed a projector since $LR^t = I_{\tilde{n}}$, but $P \neq P^t$ in general.

P2 $(Px)(i)$ is constant over each group $\forall x \in \mathbf{R}^n$.

P3 The stationary state is exactly preserved. Indeed, it is easy to verify that $p_1^t P = p_1^t$ so that $p_1^t R^t \widetilde{W} = p_1^t PWR^t = p_1^t R^t$.

P4 The structural constraint is fulfilled: $\widetilde{W}\tilde{v}_1 = \tilde{v}_1$. Furthermore, $\tilde{\lambda} = 1$ has multiplicity one since the coarse-grained graph defined from $\widetilde{W}$ is strongly connected (see Equation B.19).

By property P4, the coarse-grained matrix $\widetilde{W} = LWR^t$ is the row-stochastic matrix of the coarse-grained graph $\widetilde{G}(\widetilde{V}, \widetilde{E})$ with edge weights given by

$$\tilde{e}_{\alpha\beta} = \widetilde{A}_{\alpha\beta} = \sum_{\substack{i\in\alpha \\ j\in\beta}} \frac{p_1(i)}{\sum_{k=1}^{n} A_{ik}} A_{ij}. \tag{B.19}$$

To see this, we first write down explicitly $\widetilde{W}_{\alpha\beta}$:

$$\widetilde{W}_{\alpha\beta} = (LWR^t)_{\alpha\beta} = \sum_{\substack{i\in\alpha \\ j\in\beta}} \frac{p_1(i)}{\sum_{l\in\alpha} p_1(l)} \frac{A_{ij}}{\sum_{k=1}^{n} A_{ik}} = \frac{\widetilde{A}_{\alpha\beta}}{\sum_{l\in\alpha} p_1(l)}, \tag{B.20}$$

where we have defined $\widetilde{A}_{\alpha\beta} \equiv \sum_{\substack{i\in\alpha \\ j\in\beta}} \frac{p_1(i)}{\sum_{k=1}^{n} A_{ik}} A_{ij}$. Recall that the left eigenvector p_1 satisfies $p_1^t = p_1^t W \Leftrightarrow p_1(l) = \sum_{j=1}^{n} p_1(j) \frac{A_{jl}}{\sum_{k=1}^{n} A_{lk}}$. Substituting $p_1(l)$ in Equation B.20 gives

$$\widetilde{W}_{\alpha\beta} = \frac{\widetilde{A}_{\alpha\beta}}{\sum_{l\in\alpha} \sum_{\omega=1}^{\tilde{n}} \sum_{j\in\omega} p_1(j) \frac{A_{jl}}{\sum_{k=1}^{n} A_{lk}}} = \frac{\widetilde{A}_{\alpha\beta}}{\sum_{\omega=1}^{\tilde{n}} \widetilde{A}_{\omega\alpha}}.$$

Now, since $\widetilde{W}\tilde{v}_1 = \tilde{v}_1$, we have that $\sum_\beta \widetilde{W}_{\alpha\beta} = 1 \Leftrightarrow \sum_\beta \widetilde{A}_{\alpha\beta} = \sum_\beta \widetilde{A}_{\beta\alpha}$, and thus

$$\widetilde{W}_{\alpha\beta} = \frac{\widetilde{A}_{\alpha\beta}}{\sum_{\omega=1}^{\tilde{n}} \widetilde{A}_{\alpha\omega}}.$$

Therefore, $\widetilde{A}_{\alpha\beta}$ is the adjacency matrix of a (directed) graph $\widetilde{G}(\widetilde{V}, \widetilde{E})$ with edge weights given by Equation B.19. If $G(V, E)$ is undirected, we recall that $p_1(i) = \sum_{j=1}^{n} A_{ij}$; as a consequence

$$\tilde{e}_{\alpha\beta} = \sum_{\substack{i\in\alpha \\ j\in\beta}} A_{ij},$$

which is the (intuitive) sum of the edge weights between the two groups α and β.

Importantly, Equation B.18 implies that $\|e_P(v)\|^2$ now reads

$$\|e_P(v)\|^2 = \sum_{\alpha=1}^{\tilde{n}} \sum_{i\in\alpha} \left(v(i) - \frac{1}{\sum_{j\in\alpha} p_1(j)} \sum_{j\in\alpha} p_1(j) v(j) \right)^2. \tag{B.21}$$

Even though Equation B.21 and Equation B.9 are different in general, when the groups are composed of vertices with equal components in v (i.e. $v(i) = v(j)\ \forall i, j \in \alpha$ and $\forall \alpha$), the coarse graining is exact in both cases (i.e. $\|e_P(v)\| = 0$).

The optimal minimization of $\|e_P(v)\|$ (Algorithm B.5.3.1) can still be carried out by defining c_v accordingly. Furthermore, the fixed-size intervals method yields a similar upper bound as in Equations B.12 and B.14, so that adding the constraint $p_1^t R^t \widetilde{W} = p_1^t R^t$ to the minimization does not alter the main results of §B.5. In particular, the approximate methods presented in the previous section still lead to very accurate coarse grainings, as observed in [39].

Finally, if G is undirected, we show that the coarse-grained matrix LWR^t with L and R as in Equation B.18 is similar to the matrix $\widehat{R}M\widehat{R}^t$, where M is real symmetric and $\widehat{R}$ is a semi-orthogonal projector (see also Property P3 of the Laplacian matrix).

Let D be the diagonal matrix defined as $D_{ii} = \sum_{j=1}^{n} A_{ij}$; hence $W = D^{-1}A$. We consider the matrix M defined as

$$M = D^{1/2} W D^{-1/2} = D^{-1/2} A D^{-1/2}. \tag{B.22}$$

Clearly M and W have the same eigenvalues and if G is undirected M is symmetric. We introduce the matrix $\widetilde{M} = \widetilde{D}^{1/2}\widetilde{W}\widetilde{D}^{-1/2}$, with $\widetilde{D}$ the diagonal matrix defined as $\widetilde{D}_{\alpha\alpha} = \sum_{i\in\alpha} D_{ii}$; then $\mathrm{sp}(\widetilde{M}) = \mathrm{sp}(\widetilde{W})$. Further, $\widetilde{M}$ can be expressed as

$$\begin{aligned} \widetilde{M} &= \widetilde{D}^{1/2}\widetilde{W}\widetilde{D}^{-1/2} \\ &= \underbrace{\left(\widetilde{D}^{1/2} L D^{-1/2}\right)}_{\widehat{L}} M \underbrace{\left(D^{1/2} R^t \widetilde{D}^{-1/2}\right)}_{\widehat{R}^t}. \end{aligned}$$

It is straightforward to see that $\widehat{L}\widehat{R}^t = I_{\tilde{n}}$ as required. Finally, if the graph is undirected, we have that $\widehat{L} = \widehat{R}$ since $p_1(i) = \sum_{k=1}^{n} A_{ik}$. Therefore, although P is not orthogonal in this coarse graining, the results of the symmetric SCG apply; in particular the eigenvalues of $\widetilde{W}$ and W interlace as described by the Poincaré separation theorem.

B.7 Summary and Conclusion

Spectral Coarse Graining (SCG) is a general framework for dimension reduction of large systems. It goes beyond traditional clustering strategies by providing a coarse-grained system, and

includes Principal Component Analysis as the exact coarse graining of correlation matrices for their large eigenvalues.

In this work our first goal was to put SCG on a firm mathematical basis. To this aim, we have addressed some important theoretical issues, such as the mathematical definition of a coarse graining transformation and its connexion with projection in Linear Algebra. Then, borrowing techniques from matrix perturbation theory, we have bounded from above the minimum eigenvalue shifts caused by a coarse graining. We have extracted from the bound the quantity $\|e_P(v)\|$, whose minimization has been shown to be a necessary and sufficient condition to the preservation of the eigenpair (λ, v) in the coarse-grained matrix.

In a second part, we have defined a generic SCG problem along with the partitioning and homogeneous mixing constraints. We have solved the problem by means of an optimum algorithm and of two approximate methods—introduced do deal with very large systems—which have been further extended to the preservation of several eigenpairs. Finally, we have performed the SCG of graphs within our framework through the SCG of the adjacency, the Laplacian and the stochastic matrices. In particular, we have incorporated the conservation of the matrix structure as a constraint in all these instances.

We believe that SCG, being still in its infancy, offers a number of interesting extensions and open questions. For example, it would be interesting to refine the perturbative analysis so as to obtain upper bounds on the eigenvector shifts, as well as lower bounds on the eigenvalue shifts, in terms of $\|e_P(v)\|$. Possible extensions of the theory include the SCG of linear operators in Hilbert space (for which the spectrum is discreet and often meaningful), SCG with overlapping groups (in analogy with "soft" clustering), and the SCG of higher-rank tensors for which "eigendecompositions" have recently found application in Genetics [57].

As could be noticed, setting up a specific SCG problem involves the choice of system-dependant constraints that can make the problem delicate to solve. We hope the present framework, along with the various examples, will be helpful to anyone interested in applying SCG techniques to his or her problem. Toward this goal, a computer program will soon be released that will make the methods presented in this paper ready-to-use.

The authors thank the various reviewers of this manuscript. In particular, many thanks go to Eugenio Rodrigues at Swissquote for his helpful suggestions, to Serge Kassibrakis (Swissquote) and to Aitana Morton de Lachapelle for their careful proofreading of the text. D. Morton de

Lachapelle is truly in debt to Swissquote Bank for financially supporting his work. The work of D. Gfeller is financed by the Swiss National Science Foundation (Grant PBELA—120936).

Bibliography

[1] Alfarano, S., and Lux, T. A minimal noise traders model with realistic time series properties. In *Long memory in Economics*. Springer, Berlin, 2003.

[2] Arthur, B. W. Inductive reasoning and bounded rationality: the El Farol problem. *Am. Econ. Rev. 84* (1994), 406–411.

[3] Arthur, D., and Vassilvitskii, S. k-means++: the advantages of careful seeding. In *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (Philadelphia, PA, USA, 2007), Society for Industrial and Applied Mathematics, pp. 1027–1035.

[4] Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. Fractional integrated generalized autoegressive conditioal heteroskedasticity. *J. of Econometrics 74* (1996), 3–30.

[5] Barabási, A.-L., and Albert, R. Emergence of scaling in random networks. *Science 286*, 5439 (1999), 509–512.

[6] Barber, B. M., Lee, Y.-T., Liu, Y.-J., and Odea, T. Do individual day traders make money? evidence from taiwan.

[7] Bauer, R., Cosemans, M., and Eichholtz, P. The performance and persistence of individual investors: Rational agents or tulip maniacs? `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=965810`.

[8] Bellman, R. Dynamic programming. *Science 153*, 03731 (1966), 34–37.

[9] Bera, A. K., and Higgins, M. L. Arch models: properties, estimation and testing. *J. of Economic Surveys 7* (1993), 305–362.

[10] BERCOVICI, H., AND VOICULESCU, P. Lévy-Hinčin type theorems for multiplicative and additive free convolution. Pacific J. Math. 153(2) (1992), 217–248.

[11] BLATT, M., WISEMAN, S., AND DOMANY, E. Superparamagnetic clustering of data. *Physical Review Letters 76*, 18 (April 1996), 3251+.

[12] BOLLERSLEV, T., CHOU, R. Y., AND KRONER, K. F. Arch modelling in finance. *J of Econometrics 52* (1992), 5–59.

[13] BOUCHAUD, J.-P., FARMER, J. D., AND LILLO, F. How markets slowly digest changes in supply and demand. arXiv:0809.0822.

[14] BOUCHAUD, J.-P., GIARDINA, I., AND MÉZARD, M. On a universal mechanism for long ranged volatility correlations. *Quant. Fin. 1* (2001), 212. cond-mat/0012156.

[15] BOUCHAUD, J.-P., AND POTTERS, M. *Theory of Financial Risks*. Cambridge University Press, Cambridge, 2000.

[16] BOUCHAUD, J.-P., AND POTTERS, M. *Theory of Financial Risk and Derivate Pricing: From Statistical Physics to Risk Management*, 2nd ed. Cambridge University Press, 2003.

[17] BOUCHAUD, J. P., AND POTTERS, M. Financial applications of random matrix theory: a short review. *arXiv q-fin.ST* (Jan 2009).

[18] BRENNAN, M. The optimal number of securities in a risky asset portfolio when there are fixed costs of transacting: Theory and some empirical results. *The Journal of Financial and Quantitative Analysis 10*, 3 (1975), 483–496.

[19] BROCK, W. A., AND HOMMES, C. H. A rational route to randomness. *Econometrica 65* (1997), 1059–1095.

[20] CALDARELLI, G., MARSILI, M., AND ZHANG, Y.-C. A prototype model of stock exchange. *Europhysics Letters 50* (1997), 479–484.

[21] CHALLET, D., CHESSA, A., MARSILI, M., AND ZHANG, Y.-C. From minority games to real markets. *Quant. Fin. 1* (2000), 168. cond-mat/0011042.

[22] CHALLET, D., AND MARSILI, M. Criticality and finite size effects in a realistic model of stock market. *Phys. Rev. E 68* (2003), 036132.

[23] Challet, D., Marsili, M., and Zhang, Y.-C. *Minority Games*. Oxford University Press, Oxford, 2005.

[24] Challet, D., and Stinchcombe, R. Non-constant rates and over-diffusive prices in a simple model of limit order markets. *Quant. Fin. 3* (2001), 155.

[25] Challet, D., and Zhang, Y.-C. Emergence of cooperation and organization in an evolutionary game. *Physica A 246* (1997), 407. adap-org/9708006.

[26] Clauset, A., Shalizi, C. R., and Newman, M. E. J. Power-law distributions in empirical data. *SCIAM Review 51*, 661–703.

[27] Cleveland, W. S., and Devlin, S. J. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association 83*, 403 (1988), 596–610.

[28] Cont, R. Empirical properties of asset returns. *Quantitative Finance 1* (Mar 2001), 223–236.

[29] Cont, R., and Bouchaud, J.-P. Herd behaviour and aggregate fluctuation in financial markets. *Macroecon. Dyn. 4* (2000), 170.

[30] Dacorogna, M. M., Gencay, R., Müller, U. A., Olsen, R. B., and Pictet, O. V. *An Introduction to High-Frequency Finance*. Academic Press, London, 2001.

[31] de Lachapelle, D. M., Gfeller, D., and Rios, P. D. L. Spectral coarse graining of matrices with application to graph theory. *Swissquote QAM Journal 1* (Oct 2008), 53–78.

[32] Dea, S., Gondhib, N. R., Manglac, V., and Pochirajud, B. Success/failure of past trades and trading behavior of investors.

[33] Drineas, P., Frieze, A., Kannan, R., Vempala, S., and Vinay, V. Clustering large graphs via the singular value decomposition. *Mach. Learn. 56*, 1-3 (2004), 9–33.

[34] Efron, B., and Tibshirani, R. J. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.

[35] Engle, R. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*, 4 (1982), 987–1007.

[36] Farmer, J. D., and Lillo, F. On the origin of power law tails in price fluctuations. *Quant. Fin. 4* (2003), 7. cond-mat/0309416.

[37] GABAIX, X., GOPIKRISHNAN, P., PLEROU, V., AND STANLEY, H. A theory of power-law distributions in financial market fluctuations. *Nature 423* (2003), 267.

[38] GAVEAU, B., AND SCHULMAN, L.-S. Dynamical distance: coarse grains, pattern recognition, and network analysis. *Bulletin des Sciences Mathématiques 129*, 8 (2005), 631–642.

[39] GFELLER, D., AND DE LOS RIOS, P. Spectral coarse graining of complex networks. *Physical Review Letters 99*, 3 (2007), 038701.

[40] GFELLER, D., AND DE LOS RIOS, P. Spectral coarse graining and synchronization in oscillator networks. *Physical Review Letters 100*, 17 (2008), 174104.

[41] GIADA, L., AND MARSILI, M. Data clustering and noise undressing of correlation matrices. *Phys. Rev. E 63*, 6 (May 2001), 061101.

[42] GODSIL, C. D., AND ROYLE, G. *Algebraic Graph Theory*. Springer, 2001.

[43] GOLUB, G. H., AND VAN LOAN, C. F. *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1996.

[44] GRADSHTEYN, L., AND RYZHIK, L. Table of integrals, series, and products.

[45] GRINOLD, R. C., AND KAHN, R. N. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Selecting Superior Returns and Controlling Risk*. McGraw-Hill, October 1999.

[46] GUARANTY, M., LONGERSTAEY, J., AND SPENCER, M. Riskmetrics technical document. Tech. rep., JP Morgan Market Risk Research Group, 1996.

[47] HÄRDLE, W., AND KIRMAN, A. Nonclassical demand: A model-free examination of price-quantity relations in the Marseille fish market. *Journal of Econometrics 67*, 1 (1995), 227–257.

[48] HARTIGAN, J. A., AND WONG, M. A. A K-means clustering algorithm. *Applied Statistics 28* (1979), 100–108.

[49] HORN, R. A., AND JOHNSON, C. R. *Matrix Analysis*. Cambridge University Press, 1990.

[50] IPSEN, I. C. F. Relative perturbation results for matrix eigenvalues and singular values. *Acta Numer. 7* (1998), 151–201.

[51] IPSEN, I. C. F., AND MEYER, C. The angle between complementary subspaces. *American Mathematical Monthly 110* (1995), 904–911.

[52] Jain, A. K., Murty, M. N., and Flynn, P. J. Data clustering: a review. *ACM Comput. Surv. 31*, 3 (1999), 264–323.

[53] Jefferies, P., Hart, M., Hui, P., and Johnson, N. From market games to real-world markets. *Eur. Phys. J. B 20* (2001), 493–502. cond-mat/0008387.

[54] Jolliffe, I. T. *Principal Component Analysis*, 2nd ed. Springer-Verlag, New York, 2002.

[55] Knops, Z. F., Maintz, J. B. A., Viergever, M. A., and Pluim, J. P. W. Normalized mutual information based registration using k-means clustering and shading correction. *Medical Image Analysis 10*, 3 (2006), 432–439.

[56] Lafon, S., and Lee, A. B. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE transactions on pattern analysis and machine intelligence 28*, 9 (2006), 1393–403.

[57] Larsson, O., Golub, G. H., and Alter, O. A tensor higher-order singular value decomposition for integrative analysis of dna microarray data from different studies. *Proceedings of the National Academy of Sciences 104*, 47 (2007), 18371–18376.

[58] Lillo, F. Limit order placement as an utility maximization problem and the origin of power law distribution of limit order prices. *Eur. Phys. J. B 55*, 4 (2007), 453–459. physics/0612016.

[59] Lloyd, S. Least squares quantization in pcm. *IEEE Trans. Inform. Theory 28*, 2 (1982), 129–137.

[60] Lux, T., and Marchesi, M. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature 397* (1999), 498–500.

[61] Mantegna, R., and Stanley, H. G. *Introduction to Econophysics*. Cambridge University Press, 2000.

[62] Marčenko, V., and Pastur, L. Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics 1*, 4 (1967), 457–483.

[63] Markowitz, H. *Portfolio selection: efficient diversification of investments*. New Haven, 1959.

[64] Maslov, S., Zhang, Y.-C., and Marsili, M. Dynamical optimization theory of a diversified portfolio. *Physica A 253* (1998), 403–418.

[65] Medo, M., Yeung, C. H., and Yi-Cheng, Z. How to quantify the influence of correlations on investment diversification. Quantitative finance papers, arXiv.org, 2009.

[66] MURDOCK, W., AND SZEGÖ, G. *On the eigenvalues of certain Hermitian forms.* J. Rational Mech. Anal, 1953.

[67] NELSON, D. Stationarity and persistence in the garch (1, 1) model. *Econometric theory* (Jan 1990).

[68] PAFKA, S., POTTERS, M., AND KONDOR, I. Exponential weighting and random-matrix-theory-based filtering of financial covariance matrices for portfolio optimization. *arXiv cond-mat.stat-mech* (Feb 2004).

[69] PLEROU, V., GOPIKRISHNAN, P., ROSENOW, B., AMARAL, L. A. N., GUHR, T., AND STANLEY, H. E. Random matrix approach to cross correlations in financial data. *Phys. Rev. E 65*, 6 (Jun 2002), 1–18.

[70] POTTERS, M., BOUCHAUD, J.-P., AND LALOUX, L. Financial applications of random matrix theory: Old laces and new pieces. *Arxiv preprint physics/0507111* (2005).

[71] RAFFAELLI, G., AND MARSILI, M. Dynamic instability in a phenomenological model of correlated assets. *J. Stat. Mech.* (2006). doi:10.1088/1742-5468/2006/08/L08001.

[72] SCHWARZKOPF, Y., FARMER, J., AND SET, I. Time Evolution of the Mutual Fund Size Distribution.

[73] SHARPE, W. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance 19*, 3 (1964), 425–442.

[74] SILVERSTEIN, J., AND CHOI, S. Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis 54*, 2 (1995), 295–309.

[75] SLANINA, F., AND ZHANG, Y.-C. Dynamical spin-glass-like behavior in an evolutionary game. *Physica A 289* (2001), 290–300.

[76] SOLOMON, S., AND RICHMOND, P. Power laws of wealth, market order volumes and market returns. *Physica A: Statistical Mechanics and its Applications 299*, 1-2 (2001), 188–197.

[77] SUROWIECKI, J. *The wisdom of crowds*. Anchor Books, New York, NY, 2005.

[78] THURNER, S., FARMER, J., AND GEANAKOPLOS, J. Leverage causes fat tails and clustered volatility. *Preprint*.

[79] TSAY, R. S. *Analysis of Financial Time Series*, 1st ed. Wiley-Interscience, October 2001.

[80] Tumminello, M., Aste, T., Di Matteo, T., and Mantegna, R. N. A tool for filtering information in complex systems. *PNAS 102*, 30 (July 2005), 10421–10426.

[81] Vaglica, G., Lillo, F., Moro, E., and Mantegna, R. Scaling laws of strategic behavior and size heterogeneity in agent dynamics. *Physical Review E 77*, 3 (2008), 36110.

[82] Van de Venter, G., and Michayluk, D. A longitudinal study of financial risk tolerance. Tech. rep., 2009. `http://www.fma.org/Reno/Papers/Longitudinal_Financial_Risk_Tolerance.pdf`.

[83] Veld, C., and Veld-Merkoulova, Y. V. The risk perceptions of individual investors. *Journal of Economic Psychology 29*, 2 (April 2008), 226–252.

[84] Yakovenko, V. Econophysics, statistical mechanics approach to. *Encyclopedia of Complexity and System Science, Springer http://refworks. springer* (2007).

[85] Zovko, I., and Farmer, J. D. The power of patience: A behavioral regularity in limit order placement. *Quant. Fin. 2* (2002), 387.

[86] Zumbach, G. The empirical properties of large covariance matrices. Quantitative finance papers, arXiv.org, 2009.

[87] Zumbach, G. Inference on multivariate arch processes with large sizes. Quantitative finance papers, arXiv.org, 2009.