Jagellonian University
Faculty of Physics, Astronomy and Applied Computer Science
Marian Smoluchowski Institute of Physics
Theory of Complex Systems Department

Małgorzata Snarska

Applying Free Random Variables to the Analysis of Temporal Correlations in Real Complex Systems

Submitted in partial fulfilment of the requirements
of the degree of Doctor of Philosophy in Physics

Thesis Supervisor: prof. dr hab.Maciej A. Nowak

Cracow, 2010

Table of Contents

List of Figures

List of Tables

Acknowledgements

There are many people who contributed to my thesis and many events that influenced my work during these last few years.
I would like to express my deep appreciation and sincere gratitude to prof. Maciej A.Nowak, my supervisor for his invaluable wisdom. His patience clues and excellent comments, whenever I was excited about a new challenge, have always guided me in the right direction from the beginning to the end in this course of my research. He taught me how to ask questions and express my ideas. He showed me different ways to approach a research problem and the need to be persistent to accomplish any goal. Without him this thesis would not come to life so quickly and smoothly.
I am also indebted to prof. Zdzisław Burda who offered me a lot of friendly help. He transferred to me his wide knowledge in pursuing Monte Carlo experiments and doing numerical simulations at all. Our conversations enlightened my way of thinking about mathematical software packages like Matlab and Mathematica. His insight to numerical simulations is second to none.
I would like to extend my heartiest thanks to Andrzej Jarosz for his kind involvement and collaboration on much of the work presented in chapter 4. Our cooperation started extremely badly, but in the end it resulted with beautiful results and our conversations also helped me realize that trust should be treated as a hidden dimension in real life science.
It is a great pleasure to say thanks also to Jakub Krzych for showing me not only how to become a good programmer, but also that Stock Exchange might be an excellent playground for theoretical physics.
I have been fortunate and delighted to interact with prof. Marcel Ausloos. His perpetual energy and enthusiasm in tackling problems has motivated and inspired me during my PhD studies. Thank you for demonstrating me, the magic behind the complex systems science.
I am also grateful to prof. Jacek Osiewalski for giving me freedom and support in developing my own ideas and my colleagues and members of Department of Econometrics and Operations Research Cracow University of Economics, who improved this thesis with useful comments and always constructive criticism. The associated experience broadened my perspective on the practical aspects of this study.
I will never forget all kind people around me, who gave interesting feedback and valuable suggestions, but to whom it was not possible to give a particular mention here. I do appreciate and thank to all of them.
Last but not least I would like to thank to my family and friends for always keeping up with me and for dragging me away from my computer whenever it was possible and reminding me that PhD thesis is not the only thing that matters in life.

What we call chaos is just patterns we haven't recognized. What we call random is just patterns we can't decipher. What we can't understand we call nonsense

Survivor,Chuck Palahniuk

0 The study of real complex systems- general introduction

The analysis of complex systems in a unified framework has become recognized in the recent years as a new scientific discipline. Its development is strongly affected by the advances in many branches of science, ranging from physics, multivariate analysis to economics and quantitative finance. The existing literature on the subject is tremendous. The most common approach in complex systems research, is to try to establish universal laws and search for phenomena governing their behavior. Qualitalively to understand the behavior of a complex system we must understand not only the behavior of individual parts, but also how they act together and how they are influenced by e.g. external distortion. The dynamics of such a system is usually nontrivial.

0.1 Dynamical properties and Random Matrix Theory

Random matrix theory – a new tool in complexity studies is already changing the point of view on large random systems and how we understand their behavior([DFGZJ95],[GMG98],[Ver94]). Recent explosion of mathematical techniques in fields like quantum gravity, quantum chromodynamics etc. suddenly become crucial in handling in a neat way thousands of data sets omnipresent in so many areas of science. Nowadays it's extremely hard to find a branch of theoretical physics or any quantitative discipline, where random matrix models can be completely ignored. Surprisingly, most of the studies are limited to building models, that are able to reproduce accurately only the static (equilibrium) properties of complex structures. Complex systems in nature, on the other hand, exhibit potentially a very rich dynamical structure, represented by a mixture of eg. different stochastic processes on various time scales. It is therefore tempting to enrich random matrix model with an additional dynamical parameter (e.g. real time, temperature, area of the string, length of mesoscopic wire, finite volume etc). These fluctuating structures may capture new features and new scaling phenomena of systems affected by bewildering number of variables and observations.

0.2 Outline of the present work

Spatio-temporal structure of correlations and optimal forecasts of future correlations is a task of major importance. However, the information about cross-correlations and their temporal dynamics is usually inferred from historical (past) observations, which are inevitably contaminated by measurement noise and it is a constant challenge to unscramble signal from noise.

The main purpose of this thesis is to study the dynamical properties of real complex systems such as e.g. economy and/or financial market by looking at their spectral properties eg. density of eigenvalues under umbrella of Free Random Variables Calculus and Random Matrix Theory. The thesis can be roughly decomposed into two parts. The first one was meant as quick guide to methodology and tools used.

- In first chapter we give brief insight into methods of analyzing real complex systems, particularly focusing on correlation analysis and classical dimension reduction techniques and describe the distorting effects when the number of samples N is large and comparable to the number of observations T (Random Matrix Theory setting).
- Next we introduce Free Probability Theory and Free Random Variables as an analogue to classical probability theory and a powerful alternative to standard random matrix theory in such a complex noncommutative setting. FRV may be considered as mathematical framework for dealing with random variables that do not commute (*i.e.*, large random matrices can be regarded as free random variables). It's cornerstone is the notion of freeness, which can be viewed as non–commutative counterpart of classical independence of random variables. As such, it introduces new quality to RMT version of complex systems analysis, not only allows to extend many classical results eg. Central Limit Theorem for Gaussian Random Matrices and Marčenko–Pastur equation in terms of Free Poisson distribution, but also simplifies conceptually and technically many random matrix calculations.

Recent years have witnessed a rapid growth in data acquisition and a number of applications of large panel data sets emerged. The lack of methods and systematic description of complex systems resulted in mismatch between empirical findings and theory. The second part of this thesis is written in the spirit of applicability of Free Probability Theory to analysis data sets where the number of experimental variables is comparable to the size of the sample. While most of the work presented in the last three chapters is original, some is the result of collaborative work or the work of others and is cited accordingly in the text. Throughout the chapters we will assume, that cross–correlations of N variables can be described by the two–point covariance (correlation) function,

$$\mathcal{C}_{ia,jb} \equiv \langle X_{ia} X_{jb} \rangle \,. \tag{1}$$

For $X_{ia} \equiv x_{ia} - \langle x_{ia} \rangle$, which describe the fluctuations (with zero mean) of the returns around the trend, and collect them into a rectangular $N \times T$ matrix $\mathbf{X}$. The average $\langle \ldots \rangle$ is understood as taken according to some probability distribution whose functional shape is stable over time, but whose parameters may be time–dependent. In chapter 3 and 4 we will employ a very simplified form of the two–point covariance function (5.1), namely with cross–covariances and auto–covariances factorized, non–random, and decoupled the temporal dependence of the distribution of variable is the same, and the structure of cross–correlations does not evolve in time

$$\mathcal{C}_{ia,jb} = C_{ij} A_{ab} \tag{2}$$

With coefficient assembled into a $N \times N$ cross–covariance matrix $\mathbf{C}$ and a $T \times T$ auto–covariance matrix $\mathbf{A}$; both are taken symmetric and positive–definite). We will discover that the matrix of "temporal covariances" $\mathbf{A}$ is a way to model two temporal effects: the (weak, short–memory) lagged correlations between the returns, as well as the (stronger, long–memory) lagged correlations between the volatilities. On the other hand, the matrix of cross–covariances ("spatial

covariances," using a more physical language) $\mathbf{C}$ models the hidden factors affecting the assets, thereby reflecting the structure of mutual dependencies between variables. For our approach to be valid, both covariance matrices obviously must be finite, which is acquired by assuming the multivariate Gaussian distribution for the random variables, which displays the two–point covariances (5.2),

$$\begin{aligned} P_{\text{c.G.}}(\mathbf{X})\mathrm{D}\mathbf{X} &= \frac{1}{\mathcal{N}_{\text{c.G.}}} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{N}\sum_{a,b=1}^{T} X_{ia}\left[\mathbf{C}^{-1}\right]_{ij} X_{jb}\left[\mathbf{A}^{-1}\right]_{ba}\right)\mathrm{D}\mathbf{X} = \\ &= \frac{1}{\mathcal{N}_{\text{c.G.}}} \exp\left(-\frac{1}{2}\mathrm{Tr}\mathbf{X}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{X}\mathbf{A}^{-1}\right)\mathrm{D}\mathbf{X} \end{aligned} \tag{3}$$

where the normalization constant $\mathcal{N}_{\text{c.G.}} = (2\pi)^{NT/2}(\mathrm{Det}\mathbf{C})^{T/2}(\mathrm{Det}\mathbf{A})^{N/2}$, and the integration measure $\mathrm{D}\mathbf{X} \equiv \prod_{i,a}\mathrm{d}X_{ia}$; the letters "c.G." stand for "correlated Gaussian," and the expectation map w.r.t. this distribution will be denoted by $\langle\ldots\rangle_{\text{c.G.}}$, while "$^{\mathrm{T}}$" denotes matrix transposition.

- Chapter 3 is devoted to the analysis of dynamical properties of equal–time correlations matrices on an example of Warsaw Stock Exchange data. We extend the results from [KS06], taking as starting point the case of ordinary Wishart ensemble [Wis28], for which a spectral density is well known as Marčenko–Pastur (Bai-Silverstein) distribution [MP67, SB95]. In this setting we will develop one–factor cleaning technique and check the stability of eigenvalues spectrum over time. Furthermore we try different weighted schemes for empirical cross–correlations in order to put more importance to the more recent data.

- In chapter 4 we look more closely at large covariance matrices generated by ARMA processes. Finite order vector autoregressive moving average models (VARMA) are motivated by Wold decomposition theorem [Wol38] as an appriopriate multivariate setting for studying the dynamics of stationary time series. The main goal of chapter 5 is to show how random matrix theory can be applied to derive spectral density of sample covariance matrices generated by multivariate VMA(q), VAR(q) and VARMA(q_1, q_2) processes in a limit where the number of random variables N and the number of consecutive time measurements T are both large but the ratio N/T is fixed. In this regime the underlying random matrices are asymptotically equivalent to Free Random Variables and FRV calculus can be applied to calculate the eigenvalue density of the sample covariance for several VARMA–type processes and to explicitly solve the VARMA$(1,1)$ case. The proposed method is purely algebraic and can be easily generalized to $q_1 > 1$ and $q_2 > 1$. The results are then confirmed by a Monte - Carlo simulation. We also present application to the real data set - Polish macroeconomic data. Ideas and methods presented in this chapter were first presented in [BJNS10] and are repeated in this chapter with minor changes only.

- Chapter 5 follows slightly different approach, first presented by [BLMP07]. The idea is to divide all variables into two subsets *i.e.*, focus on N input factors X_a $a = 1, \ldots, N$ and M output factors Y_α $\alpha = 1, \ldots, M$ with the total number of observations being T and remove potential correlations inside each subset in order to avoid interferences with the out-of-sample signal. Then one builds an empirical rectangular $M \times N$ correlation matrix and compare its singular value spectrum with a benchmark obtained using Random Matrix Theory results, assuming there are no correlation between the variables. We extend the results obtained by the author [Sna08] for the data set from previous chapter.

Most of the ideas presented in here have been already published, they have been revised, completely redone or at least greatly expanded.

The sun comes up just about as often as it goes down, in the long run, but this doesn't make its motion random.

Donald E. Knuth

1 The nature of correlations

The analysis and measurement of dependence between variables, between sets of variables and between variables and sets of variables are fundamental tools of multivariate or complex systems analysis. In many real cases one has to describe the system by a large number of possible factors of large number of observations.
In the following chapter we briefly review classical methods for identifying universal behaviors in complex systems (correlation analysis, regression analysis) and main technical tools for reducing the complexity and factor analysis (i.e. dimension reduction techniques - PCA - Principal Component Analysis, FCA- Factor Component Analysis). For a more concise description of the methods c.f.[Eve06] or any book covering the subject of multivariate analysis.

1.1 Correlation and Covariance matrix estimation and analysis in classical setting

Searching for patterns, rules and universalities is the heart of any quantitative discipline. Regression and correlation analysis are certainly the most important tools at modern multivariate analysis disposal with its applications in various quantitative branches of science like e.g. physics, time series analysis. It is of particular interest in analyzing complex behavior of economy and financial markets [PGR+99, BLMP07].

1.1.1 Covariance and Correlation Matrix - Basic Notions

Dispersion Matrix

Covariance matrix or dispersion matrix is a matrix of covariances between elements of a random vector. It is the natural generalization to higher dimensions of the concept of the variance of a scalar-valued random variable. If entries in the column vector

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} \tag{1.1}$$

are random variables, each with finite variance, then the covariance matrix Σ is the matrix whose (i, j) entry is the covariance

$$\Sigma_{ij} = \operatorname{cov}(X_i, X_j) = \mathrm{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right] \tag{1.2}$$

where $\mu_i = \mathrm{E}(X_i)$ is the expected value of the ith entry in the vector X. In other words, we have

$$\Sigma = \begin{bmatrix} \mathrm{E}[(X_1 - \mu_1)(X_1 - \mu_1)] & \cdots & \mathrm{E}[(X_1 - \mu_1)(X_N - \mu_N)] \\ \mathrm{E}[(X_2 - \mu_2)(X_1 - \mu_1)] & \cdots & \mathrm{E}[(X_2 - \mu_2)(X_N - \mu_N)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_N - \mu_N)(X_1 - \mu_1)] & \cdots & \mathrm{E}[(X_N - \mu_N)(X_N - \mu_N)] \end{bmatrix}. \tag{1.3}$$

The definition above is equivalent to the matrix equality

$$\Sigma = \mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])\,(\mathbf{X} - \mathrm{E}[\mathbf{X}])^{\top}\right] \tag{1.4}$$

This form can be seen as a generalization of the scalar-valued variance to higher dimensions. Recall that for a scalar-valued random variable X

$$\sigma^2 = \mathrm{var}(X) = \mathrm{E}[(X - \mu)^2], \tag{1.5}$$

here $\mu = \mathrm{E}(X)$.

Covariance Matrix Estimator

From practical reasons one often defines an $N \times T$ matrix $\mathbf{X}$, where each row of $\mathbf{X}$ corresponds to N measurements of a particular type and each column contains a set of T measurements from particular trial (sample). Then sample covariance matrix estimator (Pearson's estimator)is often defined as:

$$C_X = \frac{1}{T}\mathbf{X}\mathbf{X}^{\mathbf{T}} \tag{1.6}$$

C_X captures the covariance between all possible sets of measurements and reflects the noise and redundancy in our measurements. In the diagonal terms, by assumption, large values correspond to interesting structure, while the off-diagonal terms large magnitudes correspond to high redundancy.

The Correlation Matrix

The correlation matrix of N random variables $X_1, \ldots, X_N$ is the $N \times N$ matrix whose (i, j) entry is is the Pearson's correlation coefficient ϱ_{X_1,X_2} between two random variables X_1 and X_2 with expected values μ_{X_1} and μ_{X_2} and standard deviations σ_{X_1} and σ_{X_1} is defined as:

$$corr(X_1, X_2) = \varrho_{X_1,X_2} = \frac{cov(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} = \frac{E[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})]}{\sigma_{X_2}\sigma_{X_1}} \tag{1.7}$$

where E is the expected value operator and *cov* means covariance.
If the measures of correlation used are product-moment coefficients, or if for simplicity one assumes, that random variables come from the distribution for which second moment - variance $\sigma(X_i)$ exists and is finite (like in Gaussian case for instance), the correlation matrix is the same as the covariance matrix of the standardized random variables $\frac{X_i}{\sigma(X_i)}$ for $i = 1, \ldots, N$.

1.1.2 Correlation analysis

The correlation coefficient ϱ measures the degree of linear association between two variables. If a pair of variables is said to be correlated, it means both variables are treated in a completely symmetric way. Thus it is not implied that changes in one variable cause changes in the second one or vice versa. Rather it is simply stated, that there is an evidence for linear relationship between the two and the movements in them are on average related to an extent given by correlation coefficient.

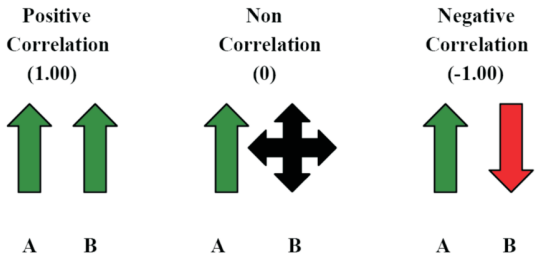

Figure 1.1: If the correlation coefficient ϱ is of order $+1$, two variables are perfectly correlated. If one variable gains in value, one would expect the other one to gain as well. Correlation coefficient of 0 simply states there is no correlation between two random variables, while the coefficient of -1 indicates that these two move in opposite directions.

1.1.3 Regression analysis

While in correlation analysis one simply identifies strength and direction of a relation between pair of random variables and we do not bother with causation, regression takes the analysis one step further, trying to fit equation to the data. In very general terms regression is concerned with describing and evaluating the relationship between given set of explained or dependent variables and one or more other variables. More specifically regression is an attempt to explain movements in a variable by reference movements of set of other variables. To choose the appropriate set of regressors and regressands one have to follow the rule, that regressands have to be at most weakly correlated with each other but strongly correlated with the set of explanatory variables.

1.2 Large N and large T limit issues

Today high amount of data is stored in the memories of computers in the form of huge matrices. Typical examples include financial markets data, wireless technology, gene expression networks etc. These data are usually blurred by the high amount of noise due to finiteness of the sample and are rapidly affected by the dimensionality curse. In this section we give some comments on possible issues when dealing when the number of possible variables present in the system is large compared to the sample size.

1.2.1 Dimensionality curse

In the absence of information on the phenomenon under study, a brute force strategy would consist in listing a large number of possible variables, and systematically look for correlations between pairs, in the hope of finding some significant signal (i.e. relevant variables). This procedure is rapidly affected by the "dimensionality curse ", also called the problem of sunspot

or dummy variables in the literature [Woo90]. Since the number of observations is always limited, it can happen that two totally unrelated phenomena appear to be correlated over a certain time interval T. More precisely, the correlation coefficient ρ, which would (presumably) be zero if very long time series could be studied, is in fact of the order of $1/\sqrt{T}$ and can be accidentally large. When one tries to correlate systematically N input variables with M output variables, the number of pairs is NM. In the absence of any true correlation between these variables, the largest of these NM empirical correlation coefficients will be, for Gaussian variables, of order $\rho_{\max} \sim \sqrt{2\ln(NM)/T}$, which grows with NM. If the input and output variables are non Gaussian and have fat-tails, this number can be even larger. If two strongly fluctuating random variable accidentally take large values simultaneously, this will contribute a lot to the empirical correlation even though ρ should be zero for large T [Bel61].

1.2.2 Spurious relationships

When dealing with real data one also has to be very careful in order to avoid so called spurious relationships. Spurious relationship (spurious regression or spurious correlation) is a mathematical relationship in which two occurrences have no causal connection, yet it may be inferred that they do, due to a certain third, unseen factor (referred to as a "confounding factor" or "lurking variable"). The spurious relationship gives an impression of a worthy link between two groups that is invalid when objectively examined. When the effects of the lurking variable are removed, they are said to have been partialed out. A spurious correlation is sometimes called an "illusory correlation". In that case, "spurious" is then reserved for the special case in which a correlation is not present in the original observations but is produced by the way the data are handled [GN74].

1.3 Classical methods for reducing the complexity- factor analysis

Availability of many time series and over very long span is today an inevitable fact. While more data at scientists disposal provide the opportunity to understand behavior of complex systems better, the researchers can also suffer from an information overload without some way to organize the data into an easy to interpret manner. The scope of this section is to describe principles of factor analysis techniques as the means of reducing the dimensionality and detecting the structure of relationships between variables.

1.3.1 General Purpose

The main applications of factor analytic techniques are to reduce the number of variables and to detect structure in the relationships between variables, that is to classify variables. Therefore, factor analysis is applied as a data reduction or structure detection method (the term factor analysis was first introduced by Thurstone, 1931). We will assume that the Reader is familiar with the basic logic of statistical reasoning and concepts of variance and correlation. There are many excellent books on factor analysis, especially Principal Component Analysis. For example, a hands-on how-to approach can be found in [Shl05] or [BN08].

1.3.2 Factor Analysis as a Data Reduction Method

Suppose we have two variables, that are highly correlated. Given a high correlation between the two items, we can conclude that they are quite redundant. One can summarize the correlation between two variables in a scatterplot. A regression line can then be fitted that represents the "best" summary of the linear relationship between the variables. If we could define a variable that would approximate the regression line in such a plot, then that variable would capture most of the "essence" of the two items. In a sense we have reduced the two variables to one factor, where the new factor is actually a linear combination of the two variables. One can easily extend the above example to multivariate case, then the computations become more involved, but the basic principle of expressing two or more variables by a single factor remains the same.

1.3.3 Principal Components

Principal Component Analysis (PCA) is quite common method in multivariate analysis. Mathematically it is defined as an orthogonal linear transformation that takes the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. With minimal effort PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified structures that often underlie it. We do not want to go into the details about the computational aspects of principal components analysis here, which can be found elsewhere. There is a huge literature like e.g.[Jol02].[1] Basically, the extraction of principal components amounts to a variance maximizing rotation of the original variable space. Suppose we have N variables $X_1, \ldots, X_N$. For a data matrix X^T with zero empirical mean (the empirical mean of the distribution has been subtracted from the data set), where each row represents a different repetition of the experiment, and each column gives the results from a particular probe, the PCA transformation is given by:

$$\mathbf{Y}^T = \mathbf{X}^T\mathbf{W} = \mathbf{V}\mathbf{\Sigma}, \tag{1.8}$$

where W, Σ, V^T is the singular value decomposition(SVD) of X and $\Sigma = (\sigma_{k,k'})$ is the covariance matrix composed of the mean-corrected second moments.

$$\sigma_{k,k'} = cov(X_k, X_{k'}) = E\left[(X_k - \mu_k)(X_{k'} - \mu_{k'})\right] \tag{1.9}$$

The goal is to reduce dimensionality by constructing a smaller number of $W = \mathbf{V}\mathbf{\Sigma}$ having the variance

$$Var(W) = \mathbf{V^T}\mathbf{\Sigma}\mathbf{V} \tag{1.10}$$

To concentrate the variation in as few consecutive factors as possible, one looks for vectors that maximize $Var(W)$. Given a set of points in Euclidean space, the first principal component (the eigenvector with the largest eigenvalue) corresponds to a line that passes through the mean and minimizes sum squared error with those points. The second principal component corresponds to the same concept after all correlation with the first principal component has been subtracted out from the points. Each eigenvalue indicates the portion of the variance that is correlated with each eigenvector. Thus, the sum of all the eigenvalues is equal to the sum squared distance of the points with their mean divided by the number of dimensions. PCA essentially rotates

[1] For a nice and pedagogical introduction see [Smi02]

Quick Summary of PCA

Organize data as $N \times T$ matrix,
where N is the number of variables and T is the number of observations
Subtract off the mean from each measurement type
Calculate the SVD or the eigenvalues and eigenvectors of Covariance Matrix
Select a subset of the eigenvectors as the new basis
Standardize the dataset and project onto the new basis

Table 1.1: Review of Principal Components Analysis

the set of points around their mean in order to align with the first few principal components. This moves as much of the variance as possible (using a linear transformation) into the first few dimensions. The values in the remaining dimensions, therefore, tend to be highly correlated and may be dropped with minimal loss of information. Successive linear combinations are sought that are orthogonal to those previously chosen. The principal component eigenvalues λ_j and principal component eigenvectors V_j are thus obtained from:

$$\lambda_j = \max V^T \Sigma V : (V^T V)_{j'} = 0; j' < j, |V| = 1 \tag{1.11}$$

As we extract consecutive factors, they account for less and less variability. The decision of

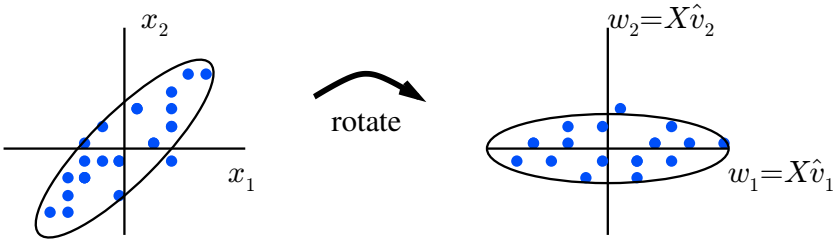

Figure 1.2: We can think of the regression line as the original X axis, rotated so that it approximates the regression line. After we have found the line on which the variance is maximal, there remains some variability around this line. In principal components analysis, after the first factor has been extracted, that is, after the first line has been drawn through the data, we continue and define another line that maximizes the remaining variability, and so on. In this manner, consecutive factors are extracted. Because each consecutive factor is defined to maximize the variability that is not captured by the preceding factor, consecutive factors are uncorrelated and orthogonal to each other. In this two-dimensional picture we might keep the first direction and discard the second

when to stop extracting factors is arbitrary and basically depends on when there is only very little "random" variability left. Without further ado, we are extracting factors that account for less and less variance. To simplify matters, one usually starts with the correlation matrix, where the variances of all variables are equal to 1.0. Therefore, the total variance in that matrix is equal to the number of variables. The eigenvalues of the correlation matrix are usually related to the variance extracted by the factors. The problem of how many factors to retain is now related to the problem of keeping the arbitrary number of eigenvalues.

1.3.4 Factor Component Analysis

One can also think of the data reduction problem in another way. First, there can exist some underlying common factors. Each item measures some part of this common factor. Second, each item also captures a unique aspect that is not addressed by any other item. Then we should not expect that the factors will extract all variance from our items; rather, only that proportion that is due to the common factors and shared by several items. In the language of factor analysis, the proportion of variance of a particular item that is due to common factors (shared with other items) is called communality. Therefore, an additional task facing us when applying this model is to estimate the communalities for each variable, that is, the proportion of variance that each item has in common with other items. The proportion of variance that is unique to each item is then the respective item's total variance minus the communality. The defining characteristic then that distinguishes between the two factor analytic models is that in principal components analysis we assume that all variability in an item should be used in the analysis, while in principal factor analysis we only use the variability in an item that it has in common with the other items. In most cases, these two methods usually yield very similar results. However, principal components analysis is often preferred as a method for data reduction, while principal factors analysis is often preferred when the goal of the analysis is to detect structure.

1.4 Asymptotic N/T limit - focus on Random Matrix Theory

The goal of PCA is to identify the most meaningful basis to re-express a data set. The hope is that this new basis will filter out the noise and reveal hidden structure. However, when extracting principal components one must use covariance matrix, which for $N \to \infty, T \to \infty, \frac{N}{T} = r, r \in [0, \infty)$ becomes unreliable [2]. The complexity of finite size distributions makes the use of asymptotic approximations appealing. Traditional statistical approach keeps the number of variables N fixed while letting the sample size $T \to \infty$. This is no longer valid in the presence of modern data. Interestingly the distorting effects of high dimensionality upon covariance matrix eigenvalue spectra and eigenvectors are well known from Random Matrix Theory [Joh07], which inform us about the expected sample covariance eigenvalue spectrum in the above limit and consequently about limits of any procedure which is based on spectral decomposition of covariance matrix.

From its origin Random Matrix Theory has been heavily influenced by its applications in physics, multivariate statistics and engineering. The landmark contributions from Wishart(1928) [Wis28], Wigner(1955)[Wig55],[Wig58] and Marčenko and Pastur(1967)[MP67] were motivated to a large extent by practical experimental problems. In this section we review a wide range of results, which are relevant to the analysis of universal statistical properties of large random matrices, that arise in the context of real complex systems. We start with some results that are relevant to the analysis of statistics of random matrices and then we simply move on to the large N and large T limit.

[2] We will often call this setting "thermodynamic limit"

1.4.1 Empirical correlations estimator - Challenge for Free Probability Theory

It is obvious, that any historical correlation matrix' estimator is inevitably marred by the measurement noise; it will reflect the true covariances only to a certain degree, with a superimposed broadening due to the finiteness of the time series. More precisely, there are $N(N+1)/2$ independent elements in $\mathbf{C}$, to be estimated from NT measured quantities $\mathbf{Y}$, hence the estimation accuracy will depend on the "rectangularity ratio",

$$r \equiv \frac{N}{T}; \tag{1.12}$$

the closer r to zero, the more truthful the estimate. This is a cornerstone of classical multivariate analysis. Unfortunately, a practical situation will typically feature a large number of variables sampled over a comparably big number of time snapshots, so that we may approximately talk about the "thermodynamical limit,"

$$N \to \infty, \qquad T \to \infty, \qquad \text{such that} \qquad r = \text{fixed}. \tag{1.13}$$

On the other hand, it is exactly this limit in which the Free Random Variables calculus (see chapter 2 for its brief elucidation) can be applied; hence, the challenge of de–noising is somewhat counterbalanced by the computationally powerful FRV techniques.

1.4.2 Gaussian and Wishart Distributions

The simplest assumption about the data is that N random variables $X_1, \ldots, X_N$ follow a N-variate Gaussian distribution $N_N(\mu, \Sigma)$[3] with probability density function given by

$$\mathbf{P}_{\beta,\mathbf{N}}(\mathbf{X}) = \det(\sqrt{2\pi}\Sigma)^{-1/2} \exp(\frac{-\beta N}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)) \tag{1.14}$$

It is common assumption, that data sample consists of T independent draws from $X \sim N_N(\mu, \Sigma)$, collected into a $N \times T$ matrix X.
The non-normalized cross - product matrix $H = XX^T$ is said to have $N-variate$ [Wis28] with T degrees of freedom, with the density function supported on the cone of non-negative definite matrices:

$$\mathbf{P}_{\beta,\mathbf{N}}(\mathbf{H}) = c^{\beta}_{T,N} \det(\Sigma)^{-\beta T/2} \det H^{\beta(T-N-1)/2} \exp\{-\frac{\beta}{2} tr(\Sigma^{-1}H)\} \tag{1.15}$$

with normalization constant $c^{\beta}_{T,N}$, positive defined Σ and $T \geq N$.

1.4.3 Joint Density of Eigenvalues

The joint distribution of eigenvalues for principal components is defined as follows:

$$\varrho_{\beta}(\lambda) = f(\lambda_1, \ldots, \lambda_N) = c_{\beta} \prod_i \lambda_i^{\beta T/2 - 1 - \beta(N-1)/2} \exp(-\lambda_i) \prod_{i<j} |\lambda_i - \lambda_j|^{\beta} \tag{1.16}$$

Where the normalization constant

$$c_{\beta} = 2^{-TN\beta/2} \prod_{i=1}^{N} \frac{\Gamma(1+\frac{\beta}{2})}{\Gamma(1+\frac{\beta}{2}i)\Gamma(\frac{\beta}{2}T - \frac{\beta}{2}(N-i))}$$

[3] In practice, when focusing on covariances we will assume for simplicity of calculations, that $\mu = 0$, since mean can always be subtracted first from real-life data.

Relation of eigendecomposition of the Wishart matrix to PCA.

Start with a Gaussian data Matrix $X \sim N_N(\mu, \Sigma)$

Form the covariance C_X, yielding a Wishart density for $H = T \cdot C_x$

The eigenvalues λ_{i_H} and eigenvectors U_i of H given by

$$HU_i = \lambda_{i_H} U_i \qquad \lambda_{1_H} \geq \ldots \lambda_{N_H} \geq 0$$

are related to principal eigenvalues λ_i and eigenvectors V_i via

$$\lambda_{i_H} = T\lambda_i \qquad U_i = V_i$$

Table 1.2: Eigendecomposition of Wishart Matrix and PCA

Then for $\beta = 1$ we have:

$$\varrho_{\beta=1}(\lambda) = f(\lambda_1, \ldots, \lambda_N) = \frac{2^{-TN/2}\pi^{N^2/2}}{\Gamma_N(N/2)\Gamma_T(T/2)} \prod_i \lambda_i^{T-N-1} \exp(-\lambda_i) \prod_{i<j}(\lambda_i - \lambda_j) \tag{1.17}$$

where $\Gamma_N(a) = \pi^{N(N-1)/4} \prod_{i=1}^{N} \Gamma(a - (i-1)/2)$ is a multivariate Gamma function. We should mention here, that general formulation of this problem takes the form:

$$f(\lambda_1, \ldots, \lambda_N) = c_\beta \prod_i^N w(\lambda_i)^{\beta/2} \prod_{i<j} |\lambda_i - \lambda_j|^\beta \tag{1.18}$$

Dyson [Dys62] showed that the scaling factor - β can be restricted to one of three values and $w(\lambda)$ is a weight function. We will restrict our deliberations in here to the real case, as

	Symmetry of Random Matrix	Matrix Entries
$\beta = 1$	orthogonal	real
$\beta = 2$	unitary	complex
$\beta = 4$	symplectic	quaternion

Table 1.3: Dyson[Dys62] symmetry classes, see eg.[Zir10] for a more deeper perspective.

$w(\lambda)=$	$e^{-\lambda^2/2}$	Gaussian
	$\lambda^a e^{-\lambda}$	Wishart
	$(1-\lambda)^a(1+\lambda)^b$	Double Wishart

Table 1.4: Three standard weight functions and corresponding probability distributions

it is a natural way to describe complex systems (economy, stock exchange etc.) in terms of distributions derived from Gaussian (see Appendix A for details).

1.4.4 Asymptotic limit for eigenvalues of a Covariance Matrix

Given the equation(1.16) it is tempting to conclude, that all eigenvalues are quite different from another. This spread is indeed an example of repulsion of eigenvalues given by Vandermonde term in (1.16). Marčenko and Pastur(1967)[MP67] first presented a systematic description of this phenomena. We consider only the case where $H \sim W_1(T, N)$. The empirical distribution

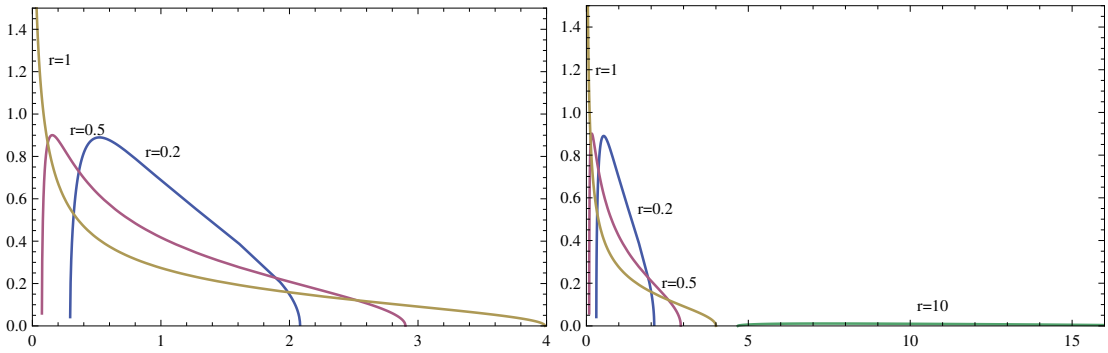

Figure 1.3: Simulation of Marčenko and Pastur limit density for different values of r. One can observe, that the actual distribution depends only on the ratio $r = N/T$. The larger the N compared to T the more spread is out in the limiting density

$F_N(\lambda) = N^{-1}\#\{\lambda_i \leq \lambda\}$, which counts how many eigenvalues fall below given value λ is related to the empirical spectral density function via

$$F'_N(\lambda) = \rho_N(\lambda) = \frac{1}{N}\left\langle \delta_{\lambda_1}(H) + \ldots + \delta_{\lambda_N}(H) \right\rangle, \tag{1.19}$$

where the expectation $\langle \ldots \rangle$ is taken w.r.t. the probability measure $P(\mathbf{H})D\mathbf{H}$ of the matrix $\mathbf{H}$. The empirical distribution has a limit density distribution if the sample size T and the number of variables N grow together $N/T \to r$

$$\varrho^{MP}(\lambda) = \left(1 - \frac{1}{r}\right)\delta(\lambda) + \frac{\sqrt{(\lambda - \lambda^-)(\lambda^+ - \lambda)}}{2\pi r\lambda} \tag{1.20}$$

where $\delta(x)$ is a Dirac delta and $\lambda_\pm = (1 \pm \sqrt{r})^2$. For $r < 1$ it is common to omit the term with the $\delta(\ldots)$ and concentrate only on the support limited to the interval $[\lambda_-, \lambda_+]$.

Things should be made as simple as possible, but not simpler.

Albert Einstein

2

Not so Short Introduction to Free Random Variables' Calculus

Free Probability Theory, sometimes referred to as "Probability Calculus of the XXI Century"[1] is a non-commutative probability theory, in which the concept of independence of classical probability theory is replaced by that of freeness. This idea incorporates both the probabilistic idea of no correlations involved and the algebraic notion of the absence of relations between eg. group generators. Furthermore Free Probability Theory, invented by Voiculescu [DNV92] in the context of operator algebras, has a very nice connection with Random Matrices as asymptotic models of free noncommutative variables (i.e in the large matrix size limit). In fact, free probability can be viewed as the theory providing concepts and notations without relying on random matrices, for dealing with the limit $N \to \infty, T \to \infty$ for large $N \times T$ random matrices but with deep connections with classical probability. The focus of this chapter is to point out main concepts of free probability and review some standard Random Matrix Theory results under umbrella of Free Random Variables calculus in an analogy to classical probability calculus of random variables avoiding rigorous mathematical proofs whenever its possible. Furthermore, we will most likely adhere to the survey provided in [BJJ+09] and the series of lectures [Now10].

2.1 Basics of Free Probability and Random Matrices

In this section we give a bunch of definitions of Free Probability theory, in analogy to classical random variables calculus. We assume the Reader possesses some standard knowledge in classical probability theory (eg. classical probability space, random variables and its distributions, moments, characteristic functions etc.) for details please refer to [Fel68].

2.1.1 Free Random Variables

We will start with some purely algebraic notions in free probability and provide a gentle and pedagogical synopsis on the subject of non–commutative probability theory. For a more detailed and accurate introduction please see [Voi97, ENV00, Bia98, Spe09]. Let's remind here that, the *empirical eigenvalue distribution* of a selfadjoint $N \times N$ matrix $\mathbf{H}$ is the probability measure on $\mathbb{R}$ which puts mass $1/N$ on each of the N eigenvalues λ_i of $\mathbf{H}$, counted with multiplicity. If $\mu_{\mathbf{H}}$ is determined by its moments m_n then it can be recovered from the knowledge of all traces of

[1] ©M.A.Nowak

powers of $\mathbf{H}$:

$$\frac{1}{N}\mathrm{Tr}(\mathbf{H}^k) = \frac{1}{N}\left\langle \lambda_1^k + \cdots + \lambda_N^k \right\rangle = \int_{\mathbb{R}} \lambda^k d\mu_{\mathbf{H}}(\lambda),$$

where by $\mathrm{tr} = \frac{1}{N}\mathrm{Tr}$ we denote the normalized trace on matrices (so that we have for the identity matrix 1 that $\frac{1}{N}\langle \mathrm{Tr}\rangle 1 = 1$), and the expectation $\langle \ldots \rangle$ is taken with respect to rotationally invariant probability measure $P(\mathbf{H})d\mathbf{H}$. This is the basis of the *moment method* which tries to understand the asymptotic eigenvalue distribution of a sequence of matrices by the determination of the asymptotics of traces of powers.

Definition 2.1.1. We say that a sequence $\{\mathbf{H}_N\}_{N\in\mathbb{N}}$ of $N \times N$ matrices *has an asymptotic eigenvalue distribution* if the limit $\lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}\left\langle \mathbf{H}_N^k \right\rangle$ exists for all $k \in \mathbb{N}$.

Notation 2.1.2. A pair $(\mathcal{H}, \frac{1}{N}\mathrm{Tr})$ consisting of a unital algebra (*unitary vector space with bilinear vector product*[2] $\mathcal{H}$) and a linear functional $\lim_{N\to\infty} \frac{1}{N}\mathrm{Tr} : \mathcal{H} \to \mathbb{C}$ with $\frac{1}{N}\langle \mathrm{Tr}1\rangle = 1$ is called a *non-commutative probability space*. Elements from $\mathcal{H}$ are addressed as *(non-commutative) random variables*, the numbers $\forall n : \{m_n = \frac{1}{N}\langle \mathrm{Tr}\mathbf{H}_i^n\rangle\}$ for such random variables $\mathbf{H}_1, \ldots, \mathbf{H}_k \in \mathcal{H}$ are called *moments*, the collection of all moments $m_1, \ldots, m_k$ is called the *joint distribution of* $\mathbf{H}_1, \ldots, \mathbf{H}_k$.

Definition 2.1.3. Let $(\mathcal{H}, \frac{1}{N}\mathrm{Tr})$ be a non-commutative probability space and let I be an index set.

1. Let, for each $i \in I$, $\mathcal{H}_i \subset \mathcal{H}$, be a unitary vector subspace (*math.unital subalgebra*). The vector subspaces $(\mathcal{H}_i)_{i\in I}$ are called *free* or *freely independent*, if $\frac{1}{N}\langle \mathrm{Tr}\mathbf{H}_1 \cdots \mathbf{H}_k\rangle = 0$ whenever we have: k is a positive integer; $H_j \in \mathcal{H}_{i(j)}$ (with $i(j) \in I$) for all $j = 1, \ldots, k$; $\frac{1}{N}\langle \mathrm{Tr}\mathbf{H}_j\rangle = 0$ for all $j = 1, \ldots, k$; and neighboring elements are from different vector subspaces, i.e., $i(1) \neq i(2), i(2) \neq i(3), \ldots, i(k-1) \neq i(k)$.

2. Let, for each $i \in I$, $\mathbf{H}_i \in \mathcal{H}$. The elements $\{\mathbf{H}_i\}_{i\in I}$ are called *free* or *freely independent*, if their generated unital vector subspaces are free, i.e., if $\{\mathcal{H}_i\}_{i\in I}$ are free, where, for each $i \in I$, $\mathcal{H}_i$ is the unital vector subspaces of $\mathcal{H}$ which is generated by $\mathbf{H}_i$.

2.1.2 Random Matrix as Free Random Variable

Mean Spectral Density

In probability theory and statistics, a random matrix $\mathbf{H}$ is in general a matrix-valued random variable drawn from some probability distribution $P(\mathbf{H})$. We will start our study of a (real symmetric $N \times N$) random matrix $\mathbf{H}$ with a fundamental question about the average values of its (real) eigenvalues $\lambda_1, \ldots, \lambda_N$, which is concisely encoded in the "mean spectral density" (see for example (see for example [Meh04, Eyn00, AGZ09, Gui09]).

$$\rho_{\mathbf{H}}(\lambda) \equiv \frac{1}{N}\sum_{i=1}^{N} \langle \delta(\lambda - \lambda_i)\rangle = \frac{1}{N}\langle \mathrm{Tr}(\lambda \mathbf{1}_N - \mathbf{H})\rangle = \frac{1}{N} E(\delta_{\lambda_1}(H) + \ldots + \delta_{\lambda_N}(H)) \tag{2.1}$$

[2] Unitary vector space $\mathcal{H}$ is a vector space which contains a multiplicative identity element (unit) *i.e.*, an element $\mathbf{1}$ with the property

$$\forall x \in \mathcal{H} \quad \mathbf{1}x = x\mathbf{1}$$

where $\delta(\lambda)$ is the real Dirac delta function and $\mathbf{1}_N$ denotes the unit $N \times N$ matrix. Here the expectation map $\langle \ldots \rangle$ is understood to be taken w.r.t. the rotationally invariant probability measure $P(\mathbf{H})\mathrm{D}\mathbf{H}$ of the random matrix -*i.e.*, $\mathbf{H} \to \mathbf{O}^{\mathrm{T}}\mathbf{H}\mathbf{O}$, with $\mathbf{O}$ orthogonal). Hence the full information about $\mathbf{H}$ resides in its eigenvalues, distributed on average according to (2.1).

Green's function (resolvent)- Stieltjes transform

The resolvent (also known as Green's function, Stieltjes transform, Cauchy transform) – complex function of a complex variable z is a primary tool in studying spectral properties of random matrix(or more for the most part operator in Hilbert space), because can capture the spectral properties of this matrix in the analytic structure.

$$G(z) = \frac{1}{N}\left\langle \mathrm{Tr}\frac{1}{z\cdot\mathbf{1}_N - H}\right\rangle = \frac{1}{N}\left\langle \sum_{i=1}^{N}\frac{1}{z-\lambda_i}\right\rangle = \int_{\mathrm{cuts}} \mathrm{d}\lambda \rho_{\mathbf{H}}(\lambda)\frac{1}{z-\lambda} \tag{2.2}$$

It is customary to write the relationship between (2.1) and (2.2) in terms of this latter,

$$\rho_{\mathbf{H}}(\lambda) = -\frac{1}{\pi}\lim_{\epsilon\to 0^+}\mathrm{Im}G_{\mathbf{H}}(\lambda+\mathrm{i}\epsilon) = -\frac{1}{2\pi\mathrm{i}}\lim_{\epsilon\to 0^+}\left(G_{\mathbf{H}}(\lambda+\mathrm{i}\epsilon) - G_{\mathbf{H}}(\lambda-\mathrm{i}\epsilon)\right). \tag{2.3}$$

resulting from a Sokhotsky's formula for generalized functions,

$$\lim_{\epsilon\to 0^+} 1/(x\pm\mathrm{i}\epsilon) = \mathrm{PV}(1/x) \mp \mathrm{i}\pi\delta(x). \tag{2.4}$$

Moments' generating function

The Green's function has another nice property - it is the function, that generates moments of a probability distribution. For finite N (2.2) is meromorphic with the poles at the λ_i's on the real axis. On the other hand, in the usually considered limit of an infinitely large random matrix ($N \to \infty$), the mean eigenvalues tend to merge into continuous intervals ("cuts"; they can be infinite or finite, connected or not), and the Green's function becomes holomorphic everywhere on the complex plane except the cuts on the real line. As such, it can typically be expanded into a power series around $z \to \infty$,

$$G_{\mathbf{H}}(z) = \sum_{n\geq 0}\frac{M_{\mathbf{H},n}}{z^{n+1}}, \qquad M_{\mathbf{H},n} \equiv \frac{1}{K}\left\langle \mathrm{Tr}\mathbf{H}^n\right\rangle = \int_{\mathrm{cuts}}\mathrm{d}\lambda\rho_{\mathbf{H}}(\lambda)\lambda^n, \tag{2.5}$$

$$\begin{array}{rcl} G(z) & = & \frac{1}{z}\left\langle \frac{1}{N}\mathrm{Tr}\frac{1}{1-\frac{\mathbf{H}}{z}}\right\rangle = \frac{1}{z} + \frac{1}{z}\left\langle \frac{1}{N}\mathrm{Tr}\mathbf{H}\right\rangle\frac{1}{z} + \ldots + \frac{1}{z^n}\left\langle \frac{1}{N}\mathrm{Tr}\mathbf{H}^n\right\rangle = \\ & = & \frac{1}{N}\left\langle \mathrm{Tr}\left[\frac{1}{z} + \frac{1}{z}\mathbf{H}\frac{1}{z} + \frac{1}{z}\mathbf{H}\frac{1}{z}\mathbf{H}\frac{1}{z}\right]\right\rangle = \frac{1}{N}\sum_n \frac{1}{z^{n+1}}\left\langle \mathrm{Tr}\mathbf{H}^n\right\rangle. \end{array} \tag{2.6}$$

This function is even for symmetric and centered $\mathbf{H}$, so odd terms vanish and the coefficients are called the "moments" of $\mathbf{H}$

$$m_n = \frac{1}{N}\left\langle \mathrm{Tr}\mathbf{H}^n\right\rangle = \int \rho_{\mathbf{H}}(\lambda)\lambda^n d\lambda \tag{2.7}$$

In particular, in the strict limit $z \to \infty$:

$$G_{\mathbf{H}}(z) \to \frac{1}{z}, \qquad \text{for} \qquad z\to\infty. \tag{2.8}$$

The above expansion (2.5) suggests working with an alternative object to the Green's function, namely the "generating function of the moments" (or the "M–transform"), simply related to the former,

$$M_{\mathbf{H}}(z) \equiv zG_{\mathbf{H}}(z) - 1 = \sum_{n\geq 1} \frac{M_{\mathbf{H},n}}{z^n}. \tag{2.9}$$

Both will be exploited, depending on convenience. We need to annotate, that even if the moments do not exist, and thus the expansions (2.5), (2.9) are not valid, the knowledge of the analytical structure of the Green's function (2.2) is sufficient to extract the statistical spectral properties of the random matrix, since the density can always be inferred from the behavior of the Green's function in the imaginary vicinity of the eigenvalues' cuts on the real axis.

Blue's function and N - transform

Finally, let us introduce the functional inverses of the Green's function (Blue's function [Zee96]) and the moments' generating function N–transform ,

$$G_{\mathbf{H}}\left(B_{\mathbf{H}}(z)\right) = B_{\mathbf{H}}\left(G_{\mathbf{H}}(z)\right) = z, \qquad M_{\mathbf{H}}\left(N_{\mathbf{H}}(z)\right) = N_{\mathbf{H}}\left(M_{\mathbf{H}}(z)\right) = z. \tag{2.10}$$

These two functions are fundamental objects within the FRV approach. Additionally, the Blue's function can be expanded into a power series around $z = 0$: it must start from a singular term $1/z$ due to (2.8) plus a regular expansion,

$$B_{\mathbf{H}}(z) = \frac{1}{z} + \sum_{n\geq 0} \kappa_{\mathbf{H},n+1} z^n, \tag{2.11}$$

where the coefficients are referred to as "free cumulants".

2.1.3 Large Matrices as Free Random Variables

It is common question in spectral analysis, to unravel the spectrum of a sum or a product of two $N \times N$ random matrices $\mathbf{H}_1$ and $\mathbf{H}_2$, knowing the spectra of each individual random matrix. As a rule the set of possible spectra of $\mathbf{H}_1 + \mathbf{H}_2$ depends in a complicated way on the spectra of $\mathbf{H}_1$ and $\mathbf{H}_2$. However when N becomes large, a stunning phenomenon occurs. For almost all choices of $\mathbf{H}_1$ and $\mathbf{H}_2$ of given eigenvalues density, the asymptotic spectrum of $\mathbf{H}_1 + \mathbf{H}_2$ can be easily computed analytically, without knowing detailed structure of the matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ (i.e. without relying on the eigenvectors). Random matrix in free probability calculus can be only a model of single noncommutative variable. The reason why random matrices play fundamental role in Free Probability is the asymptotic freeness of random matrices. Roughly speaking, when two matrices are free, there exist a rule to compute any asymptotic moment of the sum of two matrices (and thus their asymptotic spectrum) as a function of individual moments.

2.1.4 Asymptotic Freeness

The definition of asymptotic freeness is somewhat reminiscent of the concept of independent random variables. Unfortunately, defining freeness is thoroughly more complicated than defining independence. Think about four random matrices and assume

$$\begin{array}{rcl} \left\langle \frac{1}{N}\mathrm{Tr}(\mathbf{H}_1\mathbf{H}_2\mathbf{H}_3\mathbf{H}_4)\right\rangle & = & \left\langle \frac{1}{N}\mathrm{Tr}(\mathbf{H}_1\mathbf{H}_2)\right\rangle \left\langle \frac{1}{N}\mathrm{Tr}(\mathbf{H}_3\mathbf{H}_4)\right\rangle \\ \left\langle \frac{1}{N}\mathrm{Tr}(\mathbf{H}_1\mathbf{H}_3\mathbf{H}_2\mathbf{H}_4)\right\rangle & \neq & \left\langle \frac{1}{N}\mathrm{Tr}(\mathbf{H}_1\mathbf{H}_2)\right\rangle \left\langle \frac{1}{N}\mathrm{Tr}(\mathbf{H}_3\mathbf{H}_4)\right\rangle \end{array} \tag{2.12}$$

For commutative random variables both equalities would contradict each other, while for non-commutative multiplication both relations might be true at the same time. In fact, actual equation for calculating mixed moments is different for different random matrix ensembles [DNV92, ENV00], however the relation between moments remains the same *i.e.*,

Definition 2.1.4. Two Hermitian random matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ are called **free** [DNV92] with respect to $\mathbf{H} = \lim_{N\to\infty} \left\langle \frac{1}{N}\mathrm{Tr}\mathbf{H} \right\rangle_{cl}$ if for arbitrary polynomials $p_1, r_1, p_2, r_2, \ldots \geq 1$

$$\lim_{N\to\infty} \langle\ \tfrac{1}{N}\mathrm{Tr}\left(\left(\mathbf{H}_1^{p_1} - \tfrac{1}{N}\left\langle \mathrm{Tr}(\mathbf{H}_1^{p_1})\right\rangle\right)\left(\mathbf{H}_2^{r_1} - \tfrac{1}{N}\left\langle \mathrm{Tr}(\mathbf{H}_2^{r_1})\right\rangle\right)\cdot \right.$$
$$\left. \cdot \left(\mathbf{H}_1^{p_2} - \tfrac{1}{N}\left\langle \mathrm{Tr}(\mathbf{H}_1^{p_2})\right\rangle\right)\cdot \ldots\right)\ \rangle = 0$$

In other words

$$\langle p_1(\mathbf{H}_1) r_1(\mathbf{H}_2) p_2(\mathbf{H}_1) r_2(\mathbf{H}_2)\rangle = 0 \quad \text{if} \quad \langle p_i(\mathbf{H}_1)\rangle = \langle r_j(\mathbf{H}_2)\rangle = 0$$

The basic feature of this definition is that consecutive polynomials should depend on different variables. Note that, $\langle \ldots \rangle_{cl}$ is just some classical (commutative) expectation value, which we define for arbitrary (usually polynomial) potential $V(\mathbf{H})$:

$$\langle P(\mathbf{H})\rangle_{cl} \equiv \int d\mathbf{H} e^{-N\mathrm{Tr}V(\mathbf{H})} P(\mathbf{H}) \tag{2.13}$$

Freeness (2.1.4) is a rule for calculating mixed moments in $\mathbf{H}_1$ and $\mathbf{H}_2$ from the separate moments of $\mathbf{H}_1$ and of $\mathbf{H}_2$

Example 2.1.5.

$$\left\langle \frac{1}{N}\mathrm{Tr}\left(\left(\mathbf{H}_1^n - \frac{1}{N}\left\langle \mathrm{Tr}(\mathbf{H}_1^n)\cdot\mathbf{1}\right\rangle\right)\left(\mathbf{H}_2^m - \frac{1}{N}\left\langle \mathrm{Tr}(\mathbf{H}_2^m)\cdot\mathbf{1}\right\rangle\right)\right)\right\rangle = 0$$

thus

$$\left\langle \tfrac{1}{N}\mathrm{Tr}(\mathbf{H}_1^n\mathbf{H}_2^m)\right\rangle - \left\langle \tfrac{1}{N}\mathrm{Tr}\mathbf{H}_1^n\cdot\mathbf{1}\right\rangle\left\langle \tfrac{1}{N}\mathrm{Tr}\mathbf{H}_2^m\cdot\mathbf{1}\right\rangle +$$
$$+\left\langle \tfrac{1}{N}\mathrm{Tr}\mathbf{H}_1^n\right\rangle\left\langle \tfrac{1}{N}\mathrm{Tr}\mathbf{H}_2^m\right\rangle\left\langle \tfrac{1}{N}\mathrm{Tr}(\mathbf{1}\cdot\mathbf{1})\right\rangle = 0$$

and hence

$$\left\langle \frac{1}{N}\mathrm{Tr}(\mathbf{H}_1^n\mathbf{H}_2^m)\right\rangle = \left\langle \frac{1}{N}\mathrm{Tr}\mathbf{H}_1^n\right\rangle \cdot \left\langle \frac{1}{N}\mathrm{Tr}\mathbf{H}_2^m\right\rangle.$$

If the matrices are not centered $\langle \mathbf{H}_i\rangle \neq 0$, we can use the trick and rename them as $\tilde{\mathbf{H}}_i \equiv \mathbf{H}_i - \langle \mathbf{H}_i\rangle$ and since by definition $\left\langle \tilde{\mathbf{H}}_1\tilde{\mathbf{H}}_2\right\rangle = 0$, the first mixed moments are:

Example 2.1.6.

$$\begin{aligned}
\langle \mathbf{H}_1\mathbf{H}_2\rangle &= \langle \mathbf{H}_1\rangle\langle \mathbf{H}_2\rangle \\
\langle \mathbf{H}_1\mathbf{H}_2\mathbf{H}_1\mathbf{H}_2\rangle &= \left\langle \mathbf{H}_1^2\right\rangle\langle \mathbf{H}_2\rangle^2 + \langle \mathbf{H}_1\rangle^2\left\langle \mathbf{H}_2^2\right\rangle - \langle \mathbf{H}_1\rangle^2\langle \mathbf{H}_2\rangle^2 \\
\langle \mathbf{H}_1\mathbf{H}_1\mathbf{H}_2\mathbf{H}_2\rangle &= \left\langle \mathbf{H}_1^2\right\rangle\left\langle \mathbf{H}_2^2\right\rangle
\end{aligned}$$

This justifies, that freeness is also called **free independence**. One should however remember, that freeness is a more restrictive property than independence in classical probability theory *i.e.*, mixed moments are in fact combinations of products of individual moments and not just products. In other words, the mixed moments of free non–commutative random variables generally do not factorize into separate moments, as it is the case for independence. Borrowing a picture from physics, we may say that freeness is equivalent to planarity in the limit of a large number of colors in field theory [CLS82, tH74].

2.2 FRV calculus in a Nut-Shell

Free probability theory allows one to calculate the asymptotic eigenvalue distribution involving several random matrices, provided the spectral distribution of each individual random matrix is known.

2.2.1 Addition Algorithm - R -transform and Blue's function

Classical addition law

An important problem in classical probability [Fel68] is to find the probability density function ("pdf") of the sum of two random variables, $x_1 + x_2$, provided they are independent, and we are given their separate pdfs, $p(x_1)$ and $p_(x_2)$. The moments are conveniently encoded in terms of the "characteristic function,"

$$g_x(z) \equiv \sum_{n\geq 0} \frac{M_{x,n}}{n!} z^n = \langle e^{zx} \rangle. \tag{2.14}$$

which for $z = ik$ is a Fourier transform of the pdf. Expanding the characteristic function in frequency *i.e.*, applying the Newton's formula to(2.14) yields all the moments $\langle (x_1 + x_2)^n \rangle$.

$$M(x_1 + x_2, n) = \langle (x_1 + x_2)^n \rangle = \sum_{k=0}^{n} \binom{n}{k} M(x_1, k) M(x_2, n-k)$$

Thus, the problem of calculating the mixed moments $\langle (x_1 + x_2)^n \rangle$ can be simplified by taking the Fourier transforms for both pdfs $p(x_1)$ and $p(x_2)$, multiplying the resulting characteristic functions and inverting the Fourier transform to obtain the pdf for $x_1 + x_2$. In addition, if we take the logarithm of the characteristic functions,

$$r_x(z) \equiv \log g_x(z), \tag{2.15}$$

the convolution problem reduces to an additive one

$$r_{x_1+x_2}(z) = r_{x_1}(z) + r_{x_2}(z), \qquad \text{for independent } x_1,\, x_2. \tag{2.16}$$

The moments generated by the logarithm of the characteristic function are the cumulants $\kappa_{x,n}$. They are additive under the convolution of two measures, or in other words, under the addition of two independent random variables [Fel68].

FRV Addition algorithm

The above algorithm can not apparently be extended to non-commutative case. After Voiculescu *et al.* and Speicher [DNV92, Spe94] we can develop a precise answer to this question. We have already defined the moments' generating function $M_{\mathbf{H}}(z)$ and the resolvent $G_{\mathbf{H}}(z)$. The FRV calculus tells us that an analogue of the logarithm of the characteristic function is another complex function R–transform, $R_{\mathbf{H}}(z)$, defined as the generating function of *free cumulants*,

$$R_{\mathbf{H}}(z) = \sum_{n\geq 0} k_{\mathbf{H},n+1} z^n \qquad G_{\mathbf{H}}\left(R_{\mathbf{H}}(z) + \frac{1}{z}\right) = z \tag{2.17}$$

By definition, the R–transform is additive. Therefore, the R–transform of the sum of two independent random matrix ensembles $\mathbf{H}_1$ and $\mathbf{H}_2$ is a sum of the corresponding R–transforms

$$R_{\mathbf{H}_1+\mathbf{H}_2}(z) = R_{\mathbf{H}_1}(z) + R_{\mathbf{H}_2}(z). \tag{2.18}$$

Trivially, the free cumulants (2.11) are additive as well,

$$\kappa_{\mathbf{H}_1+\mathbf{H}_2,n} = \kappa_{\mathbf{H}_1,n} + \kappa_{\mathbf{H}_2,n}.$$

Without any proofs (which are not very complicated but lengthy), we will just describe the resulting procedure, using for convenience the functional inverse of the resolvent (2.10). It is related to the original R–transform by

$$R_{\mathbf{H}}(z) = B_{\mathbf{H}}(z) - \frac{1}{z}. \tag{2.19}$$

1. Since moments of the free random matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ can be obtained from Green's functions a (2.2), (2.5), we construct $G_{\mathbf{H}_1}(z)$ and $G_{\mathbf{H}_2}(z)$,
2. The Green's functions are inverted functionally to obtain the corresponding Blue's functions $B_{\mathbf{H}_1}(z)$ and $B_{\mathbf{H}_2}(z)$ (2.10)
3. We use the law of addition
$$B_{\mathbf{H}_1+\mathbf{H}_2}(z) = B_{\mathbf{H}_1}(z) + B_{\mathbf{H}_2}(z) - \frac{1}{z}, \qquad \text{for free } \mathbf{H}_1, \mathbf{H}_2 \tag{2.20}$$
to get the Blue's functions for the sum $\mathbf{H}_1 + \mathbf{H}_2$,
4. We functionally invert $B_{\mathbf{H}_1+\mathbf{H}_2}(z)$ to obtain $G_{\mathbf{H}_1+\mathbf{H}_2}(z)$ and subsequently spectral density through (2.3).

2.2.2 Multiplication Rule - S transform method

Another problem is how to deduce a composition law for the multiplication of free random matrices. The distribution of a product of independent random variables is not widely discussed in textbooks on classical probability theory, since it can always be derived from the relation $\exp x_1 \exp x_2 = \exp(x_1 + x_2)$, which reduces the multiplication problem to the addition one by a change of variables. However, this is not the case for random matrices $\mathbf{H}_1$ and $\mathbf{H}_2$, which in general do not commute and $\exp \mathbf{H}_1 \exp \mathbf{H}_2 \neq \exp(\mathbf{H}_1 + \mathbf{H}_2)$. This notwithstanding, there exists [DNV92] a multiplicative transformation (called the "S–transformation") which allows one to calculate the resolvent of a product of free random matrices $\mathbf{H}_1\mathbf{H}_2$ from the resolvents of each separate term, just like there is the R–transformation for the sum.

$$S_{\mathbf{H}_1\mathbf{H}_2}(z) = S_{\mathbf{H}_1}(z) S_{\mathbf{H}_2}(z). \tag{2.21}$$

It relates to each resolvent as follows

$$S_{\mathbf{H}}(z) = \frac{1+z}{z}\chi_{\mathbf{H}}(z), \quad \text{where} \quad \frac{1}{\chi_{\mathbf{H}}(z)} G_{\mathbf{H}}\left(\frac{1}{\chi_{\mathbf{H}}(z)}\right) - 1 = M_{\mathbf{H}}\left(\frac{1}{\chi_{\mathbf{H}}(z)}\right) = z, \tag{2.22}$$

i. e. $1/\chi_{\mathbf{H}}(z)$ is a functional inverse of the moments' generating function $M_{\mathbf{H}}(z)$. For notational convenience we will use the **N**–transform a functional inverse of the moments generating function $M_{\mathbf{H}}(z)$ which is related to the original χ–transform (2.22) by

$$\chi_{\mathbf{H}}(z) = \frac{1}{\mathbf{N}_{\mathbf{H}}(z)}. \tag{2.23}$$

Again without proofs, the algorithm for multiplication of two matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ goes now in the following steps:

1. From the knowledge of $G_{\mathbf{H}_1}(z)$ and $G_{\mathbf{H}_2}(z)$, we calculate the corresponding moments' generating functions(2.9) $M_{\mathbf{H}_1}(z)$ and $M_{\mathbf{H}_2}(z)$,

2. Then we use(2.10) to functionally invert the M–transforms and obtain respective **N**–transforms $N_{\mathbf{H}_1}(z)$ and $N_{\mathbf{H}_2}(z)$.

3. Since$N$–transforms obey the "non–commutative multiplication law", we use the multiplication law
$$N_{\mathbf{H}_1}(z)N_{\mathbf{H}_2}(z) = \frac{1+z}{z}N_{\mathbf{H}_1\mathbf{H}_2}(z) \tag{2.24}$$
we immediately get the **N**–transform for the matrix product H_1H_2, Equivalently in the original language [DNV92, Spe94], this means that"S–transforms,"
$$S_{\mathbf{H}}(z) \equiv (1+z)/(zN_{\mathbf{x}}(z)) \tag{2.25}$$
are multiplicative,
$$S_{\mathbf{H}_1\mathbf{H}_2}(z) = S_{\mathbf{H}_1}(z)S_{\mathbf{H}_2}(z) \tag{2.26}$$

4. We functionally invert $N_{\mathbf{H}_1\mathbf{H}_2}(z)$ to obtain $M_{H_1H_2}(z)$, and subsequently $G_{\mathbf{H}_1\mathbf{H}_2}(z)$ and mean spectral density.

2.3 Classical vs. Non-commutative Probability and its analogies

The parallel between classical and free probability theory is very deep. In particular, there exists free equivalent of a central limit theorem or free analogues of infinitely divisible and stable distributions. To emphasize it, here we present two instances of the FRV theory, namely Free Poisson Distribution and Wigner semicircle. These results have already been mentioned by several authors [PBL05, HP00]. We complete the picture by step-by-step rederivation of the Central Limit Theorems for Gaussian and Wishart matrices in analogy with classical results [Fel68].

2.3.1 Binomial distribution

Consider an experiment, whose results are always one of two mutually excluding possibilities *i.e.*, classical two-point distribution, which reflects the dychotomic probability p of a success and probability $q = 1 - p$ of a failure (loss). Let us define a random variable x_i, which assigns the value of 1 to each success and consequently 0 to each loss. Then Bernoulli probability distribution is well known:

$$P_{x_i}(p) = p^k(1-p)^{1-k} \quad \text{for} \quad x_i = 0, 1 \quad E(x_i) = p, Var(x_i) = p(1-p) \tag{2.27}$$

We can repeat this experiment n-times and formally examine the probability distribution of a random variable $X = x_1 + x_2 + \ldots + x_n$. Probability of k- successes in n trials is then, according to the combinatorics rules given by the binomial distribution:

$$P(k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \tag{2.28}$$

The proper normalization is due to Newton expansion of a polynomial

$$(q+p)^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k} = (1-p+p)^n = 1$$

One can simply calculate the first few moments of this probability distribution. The expectation (mean) is

$$\begin{array}{rcl} E(x_i) & = & \langle x_i \rangle = \sum_{k=0}^{n} k \frac{n!}{k!(n-k)} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \\ & = & \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} = \\ & = & np \sum_{k=1}^{n} k \frac{n!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} = \\ & = & np \sum_{k=1}^{n} k \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} p^{k-1} (1-p)^{(n-1)-(k-1)} = np \end{array} \tag{2.29}$$

To calculate the variance first it is necessary to calculate the following quantity:

$$\begin{array}{rcl} E(x_i^2) & = & \left\langle x_i^2 \right\rangle = \sum_{k=0}^{n} k^2 \frac{n!}{k!(n-k)} p^k (1-p)^{n-k} = \\ & = & np \sum_{k=1}^{n} k \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} p^{k-1} (1-p)^{(n-1)-(k-1)} = \ldots \\ & = & np((n-1)p+1) \\ Var(x_i) & = & E(x_i^2) - E(x_i)^2 = np((n-1)p+1) - n^2 p^2 = np(1-p) = npq \end{array} \tag{2.30}$$

Which for x_i/n is

$$Var\left(\frac{x_i}{n}\right) = \frac{1}{n^2} Var(x_i) = \frac{pq}{n}$$

The normal approximation of the binomial distribution

Consider now the asymptotic behavior of binomial distributions. Suppose p is fixed and let's look more closely at the distributions for different values of n. One can easily convince oneself, that for large n the distribution becomes more symmetric and concentrates around the expectations. For large n-number of trials and large k number of successes with fixed probability p. Making use of well known Stirling formula

$$\Gamma(n+1) = n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} \exp\left(-n + \frac{1}{12n} - \frac{1}{360n^2} + \ldots \right) \tag{2.31}$$

$$\begin{array}{rcl} P(k) & = & \frac{n! p^k q^{n-k}}{k!(n-k)!} \approx \frac{\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} p^k q^{n-k}}{\sqrt{2\pi} k^{k+\frac{1}{2}} e^{-k} \sqrt{2\pi} (n-k)^{n-k+\frac{1}{2}} e^{-n+k}} = \\ & = & \frac{n^{n+\frac{1}{2}} p^k q^{n-k}}{\sqrt{2\pi} k^{k+\frac{1}{2}} (n-k)^{n-k+\frac{1}{2}}} = \frac{1}{\sqrt{2\pi n}} \left(\frac{k}{n}\right)^{-k-\frac{1}{2}} \left(\frac{n-k}{n}\right)^{-n+k-\frac{1}{2}} p^k q^{n-k} \end{array} \tag{2.32}$$

One introduces an auxiliary variable x, which is the deviation of k-successes from the mean, namely $k = np + x$ and and looks for a probability distribution where x is small comparing to the mean np, which leads to:

$$\frac{k}{n} = p + \frac{x}{n} = p\left(1 + \frac{x}{np}\right) \qquad \frac{n-k}{n} = q - \frac{x}{n} = q\left(1 - \frac{x}{nq}\right) \tag{2.33}$$

and consequently

$$\begin{aligned} P(k) &\approx \frac{1}{\sqrt{2\pi n}} p^{-k+\frac{1}{2}} \left(1+\frac{x}{np}\right)^{-k-\frac{1}{2}} q^{-n+k-\frac{1}{2}} \left(1-\frac{x}{nq}\right)^{-n+k-\frac{1}{2}} p^k q^{n-k} \\ &= \frac{1}{\sqrt{2\pi npq}} \left(1+\frac{x}{np}\right)^{-np-x-\frac{1}{2}} \left(1-\frac{x}{nq}\right)^{-nq+x-\frac{1}{2}} \\ &\approx \frac{1}{\sqrt{2\pi npq}} \exp\left(-\left(np+x+\tfrac{1}{2}\right)\ln\left(1+\frac{x}{np}\right)-\left(nq-x+\tfrac{1}{2}\right)\ln\left(1-\frac{x}{nq}\right)\right) \end{aligned} \tag{2.34}$$

The number of trials n is large and $p \neq 0$, so typically the number of successes is large and we can expand ln up to the second order

$$\ln\left(1+\frac{x}{np}\right) \cong \frac{x}{np} - \frac{1}{2}\left(\frac{x}{np}\right)^2 \qquad \ln\left(1-\frac{x}{nq}\right) \cong -\frac{x}{nq} - \frac{1}{2}\left(\frac{x}{nq}\right)^2 \tag{2.35}$$

Inserting (2.35) into (2.34) up to the lowest order in $x = k - np$ we finally obtain Gaussian distribution

$$P(k) \approx \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(k-\mu)^2}{2\sigma^2}\right) \tag{2.36}$$

with $\sigma^2 = npq$ and $\mu = np$. The above derivation can be also regarded as another manifestation of central limit theorem.

The Poisson approximation of the binomial distribution

We have also another limiting form of the binomial distribution, which is usually derived under assumption, that number of trials $n \to \infty$ and $p \to 0$ in each trial, but $np = \mu$ is constant, while the number of successes k is arbitrary and finite. We will use again Stirling formula (2.31); this time however only for $n!$ and $(n-k)!$

$$\begin{aligned} P(k) &\approx \frac{\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}}{k!\sqrt{2\pi}(n-k)^{n-k+\frac{1}{2}}e^{-n+k}} \left(\frac{\mu}{n}\right)^k \left(1-\frac{\mu}{n}\right)^{n-k} = \\ &= \frac{n^{n+\frac{1}{2}}}{k!\left(1-\frac{k}{n}\right)^{n-k+\frac{1}{2}} n^{n-k+\frac{1}{2}} e^k} \left(\frac{\mu}{n}\right)^k \left(1-\frac{\mu}{n}\right)^{n-k} = \\ &= \frac{1}{k!\left(1-\frac{k}{n}\right)^{n-k+\frac{1}{2}} e^k} \mu^k \left(1-\frac{\mu}{n}\right)^{n-k} \quad \xrightarrow[n\to\infty]{} \quad \frac{\mu^k}{k!}e^{-\mu} \end{aligned} \tag{2.37}$$

2.3.2 Free Central Limit Theorem - Wigner semicircle

Similarly as in classical probability theory, there is a corresponding central limit theorem for FRV *i.e.*, Wigner semicircle. We are asking, what is the spectral distribution of the sum of $H_i \quad i = 1, 2, \ldots, N$ mutually free random matrices

$$\frac{H_1 + H_2 + \ldots + H_N}{\sqrt{N}} \quad \text{for} \quad N \to \infty \tag{2.38}$$

We know that Green's function (2.2)

$$G(z) = \frac{1}{N}\left\langle \mathrm{Tr}\frac{1}{z-H}\right\rangle$$

Then

$$G_a(z) = \frac{1}{N}\left\langle \mathrm{Tr}\frac{1}{\frac{az}{a}-\frac{H}{a}}\right\rangle = aG_a(z) \quad B_a(z) = \frac{1}{a}B\left(\frac{z}{a}\right) \quad R_a(z) = \frac{1}{a}R\left(\frac{z}{a}\right)$$

We assume, that for each H_i

$$R_i(z) = r_i(z)z \to R_{ai}(z) = \frac{z}{a^2} r_i\left(\frac{z}{a}\right) \quad a = \sqrt{N}$$

Then by substituting $z \to G(z)$ into (2.39) and making use of eq.(2.20)

$$B_{\frac{H_1+H_2+\ldots+H_N}{\sqrt{N}}}(z) = \sum_{i=1}^{N} \frac{z}{a^2} r_i\left(\frac{z}{a}\right) + \frac{1}{z} \tag{2.39}$$

we arrive at

$$z = \sum_{i=1}^{N} \frac{G(z)}{a^2} r_i\left(\frac{G(z)}{a}\right) + \frac{1}{G(z)} \tag{2.40}$$

This equation is completely insolvable for N finite. However for $N \to \infty$ we get

$$z = \sigma^2 G(z) + \frac{1}{G(z)} \quad \text{where} \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N} r_i(0) \tag{2.41}$$

and Green's function immediately reads

$$G(z) = \frac{1}{2\sigma^2}\left(z - \sqrt{z^2 - 4\sigma^2}\right) \tag{2.42}$$

From Sokhotsky formula (2.3)we arrive at Wigner semicircle with only few lines of calculations[3] [Wig58]

$$\rho(\lambda) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - \lambda^2} \tag{2.43}$$

2.3.3 Free Poisson Approximation of Wishart Distribution

There exist a very rich correspondence between classical Poisson process and Free Poisson process, whose counterpart is the eigenvalue density (1.20) for Wishart correlation matrices [MP67, SB95]. Consider an empirical correlations matrix $\mathbf{E}$ of N variables and T observiations, both very large, with $r = N/T$ finite. Suppose,that the true correlations are given in terms of two point correlations function $\langle x_{it} x_{jt'} \rangle = C_i j \delta_{tt'}$. This defines the Wishart ensemble [Wis28]. In order to find the eigenvalue density we introduce the resolvent

$$G_{\frac{XX^T}{T}}(z) = \frac{1}{N}\left\langle \mathrm{Tr}\frac{1}{z - \frac{XX^T}{T}} \right\rangle \tag{2.44}$$

w.r.t. the Gaussian measure $\int dX e^{-\frac{N}{2}\mathrm{Tr}XX^T}$ The simplest case is when $C = \mathbf{1}$. Then $\mathbf{E}$ is a sum of rotationally invariant matrices *i.e.*,

$$\mathbf{E} = \sum_t \delta E^t_{i,j} = \frac{1}{T}\left(x_{i1}x_{j1} + x_{i2}x_{j2} + \ldots + x_{iT}x_{jT}\right) = \sum\left(\frac{1}{T}|x\rangle\langle x|\right) \tag{2.45}$$

[3] When performing the simulations we have used the so called "self–averaging" property of random matrices. Namely the empirical spectrum of eigenvalues is identical for one matrix of size $N = 1000$ and average of 100 eigenvalues of size 10×10.

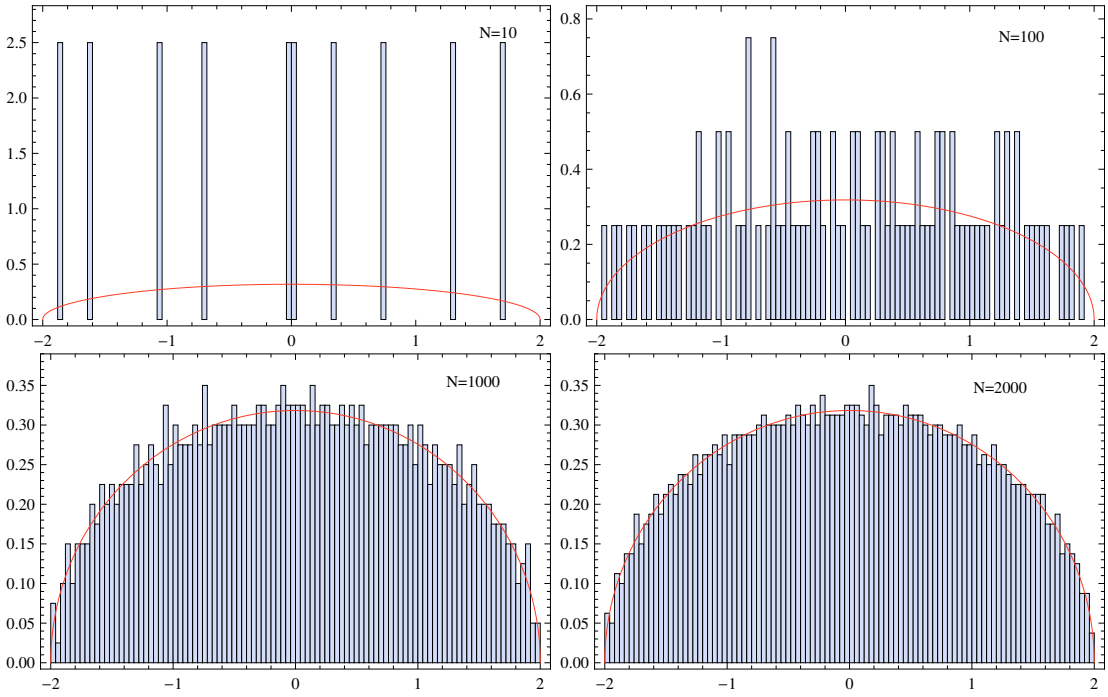

Figure 2.1: Numerical simulation of Free Central Limit Theorem for the eigenvalues Gaussian random matrices of different sizes N. The red line represents the theoretical spectral density - we notice a striking pattern- the density of eigenvalues forms a semicircle. The larger the matrix, the more deterministic the distribution.

Each of the above matrices has exactly one eigenvalue equal to λ and exactly $N-1$ zero modes *i.e.*, $N-1$ eigenvalues equal to zero.

$$\lambda = \frac{1}{T}\mathrm{Tr}\langle x|x\rangle = \frac{N}{T} = r$$

Then Green's function simply states:

$$G(z) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{z-\lambda_i} = \frac{N-1}{N}\frac{1}{z} + \frac{1}{N}\frac{1}{z-r} \tag{2.46}$$

We assume the matrices equalling the relation (2.45) are mutually free and rewrite the Green's function (2.46) in terms of its functional inverse $z = B(G(z))$

$$z = \frac{N-1}{N}\frac{1}{B_i(z)} + \frac{1}{N}\frac{1}{B_i(z)} \tag{2.47}$$

In the limit $N \to \infty$

$$B_i(z) = b(z) + \frac{1}{N}b_1(z) \quad \text{with} \quad b(z) = \frac{1}{z} \tag{2.48}$$

If we plug (2.48) into (2.47) and collect coefficient up to first order in $1/N$, then

$$b_1(z) = \frac{r}{1-rz}$$

and

$$B_i(z) = \frac{1}{z} + \frac{1}{N}\frac{r}{1-rz}$$

Since $B(z) = R(z) + \frac{1}{z}$ then

$$R_i(z) = \frac{1}{N}\frac{r}{1-rz}$$

The full R–transform (and Blue's function) for N eigenvalues

$$R(z) = \sum_{i=1}^{T} R_i(z) = \frac{T}{N}\frac{r}{1-rz} = \frac{1}{1-rz} \quad B(z) = R(z) + \frac{1}{z} = \frac{1}{1-rz} + \frac{1}{z}$$

To find the spectral density, we make an ansatz $z \to G(z)$

$$\begin{aligned} z &= \tfrac{1}{1-rG(z)} + \tfrac{1}{G(z)} = \tfrac{G(z)+1-rG(z)}{G(z)(1-rG(z))} \quad \text{then} \\ 0 &= rzG^2(z) + (1-r-z)G(z) + 1 \quad \text{and} \quad \text{finally} \\ G(z) &= \tfrac{(r+z-1)\mp\sqrt{(1-r-z)^2-4rz}}{2rz} \end{aligned} \tag{2.49}$$

If we use Sokhotsky formula (2.3) we finally arrive at the famous Marčenko-Pastur [MP67](Bai and Silverstein [SB95]) result for eigenvalue density ,

$$\rho(\lambda) = \frac{\sqrt{4\lambda r - (1-r-\lambda)^2}}{2\pi\lambda r} = \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\lambda r} \tag{2.50}$$

with $\lambda_\mp = (1 \mp \sqrt{r})^2$

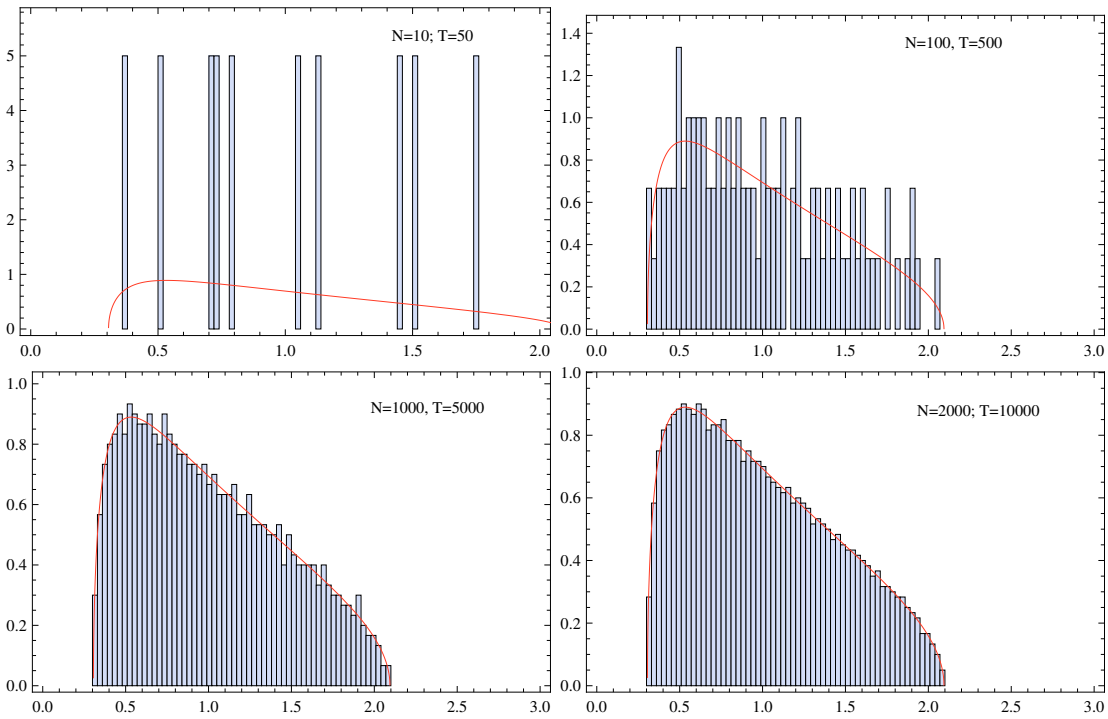

Figure 2.2: Numerical simulation of distribution of eigenvalues for sample correlations matrix with Marčenko-Pastur density superimposed. We see that even for rather small matrices, the theoretical limiting density approximates the actual density very well.

Classical Probability	**FRV**
independence	freeness
x - random variable, $p(x)$	H - random matrix, $P(H)$
pdf	spectral density $\varrho(\lambda)d\lambda$
$\downarrow$	$\downarrow$
characteristic function	Green's function $G(z) = \frac{1}{N}\left\langle \mathrm{Tr}\frac{1}{z\cdot\mathbf{1}-H}\right\rangle = \sum_{n=0}\frac{1}{z^{n+1}}\left\langle\frac{1}{N}\mathrm{Tr}H^n\right\rangle$
logarithm of characteristic function	R–transform
f.generating moments	$\left\langle\frac{1}{N}TrH^n\right\rangle = \int d\lambda\varrho(\lambda)\lambda^n$
addition law	$R_{1+2}(z) = R_1(z) + R_2(z)$ where $G\left[R(z)+\frac{1}{z}\right] = z$
multiplication law	$S_{1\cdot 2}(z) = S_1(z)\cdot S_2(z)$
Central Limit Theorem	**Free Central Limit Theorem**
In large N limit distribution of a sum of independent random variables tends to Gaussian	In large N limit distribution of a convolution of free random variables tends to Semicircle

Table 2.1: The correspondence between classical probability and FRV

2.4 Summary

We will end up this chapter with few comments

- The concept of freeness allows for a one–to–one correspondence between classical and free random variables, which in particular allows one to map probability densities of random variables into the corresponding eigenvalues' densities of large free random matrices [BP99].
- Also, one can define the analog of the concept of stability [BV93], which in the FRV calculus assumes the form of spectral stability.
- A consequence of the above two observations is that the eigenvalues' distribution of a properly normalized sum of many random matrices for which the second spectral moment is finite tends to a universal limiting distribution known in RMT as Wigner's semicircle law [Wig55]. The Wigner's distribution in the FRV calculus corresponds to the Gaussian distribution in the standard probability calculus.
- Another consequence is the equivalence between classical Poisson distribution is the Free Poisson distribution, whose counterpart is the Marčenko -Pastur [MP67] distribution for Wishart matrices [Wis28].
- Since the majority of data collected nowadays is naturally stored in the form of huge matrices, we believe that the FRV technique is the most natural candidate for the "matrix–valued probability calculus" that can provide efficient algorithms for cleaning (de–noising) large sets of data and unraveling essential but hidden spatio–temporal correlations. These features will be exploited in further chapters

- For completeness, let us also mention that FRV can also generate dynamical stochastic processes [BS01, JW04, GNJJN05], in a similar way like Gaussian distributions generate random walks in classical probability. We will not discuss them in this work, restricting ourselves to stationary properties of FRV only.

3
Equal-Time Correlations

The analysis of equal-time correlations has been actively investigated in recent years for various complex systems with major attention attracted to financial markets [LCBP99, PGR+99, LCPB00, PGR+00, JM03, GK03, UIO04, WG04, SCSR04, BJ04, KOD06, URH07, PPNK05, CRC09b, CRC09a, SH09]. The motivation behind is the modern portfolio theory [EGBG06], which heavily relies on accurate estimates of covariance matrix. Correlated moves of financial assets diminish the possibility of optimal portfolio diversification. In classical portfolio theory it is blindly assumed, that correlations are perfectly known. However, the information about correlations is typically unraveled from historical data, that are to large extent noisy due to finiteness of the samples (c.f. section1.4.1). In particular, Random Matrix Theory (RMT), has been applied to filter the relevant information from the statistical fluctuations [LCBP99, BJ04, WG04, KS06], inherent in empirical cross-correlation matrices, for various financial time series. By comparing the eigenvalue spectrum of the correlation matrix to the analytical results, obtained for random matrix ensembles, significant deviations from RMT eigenvalue predictions provide genuine information about the correlation structure of the system. This information has been used to reduce the difference between predicted and realized risk of different portfolios [PBL05]. The main aim of this chapter, given a Voiculescu FRV approach [DNV92], is to analyze more deeply and expand the results from [KS06].

3.1 Theoretical Framework

Modern Portfolio Theory (MPT) refers to an investment strategy that seeks to construct an optimal portfolio by considering the relationship between risk and return. MPT suggests that the fundamental issue of capital investment should no longer be to pick out dominant stocks but to diversify the wealth among many different assets. The success of investment does not purely depend on return, but also on the risk, which has to be taken into account. Risk itself is influenced by the correlations between different assets, thus the ability to predict future movements in prices (price changes) allows one to minimize the risk. Let us briefly remind several key tools and concepts, that MPT uses, i.e. the Markowitz's Model [Mar52], which is crucial in further analysis.

3.1.1 Mathematical Notation

Suppose one builds a portfolio of N assets with w_i being portion of wealth invested in asset i. Consider T quotations of the i -th stock and introduce a vector of returns $y_{i,1}$,where $y_{i,t}$, $t = 1, \ldots, T$ is the observed realization of a random variable y_i. Denote $S_i(t)$ - time series of

prices for a certain stock i. Then

$$y_{i,t} = \ln S_i(t+1) - \ln S_i(t) \tag{3.1}$$

and ln is a natural logarithm. Then the expected return of a single asset is given by

$$Y_i = E(y_i) = \hat{y}_i = \bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{i,t} \tag{3.2}$$

Denoting $\mathbf{y}$ as a vector of expected returns of single stocks, we see, that an expected return of a whole portfolio is a linear combination of returns of assets in a portfolio

$$Y_p = \sum_{i=1}^{N} w_i \cdot Y_i = \mathbf{w}^T \cdot \mathbf{Y}$$

The daily variance of a portfolio return is given by:

$$Y^2 = \sum_{i,j} w_i \sigma_i C_{ij} \sigma_j w_j \tag{3.3}$$

Where σ_i^2 is the daily asset i and C_{ij} is the correlation matrix. In order to measure and optimize risk of this portfolio one has to come up with reliable estimate of the correlations matrix C_{ij}. This is difficult in general [JM03] since one has to determine $N(N+1)/2$ coefficients out of N time series of length T, and in general T is not much larger than N

3.1.2 Stock exchange as a dynamic complex system - initial analysis

Throughout this chapter we will focus on the analysis cross-correlations in Polish stock market. Several similar studies have been accomplished recently [RKDO08, KDS03, DKSW02, SDG09]. The WIG index - major and the oldest index of Warsaw Stock Exchange is a sort of market indicator and economy's barometer, consisted of about 80% of all assets quoted during continuous trading. The influence of individual stock is limited to 10%, with no more than 30% of stocks belonging to one sector. WIG is calculated not only on the basis of daily prices of stocks included in it but also incorporates the income from dividends and subscription rights, though it is often treated as an income index. From our point of view, it is interesting to examine the connections (i.e. correlations) between the constituent stocks.

Sampling error

Suppose we have data set consisting of N stocks with T returns in each time series and compute a sample correlation matrix. Furthermore, let's assume for a while, that the true correlation were the identity matrix *i.e.*, the stocks are initially not correlated. This is not true in general and we will relax this statement in next sections. For normally distributed distributed returns, the median maximum correlation $\rho_{\max}$ should satisfy:

$$\ln 2 \approx \frac{N(N-1)}{2}\mathbb{N}\left(-\rho_{\max}\sqrt{T}\right) \tag{3.4}$$

With $N = 500, T = 1000$, we obtain $\rho_{\max} = 0.14$ So, sampling error induces spurious (and potentially significant) correlations between stocks.

Indentification of the stable period in financial data

Correlation analysis of financial markets indirectly assumes the stationarity of the data *i.e.*, that the expectation and the variance of the data set exist and do not change over time. We have related it with the period of the lowest volatility of the WIG index.

1. We have started with the conversion of absolute changes of the WIG time series $S(t)$ to the relative ones according to

$$y(t) = \frac{S(t+1) - S(t)}{S(t)} \tag{3.5}$$

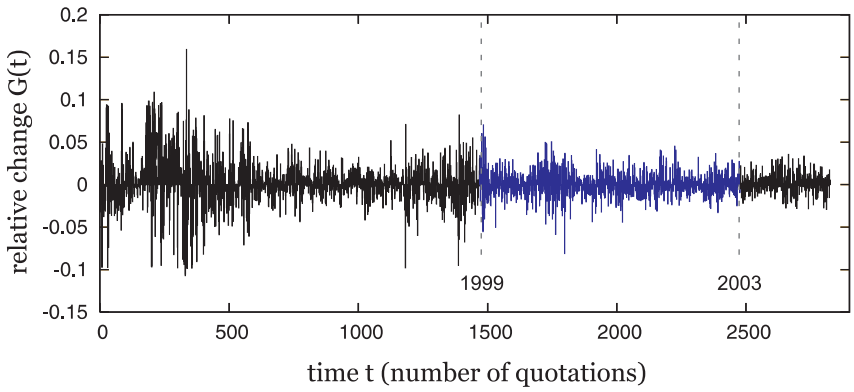

Figure 3.1: Fluctuations of relative WIG changes

2. Then for a fixed time window width $T = 990$ quotations, the volatility of the time series $y(t)$ was calculated:

$$\sigma(t_0) = \sqrt{\frac{1}{T-1}\sum_{i=0}^{T}\left(Y(t_0+i) - \overline{y(T)}\right)^2} \tag{3.6}$$

where $\overline{y(T)}$ is the average $y(t)$ over the whole time window T. This results can be presented on the diagram:

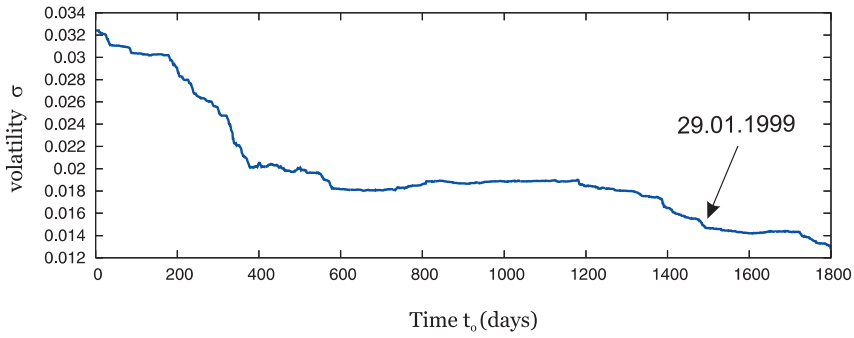

Figure 3.2: Volatility changes in time for a fixed window length

It is easy to notice, that first few years of quotations are determined by a relatively high volatility. This is related to the initial period of Polish Stock Exchange, where vast changes in WIG index dominated and investors needed to learn the rules of the stock market.

The Analysis of stocks

Another problem we have encountered during the analysis of historical data, was the incomplete information about some of 120 stocks, which may result in the infinities in relative changes $x(t)$, when the lack of information was replaced by zeros in the original $S(t)$ time series. "Zeros"appear when one is unable to settle the price of an individual stocks. The separate "zeros" were extrapolated from the future and previous relative changes of a given time series. In the case, if more information is lost in the way, one is unable to predict the further prices, then this stock is not very examined in further research.

Final data set

For the fixed period from $29:01:1999$ till $17:01:2003 = 990$ days we have finally chosen the 100 stocks in the WIG index[1] with the average standard deviation of price changes $\langle\sigma\rangle = 0,4767$ and average correlation of returns between stocks $\langle corr_{ij}\rangle = 0,0657$. Thus, in this case, $N = 100$ and $T = 990$, $r = N/T = 10/99$. There are $N(N-1)/2 = 4950$ distinct entries in the correlation matrix to be estimated from $990 \times 100 = 99000$ data points. With these parameters, we would expect the maximum error in our correlation estimates to be around 0.013.

3.2 Estimators of equal–time correlations

It is well known result, that any set of correlated Gaussian random variables can always be decomposed into a linear combination of independent Gaussian random variables. The converse is also true, since the sum of Gaussian random variables is also a Gaussian random variable. In other words, correlated Gaussian random variables are fully characterized by their correlation matrix. Which can be constructed in the simplest way via Pearson estimator (1.6).

3.2.1 Uncorrelated Wishart ensemble $C = \mathbf{1}$

Assume now that all returns comes from the Gaussian distribution

$$P_{\text{c.G.}}(\mathbf{Y})\mathrm{D}\mathbf{Y} = \frac{1}{\mathcal{N}_{\text{c.G.}}} \exp\left(-\frac{1}{2}\mathrm{Tr}\mathbf{Y}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{Y}\mathbf{A}^{-1}\right)\mathrm{D}\mathbf{Y}, \tag{3.7}$$

where the normalization constant $\mathcal{N}_{\text{c.G.}} = (2\pi)^{NT/2}(\mathrm{Det}\mathbf{C})^{T/2}(\mathrm{Det}\mathbf{A})^{N/2}$, and the integration measure $\mathrm{D}\mathbf{Y} \equiv \prod_{i=1}^{N}\prod_{a=1}^{T}\mathrm{d}Y_{ia}$, while the letters "c.G." stand for "correlated Gaussian"and we impose ($\mathbf{C} = \mathbf{A} = \mathbf{1}$).

3.2.2 Cleaning technique - one factor model

If one considers N assets, the covariance matrix need to be determined from N time series of length $T \gg N$. Typically T is not very large compared to N and one should expect that the determination of the covariances is noisy. This noise cannot be removed by simply increasing

[1] An exact of time series used is available from the author upon request

the number of independent measurement of the investigated financial market, because economic events, that affect the market are unique and cannot be repeated. Therefore the structure of the matrix estimator is dominated by "measurement" noise. In this section we will compare properties of an empirical correlation matrix to a purely random matrix, well defined in the sense of Random Matrix Theory [GMG98]. Deviations from the RMT might then suggest the presence of true information [LCBP99]. The problem is now to extract these significant eigenvalues.

Empirical spectrum of correlation matrix

First, we have computed numerically the eigenvalue spectrum and superimposed the Marčenko-Pastur [MP67] density with $r \approx 0.1$ (see Fig 3.3). An immediate observation is that, the largest eigenvalue is $\lambda_1 = 12.62$ is about 7 times larger, than the predicted $\lambda_+ = 1.737$. This largest mode reflects the strength of collective evolution of eigenvalues. We will call the largest λ_1 a "market mode", since the corresponding eigenvector has roughly equal components on all $N = 100$ stocks. Furthermore the constituents of the eigenvector are at minimum 10% for individual time series and maximum 30% if the stocks belong to one of the ordinary market sectors, which to large extent mimics the structure of WIG index.

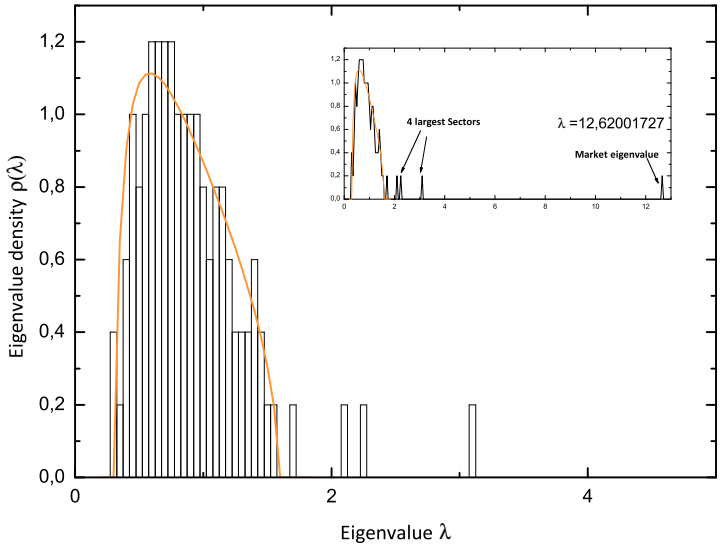

Figure 3.3: Full empirical spectrum of a covariance matrix for Polish stock market data and fit using MP distribution.The edges are predicted by (2.50). In the presence of one very large eigenvalue,corresponding to the "market mode", the fit reveals systematic deviations, suggesting a non-trivial structure of the covariance matrix.If we look closely we can observe, that there are several large eigenvalues (the largest one is labeled as *the market* one, since it consists the information about all the stocks in the market i.e. is closely related to the WIG index), however the greater part of the spectrum is concentrated between 0 and 2 (i.e. *The Wishart- fit*). We believe, that behind this *Random* part of the spectrum there exists single eigenvalue, which carries nontrivial and useful information.

The effect of non-synchronous trading

One may argue, that this unique structure is the effect of finiteness of the sample. Suppose we make an experiment and shuffle the returns in each time series (Fig. 3.4). Then again form eigenvalue spectrum for a sample correlation matrix with theoretical fit superimposed. It is easy to notice, that with this simple trick the unique spatio–temporal structure is destroyed. We relate this phenomena, that affects the dynamical structure of the eigenvalues spectrum to non–synchronous character of trading on a stock exchange. Stock exchange transactions have different activity over trading day. If we are about to consider daily data, we actually mean the close price and we treat them in a manner as if there were a $24 - hour$ period between two observations, which is not true. As a consequence we have cross correlations between returns in a portfolio of stocks and autocorrelation of individual returns mixed in an extremely non-linear structure.

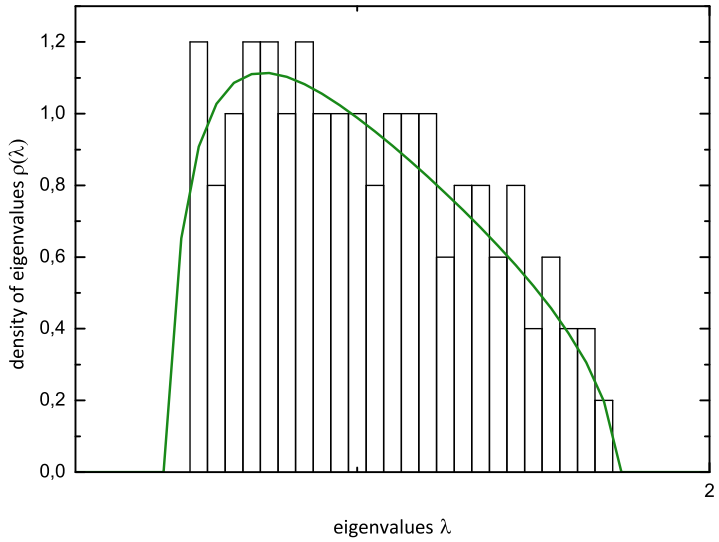

Figure 3.4: Spectrum with randomized data. The process of scrambling the returns destroys the interesting structure. The agreement between the empirical spectrum and predicted Marčenko-Pastur density is nearly perfect.

Interim conclusions

From this simple experiment, we note that:

- Even though financial return series are fat-tailed, the Marčenko-Pastur density is a very good approximation to the density of eigenvalues of the correlation matrix of the randomized returns.
- The large part of the empirical correlation matrix is considered as noise and cannot be trusted in any method involving inverse of the correlation matrix. Noise in the sample

covariance estimate leads to spurious portfolio estimates with very low or zero predicted variance.

- The Marčenko-Pastur density does not remotely fit the eigenvalue spectrum of the sample correlation matrix from which we conclude that there is non–random structure in the return data.

- The numerical results clearly show, that empirical correlation matrix constructed from financial time series has one dominant eigenvalue. This suggest that, each return can be decomposed into

$$y_t^i = \beta_i \theta_t + \epsilon_t^i \tag{3.8}$$

with θ_t - return associated with the market mode, common for all stocks, ϵ_t - idiosyncratic noise term and β_i- coefficient specific to individual asset.

Gaussian filtering

We will assume here that the only randomness in the model comes from the Gaussian Probability Distribution. Let X denotes $N \times T$ matrix, whose entries are i.i.d. random variables, which are normally distributed with zero mean and unit variance. We have already shown in chapter refch2:FRVintro, that as $N, T \to \infty$ and while $r = \frac{N}{T}$ is kept fixed, the probability density function for the eigenvalues of the Wishart matrix $\mathbf{C} = \frac{\mathbf{1}}{\mathbf{T}}\mathbf{Y} \cdot \mathbf{Y^T}$ is given by eq.(2.50). Let us just mention here that the above problem may be tackled along similar lines using the formalism of N and M–transforms and respective duality relations Exploiting the knowledge from Linear Algebra,we may rewrite our correlation matrix $\mathbf{C}$ as:

$$\mathbf{C} = \mathbf{O} \cdot \mathbf{D} \cdot \mathbf{O^T} \tag{3.9}$$

Here $\mathbf{D}$ is a diagonal matrix of eigenvalues of the original matrix $\mathbf{C}$ and $\mathbf{O}$ is an orthogonal matrix whose columns are normalized eigenvectors corresponding with proper eigenvalues, which means that $\mathbf{O}$ fulfills the equation:

$$\mathbf{O} \cdot \mathbf{O^T} = 1 = \mathbf{O} \cdot \mathbf{O^{-1}} \tag{3.10}$$

The trace is conserved, so we write:

$$\mathrm{Tr}\mathbf{C} = \mathrm{Tr}(\mathbf{O} \cdot \mathbf{D} \cdot \mathbf{O^T}) \tag{3.11}$$

Using the (3.10) and cyclic properties of the trace we get

$$\mathrm{Tr}\mathbf{D} = \mathrm{Tr}\mathbf{C} \tag{3.12}$$

Following the fact, $\mathbf{D}$ is a diagonal matrix of eigenvalues one can decompose its trace in the following way:

$$\mathrm{Tr}C = \mathrm{Tr}D = \sum_i \lambda_i + \sum_j \lambda_j \tag{3.13}$$

where $\lambda_i \in [\lambda_-, \lambda_+]$ and $\lambda_j \in [\lambda_1, \lambda_-) \cup (\lambda_+, \lambda_N]$ is set of these eigenvalues, which do not obey the RMT conditions. If we now replace $\sum_i \lambda_i$ by one eigenvalue ζ, we get

$$\zeta = \mathrm{Tr}C - \sum_j \lambda_j \tag{3.14}$$

This results in squeezing the *Random* part of the spectrum to a single degenerated eigenvalue. This amounts to subtracting the contribution of the market mode from nominal value of volatility. Several eigenvalues, that are still above λ_+ contain some potentially useful information about different economic sectors. This 5% of eigenvalues is however responsible for 20% of the total volatility. Which concurs the observations by [LCBP99, PGR+99] on the basis of $S\&P500$ index.

Unraveling eigenvalues – a hand–waiving approach

The predicted r, λ_+, λ_- follow from the equations

$$r = N/T = 100/900 \approx 0.101$$

$$\lambda_+ = (1 + \sqrt{r})^2 = 1,7367 \quad \lambda_- = (1 - \sqrt{r})^2 = 0,4654 \tag{3.15}$$

Suppose we find the values of r, λ_+, λ_- that best fit the bulk of the eigenvalue spectrum. We find

$$r = 0.1055 \pm 0.0129 \qquad \lambda_+ = 1.5238 \pm 0.1456 \qquad \lambda_- = 0.3797 \pm 0.0101 \tag{3.16}$$

and obtain the plot (Fig. 3.5). If we are to believe this estimate, a fraction 0.76 of the variance

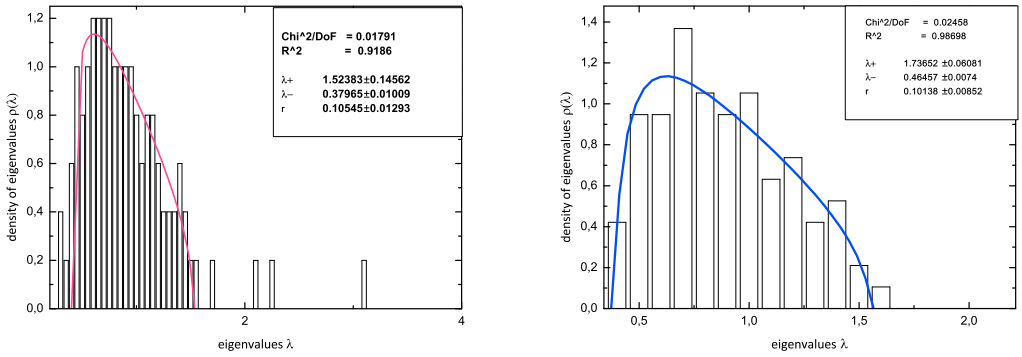

Figure 3.5: LEFT:Spectrum with fitted data.Finiteness effects could take the maximum eigenvalue to 1.67 at the most. RIGHT:Spectrum with reduced data.

is explained by eigenvalues that correspond to random noise. The remaining fraction 0.24 has information. From the (Fig. 3.5), it looks as if we should cut off eigenvalues above 1.6 or so. Summing the eigenvalues themselves, we find that 5% of the variance is explained by eigenvalues greater than 1.6. A more deeper analysis of the correlation spectrum needs the analysis of the residuals *i.e.*, for each stock, we have subtracted factor returns associated with the top 5 eigenvalues. We have found that $r \approx 0.101$ gives the best fit of the Marčenko-Pastur density (see Fig. 3.5). Maximum and minimum eigenvalues are $1,7365$ and $0,4646$. Which is consistent with the data predicted from equation (2.50). The resulting recipe for diminishing the noise effect consists of the following steps:

- Fit the Marčenko-Pastur distribution to the empirical density to determine r and λ_+, λ_-
- All eigenvalues above some number $\lambda_* \approx \lambda_+$ are considered informative; otherwise eigenvalues relate to noise.
- Replace all noise-related eigenvalues λ_i below λ_* with a constant and renormalize so that the trace is conserved.
- Undo the diagonalization of the sample correlation matrix C to obtain the denoised estimate C'.

3.2.3 Noise reduction technique in a Market + sectors model

In the first approximation we have assumed after [JM03], that the components orthogonal to the "market" mode represents pure noise. If we however analyze the spectrum more carefully we notice, that there are 4 distinct eigenvalues, that do not fit very well the predicted spectrum *i.e.*, $\lambda_2 = 3.067, \lambda_3 = 2.214, \lambda_4 = 2.0961, \lambda_5 = 1.6598$. Furthermore it is a common belief, that evolution of stock market can be decomposed into $K_{\max}$ factors associated with the eigenvalues of the correlation matrix.In our case part of these factors might also be hidden in the "bulk" of the spectrum. These small eigenvalues correspond to portfolios of stocks that have very small out–of–sample variance (risk). The key to identifying the number of latent factor lies in correctly understanding the structure of noise, namely the idiosyncratic effects in the data. Once we can separate the estimated eigenvalues of a large factor model into those due to the latent structure and those due to the noise, we can construct the procedure that will consistently estimate the number of factors.

Correlated Wishart Ensemble

We consider now the case, where the true correlations matrix is no more unity matrix *i.e.*, assets can interplay with each other, but where the temporal correlations still represent trivial structure (we assume the stationarity over time). Then the two-point correlation function is given by

$$\left\langle X_{it} X_{jt'} \right\rangle = C_{ij} \delta_{tt'}$$

FRV derivation

In order to unravel the hidden correlation structure, first we need to construct the Green's function

$$G(z) = \frac{1}{N} \int dX \exp\left(-\frac{1}{2} \mathrm{Tr} X C^{-1} X^T\right) \mathrm{Tr}\left(\frac{1}{z - \frac{1}{T} X X^T}\right) \tag{3.17}$$

We already know, that in the case of uncorrelated random variables, the respective Green's function reads

$$G(z) = \frac{1}{N} \int dX \exp\left(-\frac{1}{2} \mathrm{Tr} X X^T\right) \mathrm{Tr}\left(\frac{1}{z - \frac{1}{T} X X^T}\right) \tag{3.18}$$

Since C is positive, the trick is now to consider:

$$X C^{-1} X^T = X C^{-1/2} C^{-1/2} X^T = Y Y^T \rightarrow X = Y C^{1/2} \tag{3.19}$$

Then using the cyclic properties of the Tr

$$\frac{1}{T^k}\mathrm{Tr}(YAY^T)^k \overset{\text{cyclic}}{\overset{\downarrow}{=}} \frac{1}{T^k}\mathrm{Tr}(AY^TY)$$

$G(z)$ is just a product of $\frac{1}{T}Y^TY$ and C,and $\frac{1}{T}Y^TY$ has the same non–zero eigenvalues as $\frac{1}{T}YY^T$ and $T-N$ eigenvalues equal to zero

$$\begin{aligned} G(z) &\sim \tfrac{1}{N}\int dY \exp\left(-\tfrac{1}{2}\mathrm{Tr}YY^T\right)\mathrm{Tr}\left(\tfrac{1}{z-\frac{1}{T}(C)^{1/2}YY^T(C^T)^{1/2}}\right) \\ &= \tfrac{1}{N}\int dY \exp\left(-\tfrac{1}{2}\mathrm{Tr}YY^T\right)\mathrm{Tr}\left(\tfrac{1}{z-\frac{1}{T}Y^TCY}\right) \end{aligned} \tag{3.20}$$

Making use of (2.9) and the multiplication algorithm for N –transforms (2.24) we immediately recover

$$\begin{aligned} G(z) &= \tfrac{(r+z-1)-\sqrt{(1+r+z)^2-4rz}}{2rz} \\ M(z) &= zG(z)-1 = \tfrac{(z-r-1)-\sqrt{(1+r+z)^2-4rz}}{2r} \\ z &\to N_{\frac{1}{T}YY^T}(z) \end{aligned} \tag{3.21}$$

Then

$$N_{\frac{1}{T}YY^T}(z) = \frac{(1+z)(1+rz)}{z} \qquad N_{\frac{1}{T}Y^TY}(z) = \frac{(1+z)(r+z)}{z} \tag{3.22}$$

and

$$M_C(z) = M_C(\omega(z)) \quad \text{where} \quad \omega(z) = \frac{z}{(1+rM_C(z))} \tag{3.23}$$

One then immediately recovers the relation between the moments of the true spectrum and that of empirical one [BGJJ04]

$$\sum_{k=1}^{\infty}\frac{m_k}{z^k} = \sum_{k=1}^{\infty}\frac{M_k}{z^k}\left(1+r\sum_{j=1}\frac{m_l}{z^l}\right)^k \tag{3.24}$$

and the first few relations between the empirical moments m_k and true ones M_k are:

$$\begin{aligned} M_1 &= m_1 \\ M_2 &= m_2 - rm_1^2 \\ M-3 &= m_3 - 3rm_1m_2 + 2r^2m_1^3 \end{aligned} \tag{3.25}$$

To see it, consider again one–point Green's function and rewrite it in terms of the moments [J+10]

$$\begin{aligned} G(z) &= \tfrac{1}{N}\left\langle \mathrm{Tr}\tfrac{1}{z-H}\right\rangle \to \dots \\ zG(z) &= 1+\tfrac{1}{z}m_1+\tfrac{1}{z^2}m_2+\dots = 1+\textstyle\sum_{k=1}^{\infty}\left(\frac{1}{z}\right)^k\frac{1}{N}\sum_i(\lambda_i)^kp_i \end{aligned} \tag{3.26}$$

Then λ_i and p_i are the parameters of the spectrum, that we need, for they can be helpful in reconstructing the hidden structure. For practical reasons a finite number of $k = K_{\max}$ is assumed to carry almost all information about the underlying distribution. Cutting down the series leads to equation in the form

$$1+\sum_{k=1}^{K_{\max}}m_k\left(\frac{1}{z}\right)^k = 1+\sum_{k=1}^{K_{\max}}\left(\frac{1}{z}\right)^k\frac{1}{N}\sum_i(\lambda_i)^kp_i \tag{3.27}$$

Now on the left hand side we have the expression with the moments calculated from empirical correlation matric, while on the right hand side the true moments. Since these expression has to be satisfied for all powers of $\frac{1}{z}$, we end up with a set of $K_{\max}$ nonlinear relations in m_i and M_k. To reproduce them we will use the well known Padé approximation scheme [PTVF07] and rewrite $zG(z)$ in a rational form in terms of $\frac{1}{z}$.[2]

$$\begin{aligned} zG(z) &= z\frac{1}{N}\left\langle \mathrm{Tr}\frac{1}{z-\frac{XX^T}{T}}\right\rangle = \sum_{k=1}^{K_{\max}}\frac{zp_k}{z-\lambda_k} \\ &= \sum_{k=1}^{K_{\max}}\frac{p_k}{1-\frac{1}{z}\lambda_k} = \frac{A_{K_{\max}-1}\left(\frac{1}{z}\right)}{A_{K_{\max}}\left(\frac{1}{z}\right)} = \frac{P\left(\frac{1}{z}\right)}{Q\left(\frac{1}{z}\right)} \end{aligned} \tag{3.28}$$

Zeros of $Q\left(\frac{1}{z}\right)$ are successive λ_k, while p_k can be calculated as follows

$$p_k = \frac{1}{z}\frac{P\left(\frac{1}{z}\right)}{\frac{d}{dz}Q\left(\frac{1}{z}\right)}\Big|_{z=\lambda_k} \tag{3.29}$$

Data analysis and discussion of the results

We will use the relations between the moments to directly reproduce the structure of correlations burried in the "bulk" of the spectrum (Fig. 3.6). From practical reasons sophisticated optimal liquidation portfolio algorithms that balance risk against impact cost involve inverting the correlation matrix. Eigenvalues of the correlation matrix that are small (or even zero) correspond to portfolios of stocks that have nonzero returns but extremely low or vanishing risk; such portfolios are invariably related to estimation errors resulting from insufficient data. The above described version of Factor Analysis can be exploited to identify the hidden dimensions, which were not apparent from direct analysis. These hidden eigenvalues cannot be practically measured due to the uniqueness of financial sample *i.e.*, inferred in repeated experiments. The main achievement is the reduction of dimensionality. Large number of observable eigenvalues is now aggregated in a model to represent underlying structure. These directly unobservable factors represent shared variance or in other words, the degree to which financial time series move together. Except for 5 eigenvalues identified in the previous analysis we have discovered 6 new, that potentially carry important information. For these we have analyzed the structure of eigenvectors and identify the companies, that belong to specified sectors of economy and financial markets. We summarize the results in the (Table 3.1).

3.3 Estimators with weighting schemes

The standard Pearson estimator – an average of the realized cross-correlations over past time T allows all the past values to have an equal impact on current correlations structure. However it is common phenomenon in financial time series to include long memory and power-law tails [Eng82, EB86, Bol86, AFV10, BGW06]. It is now common to set-up updating schemes for cross-covariances like EWMA (exponentially weighted moving average [PPK04, PBL05, Sve07]. The original derivation from [PBL05] will be shown here step-by-step. Roughly speaking, the older

[2] One can reproduce the numerical results along similar lines by minimizing the function $F(\{M_i\},\{m_i\},r,\{\lambda_i\},\{p_i\})$ with respect to observed set of $\{m_i\}$ and unknown $\{M_i\},\{\lambda_i\},\{p_i\}$, which is a complicated task. The conformity of this method from statistical point of view is however not satisfactory and is the purpose of different study[J+10].

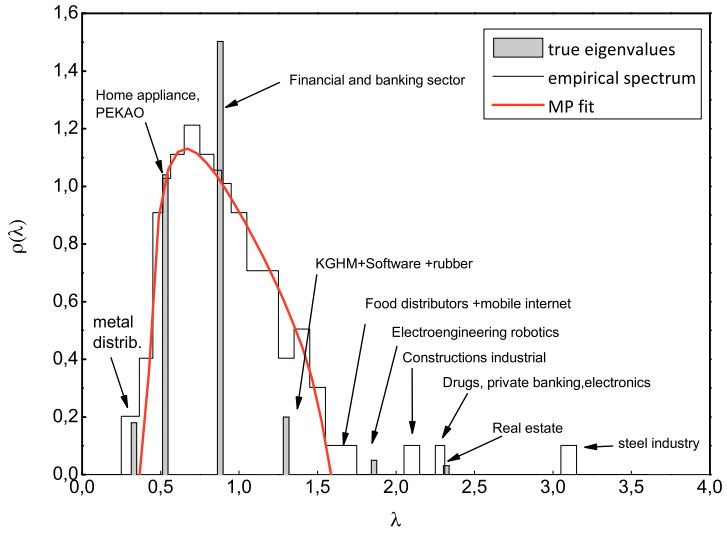

Figure 3.6: The true correlations structure vs. empirical one. The difference is pretty drastic.we have neglected the largest eigenvalue.Recall that each eigenvalue relates to the variance of a portfolio of stocks. A very small eigenvalue means that there exists a portfolio of stocks with very small out-of-sample variance.

No.	eigenvalue	sectors
1	12,62	market mode - no more than 10%, no less than 30% financial sector, software and communication, metals and mining, construction
2	3,06764	Steel, steel mines, luxuries
3	2,32	Construction of private homes and real estate
4	2,214	Pharmacy, personal investments and personal electronics
5	2,0961	Construction in industrial sector
6	1,66	Food distributors and mobile internet
7	1,86	Automatics, electroengineering and machinery
8	1,3	KGHM(copper)&Dębica(Rubber)&Sygnity(software)
9	0,88	Financial and banking sector
10	0,53	Luxury clothes, chemofarmaceuticals and home furnishing, PEKAO(largest bank in Poland)
11	0,333	Metal distributors, foundries,STALEXPORT (steel constructions)

Table 3.1: Companies and sectors that drive Stock Exchange in Poland.

the realized (portfolio) variance (3.3)), the more obsolete it is. We will consider the case where $\mathbf{C} = \mathbf{1}$, but where the empirical matrix is computed according to an exponentially weighted moving average:

$$E_{ij}^{EWMA} = \epsilon \sum_{t'=-\infty}^{t-1} (1-\epsilon)^{t-t'} x_i^{t'} x_j^{t'} \quad \text{for} \quad \epsilon = 1/T \tag{3.30}$$

Since

$$E_{ij}^{EWMA} = (1-\epsilon) E_{ij}^{EWMA} + \epsilon x_i^t x_j^t \tag{3.31}$$

We can invert the resolvent $G(z)$ to find the Blue's function:

$$\delta B_t(z) = \frac{1}{z} + \frac{r}{N(1-rz)} \tag{3.32}$$

Using the scaling properties of $G(z)$ and (2.20) allows one to write:

$$B_{EWMA}(z) = B_{(1-\epsilon)E}(z) + B_t(z) - \frac{1}{z} \tag{3.33}$$

and finally we obtain

$$B_{EWMA}(z) = \frac{1}{z}\left(1 - \frac{1}{rT\epsilon} \ln\left(1 - \frac{rT\epsilon z}{1 - \frac{rT\epsilon z}{\exp rT\epsilon - 1}}\right)\right) \tag{3.34}$$

Going back to the resolvent to find the density, we finally get the result first obtained in [PPK04]:

$$\rho(\lambda) = \frac{1}{\pi}\mathrm{Im}G(z) \quad \text{where} \quad G(z) \quad \text{solves} \quad zrG(z) = q - \ln(1 - rG(z)) \tag{3.35}$$

And the edges of the eigenvalue spectrum $(\lambda_\pm)$[Zee96]

$$\lambda_\pm = B(z_\pm) \quad \text{where} \quad B'(z_\pm) = 0 \tag{3.36}$$

In the case at hand, by evaluating $B(z)$ when $B'(z) = 0$ we can write directly an equation whose solutions are the spectrum edges $(\lambda_\pm)$

$$\lambda_\pm = \log(\lambda_\pm) + r + 1 \tag{3.37}$$

When r is zero, the spectrum is a δ in 1 as expected. But as the noise increases (or the characteristic time decreases) the lower edge approach zero very quickly as $\lambda_- \sim \exp(-r)$.

3.3.1 Evidence from the data

Although there are no exact zero eigenvalues for EWMA matrices,the smallest eigenvalue is very close to zero. This is evidently confirmed in the empirical case, where we put a stress on investigation of delayed dependencies among the stocks. Uniform weights, means that the weight factors are equally distributed, while "noweights" scheme is simply EWMA with all weight factors equal to 1. Standard EWMA gives today's measurement twice the significance, that the uniform scheme would assign. In this section we investigate the problem whether so called "implied correlations" given in EWMA scheme are useful in unraveling unobserved temporal structure. Or in other words if a forward looking structure, that puts more impact on recent observations and incorporates market expectations may provide interesting additional

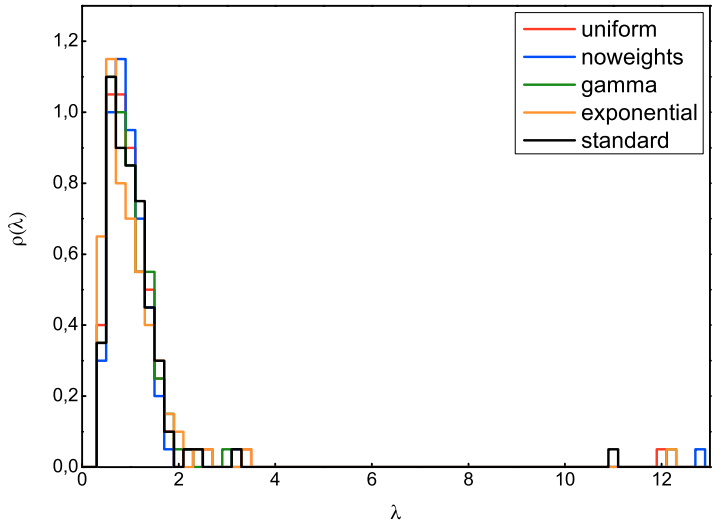

Figure 3.7: Empirical correlations calculated with EWMA schemes. The standard scheme reflect to the ordinary EWMA (3.30) weights, while exponential, gamma and uniform weights, are drawn from respective probability distribution. The case with no weights is the case, where empirical correlations matrix is calculated according to the standard Pearson estimator, though it reflects the ordinary Marčenko - Pastur case [MP67].One can clearly see, that the bulk of the spectrum for all weighted schemes, remains in an almost unchanged form,as it should [PBL05] while position of the largest eigenvalues that significantly differ from (2.50)"pure" random spectrum, changes drastically according to the calculations scheme.

information not provided by the historical Pearson estimator (1.6). Various forms of weighted moving averages have been developed to address the objection, that more recent values play major role. Instead of just adding up the measurements for a sequence of days and divide by the number of days T in EWMA scheme each measurement was first multiplied by a weight factor, which differed from day to day. The final sum is divided not by the number of days, but by the sum of all weight factors. Larger weight factors used for more recent days and consecutively smaller for measurements further back in time - the dynamics is more responsible to recent changes. We haven't however notice any significant improvements compared to the previous section – deviations are only due to the hidden spatio–temporal structure between the returns (Fig. 3.7). Different weighting schemes leave the "bulk" practically unchanged, while the largest eigenvalue is strongly affected by the parameter ϵ. This is due to the fact, that market mode created on the basis of all the stocks (mostly minor companies) is more sensitive to the short term movements in price changes. The smaller the ϵ, the more responsive the "market" to the daily price changes.The bulk, reacts much slower, since we have extracted out the "ghost" effect of a single price change – shocks die out exponentially and more smoothly than in the unweighted scheme. The approach from previous sections assumed that past is prologue. We measure history in the hope it is predictive. Factor models on the other hand hope that "market knows best" and the market price contains, even if implicitly a consensus estimates of risk. EWMA ensures, that covariance is weighted or biased towards more recent data. A more general class of "weighted estimators"of the cross–correlations for the correlated

Gaussian random variables $\mathbf{Y}$ *i.e.*, the doubly correlated Wishart ensemble, will be extensively studied in chapter 4 in the context of VARMA(q_1, q_2) processes.

$$\mathbf{c}^{\text{weight}} = \frac{1}{T}\sqrt{\mathbf{C}}\widetilde{\mathbf{Y}}\sqrt{\mathbf{A}}\mathbf{W}\sqrt{\mathbf{A}}\widetilde{\mathbf{Y}}^{\mathrm{T}}\sqrt{\mathbf{C}} \quad \text{where} \quad \mathbf{W} \equiv T\text{diag}\left(w_1, \ldots, w_T\right). \tag{3.38}$$

3.4 Empirical properties of Correlations matrix

The correlation matrix encapsulates not only the information about mutual dependencies, but also how they behave over certain period of time. Within time dependency, the eigenvalues are themselves time–series which contain the same information as the raw correlation. The analysis of temporal structure of correlation matrices for NYSE has already been proposed by [Zum09, CRC09a]. The authors analyzed projector dynamics and the empirical properties of the time–series of eigenvalues by means of sliding window technique. In this section we will stick to the latter approach.

3.4.1 Cross-correlations dynamics

The salient feature of the Random Matrix Theory [Meh04, Eyn00] is that distribution of eigenvalues are stationary over a certain period of time[3]. Though information about price changes and other important events (e.g. announcement of inflation) truly affect the correlation structure. It is then obvious, that these relations bombed by incoming information can themselves evolve in time. In this section we verify this assumption more deeply and to analyze the relation between empirical financial cross correlations and those of the pure noise we have performed the following analysis. For $N = 100$ stocks from Warsaw stock exchange, we calculate the equal-time cross-correlation matrix between time-series of stock returns using sliding time window for within time frame of $T = 200$ observations. This period was chosen to ensure the data would be close to non–stationary in each sliding window. To study the dynamics of each of the eigenvalues using a sliding window, we normalize each eigenvalue in time by adjusting the mean and dividing by the standard deviation of the eigenvalue i over a particular period τ. This normalization allows us to visually compare eigenvalues, even if their magnitudes are significantly different (Fig. 3.8). The correlations in the volatility are now represented by eigenvalues and it is easy to notice, that they remain positive over a long period of time ("bull" market). The repulsion between the largest eigenvalue and the smaller ones is evident. This is the consequence of the properties of the trace. The trace must remain constant under any transformation of the correlations matrix. Though any change of the largest eigenvalue is immediately compensated by a change in one or more other eigenvalues. Small eigenvalues move counter to the largest eigenvalue. Overall the spectrum is pretty static, in agreement with common belief, that the largest part of dynamics is captured by volatility. This is also consistent with the "central limit theorems" for large random matrices *i.e.*, the larger the matrix, the more deterministic it behaves. Yet the eigenvalues closely related to the "true" spectrum represent clear and distinct time evolution, though estimation by stationary spectrum seems inappropriate. This together shows, that reduction of a correlation matrix to just a few leading eigenmodes misses a lot of dynamics.

[3] Any stochastic process is considered strict stationary if it is a series of independent identically distributed samples. For weak stationarity we only assume, that mean and the variance for this stochastic process exist and are invariant under translations in time.

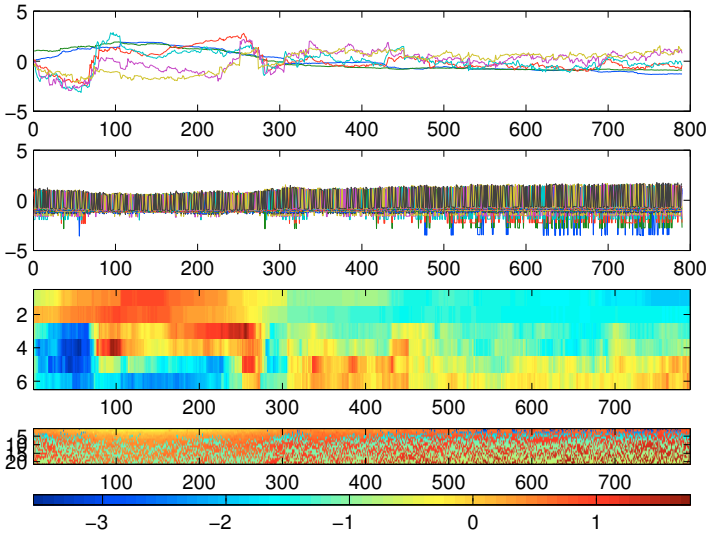

Figure 3.8: Cross-correlations dynamics over sliding window.TOP:The first 7 largest eigenvalues of a daily correlation matrix for Polish Stock Market. FIRST BELOW: the dynamics of the "bulk" of the spectrum.The last two figures represent, the normalized dynamics of the correlation spectrum for largest eigenvalues and the "bulk".

Stability of eigenvalues

Furthermore we have analyzed the stationarity of the resulting time–series of eigenvalues. For each of them standard unit-root tests were performed [Gre03]. It resulted (see Table 3.2 for details), that eigenvalues in the vicinity of "true" eigenvalues the hypothesis of non–stationarity could not be rejected, namely not all eigenvalues in the "bulk" look like stationary white noise without correlations. The empirical part of the spectrum "close" to the latent factors exhibits highly non-stationary (probably of Brownian origins) or close to non-stationary behavior which can be easily encoded in terms of Ornstein-Uhlenbeck process. This also demonstrates, that non–trivial dynamics is consequently married with the "true" correlations spectrum and is a potential result of time structure, that may affect the movement of the spectrum.

Dynamics of the top eigenvalue vs "bulk"

As we have already presented the financial covariance matrices are such that the largest eigenvalue is well separated from the "bulk", where all other eigenvalues reside. The financial interpretation of this large eigenvalue – so-called "market mode": in a first approximation, all stocks move together, up or down. One can state this more precisely in the context of the one factor model [Noh00, MS04, LM03, Sha64], where the ith stock return at time t is written[PBL05] as:

$$y_t^i = \beta_i \phi_t + \varepsilon_t^i, \tag{3.39}$$

where the market mode ϕ_t is common to all stocks through their market exposure β^i and the ε_t^i are idiosyncratic noises, uncorrelated from stock to stock. Within such a model, the correlation

p-value ADF test	empirical eigenvalues	true spectrum
0,7905	12,62	12,62
0,4607	3,06764	3,06764
0,0594	2,21402	2,32
0,1106	1,46029	2,214
0,1082	1,42788	1,86
0,1241	1,41732	
0,1197	1,38858	
0,1564	1,37285	
0,2631	1,35017	
0,2818	1,33494	
0,2833	1,30438	1,3
0,1905	1,27726	
0,166	1,25024	
0,1377	1,24385	
0,1646	1,22344	
0,127	1,19213	
0,1421	1,17177	
0,1071	1,15646	0,88
0,1123	0,41438	0,53
0,1455	0,40478	
0,1626	0,39469	
0,2051	0,39015	
0,1905	0,37672	
0,2466	0,34793	0,333
0,1746	0,29782	
0,1422	0,26703	

Table 3.2: Results from stationarity test for eigenvalues. The table shows eigenvalues for which the non-stationarity hypothesis could not be rejected.

matrix reads:

$$C_{ij} = \beta_i \beta_j \sigma_\phi^2 + \sigma_i^2 \delta_{ij}. \tag{3.40}$$

When all σ_i's are equal, this matrix is easily diagonalized; for N stocks, its largest eigenvalue is $\Lambda_0 = (\sum_j \beta_j^2)\sigma_\phi^2 + \sigma^2$ and is of order N, and all the other $N-1$ eigenvalues Λ_α are equal to σ^2. The largest eigenvalue corresponds to the eigenvector β_i. More generally, the largest eigenvalue Λ_0, normalized by the average square volatility of the stocks, can be seen as a proxy for the average interstock correlation. The empirical matrix $\mathbf{E}$ evolves in time as:

$$E_{ij,t} = (1-\epsilon)E_{ij,t-1} + \epsilon r_t^i r_t^j. \tag{3.41}$$

Denoting as λ_{0t} the largest eigenvalue of E_t associated to ψ_{0t} and using standard perturbation theory, valid for $\epsilon \ll 1$

$$\lambda_{0t} = (1-\epsilon)\lambda_{0t-1} + \epsilon\langle\psi_{0t-1}|C|\psi_{0t-1}\rangle + \epsilon\langle\psi_{0t-1}|\eta_t|\psi_{0t-1}\rangle, \tag{3.42}$$

with $\eta_{ij} = r^i r^j - \langle r^i r^j\rangle$. For Gaussian returns

$$\langle\eta_{ij}\eta_{k\ell}\rangle = C_{ik}C_{j\ell} + C_{i\ell}C_{jk}. \tag{3.43}$$

In the limit where Λ_0 becomes much larger than all other eigenvectors, the above equation simplifies to:

$$\lambda_{0t} \approx (1-\epsilon)\lambda_{0t-1} + \epsilon\cos^2\theta_{t-1}\Lambda_0\left[1+\xi_t\right], \tag{3.44}$$

leads to a Langevin equation for λ_0:

$$\frac{d\lambda_0}{dt} = \epsilon(\cos^2\theta\Lambda_0 - \lambda_0) + \epsilon\cos^2\theta\xi_t, \tag{3.45}$$

where $\cos\theta_t \equiv \langle\psi_{0t}|\Psi_0\rangle$ and ξ_t is a random noise term of mean zero and variance equal to 2, which becomes Gaussian for large matrices.

Evolution of the uninformative eigenvalues

The most popular model of stock price dynamics assumes, that the return is a Gaussian stationary process and stocks are uncorrelated. A direct analysis of all pairs of stocks is, of course, unreasonable even for moderate size of N. That is why we only dwell on the evolution of eigenvalues in a spectrum (Fig.3.9). We have especially focused on the largest eigenvalues. The analysis of the eigenvalues evolution in the sliding time window discovers a slow temporal dynamics of the statistical properties of the ensemble, which is consistent with the assumption that common economic factors affecting the time evolution of stock prices are present in financial markets. For comparison we have compared the dynamics of one of the redundant eigenvalues with the dynamics of the eigenvalues generated from the pure noise *i.e.*, that came from artificially generated correlation matrix, for which the samples had a standard Normal distribution with zero mean and unit variance. We have recovered that the redundant (uninformative) eigenvalues are in fact indistinguishable from pure noise. The result of these analyzes show that a picture based on the assumption that all stock prices are uncorrelated is not a realistic one. The stock exchange is far more "complex" than a collection of several independent random processes and the presence of cross-correlation and temporal correlations between pairs of stocks supports the assumption that common economic factors are affecting the time evolution of stock prices in financial markets.

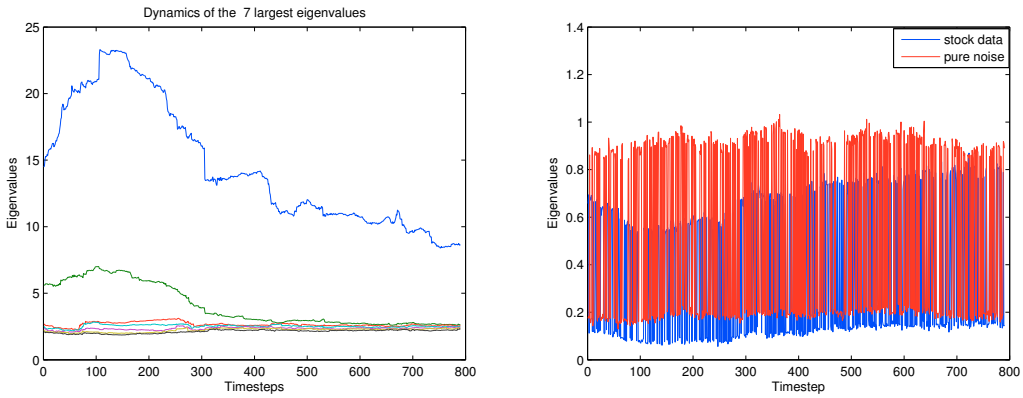

Figure 3.9: LEFT:Time evolution of the 7 largest eigenvalues.RIGHT: Sample eigenvalue from the bulk of the spectrum, far from true correlation structure versus pure white noise time series.They are almost indistinguishable, which again confirms that these eigenvalues are redundant, do not capture the dynamics of the spectrum and might be eliminated without loss of important information.

Distribution of eigenvector components

The deviations from RMT (Fig. 3.3) should also be displayed in the statistics of the corresponding eigenvector components. If there is no information in an eigenvalue, we expect the distribution of the components of its eigenvector to be a maximum entropy distribution. Gaussian distribution in that case can be easily proven. If we superimpose the empirical distribution of eigenvector components and the zero-information limiting density for various eigenvalues we see, that true eigenvalues, that do not fall into the RMT predictions are informative (*i.e.*, the components of its eigenvectors deviate systematically from Gaussian distribution), while most of the bulk represents white noise dynamics. Fig.3.10 shows that for the typical eigenvalues from the bulk the conformity is very good. The distribution of the component of the eigenvector corresponding to the largest eigenvalue is apparently uniform – mostly the components are shifted to the right (have the same sing), again confirming that significant stocks participating in the eigenvector have common component, that affects them in the same manner. The stock market, as viewed through the eigenspectrum and the eigenvectors of the correlation matrix shows that the collective dynamics of the largest eigenvalue -"market mode" cannot be approximated by pure white noise. The bulk itself is also not fully composed of the squares of white noises - some eigenvalues close to the specific factors -"true " spectrum have nonlinear dynamics both in space and time.

3.5 Summary

While factor models have been used for almost a century, standard multivariate methods were developed under assumption, that time dimension T grows large, while the number of samples N is small and fixed. We have shown the fluency of Free Random Variables calculus in the context of equal–time correlations estimators, where standard theory becomes an unreliable guide to data analysis. Even quite complex problems can be solved within not very lengthy and complicated framework. The implementation of this new tool in complexity town, which takes into account the special nature of large datasets leads us to reconsider stylized facts we have taken

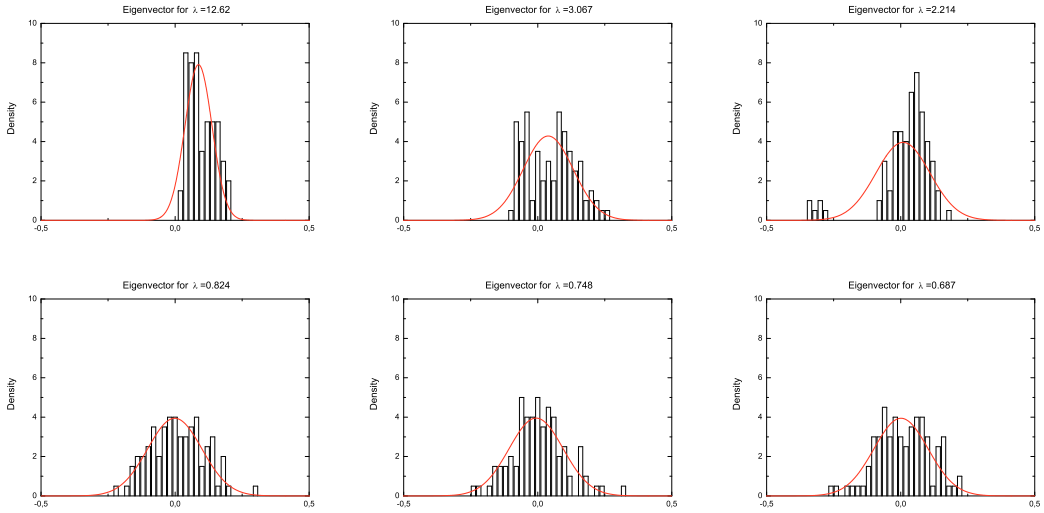

Figure 3.10: ABOVE:Three eigenvectors corresponding to the three largest informative eigenvalues.BELOW: Three eigenvectors for the uninformative eigenvalues from the bulk of the spectrum.

for granted. Applications of the above described approaches to financial correlation matrices are relatively recent [LCBP99, Zum09] but very active [DKRS01, PBL05, AFV10]. We have shown that comparing the empirical eigenvalues to the empirical correlation matrix build from Warsaw Stock Exchange assets to the theoretical upper edge of the Marčenko-Pastur [MP67] spectrum allows one to extract statistically significant factors [JM03]. We have also considered the case, when these assets are initially spatially correlated with each other *i.e.*, the case when off–diagonal terms in true correlation matrix are significantly different from zero. Then, relating via relation between true moments of the spectrum and that of empirical estimator we have immediately recovered the true correlation structure buried below the band edge, as expected [BJ04]. In our approximation scheme it is commonly assumed, that underlying random variables for the correlations matrix are stationary *i.e.*, its general properties (eg. moments) do not change over time. We have focused on the analysis of eigenvalues spectrum over time. For that we have used "sliding window" procedure to recover time series for each eigenvalues. By performing simple unit root tests we have shown, that indeed most of the spectrum behave stationary *i.e.*, it is comparable to white noise, while the eigenvalues close to the true correlations structure are relatively far from stationarity.

That which is static and repetitive is boring. That which is dynamic and random is confusing. In between lies art.

John Locke

4 Lagged Correlations from VARMA(p,q)

Finite order vector autoregressive moving average models (VARMA) motivated by Wold decomposition theorem [Wol38] as an appriopriate multivariate setting for studying the dynamics of stationary time series. Vector autoregressive (VAR) models are cornerstones in contemporary macroeconomics, being a part of an approach called the "dynamic stochastic general equilibrium"(DSGE), which is superseding traditional large–scale macroeconometric forecasting methodologies [Sim80]. The motivation behind them is based on the assertion that more recent values of a variable are more likely to contain useful information about its future movements than the older ones. On the other hand, a standard tool in multivariate time series analysis is vector moving average (VMA) models, which is really a linear regression of the present value of the time series w.r.t. the past values of a white noise. A broader class of stochastic processes used in macroeconomics comprises both these kinds together in the form of vector autoregressive moving average (VARMA) models. These methodologies can capture certain spatial and temporal structures of multidimensional variables which are often neglected in practice; including them not only results in more accurate estimation, but also leads to models which are more interpretable. The contents of this chapter were first published in [BJNS10]. The last section - conclusions is however considerably expanded.

4.1 Doubly Correlated Wishart Ensembles and Free Random Variables

VARMA models are constructed from a number of univariate ARMA (Box–Jenkins; see for example [BJR94]) processes, typically coupled with each other. Here however we investigate only a significantly simplified circumstance when there is no coupling between the many ARMA components (in fact we consider N copies of the same ARMA process). One may argue that this is too far fetched and will be of no use in describing an economic reality. However, we will treat it as a "zeroth–order hypothesis," analogously to the idea of [LCBP99, PGR+99] in finance, namely that the case with no cross–covariances is considered theoretically, and subsequently compared to some real–world data modeled by a VARMA process. One may then suppose, that any discrepancy between the two will reflect nontrivial cross–covariances present in the system, thus permitting their investigation.

4.1.1 Doubly Correlated Wishart Ensembles

A challenging and yet increasingly important problem is the estimation of large covariance matrices generated by these stationary VARMA(q_1, q_2) processes, since high dimensionality of the data as compared to the sample size is quite common in many statistical problems (the "dimen-

sionality curse"). Therefore, an appropriate "noise cleaning" procedure has to be implemented, and random matrix theory (RMT) provides a natural and efficient outfit for doing that. In particular, the mean spectral densities (a.k.a. "limiting spectral distributions," LSD) of the Pearson estimators of the cross–covariances for the VMA(1) and VAR(1) models, in the relevant high–dimensionality sector and under the full decoupling, have been derived in [JWMH09] by applying the framework proposed by [BS06]. The main aim of this chapter is to show how these calculations can be considerably simplified by resorting to a mathematical concept of the free random variables (FRV) calculus [DNV92, Spe94], succinctly introduced in sec. 4.1. Our general FRV formula [BJJ+09] allows not only to rediscover, which much less strain, the two fourth–order polynomial equations obtained in [JWMH09] in the VMA(1) and VAR(1) cases, but also to derive a sixth–order equation (B.6) which produces the mean spectral density for a more involved VARMA$(1,1)$ model. The results are verified by numerical simulations, which show a perfect agreement. Also, practical relevance of VARMA$(1,1)$ is reinforced by fitting the density retrieved from (B.6) to real macroeconomic data. This is all done in section. 4.2.

VARMA models vs.Correlated Gaussian Random Variables

VARMA(q_1, q_2) stochastic processes, fall within quite a general set–up encountered in many areas of science where a probabilistic nature of multiple degrees of freedom evolving in time is relevant, for example, multivariate time series analysis in finance, applied macroeconometrics and engineering. Namely consider N time–dependent random variables which are measured at T consecutive time moments (separated by some time interval δt); let Y_{ia} be the value od the i–th $(i = 1, \ldots, N)$ random number at the a–th time moment $(a = 1, \ldots, T)$; together, they make up a rectangular $N \times T$ matrix $\mathbf{Y}$. In what usually would be the first approximation, each Y_{ia} is supposed to be drawn from a Gaussian probability distribution. We will also assume that they have mean values zero, $\langle Y_{ia} \rangle = 0$. These degrees of freedom may in principle display mutual correlations. A set of correlated zero–mean Gaussian numbers is fully characterized by the two–point covariance function, $\mathcal{C}_{ia,jb} \equiv \langle Y_{ia} Y_{jb} \rangle$ if the underlying stochastic process generating these numbers is stationary. Linear stochastic processes, including VARMA(q_1, q_2), belong to this category. We will restrict our attention to an even narrower class where the cross–correlations between different variables and the auto–correlations between different time moments are factorized, *i.e.*,

$$\langle Y_{ia} Y_{jb} \rangle = C_{ij} A_{ab}. \tag{4.1}$$

In this setting, the inter–variable covariances do not change in time (and are described by an $N \times N$ cross–covariance matrix $\mathbf{C}$), and also the temporal covariances are identical for all the numbers (and are included in a $T \times T$ auto–covariance matrix $\mathbf{A}$; both these matrices are symmetric and positive–definite). The Gaussian probability measure with this structure of covariances is known from textbooks,

$$P_{\text{c.G.}}(\mathbf{Y}) \mathrm{D}\mathbf{Y} = \frac{1}{\mathcal{N}_{\text{c.G.}}} \exp\left(-\frac{1}{2} \sum_{i,j=1}^{N} \sum_{a,b=1}^{T} Y_{ia} \left[\mathbf{C}^{-1}\right]_{ij} Y_{jb} \left[\mathbf{A}^{-1}\right]_{ba} \right) \mathrm{D}\mathbf{Y} =$$

$$= \frac{1}{\mathcal{N}_{\text{c.G.}}} \exp\left(-\frac{1}{2} \mathrm{Tr} \mathbf{Y}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{Y} \mathbf{A}^{-1} \right) \mathrm{D}\mathbf{Y}, \tag{4.2}$$

where the normalization constant $\mathcal{N}_{\text{c.G.}} = (2\pi)^{NT/2} (\mathrm{Det}\mathbf{C})^{T/2} (\mathrm{Det}\mathbf{A})^{N/2}$, and the integration measure $\mathrm{D}\mathbf{Y} \equiv \prod_{i=1}^{N} \prod_{a=1}^{T} \mathrm{d}Y_{ia}$, while the letters "c.G." stand for "correlated Gaussian." A

standard way to approach correlated Gaussian random numbers is to recall that they can always be decomposed as linear combinations of uncorrelated Gaussian degrees of freedom

$$\mathbf{Y} = \sqrt{\mathbf{C}}\widetilde{\mathbf{Y}}\sqrt{\mathbf{A}}, \quad \text{which yields} \quad P_{\mathrm{G.}}(\widetilde{\mathbf{Y}})\mathrm{D}\widetilde{\mathbf{Y}} = \frac{1}{\mathcal{N}_{\mathrm{G.}}}\exp\left(-\frac{1}{2}\mathrm{Tr}\widetilde{\mathbf{Y}}^{\mathrm{T}}\widetilde{\mathbf{Y}}\right)\mathrm{D}\widetilde{\mathbf{Y}}, \tag{4.3}$$

where the square roots of the covariance matrices, necessary to facilitate the transition, exist due to the positive–definiteness of $\mathbf{C}$ and $\mathbf{A}$; the new normalization reads $\mathcal{N}_{\mathrm{G.}} = (2\pi)^{NT/2}$.

Estimating Equal–Time Cross–Covariances

An essential problem in multivariate analysis is to determine (estimate) the covariance matrices $\mathbf{C}$ and $\mathbf{A}$ from given N time series of length T of the realizations of our random variables Y_{ia}. For simplicity, we do not distinguish in notation between random numbers, *i.e.*, the population, and their realizations in actual experiments, *i.e.*, the sample. Since the realized cross–covariance between degrees i and j at the same time a is $Y_{ia}Y_{ja}$, the simplest method to estimate the today's cross–covariance c_{ij} is to compute the time average,

$$c_{ij} \equiv \frac{1}{T}\sum_{a=1}^{T} Y_{ia}Y_{ja}, \qquad i.e., \qquad \mathbf{c} = \frac{1}{T}\mathbf{Y}\mathbf{Y}^{\mathrm{T}} = \frac{1}{T}\sqrt{\mathbf{C}}\widetilde{\mathbf{Y}}\mathbf{A}\widetilde{\mathbf{Y}}^{\mathrm{T}}\sqrt{\mathbf{C}}. \tag{4.4}$$

This is usually named the "Pearson estimator", up to the prefactor which depending on the context is $1/(T-1)$ or $1/T$. Other estimators might be introduced, such as between distinct degrees of freedom at separate time moments ("time–delayed estimators"), or with certain decreasing weights given to older measurements to reflect their growing obsolescence ("weighted estimators"), but we will not investigate them here. Furthermore, in the last equality in (4.4), we cast $\mathbf{c}$ through the uncorrelated Gaussian numbers contained in $\widetilde{\mathbf{Y}}$, the price to pay for this being that the covariance matrices now enter into the expression for $\mathbf{c}$, making it more complicated; this will be the form used hereafter. The random matrix $\mathbf{c}$ is called a "doubly correlated Wishart ensemble" [Wis28]. Let us also mention that the auto–covariance matrix $\mathbf{A}$ can be estimated through

$$\mathbf{a} \equiv (1/N)\mathbf{Y}^{\mathrm{T}}\mathbf{Y}$$

However, it is verified that this object carries identical information to the one contained in $\mathbf{c}$ (it is "dual" to $\mathbf{c}$), and therefore may safely be discarded. Indeed, these two estimators have same non–zero eigenvalues (modulo an overall rescaling by r), and the larger one has $|T-N|$ additional zero modes.

4.1.2 The Multiplication Algorithm in terms of FRV

We have presented the original mathematical formulations [DNV92, Spe94] in chapter 2 in a slightly different language, namely in terms of S–transforms and R–transforms. For the purposes of this chapter -(multiplication of random matrices) we will employ slightly different definitions, as they serve better than the original ones.

The M–Transform and the Spectral Density

As disclosed in chapter 2 to explore the eigenvalue density of a (real symmetric $N \times N$) random matrix $\mathbf{H}$ it is convenient to work with either "Green's function" (or the "resolvent") or the

"M–transform" of $\mathbf{H}$

$$G_{\mathbf{H}}(z) \equiv \frac{1}{N}\left\langle \mathrm{Tr}\frac{1}{z\mathbf{1}_N - \mathbf{H}}\right\rangle, \quad \text{or} \quad M_{\mathbf{H}}(z) \equiv zG_{\mathbf{H}}(z) - 1, \tag{4.5}$$

The latter is also called the "moments' generating function," since if the "moments"

$$M_{\mathbf{H},n} \equiv (1/N)\langle \mathrm{Tr}\mathbf{H}^n\rangle$$

of $\mathbf{H}$ exist, it can be expanded into a power series around $z \to \infty$ as $M_{\mathbf{H}}(z) = \sum_{n\geq 1} M_{\mathbf{H},n}/z^n$. [1]

The N–Transform and Free Random Variables

The doubly correlated Wishart ensemble $\mathbf{c}$ (4.4) may be viewed as a product of several random and non–random matrices. The general problem of multiplying random matrices seems formidable. In classical probability theory, it can be effectively handled in the special situation when the random terms are independent: then, the exponential map reduces it to the addition problem of independent random numbers, solved by considering the logarithm of the characteristic functions of the respective PDFs, which proves to be additive. In matrix probability theory, a crucial insight came from D. Voiculescu and coworkers and R. Speicher [DNV92, Spe94], who showed how to parallel the commutative construction in the noncommutative world. It starts with the notion of "freeness," which basically comprises probabilistic independence together with a lack of any directional correlation between two random matrices. This nontrivial new property happens to be the right extension of classical independence, as it allows for an efficient algorithm of multiplying free random variables (FRV), which we state below:

Step 1: Suppose we have two random matrices, $\mathbf{H}_1$ and $\mathbf{H}_2$, mutually free. Their spectral properties are best wrought into the M–transforms (4.5), $M_{\mathbf{H}_1}(z)$ and $M_{\mathbf{H}_2}(z)$.

Step 2: The critical maneuver is to turn attention to the functional inverses of these M–transforms, the so–called "N–transforms,"

$$M_{\mathbf{H}}\left(N_{\mathbf{H}}(z)\right) = N_{\mathbf{H}}\left(M_{\mathbf{H}}(z)\right) = z. \tag{4.6}$$

Step 3: The N–transforms submit to a very straightforward rule upon multiplying free random matrices (the "FRV multiplication law"),

$$N_{\mathbf{H}_1\mathbf{H}_2}(z) = \frac{z}{1+z}N_{\mathbf{H}_1}(z)N_{\mathbf{H}_2}(z), \qquad \text{for free } \mathbf{H}_1,\, \mathbf{H}_2. \tag{4.7}$$

Step 4: Finally, it remains to functionally invert the resulting N–transform $N_{\mathbf{H}_1\mathbf{H}_2}(z)$ to gain the M–transform of the product, $M_{\mathbf{H}_1\mathbf{H}_2}(z)$, and consequently, all the spectral properties via formula (2.3).

With such a simple prescription (relying on the choice of the M–transform as the carrier of the mean spectral information, and the construction of its functional inverse, the N–transform, which essentially multiplies under taking the free product) one resolves the multiplication problem for free random noncommutative objects.

[1] It should however be underlined that even for probability measures disallowing such an expansion (heavy–tailed distributions), the quantities (4.5) still manage to entirely capture the spectral properties of $\mathbf{H}$; hence the name "M–transform" more appropriate, in addition to being more compact.

Doubly Correlated Wishart Ensembles from Free Random Variables

The innate potential of the FRV multiplication algorithm (4.7) is surely revealed when inspecting the doubly correlated Wishart random matrix(4.4).

$$\mathbf{c} = (1/T)\sqrt{\mathbf{C}}\widetilde{\mathbf{Y}}\mathbf{A}\widetilde{\mathbf{Y}}^{\mathrm{T}}\sqrt{\mathbf{C}}$$

This has been done in detail in [BJJ+09], so we will only accentuate the main results here, referring the reader to the original paper for a thorough explanation. The idea is that one uses twice the cyclic property of the trace (which permits cyclic shifts in the order of the terms), and twice the FRV multiplication law (4.7) (to break the N–transforms of products of matrices down to their constituents), in order to reduce the problem to solving the uncorrelated Wishart ensemble $(1/T)\widetilde{\mathbf{Y}}^{\mathrm{T}}\widetilde{\mathbf{Y}}$. This last model is further simplified, again by the cyclic property and the FRV multiplication rule applied once, to the standard **GOE** random matrix squared (and the projector $\mathbf{P} \equiv \mathrm{diag}(\mathbf{1}_N, \mathbf{0}_{T-N})$, designed to chip the rectangle $\widetilde{\mathbf{Y}}$ off the square **GOE**), whose properties are firmly established. Let us sketch the derivation,

$$N_{\mathbf{c}}(z) \overset{\text{cyclic}}{\downarrow\!=} N_{\frac{1}{T}\widetilde{\mathbf{Y}}\mathbf{A}\widetilde{\mathbf{Y}}^{\mathrm{T}}\mathbf{C}}(z) \overset{\text{FRV}}{\downarrow\!=} \frac{z}{1+z}N_{\frac{1}{T}\widetilde{\mathbf{Y}}\mathbf{A}\widetilde{\mathbf{Y}}^{\mathrm{T}}}(z)N_{\mathbf{C}}(z) \overset{\text{cyclic}}{\downarrow\!=}$$

$$\overset{\text{cyclic}}{\downarrow\!=} \frac{z}{1+z}N_{\frac{1}{T}\widetilde{\mathbf{Y}}^{\mathrm{T}}\widetilde{\mathbf{Y}}\mathbf{A}}(rz)N_{\mathbf{C}}(z) \overset{\text{FRV}}{\downarrow\!=} \frac{z}{1+z}\frac{rz}{1+rz}N_{\frac{1}{T}\widetilde{\mathbf{Y}}^{\mathrm{T}}\widetilde{\mathbf{Y}}}(rz)N_{\mathbf{A}}(rz)N_{\mathbf{C}}(z) =$$

$$= rzN_{\mathbf{A}}(rz)N_{\mathbf{C}}(z). \tag{4.8}$$

This is the basic formula. Since the spectral properties of $\mathbf{c}$ are given by its M–transform, $M \equiv M_{\mathbf{c}}(z)$, it is more pedagogical to recast (4.8) as an equation for the unknown M,

$$z = rMN_{\mathbf{A}}(rM)N_{\mathbf{C}}(M). \tag{4.9}$$

It provides a means for computing the mean spectral density of a doubly correlated Wishart random matrix once the "true" covariance matrices $\mathbf{C}$ and $\mathbf{A}$ are given. In this study, only a particular instance of this fundamental formula is applied, namely with an arbitrary auto–covariance matrix $\mathbf{A}$, but with trivial cross–covariances, $\mathbf{C} = \mathbf{1}_N$. Using that $N_{\mathbf{1}_K}(z) = 1 + 1/z$, equation (4.9) thins out to

$$rM = M_{\mathbf{A}}\left(\frac{z}{r(1+M)}\right), \tag{4.10}$$

which will be strongly exploited below. Let us mention that these equalities (4.9), (4.10) have been derived through other, more tedious, techniques (the planar Feynman–diagrammatic expansion, the replica trick) in [BGJJ04, BJ04, BJW05b, BGJW06, BJW05a].

4.2 VARMA from Free Random Variables

In what follows, we will assume that the VMA(q), VAR(q), or VARMA(q_1, q_2) stochastic processes are covariance (weak) stationary; for details, we refer to [Lü05]. It implies certain restrictions on their parameters, but we will not bother with this issue in the current work. Another consequence is that the processes display some interesting features, such as invertibility. For all this, we must in particular take both N and T large from the start, with their ratio $r \equiv N/T$ fixed (1.13). More precisely, we stretch the range of the a–index from minus to plus infinity.

This means that all the finite–size effects (appearing at the ends of the time series) are readily disregarded. In particular, there is no need to care about initial conditions for the processes, and all the recurrence relations are assumed to continue to the infinite past.

4.2.1 The VMA(q) Process

The Definition of VMA(q)

We consider a situation when N stochastic variables evolve according to identical independent VMA(q) (vector moving average) processes, which we sample over a time span of T moments. This is a simple generalization of the standard univariate weak–stationary moving average MA(q). In such a setting, the value Y_{ia} of the i–th ($i = 1, \ldots, N$) random variable at time moment a ($a = 1, \ldots, T$) can be expressed as

$$Y_{ia} = \sum_{\alpha=0}^{q} a_\alpha \epsilon_{i,a-\alpha}. \tag{4.11}$$

Here all the ϵ_{ia}'s are independent identically distributed standard (mean zero, variance one) Gaussian random numbers (white noise), $\langle \epsilon_{ia}\epsilon_{jb}\rangle = \delta_{ij}\delta_{ab}$. The a_α's are some $(q+1)$ real constants; importantly, they do not depend on the index i, which reflects the fact that the processes are identical and independent (no "spatial" covariances among the variables). The rank q of the process is a positive integer.

The Auto–Covariance Matrix

In order to handle such a process (4.11), notice that the Y_{ia}'s, being linear combinations of uncorrelated Gaussian numbers, must also be Gaussian random variables, albeit displaying some correlations. Therefore, to fully characterize these variables, it is sufficient to calculate their two–point covariance function; this is straightforwardly done (see appendix B.1 for details),

$$\langle Y_{ia} Y_{jb}\rangle = \delta_{ij} A^{(1)}_{ab}, \tag{4.12}$$

where

$$\begin{array}{c} A^{(1)}_{ab} = \kappa^{(1)}_0 \delta_{ab} + \sum_{d=1}^{q} \kappa^{(1)}_d \left(\delta_{a,b-d} + \delta_{a,b+d}\right), \\ \text{with} \quad \kappa^{(1)}_d \equiv \sum_{\alpha=0}^{q-d} a_\alpha a_{\alpha+d}, \quad d = 0, 1, \ldots, q. \end{array} \tag{4.13}$$

In other words, the cross–covariance matrix is trivial, $\mathbf{C} = \mathbf{1}_N$ (no correlations between different variables), while the auto–covariance matrix $\mathbf{A}^{(1)}$, responsible for temporal correlations, can be called "$(2q+1)$–diagonal." In the course of this article, we will use several different auto–covariance matrices, and for brevity, we decide to label them with superscripts; their definitions are all collected in appendix B.1.1. For example, in the simplest case of VMA(1), it is tri–diagonal,

$$A^{(1)}_{ab} = \left(a_0^2 + a_1^2\right)\delta_{ab} + a_0 a_1 \left(\delta_{a,b-1} + \delta_{a,b+1}\right). \tag{4.14}$$

The Fourier Transform and the M–Transform of the Auto–Covariance Matrix

Such an infinite matrix (4.13) is translationally invariant (as announced, it is one of the implications of the weak stationarity), *i.e.*, the value of any of its entries depends only on the

distance between its indices, $A^{(1)}_{ab} = A^{(1)}(a-b)$; specifically, $A^{(1)}(\pm d) = \kappa^{(1)}_d$, for $d = 0, 1, \ldots, q$, and $A^{(1)}(|d| > q) = 0$. Hence, it is convenient to rewrite this matrix in the Fourier space,

$$\widehat{A^{(1)}}(p) \equiv \sum_{d\in\mathbf{Z}} \mathrm{e}^{\mathrm{i}dp} A^{(1)}(d) = \kappa^{(1)}_0 + 2\sum_{d=1}^{q} \kappa^{(1)}_d \cos(dp). \tag{4.15}$$

In this representation, the M–transform of $\mathbf{A}^{(1)}$ is readily obtained [BJJ+09],

$$M_{\mathbf{A}^{(1)}}(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}p \frac{\widehat{A^{(1)}}(p)}{z - \widehat{A^{(1)}}(p)}. \tag{4.16}$$

This integral can be evaluated by the method of residues for any value of q, which we do in appendix B.1.2, where also we print the general result (B.3). In particular, for $q = 1$,

$$M_{\mathbf{A}^{(1)}}(z) = \frac{z}{\sqrt{z-(a_0+a_1)^2}\sqrt{z-(a_0-a_1)^2}} - 1, \tag{4.17}$$

where the square roots are principal.

The Pearson Estimator of the Covariances from Free Random Variables

We will be interested in investigating the spectral properties of the Pearson estimator(4.4) $\mathbf{c} = (1/T)\mathbf{Y}\mathbf{Y}^{\mathrm{T}} = (1/T)\widetilde{\mathbf{Y}}\mathbf{A}^{(1)}\widetilde{\mathbf{Y}}^{\mathrm{T}}$. The M–transform of this correlated Wishart random matrix, $M \equiv M_{\mathbf{c}}(z)$, can be retrieved from equation (4.10). We could write it for any q using (B.3), but we will restrict ourselves only to $q = 1$, in which case the substitution of (4.17) leads to a fourth–order polynomial (Ferrari) equation for the unknown M,

$$\begin{aligned}
&r^4\left(a_0^2-a_1^2\right)^2 M^4 + 2r^3\left(-\left(a_0^2+a_1^2\right)z + \left(a_0^2-a_1^2\right)^2(r+1)\right)M^3 + \\
&+r^2\left(z^2 - 2\left(a_0^2+a_1^2\right)(r+2)z + \left(a_0^2-a_1^2\right)^2\left(r^2+4r+1\right)\right)M^2 + \\
&+2r\left(z^2 - \left(a_0^2+a_1^2\right)(2r+1)z + \left(a_0^2-a_1^2\right)^2 r(r+1)\right)M + \\
&+r\left(-2\left(a_0^2+a_1^2\right)z + \left(a_0^2-a_1^2\right)^2 r\right) = 0.
\end{aligned} \tag{4.18}$$

The FRV technique allowed us therefore to find this equation in a matter of a few lines of a simple algebraic computation. It has already been derived in [JWMH09], and (4.18) may be verified to coincide with the version given in that paper. In [JWMH09], the pertinent equation is printed before (A.6), and to compare the two, one needs to change their variables into ours according to $y \to 1/r$, $x \to z/r$, and $\underline{m} \to -r(1+M)/z$. The last equality means that $\underline{m}$ and m of [JWMH09] correspond in our language to the Green's functions $-rG_{\mathbf{c}}(z)$ and $-G_{\mathbf{a}}(z/r)$, respectively, where $\mathbf{a} = (1/N)\mathbf{Y}^{\mathrm{T}}\mathbf{Y}$ is the Pearson estimator dual to $\mathbf{c}$. As mentioned, a quick extension to the case of arbitrary q is possible, however the resulting equations for M will be significantly more complicated; for instance, for $q = 2$, a lengthy ninth–order polynomial equation is discovered.

4.2.2 The VAR(q) Process

The Definition of VAR(q)

A set–up of N identical and independent VAR(q) (vector auto–regressive) processes is somewhat akin to (4.11), *i.e.*, we consider N decoupled copies of a standard univariate AR(q) process,

$$Y_{ia} - \sum_{\beta=1}^{q} b_\beta Y_{i,a-\beta} = a_0 \epsilon_{ia}. \tag{4.19}$$

It is again described by the demeaned and standardized Gaussian white noise ϵ_{ia} (which triggers the stochastic evolution), as well as $(q+1)$ real constants a_0, b_β, with $\beta = 1, \ldots, q$. As announced before, the time stretches to the past infinity, so no initial condition is necessary. Although at first sight (4.19) may appear to be a more involved recurrence relation for the Y_{ia}'s, it is actually easily reduced to the VMA(q) case: It remains to remark that if one exchanges the Y_{ia}'s with the ϵ_{ia}'s, one precisely arrives at the VMA(q) process with the constants $a_0^{(2)} \equiv 1/a_0$, $a_\beta^{(2)} \equiv -b_\beta/a_0$, $\beta = 1, \ldots, q$. In other words, the auto–covariance matrix $\mathbf{A}^{(3)}$ of the VAR(q) process (4.19) is simply the inverse of the auto–covariance matrix $\mathbf{A}^{(2)}$ of the corresponding VMA(q) process with the described modification of the parameters,

$$\mathbf{A}^{(3)} = \left(\mathbf{A}^{(2)}\right)^{-1}. \tag{4.20}$$

This inverse exists thanks to the weak stationarity supposition.

The Fourier Transform and the M–Transform of the Auto–Covariance Matrix

The Fourier transform of the auto–covariance matrix $\mathbf{A}^{(3)}$ of VAR(q) is therefore a (number) inverse of its counterpart for VMA(q) with its parameters appropriately changed,

$$\widehat{A^{(3)}}(p) = \frac{1}{\widehat{A^{(2)}}(p)} = \frac{1}{\kappa_0^{(2)} + 2\sum_{d=1}^{q} \kappa_d^{(2)} \cos(dp)}, \tag{4.21}$$

where

$$\kappa_d^{(2)} = \frac{1}{a_0^2} \sum_{\alpha=0}^{q-d} b_\alpha b_{\alpha+d}, \qquad d = 0, 1, \ldots, q, \tag{4.22}$$

and where we define $b_0 \equiv -1$. In order to find the M–transform of the inverse matrix, $\mathbf{A}^{(3)} = (\mathbf{A}^{(2)})^{-1}$, one employs a general result, true for any (real symmetric) random matrix $\mathbf{H}$, and obtainable through an easy algebra,

$$M_{\mathbf{H}^{-1}}(z) = -M_{\mathbf{H}}(1/z) - 1. \tag{4.23}$$

Since the quantity $M_{\mathbf{A}^{(2)}}(z)$ is known for any q (B.3), hence is $M_{\mathbf{A}^{(3)}}(z)$ via (4.23), but we will not print it explicitly. Let us just give it for $q = 1$, in which case (4.23) and (4.17) yield

$$M_{\mathbf{A}^{(3)}}(z) = -\frac{1}{\sqrt{1 - \frac{(1-b_1)^2}{a_0^2} z}\sqrt{1 - \frac{(1+b_1)^2}{a_0^2} z}}. \tag{4.24}$$

The Auto–Covariance Matrix

Despite being somewhat outside of the main line of thought of this article, an interesting question would be to search for an explicit expression for the auto–covariance matrix $\mathbf{A}^{(3)}$ from its Fourier transform (4.21),

$$A^{(3)}(d) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}p \mathrm{e}^{-\mathrm{i}dp} \frac{1}{\kappa_0^{(2)} + 2\sum_{l=1}^{q}\kappa_l^{(2)}\cos(lp)}, \tag{4.25}$$

where we exploited the fact that $\mathbf{A}^{(3)}$ must be translationally invariant, $A^{(3)}_{ab} = A^{(3)}(a-b)$. This computation would shed light on the structure of temporal correlations present in a VAR setting. This integral is evaluated by the method of residues in a very similar manner to the one shown in appendix B.1.2, and we do this in appendix B.2. We discover that the auto–covariance matrix is a sum of q exponential decays,

$$A^{(3)}(d) = \sum_{\gamma=1}^{q} C_\gamma \mathrm{e}^{-|d|/T_\gamma}, \tag{4.26}$$

where C_γ are constants, and T_γ are the characteristic times (B.5), $\gamma = 1, \ldots, q$; these constituents are given explicitly in (B.4). This is a well–known fact, nevertheless we wanted to establish it again within our approach. For example, for $q = 1$, the auto–covariance matrix of VAR(1) is one exponential decay,

$$A^{(3)}(d) = \frac{a_0^2}{1-b_1^2} b_1^{|d|}, \tag{4.27}$$

where we assumed for simplicity $0 < b_1 < 1$ (the formula can be easily extended to all values of b_1).

The Pearson Estimator of the Covariances from Free Random Variables

Having found an expression for the M–transform of the auto–covariance matrix $\mathbf{A}^{(3)}$ of a VAR(q) (4.23), (B.3), we may proceed to investigate the equation (4.10) for the M–transform $M \equiv M_{\mathbf{c}}(z)$ of the correlated Wishart random matrix $\mathbf{c} = (1/T)\mathbf{Y}\mathbf{Y}^{\mathrm{T}} = (1/T)\widetilde{\mathbf{Y}}\mathbf{A}^{(3)}\widetilde{\mathbf{Y}}^{\mathrm{T}}$ (4.4). We will do this explicitly only for $q = 1$, when (4.24) leads to a fourth–order (Ferrari) polynomial equation for the unknown M,

$$\begin{aligned} &a_0^4 r^2 M^4 + 2a_0^2 r\left(-\left(1+b_1^2\right)z + a_0^2 r\right)M^3 + \\ &+\left(\left(1-b_1^2\right)^2 z^2 - 2a_0^2 r\left(1+b_1^2\right)z + \left(r^2-1\right)a_0^4\right)M^2 - 2a_0^4 M - a_0^4 = 0. \end{aligned} \tag{4.28}$$

This equation has been derived by another method in [JWMH09], and our result confirms their equation (A.8), with the change in notation, $y \to 1/r$, $x \to z/r$, $z \to rM$.

4.2.3 The VARMA(q_1, q_2) Process

The Definition of VARMA(q_1, q_2)

The two types of processes which we elaborated on above, VAR(q_1) and VMA(q_2), can be combined into one stochastic process called VARMA(q_1, q_2),

$$Y_{ia} - \sum_{\beta=1}^{q_1} b_\beta Y_{i,a-\beta} = \sum_{\alpha=0}^{q_2} a_\alpha \epsilon_{i,a-\alpha}. \tag{4.29}$$

Now it is a straightforward and well–known observation (which can be verified by a direct calculation) that the auto–covariance matrix $\mathbf{A}^{(5)}$ of this process is simply the product (in any order) of the auto–covariance matrices of the VAR and VMA pieces; more precisely,

$$\mathbf{A}^{(5)} = \left(\mathbf{A}^{(4)}\right)^{-1} \mathbf{A}^{(1)}, \tag{4.30}$$

where $\mathbf{A}^{(1)}$ corresponds to the generic VMA(q_2) model (4.13), while $\mathbf{A}^{(4)}$ denotes the auto–covariance matrix of VMA(q_1) with a slightly different modification of the parameters compared to the previously used, namely $a_0^{(4)} \equiv 1$, $a_\beta^{(4)} \equiv -b_\beta$, for $\beta = 1, \ldots, q_1$. We have thus already made use here of the fact that the auto–covariance matrix of a VAR process is the inverse of the auto–covariance matrix of a certain corresponding VMA process (4.20), but the new change in parameters necessary in moving from VAR to VMA has effectively $a_0 = 1$ w.r.t. what we had before (4.20); it is understandable: this "missing" a_0 is now included in the matrix of the other VMA(q_2) process.

The Fourier Transform and the M–Transform of the Auto–Covariance Matrix

The Fourier transform of the auto–covariance matrix $\mathbf{A}^{(5)}$ of VARMA(q_1, q_2) (4.30) is simply the product of the respective Fourier transforms (4.15) and (4.21),

$$\widehat{A^{(5)}}(p) = \frac{\kappa_0^{(1)} + 2\sum_{d_2=1}^{q_2} \kappa_{d_2}^{(1)} \cos(d_2 p)}{\kappa_0^{(4)} + 2\sum_{d_1=1}^{q_1} \kappa_{d_1}^{(4)} \cos(d_1 p)}, \tag{4.31}$$

where

$$\kappa_{d_1}^{(4)} = \sum_{\alpha_1=0}^{q_1-d_1} b_{\alpha_1} b_{\alpha_1+d_1}, \quad \kappa_{d_2}^{(1)} = \sum_{\alpha_2=0}^{q_2-d_2} a_{\alpha_2} a_{\alpha_2+d_2},$$
$$d_1 = 0, 1, \ldots, q_1, \quad d_2 = 0, 1, \ldots, q_2, \tag{4.32}$$

where we recall $b_0 = -1$. For instance, for VARMA(1, 1) (it is described by three constants, a_0, a_1, b_1), one explicitly has

$$\widehat{A^{(5)}}(p) = \frac{a_0^2 + a_1^2 + 2a_0 a_1 \cos p}{1 + b_1^2 - 2b_1 \cos p}. \tag{4.33}$$

The M–transform of $\mathbf{A}^{(5)}$ can consequently be derived from the general formula (4.16). We will evaluate here the pertinent integral only for the simplest VARMA(1, 1) process, even though an arbitrary case may be handled by the technique of residues,

$$\begin{aligned} M_{\mathbf{A}^{(5)}}(z) &= \frac{1}{a_0 a_1 + b_1 z}\left(-a_0 a_1 + \right. \\ &+ \left. \frac{z\left(a_0 a_1 + \left(a_0^2 + a_1^2\right) b_1 + a_0 a_1 b_1^2\right)}{\sqrt{(1-b_1)^2 z - (a_0+a_1)^2}\sqrt{(1+b_1)^2 z - (a_0-a_1)^2}}\right). \end{aligned} \tag{4.34}$$

The Auto–Covariance Matrix

One might again attempt to track the structure of temporal covariances in a VARMA process. This can be done either by the inverse Fourier transform of (4.31), or through a direct computation based on the recurrence relation (4.29) (importantly, adhering to the assumption that it stretches to the past infinity). Let us print the result just for VARMA(1, 1),

$$A^{(5)}(d) = -\frac{a_0 a_1}{b_1}\delta_{d,0} + \frac{(a_1 + a_0 b_1)(a_0 + a_1 b_1)}{b_1\left(1 - b_1^2\right)} b_1^{|d|}, \tag{4.35}$$

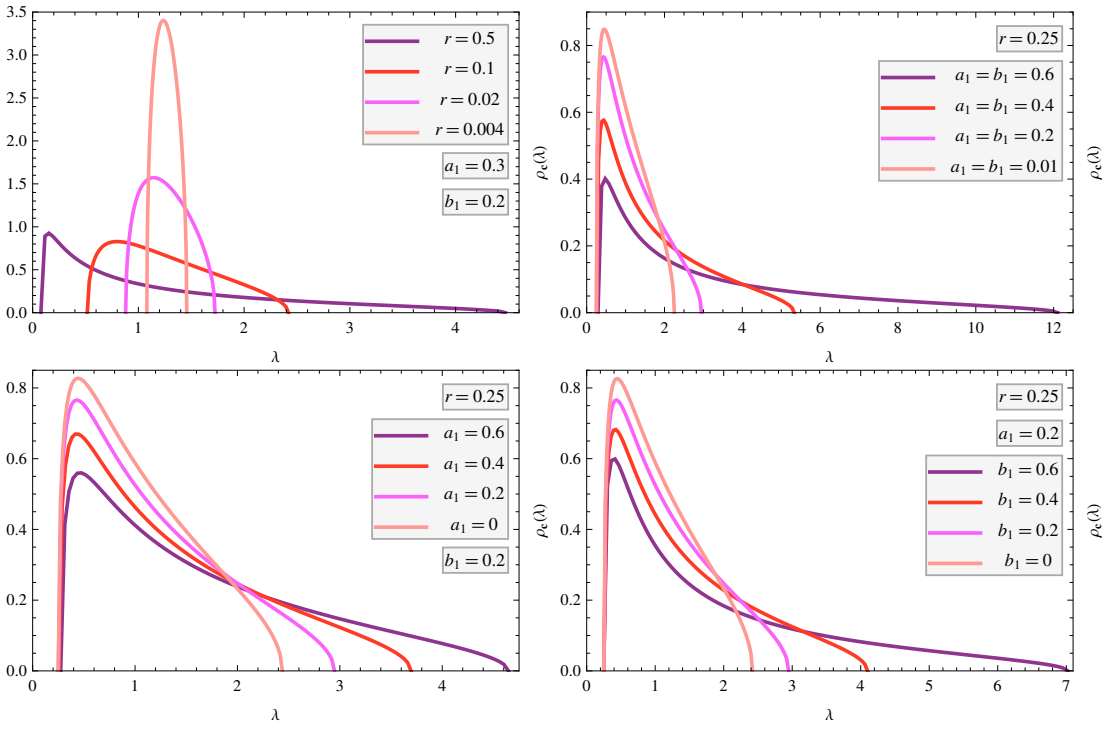

Figure 4.1: The mean spectral density $\rho_{\mathbf{c}}(\lambda)$ of the Pearson estimator $\mathbf{c}$ of the cross–covariances in the VARMA$(1,1)$ process computed numerically from the sixth–order polynomial equation (B.6), for various values of the process' parameters. The scale of these parameters is determined by choosing $a_0 = 1$ everywhere. Recall that the theoretical formula (B.6) is valid in the thermodynamical limit (1.13) of $N, T \to \infty$, with $r = N/T$ kept finite.
UP LEFT: We set the remaining VARMA parameters to $a_1 = 0.3$, $b_1 = 0.2$, while the rectangularity ratio takes the values $r = 0.5$ (the purple line), 0.1 (red), 0.02 (magenta), 0.004 (pink); each one is 5 times smaller than the preceding one. We observe how the graphs become increasingly peaked (narrower and taller) around $\lambda = 1$ as r decreases, which reflects the movement of the estimator $\mathbf{c}$ toward its underlying value $\mathbf{C} = \mathbf{1}_N$.
UP RIGHT: We fix $r = 0.25$ and consider the two VARMA parameters equal to each other, with the values $a_1 = b_1 = 0.6$ (purple), 0.4 (red), 0.2 (magenta), 0.01 (pink).
DOWN LEFT: We hold $r = 0.25$ and $b_1 = 0.2$, and modify $a_1 = 0.6$ (purple), 0.4 (red), 0.2 (magenta), 0.0 (pink); for this last value, the VARMA$(1,1)$ model reduces to VAR(1).
DOWN RIGHT: Similarly, but this time we assign $r = 0.25$ and $a_1 = 0.2$, while changing $b_1 = 0.6$ (purple), 0.4 (red), 0.2 (magenta), 0.0 (pink); this last value corresponds to VMA(1).

where for simplicity $0 < b_1 < 1$. This is an exponential decay, with the characteristic time of the VAR piece, with an additional term on the diagonal.

The Pearson Estimator of the Covariances from Free Random Variables

Expression (4.34), along with the fundamental FRV formula (4.10), allow us to write the equation satisfied by the M–transform $M \equiv M_{\mathbf{c}}(z)$ of the Pearson estimator $\mathbf{c} = (1/T)\mathbf{Y}\mathbf{Y}^{\mathrm{T}} = (1/T)\widetilde{\mathbf{Y}}\mathbf{A}^{(5)}\widetilde{\mathbf{Y}}^{\mathrm{T}}$ (4.4) of the cross–covariances in the VARMA$(1,1)$ process; it happens to be polynomial of order six, and we print it (B.6) in appendix B.2.1. It may be solved numerically, a proper solution chosen (the one which leads to a sensible density: real, positive–definite, normalized to unity), and finally, the mean spectral density $\rho_{\mathbf{c}}(\lambda)$ derived from (2.3). We show the shapes of this density for a variety of the values of the parameters r, a_0, a_1, b_1 in fig. 4.1. Moreover, in order to test the result (B.6), and more broadly, to further establish our FRV framework

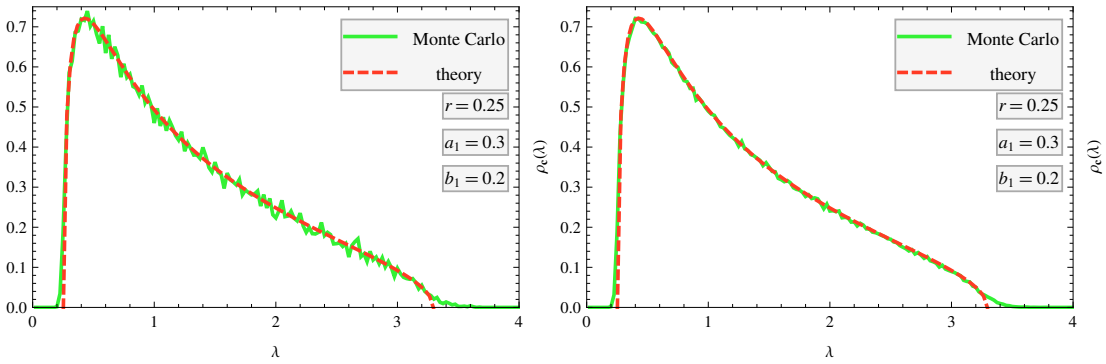

Figure 4.2: Monte Carlo simulations of the mean spectral density $\rho_{\mathbf{c}}(\lambda)$ (the green plots) compared to the theoretical result obtained numerically from the sixth–order equation (B.6) (the dashed red lines). The conformity is nearly perfect. We generate the matrices $\mathbf{Y}$ of sizes $N = 50$, $T = 200$ (*i.e.*, $r = 0.25$) from the VARMA$(1,1)$ process with the parameters $a_0 = 1$, $a_1 = 0.3$, $b_1 = 0.2$. The Monte Carlo simulation is repeated $1,000$ (LEFT) or $10,000$ (RIGHT) times; in this latter case, a significant improvement in the quality of the agreement is seen. One notices finite–size effects at the edges of the spectrum ("leaking out" of eigenvalues): in the numerical simulations, N and T are obviously finite, while equation (B.6) is legitimate in the thermodynamical limit (1.13) only, hence the small discrepancies; by enlarging the chosen dimensions 50×200 one would diminish this fallout.

in the guise of formula (4.10), the theoretical form of the density is compared to Monte Carlo simulations in fig. 4.2; they remain in excellent concord. These are the main findings of this article.

4.2.4 Example from Macroeconomic Data

We apply the method described above to Polish macroeconomic data. The motivation behind is twofold. First of all economic theory rarely has any sharp implications about the short-run dynamics of economic variables (so called scarcity of economic time series). Secondly in these very rare situations, where theoretical models include a dynamic adjustment equation, one has to work hard to exclude the moving average terms from appearing in the implied dynamics of the variables of interest.

An Application to Macroeconomic Data

Let us pursue further the above analysis of the VARMA$(1,1)$ model on a concrete example of real data. Economists naturally think of comovement in economic time series as arising largely from relatively few key economic factors like productivity, monetary policy and so forth. Classical way of representing this notion is in terms of statistical factor model, for which we allow the limited spatio–temporal dependence expressed via our VARMA$(1,1)$ model. Two empirical questions are addressed in this section:

1. How the empirical data behave? Does the eigenvalues represent similar structure to financial market data? In other words, does the macroeconomic data represent collective response to the shock expressed in term of VARMA$(1,1)$ model or are there any apparent outliers.

2. Should then the forecasts be constructed using small factor models or are there any non-

zero, but perhaps small coefficients. If so, then the large scale model framework is appropriate.

We investigate $N = 52$ various macroeconomic time series for Poland of length $T = 118$. They have been selected on a monthly basis in such a manner so as to cover most of the main sectors of the Polish economy, *i.e.*, the money market, domestic and foreign trade, labor market, balance of payments, inflation in different sectors, *etc.* (their full list is attached in the appendix C.1). The time series were taken directly from the Reuters©3000Xtra database. Although longer time series for Poland are accessible, we restrict ourselves to the last ten years in order to avoid the effects of structural change. We assume that each economic variable is affected by the same shock (*i.e.*, the "global market shock") of an ARMA$(1, 1)$ type with unknown parameters, which we are to estimate; the AR part implies that the shock dies quite quickly, while the MA part is responsible for the persistency of the shock. To preserve the proper VARMA representation, the original time series were transformed using one of the following methods:

- First, many of the series are seasonally adjusted by the reporting agency.
- Second, the data were transformed to eliminate trends and obvious nonstationarities. For real variables, this typically involved transformation to growth rates (the first difference of logarithms), and for prices this involved transformation to changes in growth rates (the second difference of logarithms).
- Interest rates were transformed to first differences.
- Finally, some of the series contained a few large outliers associated with events like labor disputes, other extreme events, or with data problems of various sorts. These outliers were identified as observations that differed from the sample median by more than 6 times the sample interquartile range, and these observations were dropped from the analysis.

In fig. 4.3 we plot these time series (LEFT), and we make (RIGHT) a histogram of the mean spectral density $\rho_{\mathbf{c}}(\lambda)$, which we compare to a theoretical prediction from our FRV equation (B.6) with the estimated values of the parameters a_0, a_1, b_1. We have also plotted the standard Marčenko-Pastur (Bai-Silverstein) [SB95, MP67] as of eq.(2.50)

Discussion

The result is important for forecast design, but more importantly, it provides information about the way macroeconomic variable interact. The empirical data as compared to the spectral density given by eq.(2.50) suggest that a lot of eigenvalues, similarly to stock market data express marginal predictive content. One can suppose, that each of the economic time series contains important information about the collective movements, that cannot be gleaned from other time series. Alternatively, if we suppose that macro variables interact in the simplest low-dimensional way suggested by VARMA$(1, 1)$ model,the conformity is nearly ideal (modulo the finite–size effects at the right edge of the spectrum). The economic time series express common response to the "global shock" process *i.e.*, each eigenvalue now contains useful information about the values of the factors, that affect comovement and hence useful information about the future behavior of economy. Thus, while many more eigenvalues appear to be useful, the predictive component is apparently common to many series in a way suggested by our simplified VARMA$(1, 1)$ model.

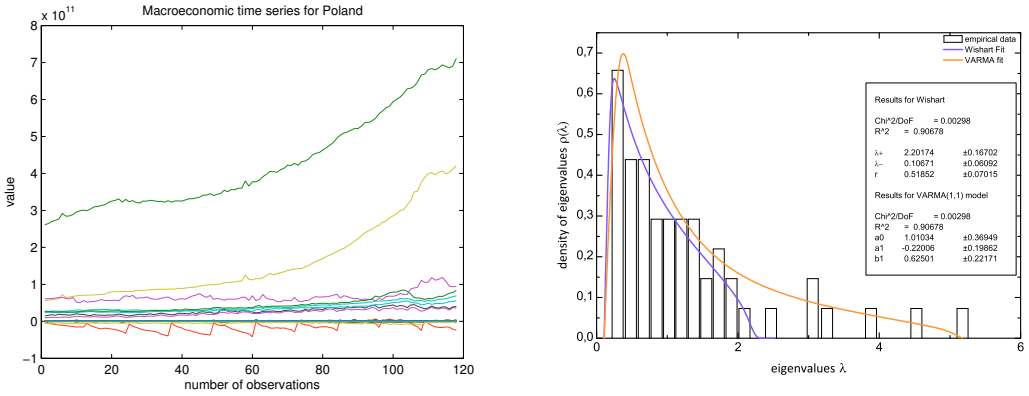

Figure 4.3: LEFT: The original $N = 52$ time series of length $T = 118$; they are non–stationary, with the seasonal components.
RIGHT: The histogram of the mean spectral density $\rho_{\mathbf{c}}(\lambda)$ (the solid black line) compared to the theoretical result obtained by numerically solving the sixth–order equation (B.6) (the solid orange line) and the "Wishart-fit" (purple line).

4.3 Conclusions

In this chapter we have again shown, how powerful and flexible the Free Random Variables calculus can be.

- The FRV calculus is ideally suited for multidimensional time series problems (e.g. multivariate stochastic processes of the VARMA type), provided the dimensions of the underlying matrices are large. The operational procedures are simple, algebraic and transparent.

- The structure of the final formula which relates the moments' generating function of the population covariance and the sample covariance allows one to easily derive eigenvalue density of the sample covariance.

- We in detail illustrated how this procedure works for VARMA$(1, 1)$, confronted the theoretical prediction with numerical data obtained by Monte Carlo simulations of the VARMA process and observed a perfect agreement.

- We have also pursued the above analysis on a real - complex systems example - i.e. Economy of Poland, for which we have assumed, that each of the time series under study is generated by the same type of univariate VARMA(q_1, q_2) process. A stunning fact, is that again the flawless correspondence between theoretical spectral density and empirical data is found.

- We are in the position, where we cannot reject the hypothesis, there are indeed no autocorrelations among macroeconomic time series. One may also argue, that all these time series are closely bounded with the process, which we will identify as "global shock-process" *i.e.*, all time series represent the global response complex system under study, to a distortion and its adjustment to equilibrium state. This process is of univariate VARMA(q_1, q_2) *i.e.*, $ARMA(1, 1)$ type with hidden structure, that has to be revealed based on historical time series data.

- This crude empirical study allows potentially for variety of extensions. At least two are possible. First, the approach used here identifies the underlying factors only up to a linear transformation, making economic interpretation of the factors themselves difficult. It would be interesting to be able to relate the factors more directly to fundamental economic forces in the spirit of DSGE models. Secondly, our theoretical result covers only stationary models, but say nothing about integrated, cointegrated and cotrending variables. We know that common long-run factors are important for describing macroeconomic data, and theory needs to be developed to handle these features in a large model framework.

5

Temporal Correlations from Singular Value Spectra

In order to investigate the temporal properties of the correlations between two data sets one is often interested not only in the analysis of it's static properties given by Pearson estimator (1.6), $\mathbf{C}_X = \frac{1}{T}\mathbf{X}\mathbf{X}^{\mathbf{T}}$ but more likely how this features behave over a certain period of time. We have already seen in previous chapters, that the primary way to describe cross–correlations in a Gaussian framework is through the two–point covariance function,

$$\mathbf{C}_{ia,jb} \equiv \langle X_{ia} X_{jb} \rangle \,. \tag{5.1}$$

Where $X_{ia} \equiv x_{ia} - \langle x_{ia} \rangle$ are mean adjusted data, that can be further collected into a rectangular $N \times T$ matrix $\mathbf{R}$. The average $\langle \ldots \rangle$ is understood as taken according to some probability distribution whose functional shape is stable over time, but whose parameters may be time–dependent. So far (see. chapter 3 and chapter 4) we have used a very simplified form of the two–point covariance function(5.1), namely with cross–covariances and auto–covariances factorized and non–random,

$$\mathbf{C}_{ia,jb} = C_{ij} A_{ab} \tag{5.2}$$

(we have assembled coefficients into an $N \times N$ cross–covariance matrix $\mathbf{C}$ and a $T \times T$ auto–covariance matrix $\mathbf{A}$; both are taken symmetric and positive–definite). The matrix of "temporal structure" $\mathbf{A}$ is a way to model two temporal effects: the (weak, short–memory) lagged correlations between the returns (see chapter 4), as well as the (stronger, long–memory) lagged correlations between the volatilities (weighting schemes, eg.EWMA [PPK04]; see chapter 3 for details). On the other hand, the matrix of spatial correlations $\mathbf{C}$ models the hidden factors affecting the variables, thereby reflecting the structure of mutual dependencies of the complex system(see chapter 3). The salient feature assumed so far, these two matrices were decoupled and the assumption about the Gaussianity of random variables provides crude approximation, that variances of all random variables always exist. This was sufficient to fully characterize the dependencies of the X_{ia}'s. However, in more realistic circumstances (*i.e.*, building efficient multivariate models,which help understanding the relation between a large number of possible causes and resulting effects) one is more interested in the situations, where the spatio–temporal structure does not factorize. Cross-correlations technique (sometimes alluded as "time–lagged correlations technique") is most likely meets these critical requirements.

$$\mathcal{C}_{ia,ja+\Delta}(\Delta) = \frac{1}{T} \sum_{a=1}^{T} X_{ia} X_{ja+\Delta} \tag{5.3}$$

The precise answer boils down to how to separate the spectrum of such a covariance matrix in the large N, large T limit (*i.e.*, thermodynamical limit), when one can make use of the power of FRV calculus (see [TB07] for a solution based on the circular symmetry of the problem and Gaussian approximation). In this chapter we will very closely follow the method presented in [BLMP07], where the authors suggested to compare the singular value spectrum of the empirical rectangular $M \times N$ correlation matrix with a benchmark obtained using Random Matrix Theory results (c.f. [ERR05]), assuming there are no correlation between the variables. For $T \to \infty$ at N, M fixed, all singular values should be zero, but this will not be true if T is finite. The singular value spectrum of this benchmark problem can in fact be computed exactly in the limit where $N, M, T \to \infty$, when the ratios $m = M/T$ and $n = N/T$ fixed. Since the original description is veiled, for pedagogical purposes we rederive all these results in the language of FRV presented in chapter 2. Furthermore we extend the results obtained in [Sna08] to meet encounter time-lagged correlations.

5.1 Mathematical Formulation of a Problem

Due to the works [Gra01, BEW88, Sim80] it is believed that, the system itself should determine the number of relevant input and output factors. In the simplest approach one would take all the possible input and output factors and systematically correlate them, hoping to unravel the hidden structure. This procedure swiftly blow up with just few variables (see sec. 1.2.1). The cross - equation correlation matrix contains all the information about contemporaneous correlation in a Vector model and may be its greatest strength and its greatest asset. Since no questionable a priori assumptions are imposed, fitting a Vector model allows data–set to speak for itself *i.e.*, find the relevant number of factors. Still without imposing any restrictions on the structure of the correlation matrix one cannot make a causal interpretation of the results. The theoretical study of high dimensional factor models is indeed actively pursued in literature [Gew97, SW05, SW02b, SW02a, FHLR00, FHLR04, Bai03, BN02]. The main aim of this chapter is to present a method, which helps extract highly non-trivial spatio–temporal correlations between two samples of non-equal size (i.e. input and output variables of large dimensionality), for these can be then treated as "natural" restrictions for the correlations matrix structure.

5.1.1 Basic framework and notation

We will divide all variables into two subsets *i.e.*, focus on N input factors X_a $(a = 1, \ldots, N)$ and M output factors Y_α $(\alpha = 1, \ldots, M)$ with the total number of observations being T. All time series are standardized to have zero mean and unit variance. The data can be completely different or be the same variables but observed at different times. First one has to remove potential correlations inside each subset, otherwise it may interfere with the out-of-sample signal. To remove the correlations inside each sample we form two correlation matrices,which contain information about in-the-sample correlations.

$$\mathbf{C_X} = \frac{1}{T} X X^T \qquad \mathbf{C_Y} = \frac{1}{T} Y Y^T \tag{5.4}$$

The matrices are then diagonalized,provided $T > N, M$, and the empirical spectrum is compared to the theoretical Marčenko-Pastur spectrum [MP67, LCBP99, BJ04, BGJJ04]in order to unravel

statistically significant factors.[1] Having found all eigenvectors and eigenvalues, one can then construct a set of uncorrelated unit variance input variables $\hat{X}$ and output variables $\hat{Y}$.

$$\hat{X}_{at} = \frac{1}{\sqrt{T\lambda_a}} V^T X_t \qquad \hat{Y}_{\alpha t} = \frac{1}{\sqrt{T\lambda_\alpha}} U^T Y_t \tag{5.5}$$

where V, U, λ_a, λ_α are the corresponding eigenvectors and eigenvalues of C_X , C_Y respectively. It is obvious, that $C_{\hat{X}} = \hat{X}\hat{X}^T$ and $C_{\hat{Y}} = \hat{Y}\hat{Y}^T$ are identity matrices, of dimension, respectively, N and M. Using general property of diagonalization, this means that the $T \times T$ matrices $D_{\hat{X}} = \hat{X}^T\hat{X}$ and $D_{\hat{Y}} = \hat{Y}^T\hat{Y}$ have exactly N (resp. M) eigenvalues equal to 1 and $T - N$ (resp. $T - M$) equal to zero. These non-zero eigenvalues are randomly arranged on a diagonal. Finally we can reproduce the asymmetric $M \times N$ cross-correlation matrix G between the $\hat{Y}$ and $\hat{X}$:

$$G = \hat{Y}\hat{X}^T \tag{5.6}$$

which includes only the correlations between input and output factors. In general the spectrum of such a matrix is complex, but we will use the singular value decomposition (SVD) technique (c.f. [FIS02]) to find the empirical spectrum of eigenvalues.

The Singular Value Decomposition

The singular value spectrum represent the strength of cross-correlations between input and output factors. Suppose G is an $M \times N$ matrix whose entries are either real or complex numbers. Then there exists a factorization of the form

$$G = U\Sigma V^\dagger \tag{5.7}$$

where U is an $M \times M$ unitary matrix. The columns of U form a set of orthonormal "output" basis vector directions for G - these are the eigenvectors of $G^\dagger G$. Σ is $M \times N$ diagonal matrix with nonnegative real numbers on the diagonal,which can be thought of as scalar "gain controls" by which each corresponding input is multiplied to give a corresponding output. These are the square roots of the eigenvalues of $GG^\dagger$ and $G^\dagger G$ that correspond with the same columns in U and V. and $V^\dagger$ denotes the conjugate transpose of V, an $N \times N$ unitary matrix,whose columns form a set of orthonormal "input" or vector directions for G. These are the eigenvectors of $GG^\dagger$. A common convention for the SVD decomposition is to order the diagonal entries $\Sigma_{i,i}$ in descending order. In this case, the diagonal matrix Σ is uniquely determined by G (though the matrices U and V are not). The diagonal entries of Σ are known as the singular values of G.

5.2 Singular values from free random matrix theory

In order to evaluate these singular eigenvalues, assume without loss of generality $M < N$. The trick is to consider the matrix $M \times M$ matrix GG^T (or the $N \times N$ matrix G^TG if $M > N$), which is symmetrical and has M positive eigenvalues, each of which being equal to the square of a singular value of G itself. Furthermore use the cyclic properties of the trace. Then non-zero eigenvalues of

$$GG^T = \hat{Y}\hat{X}^T\hat{X}\hat{Y}^T$$

[1] The eigenvalues,which lie much below the lower edge of the Marčenko-Pastur spectrum represent the redundant factors, rejected by the system, so one can exclude them from further study and in this manner reduce somewhat the dimensionality of the problem, by removing possibly spurious correlations.

are then the same (up to the zero modes) as those of the $T \times T$ matrix

$$\mathbf{D} = D_{\hat{X}} D_{\hat{Y}} = \hat{X}^T \hat{X} \hat{Y}^T \hat{Y}$$

obtained by swapping the position of $\hat{Y}$ from first to last. In the limit $N, M, T \to \infty$ where the $\hat{X}$'s and the $\hat{Y}$'s are independent from each other, the two matrices $D_{\hat{X}}$ and $D_{\hat{Y}}$ are mutually free [Voi91], and we can use the results from FRV, where given the spectral density of each individual matrix, one is able to construct the spectrum of the product or sum of them.

5.2.1 FRV Algorithm for Cross-correlation matrix

As usual we will start with constructing the Green's function for matrices $\mathbf{D}_{\hat{X}}$ and $\mathbf{D}_{\hat{Y}}$. Each of these matrices, have off-diagonal elements equal to zero, while on diagonal a set of M (or N respectively) randomly distributed eigenvalues equal to 1

$$\begin{aligned} G_{D_{\hat{X}}} &= \frac{1}{T}\left(\frac{M}{z-1} + \frac{T-M}{z}\right) = \frac{m}{z-1} + \frac{1-m}{z} \quad m = \frac{M}{T}; \\ G_{D_{\hat{Y}}} &= \frac{1}{T}\left(\frac{N}{z-1} + \frac{T-N}{z}\right) = \frac{n}{z-1} + \frac{1-n}{z} \quad n = \frac{N}{T}; \end{aligned} \tag{5.8}$$

From chapter 2 we know, that

$$\mathbf{S}_{D_{\hat{X}} \cdot D_{\hat{Y}}}(z) = \mathbf{S}_{D_{\hat{X}}}(z) \cdot \mathbf{S}_{D_{\hat{Y}}}(z) \tag{5.9}$$

or equivalently

$$\frac{1+z}{z} \mathbf{N}_{D_{\hat{X}} \cdot D_{\hat{Y}}}(z) = \mathbf{N}_{D_{\hat{X}}}(z) \cdot \mathbf{N}_{D_{\hat{Y}}}(z), \tag{5.10}$$

where:

$$\mathbf{S}_x(z) = \frac{1+z}{z}\chi_x(z) \qquad \mathbf{N}_x(z) = \frac{1}{\chi_x(z)} \qquad \mathbf{N_x}(\mathbf{z}) G\left(\mathbf{N_x}(\mathbf{z})\right) - 1 = z$$

From this, one easily obtains:

$$\begin{aligned} \mathbf{N}_{D_{\hat{X}}(z)}\left(\frac{m}{\mathbf{N}_{D_{\hat{X}}(z)}-1} + \frac{1-m}{\mathbf{N}_{D_{\hat{X}}}(z)}\right) - 1 &= z \\ \frac{\mathbf{N}_{D_{\hat{X}}}(z) m}{\mathbf{N}_{D_{\hat{X}}}(z)-1} + 1 - m - 1 &= z \end{aligned} \tag{5.11}$$

$$\mathbf{N}_{D_{\hat{X}}}(z) = \frac{m+z}{z} \qquad \mathbf{N}_{D_{\hat{Y}}}(z) = \frac{n+z}{z} \tag{5.12}$$

$$\mathbf{N}_{D_{\hat{X}}} \mathbf{N}_{\cdot D_{\hat{Y}}}(z) = \frac{(m+z)(n+z)}{z^2} \tag{5.13}$$

and one readily gets the N–transform for the matrix $\mathbf{D_{\hat{X}} \cdot D_{\hat{Y}}}$

$$\mathbf{N}_{D_{\hat{X}} \cdot D_{\hat{Y}}}(z) = \frac{(m+z)(n+z)}{z(1+z)} \tag{5.14}$$

Inverting functionally (5.14)

$$\mathbf{N}_{D_{\hat{X}} \cdot D_{\hat{Y}}}(z) \mathbf{G_D}\left(N_{D_{\hat{X}} \cdot D_{\hat{Y}}}(z)\right) = z + 1 \tag{5.15}$$

i.e., solving the second order equation in z, one is able to find the Green's function of a product $D_{\hat{X}} \cdot D_{\hat{Y}}$

$$\begin{aligned} 0 &= z^2(1 - N(z)) + (n + m - N(z))z + mn \\ G(N(z)) &= \frac{2-(n+m+N(z)) - \sqrt{(n+m-N(z))^2 - 4(1-N(z))mn}}{2N(z)(1-N(z))}, \end{aligned} \tag{5.16}$$

where we have omitted the subscripts for brevity. Subsequently mean spectral density is obtained from the standard relation (2.3), which yields

$$\lim_{\epsilon\to 0^+}\frac{1}{\lambda+i\epsilon}=PV\left(\frac{1}{\lambda}\right)-i\pi\delta(\lambda)\Rightarrow\rho_{\mathbf{D}}(\lambda)=-\frac{1}{\pi}\mathrm{Im}G_{\mathbf{D}}(\lambda+i\epsilon). \tag{5.17}$$

The final result *i.e.*, the benchmark case where all (standardized) variables X and Y are uncorrelated, meaning that the ensemble average $E(C_X)=E(XX^T)$ and $E(C_Y)=E(YY^T)$ are equal to the unit matrix, whereas the ensemble average cross-correlation $E(G)=E(YX^T)$ is identically zero, reads as in original paper [BLMP07]:

$$\begin{aligned}\rho_{\mathbf{D}}(\lambda)&=\max(1-n,1-m)\delta(\lambda)+\max(m+n-1,0)\delta(\lambda-1)+\\&+\frac{\mathrm{Re}\sqrt{(\lambda^2-s_-)(\lambda_+-s^2)}}{\pi\lambda(1-\lambda^2)}\end{aligned} \tag{5.18}$$

where $s_\pm=n+m-2mn\pm 2\sqrt{mn(1-n)(1-m)}$ are the two positive roots of the quadratic expression under the square root It is easy to discover the fact, that in the limit $T\to\infty$ at

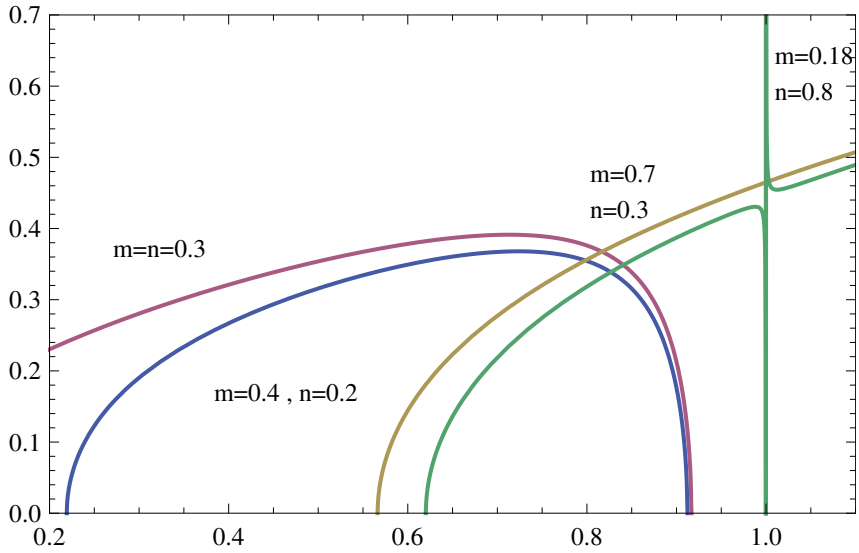

Figure 5.1: Simulation of a continuous part of the theoretical random singular value spectrum $\rho(\lambda)$ for different values of n and m.It is obvious to see that $\lambda_+\leq 1$ for all values of $n,m<1$. The upper bound is reached only when $n+m=1$, in which case the upper edge of the singular value band touches $\lambda=1$ *i.e.*, for $n=m$ the spectrum extends down to $\lambda=0$, whereas for $n+m\to 1$, the spectrum develops a $(1-\lambda)^{-1/2}$ singularity, just before the appearance of a δ peak at $\lambda=1$ of weight $n+m-1$.

fixed N, M, all singular values collapse to zero, as they should since there is no true correlations between X and Y; the allowed band in the limit $n,m\to 0$ becomes:

$$\lambda\in\left[|\sqrt{m}-\sqrt{n}|,\sqrt{m}+\sqrt{n}\right]. \tag{5.19}$$

When $n\to m$, the support becomes $\lambda\in[0,2\sqrt{m(1-m)}]$ (plus a δ function at $\lambda=1$ when $n+m>1$), while when $m=1$, the whole band collapses to a δ function at $\lambda=\sqrt{1-n}$. For $n+m\to 1^-$ there is an initial singularity of $\rho(\lambda)$ $\lambda=1$ diverging as $(1-\lambda)^{-1/2}$. Ultimately $m\to 0$ at fixed n, one finds that the whole band collapses again to a δ function at $\lambda=\sqrt{n}$.

5.3 SVD cleaning technique and the MP^2 case

The results from the previous section were obtained under belief there were no correlations between input and output samples of infinite sizes. However, for a given finite size sample, the eigenvalues of C_X and C_Y will differ from unit, and the singular values of G will not be zero and instead *cross*-correlations between input and output variables are involved. The SVD spectrum in that case is the convolution of two Marčenko-Pastur [MP67] distributions with parameters m and n, respectively, which reads, for $r = n, m < 1$:

$$\rho_{MP}(\lambda) = \frac{1}{2\pi\beta\lambda}\mathrm{Re}\sqrt{(\lambda - \lambda_-)(\lambda_+ - \lambda)} \tag{5.20}$$

with $\lambda_\pm = (1 \pm \sqrt{r})^2$ The N-transform of this density takes a particularly simple form (cf. [BJJ+09] for an exact derivation)

$$\mathbf{N}_{MP}(z) = \frac{1+z}{1+rz} \tag{5.21}$$

The singular values of G are obtained as the square-root of the eigenvalues of $D = X^T X Y^T Y$. Under assumption, that $X^T X$ and $Y^T Y$ are mutually free, after having noted that the N-transform of the $T \times T$ matrices $X^T X$ and $Y^T Y$ are now given by:

$$\mathbf{N}(z) = \frac{(1+z)(1+rz)}{rz} \tag{5.22}$$

one can again use the multiplication rule of N-transforms and finds the Green's function of D by solving the following cubic equation for z:

$$(1+z)(1+nz)(1+mz)\mathbf{N}(z) - mnz = 0 \tag{5.23}$$

which with little effort can be solved analytically. Then Green's function is readily obtained by inserting the solution of the eq.(5.23)

$$G(\mathbf{N}(z)) = \frac{z(\mathbf{N}(z)) + 1}{\mathbf{N}(z)} \Rightarrow \rho(\lambda^2) = -\frac{1}{\pi}\mathrm{Im}G(\lambda + i\epsilon) \tag{5.24}$$

This will lead to a rather complicated form of the final function

$$\rho(\lambda) = \left(\frac{2}{\theta(\lambda^2)}\right)^{1/3} \frac{3^{-1/2}}{\pi\lambda}\left(2^{-2/3} + \varphi(\lambda^2)\right) \tag{5.25}$$

where

$$\varphi(\lambda^2) = 2 - 3m(1-m) - 3n(1-n) - 3mn(n+m-4) + 2(m^3+n^3) + 9\lambda^2(1+m+n)$$

$$\theta(\lambda^2) = \varphi(\lambda^2) - \sqrt{\varphi(\lambda^2) - 4(1+m^2+n^2-mn-m-n+3\lambda^2)^3}$$

5.4 Example from the Data

The last decade has been a witness of an enormous progress in the development of small-scale macroeconomic models. It's not too much an overstatement to say, that the statistical analysis of VAR models, Kalman filter models etc. is nowadays complete. The major issue with these models is that they can accurately approximate small number of time series only. On

the other hand Central Banks must construct their forecasts in rich data environment [BB03]. This mismatch between standard macroeconometric models and real world practice has led to unfortunate consequences. Forecasters have had to rely on informal methods to distill information from the available data, and their published forecasts reflect considerable judgement in place of formal statistical analysis. Forecasts are impossible to reproduce, and this makes economic forecasting a largely non-scientific activity *i.e.*, formal small-scale models have little effect on day-to-day policy decisions, making these decisions more ad hoc and less predicable than if guided by the kind of empirical analysis that follows from careful statistical modeling. The goal of this research is to use the wide range of economic variables that practical forecasters and macroeconomic policymakers have found useful, and establish a direction that explicitly incorporates information from a large number of macroeconomic variables into formal statistical models. We have focused on two different data sets, namely Polish macroeconomic data and generated set of data, where temporal cross - correlations are introduced by definition.The full data set is the same as it was used in previous chapter.

5.4.1 Polish Macroeconomic data

Poland is and interesting emerging market with unique social and business activity in the process of rapid growth and industrialization. We hope our analysis might be helpful in understanding the factors that helped Poland to survive during the 2008 crisis. The main problem to be solved is to choose the correct variables to include. This is the familiar problem of variable selection in regression analysis. Economic theory is of some help, but usually suggests large categories of variables (money, interest rates, wages, stock prices, etc.) and the choice of a specific subset of variables then becomes an open problem. The analysis began with checking, whether the method described in [BLMP07] is relevant for describing the relation between the inflation indexes for Polish macroeconomic indexes and other Polish macroeconomic data published by different government and non-government agencies. A consumer price index (CPI) is a measure estimating the average price of consumer goods and services purchased by households. A consumer price index measures a price change for a constant market basket of goods and services from one period to the next within the same area (city, region, or nation). It is a price index determined by measuring the price of a standard group of goods meant to represent the typical market basket of a typical urban consumer. The percent change in the CPI is a measure estimating inflation. It is commonly viewed as the indicator not only the measure of inflation, but rather the indicates the change of costs of maintenance. The data set represent a wide range of macroeconomic activity and were initially transformed to ensure stationarity and diminish the effects of seasonal components. The same data set we have already analyzed in chapter 4 and the detailed list of all time series is attached in the appendix C.1. This time, the whole set of 52 time series,observed on a monthly basis between $Jan - 2000$ and $Oct - 2009$ ($T = 118$) was divided into two subsets *i.e.*,

- We have used monthly $M = 15$ changes of different CPI indicators as our predicted variables (i.e. output sample Y)

- The input sample X consisted of $N = 37$ monthly changes of economic indicators (eg. sectoral employment, foreign exchange reserves, PPI's) as explanatory variables.

The data were standardized and mean adjusted, but following the general idea of [BLMP07] the input and output samples' factors were not selected very carefully, so the data could speak for themselves and system could be able to select the optimal combination of variables. The resulting

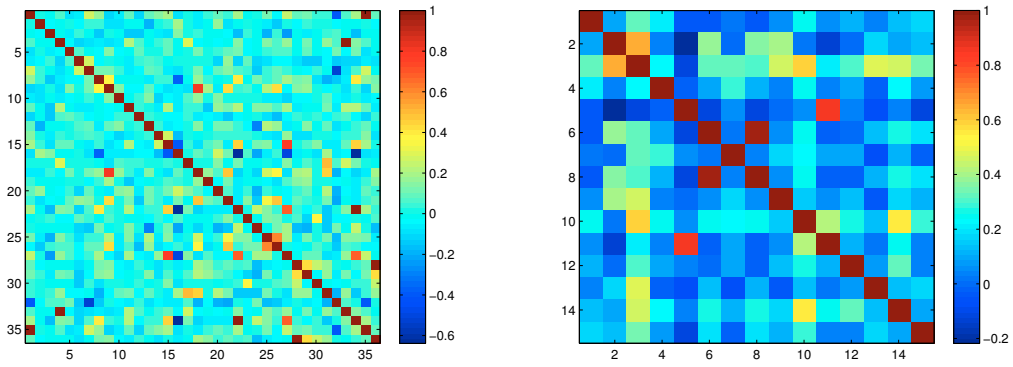

Figure 5.2: Correlation matrices representing generic in-the-sample correlations.The data were mean adjusted and standardized. In a perfect situation, one is expecting that cross–correlations tends to zero, however still nontrivial correlations are present. LEFT:Matrix with 37 input variables X.RIGHT: Matrix with 15 input variables Y - components of CPI.

diagrams (see Fig.5.2) now demonstrate, that even standardized and converted to stationary time series may represent nontrivial in-the-sample correlations. Short – term economic forecasts build from these type data in consequence may be poor and show no sign of improving over time. The next step involved cleaning internal correlations in each sample. To do it, we have used equation (5.4). The effective matrices were then diagonalized and two sets of internally uncorrelated data were prepared.

Results for Equal-time spectra

From the uncorrelated data we create the rectangular matrix G and diagonalize it to calculate singular eigenvalues. Finally we have used the benchmark calculated in equation (5.5) to compare the data with the predicted eigenvalue density. For the same data sets we have also created the set of correlated samples *i.e.*, the set, where internal spatial cross–correlations were not a-priori removed (see Fig. 5.3). Apparently there is enough idiosyncratic variation in standard activity measures like the unemployment rate and capacity utilization, that removing noisy components from these might provide a clearer picture of factor models affecting inflation. We have excluded from further analysis series responsible for reference NBP bill rate balance of payments, and from the set of explanatory variables ordinary measures of inflation - CPI in food sector, beverages and tobacco ans services. This approach allows us to directly reproduce temporal cross–correlations.

The lack of symmetry condition allows us to focus only on out-of-the-sample correlations without mixing them with inner ones and to study temporal properties of such matrix. The results show, that there exists some singular eigenvalues, which do not fit the benchmark. Among them, the highest singular eigenvalue $s_1 = 2.5$ and the corresponding singular eigenvector, represent standard negative correlation between expenses for electricity and net balance of payments in the energy and positive correlation between CPI in health sector and unemployment. In our ap-

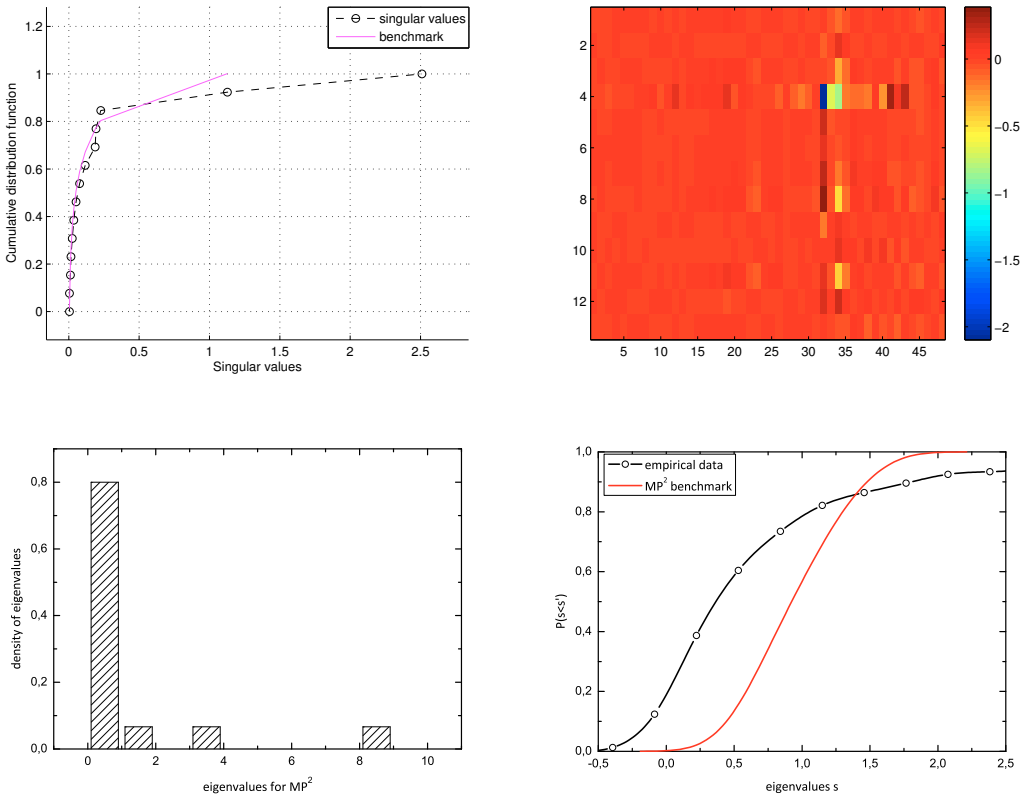

Figure 5.3: Comparison of cleaning techniques for equal–time SVD spectra. ABOVE: Cumulative singular density and "heat map" for the cleaned problem. BELOW LEFT: Empirical density of eigenvalues in the MP^2 framework.BELOW RIGHT: Benchmark calculated according to MP^2 spectrum.

proach we can not only observe this relations, but also interpret them in terms of causality. That is, larger unemployment rate causes increase in the CPI. There are other non-trivial relations between eg. CPI in telecommunication sector and foreign exchange reserves. All of these correlations are however well known from textbooks or can be easily explained by means of classical economic theory. When some of the eigenvalues become strongly related, zero modes emerge - clearly the majority (around 80% of all eigenvalues are concentrated close to zero, meaning that there are strong spatial correlations inside X and Y data set. If we are to use the MP^2 benchmark then it is clear, that empirical spectrum is affected by idiosyncratic components, again confirming, that spatial structure strongly interferes with temporal, and it is crucial to "remove" redundant factors to avoid spurious (confunding) correlations.

Solution for lagged spectra

A natural way to examine macroeconomic data is via factor models. In previous section we have assumed that the inflation can be accurately explained by the factor model using relatively small number of latent variables. Pushing the factor model one step further, these latent factors might also explain the predictive relationship between current values of variables and those of the previous month. The next step of our study involved shifting the input and output

data set by one observation (one month). The Y were calculate from $t = 2, \ldots, 118$ and X's for $(t = 1, \ldots, 117)$. We were motivated by the common belief, that it is the "yesterday" shock, that affects "today" change *i.e.*, there is some persistency and long memory within this system. This is the same approach, that underlies the idea of VARMA(q_1, q_2) processes [LÖ5]. The temporal structure (Fig.5.4) manifests itself via the existence of significant non–

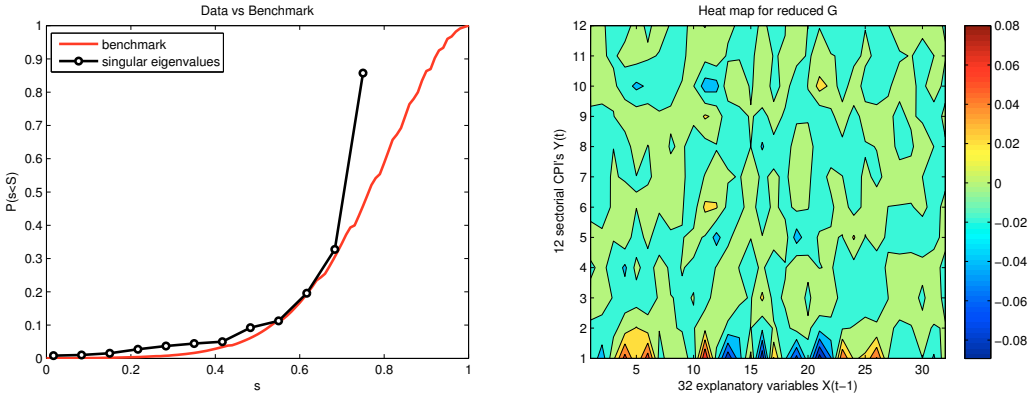

Figure 5.4: Results for lagged-by-one-month spectra.LEFT: There is a bunch of singular values that do not fit the benchmark. The largest $s \approx 0.77$ represents the same correlation structure as within unshifted framework.RIGHT: Note, that now we can see whole bunch of islands of non-trivial factors, that affect CPI's.

symmetric relation (represented by one singular eigenvalue, that does not fit the benchmark) between data sets X and Y, that are shifted by one month. It is easy to notice, that only few factors are responsible for the model's performance. CPI in telecommunication sector is affected by the largest number of possible explanatory variables (c.f. Table 5.1). Among them the most unexpected is the correlation with heavy trucks. Two or three factors are useful for some categories of series, but only a single factor is responsible for the predictability of prices in all sectors. Apparently, the first factor is foreign exchange reserves level, and the results say that it is an important predictor of future prices in telecommunication, manufacturing and transport sector. We can say that when forecasting inflation a large model might be a clue, but if we remove redundant factors the inflation can be forecasted by using simple measures of real activity like the unemployment rate, industrial production or capacity utilization. While the first factor is easy to interpret, a complete understanding of the results requires an understanding of other factors as well. Unfortunately, their interpretation and role in explaining future changes in the consumer prices is an open question.

5.5 Conclusions

We will now recap this illustrative study with few comments:

- In general both input and output data sets may represent highly complex correlation structure strongly interfered by redundant noisy factors. This significant amount of noise need to be carefully eliminated by performing initial decoupling of spatial correlations, so these large matrices become mutually free.

- This is again precisely the case when FRV approach "takes the stage" and reduces the

Y	X	Type of correlation
CPI in communication sector	completed dwellings	negative
	net balance of payments of goods	
	$M3$ money aggregate	
	Employment in manufacturing sector	
	employment in enterprize sector	positive
	Direct investments	
	Foreign exchange reserves	
	Official reserve assets	
	New heavy trucks registration	
	Balance of payments - services	
CPI in clothing sector	Total export	negative
CPI in restaurants and hotels sector	Foreign exchange reserves	positive
CPI in transport sector	Foreign exchange reserves	positive
	Total production in manufacturing sector	negative
	Total export	

Table 5.1: Factors affecting the set of output variables for lagged spectra.

solution to few lines.

- The procedure tested on real data within the case of unshifted variables hasn't show any significant improvement in comparison to standard factor analysis known in econometric literature for similar data sets[SW99]. For data lagged by one observation we have however recovered the sea of different non–trivial relations, and it might be interesting to compare these results from a more general perspective of factor models, however no implicitly close approach was found in the literature.

Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away.

A. de Saint-Exupery

6 Conclusions

Cross-correlations analysis lies at the confluence of many branches of science, where unraveling dependencies from data sets is fundamental. The ability to find good estimates however heavily relies on spatio–temporal structure and becomes quite a complex task, when the dimension of the data set grows. At the same time Random Matrix Theory may hold the key to solving critical problems for a broad range of complex systems from biophysics to quantum chaos to signals and communication theory to machine learning to finance to geoscience modeling. The Voiculescu Free Random Variables (FRV) Technique [DNV92], is the most natural candidate for the "matrix–valued probability calculus", that can provide efficient yet elegant algorithms for cleaning (de–noising) large sets of data and unraveling essential but hidden correlations and in this way promote new insights into classical methods of data analysis and reduction. The primary goal of this thesis, was to show the fluency of FRV approach in solving complex problems of unwinding spatio–temporal correlations omnipresent in large covariance matrices generated by various complex systems, like eg. financial markets or economies. We have showed how complex and almost unsolvable problems on a basis of ordinary Random Matrix Theory can be easily tackled and solved analytically within just few lines of not very complicated calculations The salient feature was the larger the matrix, the more deterministic it becomes.

- First chapter was generally devoted to give a quick summary of classical methods of data analysis and data reduction techniques , like PCA of Factor Analysis for instance. We have stressed not only the important aspects, but also main weakness, also known as "dimensionality curse".This weakness is however challenge for our FRV approach, where large dimension is an innate assumption.

- In chapter 2 we have presented a short survey of Free Probability Theory, whose crucial part was the notion of freeness *i.e.*, the counterpart of classical independence. Furthermore have explored the interplay of classical probabilistic aspects to FRV version random matrix theory and construct a short dictionary of matrix–valued probability calculus in a correspondence with classical results. In particular we have enhanced the conformity by deriving the Free Central limit theorem and Free Poisson process in analogy with asymptotic limits of classical binomial distribution.

Over the recent years the practical environment has changed dramatically withe the spectacular evolution of data acquisition technologies and computing facilities and many applications have emerged in which the number of experimental variables is comparable to the underlying dimension. At the same time methodology hasn't responded vigorously to these challenges. There is still need of consolidation in the form of systematic and critical assessment of the

new approaches as well as development of an appropriate theoretical underpinning for analysis dynamical properties of these data. The rest of the thesis represents original work tailored to specific problems. We have used Free Random Variables approach to the analysis of spectral densities that accurately represent the problem's dynamics.

- Chapter 3 deals with estimators of equal-time cross correlations in large N, T limit. The complex system under study was Warsaw Stock Exchange. We have analyzed spectral density function under assumption of uncorrelated and correlated assets. The original idea of one factor model comes from [JM03]. We have shown, that the unique structure of financial market data is an effect of non-synchronous character of the data and vanishes if we shuffle the returns. For the latter case of correlated assets we have used the version of factor analysis developed by [J+10] to identify hidden structure, that drives the behavior of a portfolio.We have identified 6 new sectors of activity. Hasty assumption is that eigenvalues are stationary over time. We have focused on the dynamics of the eigenvalues by means of sliding windows and discovered, that indeed noisy eigenvalues are stationary over time whereas eigenvalues that represent nontrivial correlations are far from stationarity. Their dynamics can be fully characterized by Langevin equation. In this spirit we have reconsidered the one factor model [JM03] and found that previously identified true eigenvalues have informative eigenvectors. Furthermore, to analyze long-range persistency in temporal correlations, we have introduced weighting schemes into correlations estimator *i.e.*, Exponentially Weighted Moving Average (EWMA) estimator and again we have used flexible FRV rules to calculate the spectral density in this regime afterwards applied to the empirical eigenvalues spectrum.

- Chapter 4 studied the correlations matrices generated by widely known, and widely used VARMA(q_1, q_2) processes. The correlations estimators are symmetric. Since hey are built of lagged data, may still carry some information about delayed temporal correlations. It resulted, that these matrices can be easily treated in a framework of doubly correlated Wishart ensembles. We have derived respective spectral densities for VAR(q),VMA(q) and VARMA(q_1, q_2) processes. The results were then confirmed by numerical simulations. The empirical data as compared to the spectral density given by eq.(2.50) suggest that a lot of eigenvalues, similarly to stock market data express marginal predictive content. Tests performed on real–life macroeconomic data showed surprisingly "perfect agreement". This suspected to be the manifestation of the collective response to "global" shocks and the identification of these is crucial for forecast constructions.

- Finally in chapter 5 we attracted our attention to time–delayed cross–correlations. The correlations matrix is then nonsymmetric and the eigenvalues are complex numbers. Therefore we have used singular eigenvalues instead. We have compared this spectrum with a benchmark obtained under assumption, there are no correlations. This results applied again to Polish macroeconomic data showed, that non-trivial temporal structure survives when the input and output matrices are shifted at least by one observation ("month") and are not visible for data in an non–delayed scheme. The lack of symmetry condition allowed us to focus only on out-of-the-sample correlations without mixing them with inner ones and to study temporal properties of such matrix. The results show, that there exists some singular eigenvalues, which do not fit the benchmark. Among them, for non–delayed

spectra the highest singular eigenvalue $s_1 = 2.5$ and the corresponding singular eigenvector, represent standard negative correlation between expenses for electricity and net balance of payments in the energy and positive correlation between CPI in health sector and unemployment. Apparently, the major factor was foreign exchange reserves level, and the results say that it is an important predictor of future prices in telecommunication, manufacturing and transport sector.

- The last two chapters opens a door to the macroeconomic studies and are of certain importance for central banks, which operate on large data sets, but where striking mathematical procedures are so far unknown and the prominent forecasts are usually intuitive and subjective.

The FRV approach opens new possibilities to quantify the character of temporal fluctuations of eigenvalue spectrum in an analytic way. It is tempting to conclude, that various potentially complex problems does not really need sophisticated numerical methods, but may be tackled within few lines of calculations instead. My hope is, that dissertation would be a small contribution to a continuing and fruitful influence of Free Probability Theory developments in large Random Matrix Theory in specific frontier fields and application of methods to analyze complex systems, high dimensional data analysis and variety of intriguing applications, furthering our understanding of the spectral properties of random matrices, behavior of eigenvalues of matrix valued stochastic processes and recovering the information from an observed data set.

A

A.1 Notational aspects of RMT

The most important distribution is the $N(\mu, \sigma^2)$ normal distribution with mean μ and variance σ^2.(For complex random variables $z = x + iy$, $\tilde{N}(\mu, \sigma^2)$ refer to a distribution ,whose entries x and y are *iid* from $N(\mu, \sigma^2)$). It is worth this special place, due to the central limit theorem, which loosely speaking, states that a sum of large number of random variables quite frequently behave as if they were drawn from a normal distribution. Furthermore most of the distributions useful in complex systems analysis are derived from normal distribution. Following this rule, we will be interested in the matrix ensembles derived on the basis of normal distribution, while they are invariant under orthogonal transformations [ERR05].

- **Gaussian ensemble** $X_\beta(T, N)$ - $T \times N$ random matrix, whose entries are (real for $\beta = 1$, complex for $\beta = 2$, quaternion for $\beta = 4$)independent and identically distributed *iid* from $N_\beta(0, 1)$.

- **Gaussian Orthogonal Ensemble (GOE)=$\frac{X_\beta + X_\beta^T}{2}$:** Symmetric $N \times N$ matrix, whose entries are *iid* elements, $N_1(0, \frac{1}{2})$ on the upper triangle and *iid* $N_1(0, 1)$ on the diagonal and X_β is an $N \times N$ Gaussian matrix with real entries and $\beta = 1$.

- **Gaussian Unitary Ensemble (GUE)=$\frac{X_\beta + X_\beta^\dagger}{2}$** Hermitian $N \times N$ matrix, whose entries are iid elements, that are complex $N_2(0, \frac{1}{2})$ on the upper triangle and *iid* $N_1(0, 1)$ on the diagonal.X_β is an $N \times N$ Gaussian matrix with complex entries and $\beta = 2$.

- **Gaussian Symplectic Ensemble (GSE)=$\frac{X_\beta + X_\beta^D}{2}$,** $\beta = 4$ is self–dual $N \times N$ matrix. Where D denotes the dual transpose of a quaternion matrix. The diagonal elements are *iid* with the normal distribution $N_1(0, 1)$ and the off-diagonal entries subject to being self–dual are *iid* with distribution $N_4(0, \frac{1}{2})$

- **Wishart ensemble** $W_\beta(T, N) = X_\beta X'_\beta, \quad T \geq N$- symmetric (for $\beta = 1$), Hermitian (for $\beta = 2$) or self–dual (for $\beta = 4$) $N \times N$ random matrix, where X' denotes X^T, $X^\dagger$,X^D depending on whether the entries of $X_\beta(T, N)$ are real, complex, quaternion, respectively.

B

B.1 The Auto–Covariance Matrix for VMA(q)

In this appendix, we sketch a proof of the formula (4.13) for the auto–covariance matrix of the VMA(q) process. As mentioned, since the random variables are centered Gaussian, this matrix alone suffices to completely capture all their properties. We set $i = j$; the dependence on this index may be dropped as there are no correlations here. We use the definition (4.11) of VMA(q), as well as the auto–covariance structure of the white noise, $\langle \epsilon_{ia} \epsilon_{jb} \rangle = \delta_{ij} \delta_{ab}$. This leads to

$$A^{(1)}_{ab} = \langle Y_{ia} Y_{ib} \rangle = \sum_{\alpha=0}^{q} \sum_{\beta=0}^{q} a_\alpha a_\beta \, \langle \epsilon_{i,a-\alpha} \epsilon_{i,b-\beta} \rangle = \sum_{\alpha=0}^{q} \sum_{\beta=0}^{q} a_\alpha a_\beta \delta_{a-\alpha,b-\beta} = \dots .$$

The double sum is symmetrized, the index β replaced by $d \equiv \beta - \alpha$,

$$\dots = \frac{1}{2} \sum_{\alpha=0}^{q} \sum_{d=-\alpha}^{q-\alpha} a_\alpha a_{\alpha+d} \, (\delta_{b,a+d} + \delta_{b,a-d}) = \dots ,$$

and the order of the sums interchanged (an elegant method for this is explained in [GKP94]),

$$\dots = \frac{1}{2} \sum_{d=-q}^{q} \left(\sum_{\alpha=\max(0,-d)}^{q-\min(0,d)} a_\alpha a_{\alpha+d} \right) (\delta_{b,a+d} + \delta_{b,a-d}) \, ,$$

which, upon splitting the sum over d into three pieces (from $-q$ to -1, $d = 0$, and from 1 to q), is quickly seen to coincide with (4.13).

B.1.1 A List of the Various Auto–Covariance Matrices Used

For the reader's convenience, let us collect in this appendix the five auto–covariance matrices which are defined throughout chapter 4:

- By $\mathbf{A}^{(1)}$ we denote the auto–covariance matrix of the VMA(q) process with the generic constants a_α, with $\alpha = 0, 1, \dots, q$, as defined in (4.11).
- By $\mathbf{A}^{(2)}$ we denote the auto–covariance matrix of the VMA(q) process with the constants $a^{(2)}_0 \equiv 1/a_0$, $a^{(2)}_\beta \equiv -b_\beta / a_0$, where $\beta = 1, \dots, q$.
- By $\mathbf{A}^{(3)}$ we denote the auto–covariance matrix of the VAR(q) process with the generic constants a_0, b_β, with $\beta = 1, \dots, q$, as defined in (4.19). There holds $\mathbf{A}^{(3)} = (\mathbf{A}^{(2)})^{-1}$ (4.20).
- By $\mathbf{A}^{(4)}$ we denote the auto–covariance matrix of the VMA(q_1) process with the constants $a^{(4)}_0 \equiv 1$, $a^{(4)}_\beta \equiv -b_\beta$, where $\beta = 1, \dots, q_1$.

- By $\mathbf{A}^{(5)}$ we denote the auto–covariance matrix of the VARMA(q_1, q_2) process with the generic constants b_β, a_α, with $\beta = 1, \ldots, q_1$ and $\alpha = 0, 1, \ldots, q_2$, according to the definition (4.29). There is $\mathbf{A}^{(5)} = (\mathbf{A}^{(4)})^{-1}\mathbf{A}^{(1)}$ (4.30), where in the latter piece $q = q_2$.

B.1.2 The M–Transform of the Auto–Covariance Matrix for VMA(q)

We will derive here the M–transform (4.16) of the auto–covariance matrix $\mathbf{A}^{(1)}$ of an arbitrary VMA(q) process, using the expression for its Fourier transform (4.15). It is a little simpler to consider the Green's function,

$$G_{\mathbf{A}^{(1)}}(z) = \frac{1 + M_{\mathbf{A}^{(1)}}(z)}{z} = \frac{1}{\pi}\int_0^\pi \mathrm{d}p \frac{1}{z - \widehat{A^{(1)}}(p)}, \tag{B.1}$$

where the integration range has been halved due to the evenness of the integrand.

This integral is performed with help of the change of variables $y \equiv 2\cos p$. The measure, when $p \in [0, \pi]$, reads $\mathrm{d}p = -\mathrm{d}y/\sqrt{4 - y^2}$. A basic observation is that the denominator of the integrand is a linear combination of $\cos(dp)$, for $d = 1, \ldots, q$, and each such a cosine can be cast as a polynomial of order d in y through the de Moivre formula. Hence, the denominator is a polynomial of order q in y,

$$\widehat{A^{(1)}}(p) - z = \kappa_0^{(1)} - z + 2\sum_{d=1}^{q} \kappa_d^{(1)} \cos(dp) = \psi \prod_{\beta=1}^{q} (y - y_\beta)\,, \tag{B.2}$$

where the y_β's are the q roots (which we assume to be single), and ψ is the coefficient at y^q. Using the method of residues, one readily finds

$$\begin{aligned} G_{\mathbf{A}^{(1)}}(z) &= -\tfrac{1}{\pi}\tfrac{1}{\psi}\int_{-2}^{2} \mathrm{d}y \tfrac{1}{\sqrt{4-q^2}} \tfrac{1}{\prod_{\beta=1}^{q}(y-y_\beta)} = \\ &= \tfrac{1}{\psi}\sum_{\gamma=1}^{q} \tfrac{1}{\prod_{\substack{\beta=1 \\ \beta\neq\gamma}}^{q}(y_\gamma - y_\beta)} \tfrac{1}{\sqrt{y_\gamma - 2}\sqrt{y_\gamma + 2}}, \end{aligned} \tag{B.3}$$

where the two square roots on the r.h.s. are principal. This is an explicit formula for the Green's function of $\mathbf{A}^{(1)}$, provided one has factorized the order–q polynomial (B.2).

B.2 The Auto–Covariance Matrix for VAR(q)

Let us argue now that the Fourier transform (4.21) leads to the auto–covariance matrix of VAR(q) (4.25) of the form of a sum of exponential decays (4.26), and let us give precise expressions for the constants C_γ and the characteristic times T_γ, $\gamma = 1, \ldots, q$.

We proceed by the technique of residues, analogously to (B.1.2), however this time with aid of another variable, $x \equiv \mathrm{e}^{-\mathrm{i}p}$, related to the previously used through $y = 2\cos p = x + 1/x$. The integration measure is $\mathrm{d}p = \mathrm{i}\mathrm{d}x/x$, and the integration path is counterclockwise around the centered unit circle. The denominator of the integrand is a polynomial of order q in the variable y, having thus some q roots $\tilde{y}_\beta$, $\beta = 1, \ldots, q$. Therefore, there are $2q$ corresponding solutions for the variable x, with a half of them inside the integration path and a half outside; let $\tilde{x}_\beta$ be the solutions to $x + 1/x = \tilde{y}_\beta$ with the absolute values less than 1. Only them contribute to the integral, and their residues straightforwardly give

$$A^{(3)}(d) = \frac{1}{\psi}\sum_{\gamma=1}^{q} \frac{(\tilde{x}_\gamma)^{|d|+q-1}}{\prod_{\substack{\beta=1 \\ \beta\neq\gamma}}^{q}(\tilde{x}_\gamma - \tilde{x}_\beta)\prod_{\beta=1}^{q}\left(\tilde{x}_\gamma - \frac{1}{\tilde{x}_\beta}\right)}. \tag{B.4}$$

This is indeed q exponents $(\tilde{x}_\gamma)^{|d|}$, $\gamma = 1, \ldots, q$. Remark that the solutions may be complex, hence this is really q different exponential decays $\exp(-|d|/T_\gamma)$, with the characteristic times

$$T_\gamma \equiv -\frac{1}{\log|\tilde{x}_\gamma|} \tag{B.5}$$

(these times are positive as the roots have the absolute values less than 1), possibly modulated by sinusoidal oscillations when a root has an imaginary part. For example, for $q = 1$ there is one exponential decay (4.27), while for $q = 2$, one obtains either two exponential decays (the two roots are real and different), or one exponential decay modulated by oscillations (the two roots are complex and mutually conjugate), *etc.*

B.2.1 The Equation for the M–Transform of the Pearson Estimator of the Covariances for VARMA$(1,1)$

The sixth–order polynomial equation obeyed by $M \equiv M_{\mathbf{c}}(z)$ in the case of VARMA$(1,1)$ reads,

$$r^4 a_0^2 a_1^2 \left(a_0^2 - a_1^2\right)^2 M^6 + 2r^3 a_0 a_1 \Big(\Big(\left(a_0^4 - 6a_0^2 a_1^2 + a_1^4\right) b_1 -$$

$$+ a_0 a_1 \left(a_0^2 + a_1^2\right)\left(b_1^2 + 1\right) \Big) z + (1+2r)\, a_0 a_1 \left(a_0^2 - a_1^2\right)^2 \Big) M^5 +$$

$$+ r^2 \Big(\Big(\left(a_0^4 - 20 a_0^2 a_1^2 + a_1^4\right) b_1^2 - 4 a_0 a_1 \left(a_0^2 + a_1^2\right) b_1 \left(b_1^2 + 1\right) +$$

$$+ a_0^2 a_1^2 \left(b_1^4 + 1\right) \Big) z^2 + 2 a_0 a_1 \Big(\left((1+3r)\left(a_0^4 + a_1^4\right) - 2\left(5 + 9r\right) a_0^2 a_1^2\right) b_1 -$$

$$+ (2+3r)\, a_0 a_1 \left(a_0^2 + a_1^2\right)\left(b_1^2 + 1\right) \Big) z + \left(1 + 8r + 6r^2\right) a_0^2 a_1^2 \left(a_0^2 - a_1^2\right)^2 \Big) M^4 +$$

$$+ 2r \Big(b_1 \Big(-6 a_0 a_1 b_1^2 - \left(a_0^2 + a_1^2\right) b_1 \left(b_1^2 + 1\right) + a_0 a_1 \left(b_1^4 + 1\right) \Big) z^3 +$$

$$+ \Big(\left(-10\left(1 + 2r\right) a_0^2 a_1^2 + r\left(a_0^4 + a_1^4\right)\right) b_1^2 - 2\left(1 + 2r\right) a_0 a_1 \left(a_0^2 + a_1^2\right) b_1 \left(b_1^2 + 1\right) +$$

$$+ (1+r)\, a_0^2 a_1^2 \left(b_1^4 + 1\right) \Big) z^2 + a_0 a_1 \Big(\left(3\left(1 + r\right)\left(a_0^4 + a_1^4\right) - 2\left(2 + 15r + 9r^2\right) a_0^2 a_1^2\right) b_1 -$$

$$+ \left(1 + 6r + 3r^2\right) a_0 a_1 \left(a_0^2 + a_1^2\right)\left(b_1^2 + 1\right) \Big) z +$$

$$+ 2r\left(1 + 3r + r^2\right) a_0^2 a_1^2 \left(a_0^2 - a_1^2\right)^2 \Big) M^3 + \Big(b_1^2 \left(1 - b_1^2\right)^2 z^4 +$$

$$+ 2 b_1 \Big(-2\left(1 + 3r\right) a_0 a_1 b_1^2 - r\left(a_0^2 + a_1^2\right) b_1 \left(b_1^2 + 1\right) + (1+r)\, a_0 a_1 \left(b_1^4 + 1\right) \Big) z^3 +$$

$$+ \Big(- \left(\left(1 - r^2\right)\left(a_0^4 + a_1^4\right) + 2\left(3 + 20r + 10r^2\right) a_0^2 a_1^2\right) b_1^2 -$$

$$
\begin{aligned}
&-2\left(1+4r+2r^2\right)a_0a_1\left(a_0^2+a_1^2\right)b_1\left(b_1^2+1\right)+r\left(4+r\right)a_0^2a_1^2\left(b_1^4+1\right)\Big)z^2+\\
&+2ra_0a_1\Big(\left(r\left(3+r\right)\left(a_0^4+a_1^4\right)-6\left(2+5r+r^2\right)a_0^2a_1^2\right)b_1-\\
&+\left(3+6r+r^2\right)a_0a_1\left(a_0^2+a_1^2\right)\left(b_1^2+1\right)\Big)z+\\
&+r^2\left(6+8r+r^2\right)a_0^2a_1^2\left(a_0^2-a_1^2\right)^2\Big)M^2+2\Big(a_0a_1b_1\left(1-b_1^2\right)^2z^3+\\
&+\Big(-\left(a_0^4+a_1^4+2\left(3+5r\right)a_0^2a_1^2\right)b_1^2-2\left(1+r\right)a_0a_1\left(a_0^2+a_1^2\right)b_1\left(b_1^2+1\right)+\\
&+ra_0^2a_1^2\left(b_1^4+1\right)\Big)z^2+ra_0a_1\Big(\left(r\left(a_0^4+a_1^4\right)-2\left(6+5r\right)a_0^2a_1^2\right)b_1-\\
&+\left(3+2r\right)a_0a_1\left(a_0^2+a_1^2\right)\left(b_1^2+1\right)\Big)z+r^2\left(2+r\right)a_0^2a_1^2\left(a_0^2-a_1^2\right)^2\Big)M+\\
&-b_1\Big(\left(a_0^4+6a_0^2a_1^2+a_1^4\right)b_1+2a_0a_1\left(a_0^2+a_1^2\right)\left(b_1^2+1\right)\Big)z^2-\\
&-2ra_0^2a_1^2\Big(4a_0a_1b_1+\left(a_0^2+a_1^2\right)\left(b_1^2+1\right)\Big)z+r^2a_0^2a_1^2\left(a_0^2-a_1^2\right)^2=0.
\end{aligned}
\tag{B.6}
$$

C

C.1 A detailed list of macroeconomic time series

No.	Name
1	Reference Rate (7-Day NBP Bill Rate)
2	Overall balance of payment on a balance basis
3	Domestic Budget balance (public finances)
4	Net Balance of Payment on Capital Account basis
5	Inflation in Clothing and footwear sector
6	Inflation in Communication's sector
7	Inflation excluding food and energy prices
8	Inflation in Education's sector
9	Inflation - Electricity, gas and other fuels sector
10	Inflation in Food sector
11	Inflation al. beverages and tobacco
12	Inflation food and non-alcoholic beverages
13	Inflation Health
14	Inflation Furnishings and household equipment
15	Inflation Housing, water, electricity, gas and other fuels
16	Inflation Miscellaneous goods and services
17	Inflation Recreation and culture
18	Inflation Restaurants and hotels
19	Inflation Transport, Total
20	Inflation Credit, households
21	Total credits
22	Total Current Account Balance
23	Net Current transfers
24	Completed Dwellings in Construction sector
25	dwellings Under construction

No.	Name
26	Total Employment in Enterprise sector
27	Total Exports
28	Total Direct Investment
29	Total Financial Account Balance
30	Net Investments in financial derivatives
31	Net Investments financial account
32	Total Other Investments financial account
33	Foreign exchange rate
34	Credit, general government, net
35	Net Goods - current account
36	Total Imports
37	Net Current Account income
38	M3 financial aggregate
39	Total New Heavy Trucks Registration
40	Total Portfolio Investments on financial account balance of payments
41	Employment in Manufacturing Sector
42	Total Production in Manufacturing Sector
43	Official reserve assets
44	Total Domestic trade
45	Retail sales Solid, liquid and gaseous fuels
46	Net Balance of Payment/Current Account/Services
47	Total foreign trade balance
48	Total unemployment rate
49	Central Bank Forex Reserves
50	General government credit
51	Total retail sales
52	Base money market rate

References

[AFV10] G. Akemann, J. Fischmann, and P. Vivo. Universal correlations and power-law tails in financial covariance matrices. *Physica A: Statistical Mechanics and its Applications*, 389:2566 – 2679, 2010. [arXiv:0906.5249v1].

[AGZ09] G.W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics (No. 118). Cambridge University Press, 2009. [e-print].

[Bai03] J. Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71:135–171, 2003. [DOI:10.1111/1468-0262.00392].

[BB03] B. Bernanke and J. Boivin. Monetary policy in a data rich environment. *Journal of Monetary Economics*, 50:525, 2003. [DOI:10.1016/S0304-3932(03)00024-2].

[Bel61] R. Bellman. *Adaptive control processes: a guided tour*. Princeton University Press, Princeton, N.J.,, 1961.

[BEW88] T. Bollerslev, R.F. Engle, and J.M. Wooldridge. A capital asset pricing model with time-varying covariances. *The Journal of Political Economy*, 96:116–131, 1988. [JSTOR:1830713].

[BGJJ04] Z. Burda, A. Görlich, A. Jarosz, and J. Jurkiewicz. Signal and noise in correlation matrix. *Physica A: Statistical Mechanics and its Applications*, 343:295 – 310, 2004. [arXiv:cond-mat/0305627].

[BGJW06] Z. Burda, A. Görlich, J. Jurkiewicz, and B. Wacław. Correlated wishart matrices and critical horizons. *The European Physical Journal B - Condensed Matter and Complex Systems*, 49(3):319–323, 2006. [arXiv:cond-mat/0508341].

[BGW06] Z. Burda, A. Görlich, and B. Wacław. Spectral properties of empirical covariance matrices for data with power–law tails. *Phys. Rev.* ***E***, 74:041129, 2006. [arXiv:physics/0603186].

[Bia98] P. Biane. Free probability for probabilists. *MSRI report*, 0040:16, 1998. [arXiv:math/9809193v1].

[BJ04] Z. Burda and J. Jurkiewicz. Signal and noise in financial correlation matrices. *Physica A: Statistical Mechanics and its Applications*, 344(1-2):67 – 72, 2004. [arXiv:cond-mat/0312496].

[BJJ+09] Z. Burda, A. Jarosz, J. Jurkiewicz, M.A. Nowak, G. Papp, and I. Zahed. Applying free random variables to random matrix analysis of financial data. *submitted to Quantitative Finance*, 2009. [arXiv:physics/0603024].

[BJNS10] Z. Burda, A. Jarosz, M.A. Nowak, and M. Snarska. A random matrix approach to varma processes. *New Journal of Physics*, 12:075036, 2010. [arXiv:1002.0934v1].

[BJR94] G. Box, G.M. Jenkins, and G. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice–Hall, Englewood Cliffs NJ, 3 edition, 1994. [book-info].

[BJW05a] Z. Burda, J. Jurkiewicz, and B. Wacław. Eigenvalue density of empirical covariance matrix for correlated samples. *Acta Physica Polonica* ***B***, 36:2641, 2005. [arXiv:cond-mat/0508451].

[BJW05b] Z. Burda, J. Jurkiewicz, and B. Wacław. Spectral moments of correlated wishart matrices. *Phys. Rev. E*, 71(2):026111, Feb 2005. [arXiv:cond-mat/0405263].

[BLMP07] J.P. Bouchaud, L. Laloux, M.A. Miceli, and M. Potters. Large dimension forecasting models and random singular value spectra. *The European Physical Journal B - Condensed Matter and Complex Systems*, 55(2):201 – 207, 2007. [arXiv:physics/0512090v1].

[BN02] J. Bai and S. Ng. Determining the number of factors in approximate factor model. *Econometrica*, 70:191–221, 2002. [DOI:10.1111/1468-0262.00273].

[BN08] J. Bai and S. Ng. *Large Dimensional Factor Analysis*, volume 3:No.2 of *Foundations and Trends in Econometrics*. Now Publishers - World Scientific, 2008. [book-info].

[Bol86] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307, 1986. [DOI:10.1.1.161.7380].

[BP99] H. Bercovici and V. Pata. Stable laws and domains of attraction in free probability theory. *Annals of Mathematics*, 149:1023, 1999. appendix by P.Biane [DOI:10.1.1.27.3736].

[BS01] P. Biane and R. Speicher. Free diffusions, free entropy and free Fisher information. *Annales de l'Institut Henri Poincaré* ***B***, 37:581, 2001. [DOI:10.1016/S0246-0203(00)01074-8].

[BS06] Z.D. Bai and J.W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Science Press, Beijing, 2006. [DOI:10.1007/978-1-4419-0661-8].

[BV93] H. Bercovici and D.V. Voiculescu. Free convolution of measures with unbounded support. *Indiana University Mathematics Journal*, 42:733, 1993.

[CLS82] P. Cvitanović, P.G. Lauwers, and P.N. Scharbach. The planar sector of field theories. *Nuclear Physics* ***B***, 203:385, 1982. [DOI:10.1016/0550-3213(82)90320-0].

[CRC09a] T. Conlon, H.J. Ruskin, and M. Crane. Cross-correlation dynamics in financial time series. *Physica A*, 388(5):705–714, 2009. [arXiv:1002.0321v1].

[CRC09b] T. Conlon, H.J. Ruskin, and M. Crane. Multiscaled cross-correlation dynamics in financial time-series. *Advances in Complex Systems*, 12(4-5):439–454, 2009. [arXiv:1001.0497v1].

[DFGZJ95] P. Di Francesco, P. Ginsparg, and J. Zinn-Justin. 2d gravity and random matrices. *Physics Reports*, 254:1–133, 1995. [arXiv:hep-th/9306153v2].

[DKRS01] S. Drożdż, F. Kwapień, J.and Grümmer, F. Ruf, and P. Speth. Quantifying dynamics of the financial correlations. *Physica* ***A***, 299:144, 2001. [arXiv:cond-mat/0102402].

[DKSW02] S. Drożdż, J. Kwapień, P. Speth, and M. Wójcik. Identifying complexity by means of matrices. *Physica* ***A***, 314:355–361, 2002. [arXiv:cond-mat/0112271].

[DNV92] K.J. Dykema, A. Nica, and D.V. Voiculescu. *Free Random Variables*. CRM Monograph Series. American Mathematical Society, Centre de Recherches Mathematiques, Providence, new edition, 12 1992. [book-info].

[Dys62] F.J. Dyson. The threefold way.algebraic structure of symmetry groups and ensembles in quantum mechanics. *Journal of Mathematical Physics*, 3(6):1119–1215, 1962. [DOI:10.1063/1.1703863].

[EB86] R.F. Engle and T. Bollerslev. Modelling the persistence of conditional variances. *Econometric Reviews*, 5:1, 1986. [DOI:10.1080/07474938608800095].

[EGBG06] E.J. Elton, M.J. Gruber, S.J. Brown, and W. Götzmann. *Modern Portfolio Theory and Investment Analysis*. Wiley, 7^{th} edition, 2006.

[Eng82] R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of variance of united kingdom inflation. *Econometrica*, 50:987, 1982. [JSTOR:1912773].

[ENV00] M. Emery, A. Nemirovski, and D.V. Voiculescu. *Lectures on probability theory and statistics. Lectures from the 28th Summer School on Probability Theory held in Saint-Flour, August 17–September 3, 1998, Edited by Pierre Bernard*, volume 1738 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2000. [book-info].

[ERR05] A Edelman and N. Raj Rao. Random matrix theory. *Acta Numerica*, 14:233–297, 2005. [e-print].

[Eve06] B.S. Everitt. *The Cambridge dictionary of statistics*. Cambridge Univ. Press, Cambridge, 3 edition, 2006. [DOI:10.2277/0521690277].

[Eyn00] B. Eynard. Random matrices. *Saclay-T01/014*, CRM-2708, 2000. lecture notes [e-print].

[Fel68] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume Vols. 1 and 2. John Wiley & Sons, Inc, 1968.

[FHLR00] M. Forni, M. Hallin, M. Lippi, and R. Reichlin. The generalized dynamic factor model:identification and estimation. *The Review of Economic and Statistics*, 82:540–554, 2000. [DOI:10.1162/003465300559037].

[FHLR04] M. Forni, M. Hallin, M. Lippi, and R. Reichlin. The generalized dynamic factor model: Consistency and rates. *Journal of Econometrics*, 119:231–255, 2004. [DOI:10.1016/S0304-4076(03)00196-9].

[FIS02] S.H. Friedberg, A.J. Insel, and L.E. Spence. *Linear Algebra*. Prentice Hall, 2002.

[Gew97] J. Geweke. The dynamic factor analysis of economic time series. In D.J. Aigner and A.S. Goldberger, editors, *Latent Variables in Social Economic Models*. North Holland, Amsterdam, 1997.

[GK03] T. Guhr and B. Kälber. A new method to estimate the noise in financial correlation matrices. *Journal of Physics* ***A***, 36:3009, 2003. [arXiv:cond-mat/0206577v1].

[GKP94] R. Graham, D. Knuth, and O. Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison–Wesley, 1994.

[GMG98] T. Guhr and A. Hans A. Weidenmüller Müller-Gröling. Random matrix theories in quantum physics: Common concepts. *Physics Reports*, 299:189–425, 1998. [arXiv:cond-mat/9707301v1].

[GN74] C.W.J Granger and P. Newbold. Spurious regressions in econometrics. *Journal of Econometrics*, 2(2):111–120, 1974. [DOI:10.1016/0304-4076(74)90034-7].

[GNJJN05] E. Gudowska-Nowak, R.A. Janik, J. Jurkiewicz, and M.A. Nowak. On diffusion of large matrices. *New Journal of Physics*, 7:54, 2005. [DOI:10.1088/1367-2630/7/1/054].

[Gra01] C.W.J. Granger. Macroeconometrics - past and future. *Journal of Econometrics*, 100(1):17 – 19, 2001. [DOI:10.1016/S0304-4076(00)00047-6].

[Gre03] W.F. Greene. *Econometric Analysis*. Prentice Hall, Upper Saddle River New Jersey, 5th edition edition, 2003. [book-info].

[Gui09] A. Guionnet. *Large random matrices: Lectures on macroscopic asymptotics. École d'Été des Probabilités de Saint-Flour XXXVI – 2006.* Lecture Notes in Mathematics 1957. Springer Verlag, Berlin, 2009. [e-print].

[HP00] F. Hiai and D. Petz. *The Semicircle Law, Free Random Variables and Entropy*, volume 77 of *Math. Surveys and Monogr.* AMS, 2000. [book-info].

[J+10] J. Jurkiewicz et al. *to appear*, 2010.

[JM03] J.P.Bouchaud and M.Potters. *Theory of financial risks : from statistical physics to risk management*. Cambridge University Press, 2 edition, Feb 2003. [book-info].

[Joh07] I.M. Johnstone. High dimensional statistical inference and random matrices. In *Proceedings of the International Congress of Mathematicians*, volume 1 of *ICM2006*. EMS Publishing House, 2007. [arXiv.org:math/0611589].

[Jol02] I.T. Jolliffe. *Principal Component Analysis*. Springer, second edition, 2002. [book-info].

[JW04] R.A. Janik and W. Wieczorek. Multiplying unitary random matrices — universality and spectral properties. *Journal of Physics* ***A***, 37:6521, 2004. [arXiv:math-ph/0312043].

[JWMH09] B. Jin, C. Wang, B. Miao, and M.N.L. Huang. Limiting spectral distribution of large–dimensional sample covariance matrices generated by varma. *Journal of Multivariate Analysis*, 100:2112, 2009. [DOI:10.1016/j.jmva.2009.06.011].

[KDS03] J. Kwapień, S. Drożdż, and P. Speth. Alternation of different fluctuation regimes in the stock market dynamics. *Physica* ***A***, 330:605–621, 2003. [arXiv:cond-mat/0306608v1].

[KOD06] J. Kwapień, P. Oświecimka, and S. Drożdż. The bulk of the stock market correlation matrix is not pure noise. *Physica* ***A***, 359:589–606, 2006. [arXiv:physics/0505074v2].

[KS06] J. Krzych and M. Snarska. Automatic trading agent. rmt based portfolio theory and portfolio selection. *Acta Physica Polonica B*, 37(11):3145, 2006. [arXiv:physics/0608293].

[LÖ5] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer Verlag, Berlin, 2005. [book-info].

[LCBP99] L. Laloux, P. Cizeau, J.P. Bouchaud, and M. Potters. Noise dressing of financial correlation matrices. *Phys. Rev. Lett.*, 83(7):1467–1470, Aug 1999. [arXiv:cond-mat/9810255].

[LCPB00] L. Laloux, P. Cizeau, M. Potters, and J.P. Bouchaud. Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, 3:391, 2000. [e-print].

[LM03] F. Lillo and R.N. Mantegna. Noise dressing of the correlation matrix of factor models. [arXiv:cond-mat/0305546], 2003.

[Mar52] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. [JSTOR:2975974].

[Meh04] M.L. Mehta. *Random Matrices, Volume 142, (Pure and Applied Mathematics)*. Academic Press, 3 edition, 2004. [book-info].

[MP67] V.A. Marcenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math. USSRSbornik*, 1(4):457–483, 1967. [DOI:10.1070/SM1967v001n04ABEH001994].

[MS04] Y. Malevergne and D. Sornette. Collective origin of the coexistence of apparent random matrix theory noise and of factors in large sample correlation matrices. *Physica A: Statistical Mechanics and its Applications*, 331(3-4):660 – 668, 2004. [arXiv:cond-mat/0210115].

[Noh00] J.D. Noh. A model for stock correlations for stock markets. *Phys. Rev. E*, 61:5981–5982, 2000. [arXiv:cond-mat/9912076v1].

[Now10] M.A. Nowak. Random matrix theory and applications. course provided for Jagellonian University International PhD Studies in Physics of Complex Systems-Kraków, February–June 2010.

[PBL05] M. Potters, J.P. Bouchaud, and L. Laloux. Financial applications of random matrix theory: Old laces and new pieces. *Acta Physica Polonica* ***B***, 36:2767, 2005. [arXiv:physics/0507111].

[PGR+99] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, and H.E. Stanley. Universal and nonuniversal properties of cross correlations in financial time series. *Phys. Rev. Lett.*, 83(7):1471–1474, Aug 1999. [arXiv:cond-mat/9902283].

[PGR+00] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, and H.E. Stanley. A random matrix theory approach to financial cross-correlations. *Physica A: Statistical Mechanics and its Applications*, 287(3-4):374 – 382, 2000. [DOI:10.1016/S0378-4371(00)00376-9].

[PPK04] S. Pafka, M. Potters, and I. Kondor. Exponential weighting and random–matrix–theory–based filtering of financial covariance matrices for portfolio optimization. [arXiv:cond-mat/0402573], 2004.

[PPNK05] G. Papp, S. Pafka, M.A. Nowak, and I. Kondor. Random matrix filtering in portfolio optimization. *Acta Physica Polonica* ***B***, 36:2757, 2005. [arXiv:physics/0509235].

[PTVF07] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 3 edition, 2007. [book-info].

[RKDO08] R. Rak, J. Kwapień, S. Drożdż, and P. Oświecimka. Cross-correlations in warsaw stock exchange. *Acta Physica Polonica* ***A***, 114(3):561–568, 2008. [arXiv:0803.0057v1].

[SB95] J.W. Silverstein and Z.D. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):175 – 192, 1995. [DOI:10.1006/jmva.1995.1051].

[SCSR04] S. Sharifi, M. Crane, A. Shamaie, and H. Ruskin. Random matrix theory for portfolio optimization: a stability approach. *Physica A: Statistical and Theoretical Physics*, 335(3-4):629 – 643, 2004. [DOI:10.1016/j.physa.2003.12.016].

[SDG09] J Speth, S Drożdż, and F. Grümmer. Complex systems: From nuclear physics to financial markets. Technical report, FZJ-IKP-TH-2009-31, 2009. [arXiv:0910.4348].

[SH09] P. Sieczka and J.A. Hołyst. Correlations in commodity markets. *Physica A*, 388:1621–1630, 2009. [arXiv:0803.3884].

[Sha64] W.F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19:425, 1964. [JSTOR:2977928].

[Shl05] J. Shlens. A tutorial on principal component analysis. Technical report, Systems Neurobiology Laboratory, Salk Insitute for Biological Studies, 2005. [DOI:10.1.1.115.3503].

[Sim80] C.A. Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980. [JSTOR:1912017].

[Smi02] L.I. Smith. A tutorial on principal components analysis. [e-print], 2002.

[Sna08] M. Snarska. Toy model for large non-symmetric random matrices. *Acta Physica Polonica A*, 114(3), 2008. [arXiv:1004.4522].

[Spe94] R. Speicher. Multiplicative functions on the lattice of non–crossing partitions and free convolution. *Math. Ann.*, 298:611, 1994. [DOI:10.1007/BF01459754].

[Spe09] R. Speicher. *to appear in* The Oxford Handbook of Random Matrix Theory, chapter Free Probability Theory. Oxford University Press, 2009. [arXiv:0911.0087].

[Sve07] J. Svensson. The asymptotic spectrum of the ewma covariance estimator. *Physica* ***A***, 385:621, 2007. [Science Direct:j.physa.2007.07.030].

[SW99] J.H. Stock and M.W Watson. Forecasting inflation. *Journal of Monetary Economics*, 44:293–335, 1999. [SSRN:155850].

[SW02a] J.H. Stock and M.W Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179, 2002. [JSTOR:3085839].

[SW02b] J.H. Stock and M.W Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20:147–162, 2002. [DOI:10.1198/073500102317351921].

[SW05] J.H. Stock and M.W. Watson. Implications of dynamical factor models for var analysis. working paper [NBER-eprint], 2005.

[TB07] S. Thurner and C. Biely. The eigenvalue spectrum of lagged correlation matrices. *Acta Physica Polonica* ***B***, 38:4111, 2007. [arXiv:physics/0609053].

[tH74] G. 't Hooft. A planar diagram theory for strong interactions. *Nuclear Physics* ***B***, 72:461, 1974. [DOI:10.1016/0550-3213(74)90154-0].

[UIO04] A. Utsugi, K. Ino, and M. Oshikaw. Random matrix theory analysis of cross correlations in financial markets. *Phys. Rev. E Stat Nonlin Soft Matter Phys.*, 70(2 Pt. 2):026110, 2004. [arXiv:cond-mat/0312643].

[URH07] K. Urbanowicz, P. Richmond, and J.A. Hołyst. Risk evaluation with enhanced covariance matrix. *Physica A*, 384:468–474, 2007. [arXiv:physics/0612059].

[Ver94] J. Verbaarschot. The spectrum of the dirac operator near zero virtuality for nc = 2 and chiral random matrix theory. *Nuclear Physics B*, 426(3):559 – 574, 1994. [arXiv:hep-th/9401092].

[Voi91] D. V. Voiculescu. Limit laws for random matrices and free products. *Invent. Math.*, 104:201, 1991. [DOI:10.1007/BF01245072].

[Voi97] D.V. Voiculescu, editor. *Free probability theory*, volume 12 of *Fields Institute Communications*. American Mathematical Society, Providence, RI, 1997. Papers from the Workshop on Random Matrices and Operator Algebra Free Products held during the Special Year on Operator Algebra at the Fields Institute for Research in Mathematical Sciences, Waterloo, ON, March 1995 [book-info].

[WG04] D. Wilcox and T. Gebbie. An analysis of cross-correlations in South African Market data. *Physica A*, 344(1-2):294 – 298, 2004. [arXiv:cond-mat/0402389v3].

[Wig55] E.P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *The Annals of Mathematics*, 62(3):548–564, 1955. [JSTOR:1970079].

[Wig58] E.P. Wigner. On the distribution of the roots of certain symmetric matrices. *The Annals of Mathematics*, 67(2):325–327, 1958. [JSTOR:1970008].

[Wis28] J. Wishart. The generalized product of moment distribution in samples from a multivariate normal population. *Biometrika*, 20 A(1-2):32–52, 1928. [DOI:10.1093/biomet/20A.1-2.32].

[Wol38] H. Wold. A study in the analysis of stationary time series. Technical report, Almquist and Wiksell, Uppsala, 1938. [DOI:99999990012807].

[Woo90] M. Woodford. Learning to believe in sunspots. *Econometrica*, 58(2):277–307, 1990. [DOI:10.2307/2938205].

[Zee96] A. Zee. Law of addition in random matrix theory. *Nucl. Phys.**B***, 474:726, 1996. [arXiv:cond-mat/9602146].

[Zir10] M.R. Zirnbauer. *to appear in* The Oxford Handbook of Random Matrix Theory, chapter Symmetry classes. Oxford University Press, 2010. [arXiv:1001.0722v1].

[Zum09] G. Zumbach. The empirical properties of large covariance matrices. *The RiskMetrics Journal*, 9:31, 2009. [SSRN:1346412].

Index