# Factor Models in Finance

Zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für Wirtschaftswissenschaften
der Universität Fridericiana zu Karlsruhe

Genehmigte

DISSERTATION

von
Ing Sergio Focardi

| | |
|---|---|
| Tag der mündlichen Prüfung: | 26 June 2009 |
| Referent: | Prof. Dr. S.T. Rachev |
| Koreferent: | Prof. Dr. Andreas Geyer-Schulz |

Karlsruhe (2009)

# Acknowledgments

I want to thank my Supervisor, Prof. Dr. Svetlozar T. Rachev, for his continuous support and for sharing his insights with me. I consider it a great privilege to have had the opportunity to study under the guidance of one of the world's leading authorities in mathematical finance and financial econometrics.

I also take this occasion to renew my thanks to Prof Rachev for having previously offered me the opportunity to collaborate in the writing of the book *Financial Econometrics: From Basics to Advanced Modeling Techniques* . Collaborating with Prof Rachev was both a pleasure and a challenge.

I also want to thank Prof. Dr. Andreas Geyer-Schulz for the rigor with which he reviewed my Dissertation and for the many important suggestions he kindly made.

Finally, I want to thank Prof. Dr. Frank Fabozzi for providing many valuable feedbacks during the course of this work.

# Extended Abstract

I will succinctly state the problems that motivate this dissertation and outline the key findings. Though widely used in many disciplines ranging from psychometrics to economics and finance, there are two reasons why classical factor analysis cannot be correctly applied to large panels of data such as econometric time series or financial returns. First, for large samples such as the Russell 1000 it is very onerous to carry out the computation of maximum likelihood of classical factor analysis estimation. Second, the assumption of uncorrelated residuals is not empirically tenable.

Allowing correlated residuals poses several problems. In fact, consider a factor model with a finite number of observed variables $N$ and a possibly infinite number of observations $x = \beta f + \varepsilon$ where $f$ denotes common factors and $\varepsilon$ denotes possibly correlated residuals. The fact that the residuals are possibly correlated implies that residuals themselves can have a factor structure $\varepsilon = \chi g + \eta$ . As a consequence, there is the need for criteria to determine factors and to separate common factors from residual factors.

It proved to be very difficult to establish theoretically sound and empirically meaningful criteria to separate common factors from correlated residuals. The literature on factor models has taken a different path, proposing the paradigm of approximate factor models that are infinite in both the number $N$ of observed variables and the number $T$ of observations. In approximate factor models, the separation between common factors and residuals is achieved in a natural way by assuming that the eigenvalues of the covariance matrix corresponding to the common factors diverge while the remaining eigenvalues remain bounded. Under these assumptions, factors are unique (up to a rotation) and can be estimated with principal components.

In the literature, it is generally assumed that large samples including stock returns and macroeconomic variables can be analyzed with approximate factor models. The practice of asset management, however, has taken a different approach in applying factor models, proposing a number of factor models constructed with criteria based on identifying fundamental parameters, sectors, countries, or exogenous variables. Considerable effort is currently dedicated to identifying proprietary factors that might differentiate one portofolio or asset management process from another. Events following the market turmoil of the summer of 2007 heightened the attention to the problem of differentiating portfolios and strategies to reduce the risk of catastrophic events. (See Fabozzi, Focardi, Jonas (2008) for a recent study from the perspective of the quantitative asset manager.)

The problem under discussion can be stated as follows. If the approximate factor model paradigm is generally applicable in finance and asset management, then efforts to identify proprietary factors are futile: factor models are unique and factors can be determined with principal components. If, on the other hand, the approximate factor model paradigm is not always applicable, we need 1) criteria to determine when the paradigm of approximate factor models is indeed applicable and 2) criteria to select factor models created with methods different from both factor analysis and principal components analysis (PCA).

The theoretical and practical solution I propose in this dissertation and that I discuss in Section 4 can be summarized in the following six points:

1. The theoretical paradigm of approximate factor models cannot be applied blindly to large samples. There must be a neat and natural separation between "large" and "small" eigenvalues of the covariance matrix of data. In cases in which the eigenvalues of the covariance matrix of data decay smoothly, approximate factor models cannot be applied.
2. Factor models must be "learnable", that is, the number of model parameters (and thus of factors) cannot exceed the limits that can be learned given the size of the sample. Empirical samples might be too small to allow the estimation of all the true factors of the relative population.
3. Criteria to separate "large" and "small" eigenvalues are largely arbitrary. I propose to sidestep the problem by defining directly criteria that determine the ability to approximately identify factors and principal components.
4. When factors cannot be approximately identified with principal components, I propose to look at factor models as multiple communication channels and to use the channel capacity, that is, the average mutual information, as criterion for selecting among the different models.
5. As regards financial returns, I found that in large panels of returns the eigenvalues of the covariance matrix of returns decay smoothly and the conditions for identification of factors and principal components do not apply. Returns cannot be faithfully represented with unique approximate static factor models.
6. I try to prove that the inability to find unique static factor models of returns might be due to the presence of dynamic factors in models of asset returns and, especially, of cointegration of price processes. I try to prove that, if returns can be represented by dynamic factor models or prices can be represented by cointegration-based models, the distribution of eigenvalues of the covariance matrix is likely to exhibit a smooth decay. In other words, dynamic effects of returns and mean reversion of prices are the reasons why static factor models are not unique and give only a partial representation of the correlation structure.

# Contents

**1. STATEMENT OF THE PROBLEM AND CONTRIBUTION OF THIS DISSERTATION................................7**

1.1. Practical and theoretical implications of approximate factor models of returns....................................................7
1.2. A critical look at the assumptions needed to ensure that approximate factor models of returns hold..................9
1.3. What changes if we assume that models are dynamical and not static..................................................................11
1.4 The contribution of this Dissertation..........................................................................................................................12
*1.4.1 Static factor models of stationary processes (returns)....................................................................................13*
*1.4.2 Dynamic factor models of stationary processes (returns)...............................................................................15*
*1.4.3 Dynamic factor models of integrated processes (prices)................................................................................16*

**2. SURVEY OF THE LITERATURE..............................................................................................................19**

2.1 Random Matrix Theory.............................................................................................................................................19
*2.1.1. Gaussian matrices.............................................................................................................................................20*
*2.1.2. Wigner matrices...............................................................................................................................................21*
*2.1.3. Wishart matrices..............................................................................................................................................21*
2.2. Asymptotic distribution of the bulk of eigenvalues...............................................................................................22
2.3. Asymptotic distribution of the largest eigenvalues................................................................................................24
2.4 Asymmetric matrices.................................................................................................................................................29
2.5. Static Factor Models.................................................................................................................................................30
2.6. Estimation of static factor models...........................................................................................................................34
2.7. Application of RMT to static factor models............................................................................................................36
2.8. Dynamic Factor Models............................................................................................................................................38
2.9 Dynamic factor models of integrated processes........................................................................................................41
2.10. Application of RMT to dynamic factor models......................................................................................................42
2.11. Application of information theory to factor models..............................................................................................43
2.12. Sparse Principal Components Analysis...................................................................................................................44

**3. MATHEMATICAL METHODS..................................................................................................................46**

3.1 Classical methods of factor analysis..........................................................................................................................46
3.2 Factor uniqueness.......................................................................................................................................................48
3.3 Spectral analysis.........................................................................................................................................................50
3.4 Principal components................................................................................................................................................57
3.5 Dynamic principal components................................................................................................................................59
3.6 Sparse principal components.....................................................................................................................................59
3.7 Independent Components..........................................................................................................................................60
3.8 Random Matrix Models (RMM)...............................................................................................................................61
3.9 Transforms..................................................................................................................................................................64
*3.9.1 The Stieltjes transform.......................................................................................................................................65*
*3.9.2 The η-transform.................................................................................................................................................65*
*3.9.3 The R-transforms...............................................................................................................................................66*
3.10 Free Probability........................................................................................................................................................66
*3.10.1 Algebraic probability theory............................................................................................................................67*
*3.10.2 Freeness............................................................................................................................................................69*
*3.10.3 Asymptotic freeness of random matrices........................................................................................................70*
3.11 The method of the Resolvent...................................................................................................................................71

**4. FORECASTING WITH FACTORS: NEW RESULTS................................................................................74**

4.1 Key findings.................................................................................................................................................................74
4.2 Conditions for applying the paradigm of approximate factor models.....................................................................75
*4.2.1 Pitfalls in determining the number of factors..................................................................................................77*
*4.2.2 Pitfalls in estimating factors..............................................................................................................................80*
*4.2.3 Monte Carlo simulation of approximate factor models....................................................................................81*
*4.2.4 Simulation results for confidence bands...........................................................................................................84*
4.3 When can factors be estimated with principal components?....................................................................................88
4.4 Factor models as noisy multiple communication channels........................................................................................97
*4.4.1 Channel Capacity...............................................................................................................................................98*
*4.4.2 Criteria for model choice...................................................................................................................................99*
4. 5. Empirical results relative to the Russell 1000.......................................................................................................100

4. 6. Dynamic factor models of returns....106
4. 7. Dynamic factor models of prices....108
*4.7.1 Empirical evidence of cointegration in equity prices....109*
*4.7.2 Cointegration and correlation....111*

**SUMMARY AND FUTURE DIRECTION OF RESEARCH....114**

**NOTATION....116**

**REFERENCES....117**

**STATUTORY DECLARATION....125**

# 1. Statement of the problem and contribution of this dissertation

## *1.1. Practical and theoretical implications of approximate factor models of returns*

Factor models are fundamental tools in finance at both the theoretical and the practical levels. To mention a few theoretical applications, factor models embody the capital asset pricing model (CAPM), the Arbitrage Pricing Theory (APT) of Ross (1976), and various fund-separation theorems. In the practice of asset management, factor models are used, for example, to construct portfolios, to assess risk exposures to "factors", to forecast returns, and to measure the performance of asset managers. The term structure of interest rates can also be parsimoniously described through factor models. It is fair to say that factor models are important tools in asset management, especially equity portfolio management, where they are used to build stock ranking systems or to run (quasi) automated portfolios of equities.

Therefore, understanding whether or not asset returns can be represented through factor models (either in a static or dynamic form), and, if they can, estimating the model is of high theoretical and practical importance. Indeed, many efforts have been devoted to this endeavour. Determining the number of static and/or dynamic factors of equity returns has been the subject of extensive research. While the CAPM is a one-factor static model, the APT of Ross (1976) is a theory that makes use of a multifactor static model where the number of factors is left unspecified. Fama and French (1993) proposed what is now a widely used fundamental three-factor model. Other fundamental models with multiple factors have been proposed while a variety of sector-country models exhibit a number of factors that reflect sector and country segmentation. Commercial suppliers such as MSCI Barra offer models based on a large number of factors. Many of these models are often employed in a dynamic context, using lagged factors in order to allow returns forecasting. See Fabozzi, Focardi, and Kolm (2006) for a review of current approaches to factor models.

From a theoretical point of view, the state-of-the-art approach to factor models in financial econometrics ─as well as in modern macroeconomics─ is the approximate factor model of Chamberlain and Rothshild (1983) and its recent static and dynamic extensions by several researchers (see Chapter 2 for a detailed review of the literature on factor models). I will use the term "approximate factor models" to indicate both the original model proposed in Chamberlain and Rothshild (1983) and its generalization unless the differences between the models need to be made explicit. Approximate static factor models differ from classical static factor models in three principal ways:

- In approximate factor models, both the number of observations and the number of time series tend to infinity while classical factor models assume a finite number of time series (assets).
- In approximate factor models, both the model parameters and the factors are uniquely identified (factors only up to a rotation) while in classical factor models, the model parameters are identifiable but factors are undetermined (in addition to rotation).
- In approximate factor models, factors can be identified (up to a rotation) with principal components; this implies that factors can be estimated with portfolios of returns. In classical factor models, factors are not necessarily portfolios.

There are strong theoretical reasons in favour of the adoption of the approximate factor model paradigm. First, because they assume uncorrelated residuals, classical factor models are too restrictive for applications in financial econometrics. However, dropping the assumption of uncorrelated residuals in finite models creates the theoretical problem of how to apportion correlations between factors and residuals. In fact, if we allow residuals to be correlated in a finite model, factor models become the sum of two factor models because residuals can also be represented as a factor model. In order to distinguish between common factors and the factors of residuals, we have to introduce special conditions and/or quantitative thresholds.

Infinite models allow one to assume correlated residuals in a natural way as common factors are associated with those eigenvalues of the covariance matrix which diverge when the number of time series tends to infinity (see Chapter 2). In financial econometrics, models that are infinite in both the number of time series and the number of observations are required to properly state the arbitrage pricing theory *in a static form*. In fact, the APT theorem of Ross (1976) could not have been established in finite markets described by static factor models. Chamberlain and Rothshild developed infinite approximate factor models to prove that the APT theorem still holds in the approximate factor model framework, thereby avoiding the limitations of the strict factor models with uncorrelated residuals as used by Ross. There are therefore strong theoretical reasons behind the adoption of models that are infinite in both the number of time series $N$ and the number of observations $T$.

In addition, there are practical motivations in favour of approximate factor models. Classical factor analysis based on maximum likelihood is a complicated, numerically delicate process while the ability to estimate factors with principal components makes approximate factor models a more robust methodology. In fact, the extraction of principal components can be performed with robust techniques such as singular value decomposition. Classical factor models can be estimated with maximum likelihood when the number of variables is at most of the order of a few tens. As financial markets are much larger (it is possible to build matrices in the range of hundreds of observations of hundreds of stock returns), classical factor models cannot be realistically estimated. In macroeconomics, hundreds of variables can now be observed with the consequent need for dimensionality reduction. But again factors cannot in practice be estimated with maximum likelihood.

However, if approximate factor models do indeed represent returns, there are very important consequences for both the theory and practice of asset management. In fact, the current proliferation of static models would be illusory as it would be possible to identify, albeit up to a rotation, a unique, true factor model. In addition, all factors could be represented as portfolios of returns, implying that, at least theoretically, no information outside the history of returns would be needed to determine factors. Hence, it is interesting at both the theoretical and practical levels to determine if returns can be represented with static approximate factor models.

If this conclusion is correct, the search for factors would consist in determining principal components of returns; equity portfolio management would become an almost exact science and portfolios would be constructed with well-defined optimization rules, creating highly homogeneous portfolios. As Fabozzi, Focardi, and Jonas (2008) discuss, the market turmoil of the summer of 2007 raised the concern, among asset managers, that the uniformity of factors and portfolios creates large unforeseeable risk.

### *1.2. A critical look at the assumptions needed to ensure that approximate factor models of returns hold*

When analyzing the factor structure of returns, the following questions must be addressed:

1. Can approximate factor models represent returns?
2. Can factors be uniquely determined by principal components following the paradigm of approximate factor models?
3. Given that learning theory establishes that there are trade-offs that constrain model complexity in function of the size of the sample, how can we ensure that the true number of factors is learnable?
4. If factors are not unique, what are the model selection criteria appropriate for factor models of returns?

The basis for applying generalized approximate factor models is that markets are large and therefore asymptotic results approximate empirical results. In standard statistics, asymptotic results are applied to large samples because samples are considered independent samples from the same distribution. Under this assumption, various central limit theorems and results on the asymptotic distribution of estimators prove that it makes sense to consider asymptotic results. But in the case of generalized factor models, samples of increasing size are samples from different populations and finite samples might have characteristics that in no way approximate infinite samples. In fact, approximate factor models depend critically on separating a finite number of eigenvalues that grow indefinitely from an infinite number of eigenvalues that remain bounded.

In many cases, the actual samples are the largest possible samples; it does not therefore make sense to speculate on how a much larger sample would look. For example, how can the market for equity returns grow to infinity? Adding industrial sectors? Adding countries? Adding stocks but keeping the number of sectors and countries fixed? Or adding stocks, sectors, and countries at the same time? Clearly any assumption is arbitrary. However, different assumptions would produce different infinite factor models. Therefore, applying asymptotic results from generalized approximate factor model theory implies additional assumptions and criteria on how the current sample might evolve to infinity.

There is another important consideration. The number of factors determined by estimation criteria is supposed to reveal the true number of factors of the population. In practice, we have only a finite sample where the ratio between the number of observations and the number of variables is close to 1. Learning theory tells us that model selection criteria from a sample must include a trade-off between model complexity and in-sample performance. In fact, a fundamental tenet of modern learning theory in its many forms up to the Vapnik-Chervonenkis theory of statistical learning (see Vapnik, 1998) is that there is a trade-off between model complexity (which improves in-sample performance) and model generalization ability (out-of-sample performance) for each sample size.

In practice, it might happen that this trade-off is optimal for a number of factors that is inferior to the true number of factors of the population. In other words, given a finite sample extracted from a population, we might not be able to identify the population factor structure because the sample is insufficient to estimate a factor model with the level of complexity required by the true number of factors. One has therefore to assume that the population model is simple enough to be learnable from available empirical data.

All criteria introduced in the literature to assess if a universe of returns possesses a factor structure and to determine the eventual number of factors are *asymptotic criteria*. The criteria proposed thus far to determine the number of factors work only asymptotically due to the fact that the eigenvalues relative to the factors grow without limit. That is, criteria pick the right number of factors when both the number of asset returns and the number of observations tend to infinity essentially because, loosely speaking, the contribution of non factor eigenvalues goes to zero. However, the same criteria are not designed to implement a trade-off in finite samples. This holds true both for criteria based on information theory and for criteria based on random matrix theory (see Chapter 4 for a detailed discussion).

### *1.3. What changes if we assume that models are dynamic and not static*

Thus far I have stated the problem of determining the static factor structure of returns. I will next discuss whether or not the adoption of a dynamic factor framework for returns will allow the identification of  the number of factors. As concisely stated in Amengual and Watson (2006), identification of static factors is obtained by fitting the covariance matrix of observed variables while identification of dynamic factors is obtained by fitting the spectral density matrix of the observed variables.

The theory of dynamic factor models has many parallels with the theory of static factor models. In particular, identifying the number of common factors requires identifying a clear-cut criterion that separates the common factors from other factors. For dynamic factor models, one such criterion is given by the condition of the presence/absence of autocorrelation of residuals, while for static factor models the key criterion is the presence/absence of crosscorrelations of residuals. In fact, if we assume that residuals are not autocorrelated, as described in Peña and Box (1987), the autocovariance matrices at different lags exhibit the same rank, which is equal to the number of dynamic factors. Therefore if we assume that residuals are not autocorrelated, identification of the number of *dynamic* factors can be achieved exactly in finite samples. However, if residuals are allowed to be autocorrelated, then the identification of factors can be achieved only in infinite markets (see Chapter 2 for a review of the literature).

It might therefore seem that there is not much difference between static and factor models. However, there is a major difference empirically: the assumption of no autocorrelation of residuals of returns is more reasonable than the assumption of no correlation of residuals. Returns are cross autocorrelated but only weakly autocorrelated. It is therefore reasonable to assume that dynamic effects are due to common lagged factors.

Again, this finding has important implications both for the theory and practice of asset pricing. It suggests that discussions on the number of factors needed to describe returns must take into account dynamic phenomena. There are theoretical reasons that support this view. In fact, asset pricing theories are dynamic theories such as the Merton intertemporal model. It is ultimately not surprising that returns cannot be uniquely described by a static factor model.

One of the reasons why returns cannot be uniquely decomposed into factors is due to cointegration, a fact already implied by the Merton model. Johansen and Lando (2006) prove that arbitrage restrictions in a multiperiod model induce cointegration. Because of cointegration, the covariance matrix exhibits a slow decay of the eigenvalues. This brings us to the last problem discussed in this Dissertation: Are prices a better choice than returns for factor models? That is, this Dissertation discusses whether or not dynamic factor models of *prices*, which imply cointegration, are more stable, more identifiable, and offer better forecasting capability than factor models of *returns*.

As prices are integrated processes, factor models of prices are based on different estimation principles than factor models of returns. To my knowledge, there is no "large model" theory of integrated factor models. The theory of factor models of integrated processes can be found in Stock and Watson (1988), Plilips and Ouliaris (1988), and Peña and Poncela (2004a,b). This theory is applied to models with a finite number of variables.

The determination of the number of common factors in models of integrated processes is indeed based on a clear-cut criterion. In fact, the distinction between common and idiosyncratic factors is based on the distinction between integrated processes and stationary processes. The question we have to answer is: Are finite dynamic factor models of prices a better choice than finite dynamic factor models of returns in terms of their ability to generalize, that is, in terms of out-of-sample performance? If factor models of prices apply, what are the consequences for factor models of returns?

### *1.4 The contributions of this Dissertation*

The contributions of this Dissertation are to: 1) show that factor models of asset returns are not unique and do not fit the approximate factor model paradigm, 2) introduce criteria for factor model selection and, 3) try to prove that if asset prices can be represented through dynamic factor models with integrated factors, factor models of returns cannot be represented through the approximate factor model paradigm.

The discussion is divided into three parts: 1) the discussion of *static* factor models of stationary processes (returns), 2) the discussion of *dynamic* factor models of stationary processes (returns) and, 3) the discussion of dynamic factor models of integrated processes (prices). I will first consider static factor models of stationary processes.

### *1.4.1 Static factor models of stationary processes (returns)*

I first observe that if the paradigm of approximate factor models is adopted, the factors and the model's parameters suffer from an essential indeterminacy insofar as one needs to make assumptions as regards the way the model evolves to infinity. In particular, one needs to make assumptions not only as regards the function $N=N(T)$ but also as regards the behaviour of the eigenvalues of the covariance matrix of the variables. The latter assumptions, in practice, determine the entire model. In particular one needs to decide the number of eigenvalues that will diverge and the number of eigenvalues that will remain bounded. This dichotomization cannot be performed from a finite sample without making specific assumptions. As observed above, asymptotic results relative to approximate factor models imply assumptions as to how the market will evolve to infinity. These assumptions are arbitrary.

The perspective of this Dissertation is that the paradigm of approximate factor models cannot be applied blindly without making specific assumptions on how the model will evolve to infinity. Assuming that the number of observed time series goes to infinity implies specific assumptions on the "physical" or "economic" structure of the phenomena under study. We observed that an infinite market for returns implies economic assumptions on the structure of sectors and of countries when the number of returns increases. However, similar considerations can be made for macroeconomic variables. What is the meaning of an economy with an infinite number of observed variables? What is observed? Is it the same phenomena from slightly different points of view? Or, again, is it a different economy with an infinite number of different industrial sectors?

To substantiate these claims, I first show through a Monte Carlo study that in generalized approximate factor models such as the Bai and Ng model, the theoretical asymptotic distribution of factor estimators can be grossly violated if the signal-to-noise ratio is too small ─ even if the sample has the size of a problem typical in financial econometrics. Therefore the assumption that we can apply the asymptotic conclusions to the large samples of financial econometrics is not warranted.

The economic perspective of my Dissertation is that financial markets are dynamic phenomena that can only be approximately represented by static factor models. The first step is to define the limits of applicability of (generalized) approximate factor models. Approximate factor models might not be a sound theory to apply to finite samples even if samples are large. Therefore, I propose to first determine if samples support the essential features of approximate factor models which I state as follows:

1. It must be possible to represent factors with principal components.
2. Factors must be approximately unique.
3. Factors must be learnable.

Uniqueness of factors means that we should be able to define a “distance” between factors and that the distance between principal components and factors should be small. As approximate factor models are estimated with asymptotic principal components analysis, in order to assess the applicability of the approximate factor model paradigm I propose that factor analysis should start with the assessment of the generalization ability of PCA. This problem has been extensively researched in the pattern recognition literature where PCA is widely used in tasks such as speech and face recognition. I propose to apply the standard Akaike Information Criterion (AIC) or Schwarz’s Bayesian Information Criterion (BIC) criteria to choose the optimal number of principal components. Optimality means the best compromise between accuracy and generalization capabilities, i.e., out-of-sample performance.

Next I determine the conditions under which the AIC and BIC choose a number of principal components corresponding to the minimum of the noise-to-signal ratio. Building on results from Schneeweiss and Mathes (1995), I show that, if the signal-to-noise ratio is particularly high for a given number $q$ of principal components, the factors of any other factor model with $q$ factors will be close to the principal components. I prove this result using two complementary measures of factor distance that are widely used in the literature. The two measures are 1) the sum of the squares of the canonical correlation coefficients obtained performing the canonical correlations analysis between factors and principal components and 2) the solution of the Procrustes problem between factors and principal components. In this case, it is shown in Section 4.3 that the asymptotic results from approximate factor theory approximately hold.

Therefore, the first result is that if the signal-to-noise ratio assumes a particularly high value for a given number $q$ of principal components, and if the AIC and BIC choose the same number of principal components, then if data can be represented by a factor model with $q$ factors, the factors and the first $q$ principal components are very close. If the noise-to-signal ratio always assumes a high value, then different factor models are able to represent the data. The intuition of this fact is that for the number of principal components chosen by the AIC or the BIC, there are possibly other principal components that are not close to the first $q$ principal components but that will nevertheless have a similar noise-to-signal ratio.

In the above situation, I propose to choose models maximizing the mutual information from factors to observed variables. Borrowing concepts from communication theory, the general perspective of my Dissertation is to look at factor models as multiple communication channels subject to noise, where a small number of emitters (the factors) broadcast to a large number of receivers (stock returns or observed variables) subject to random disturbance. Using this analogy, maximizing mutual information corresponds to maximizing the capacity of the channel (see Tulino and Verdù, 2004 and Silverstein and Tulino, 2006).

In Section 4.5, I show that the covariance matrix of a large universe of returns such as the Russell 1000 indeed shows a smooth decay of the eigenvalues. I therefore deduce that there is no solid empirical basis for applying the paradigm of approximate factor models to financial returns. This result is justifiable on the basis of dynamic asset pricing theories which make it unlikely that returns exhibit a stable covariance matrix of returns.

### *1.4.2 Dynamic factor models of stationary processes (returns)*

I will now discuss dynamic factor models, which are widely used in equity portfolio management to make forecasts. The general framework is the following. Static factor models of stationary processes are models where factors and observed variables are considered simultaneously, while dynamic factor models also consider lagged factors. In general, each dynamic factor model can be cast in a static factor model form.

Given a static factor model representation, it is important to disentangle lagged factors from same-time factors. Different procedures have been proposed, among which the canonical correlation analysis of factors and the principal components analysis of the spectral matrix of the factors. Peña and Box (1987) propose a methodology for estimating the number of factors and the factors themselves that work in finite models (see Chapter 4 for a detailed discussion). According to this methodology, dynamic factors can be uniquely determined in factor models even if static factors cannot be uniquely determined. The identification of dynamic factors is possible if residuals are not serially correlated. Under this assumption, the autocovariance matrices all have the same rank equal to the number of factors and the same eigenvalues.

In practice, the autocovariance matrices are too large to be estimated directly. Autocovariance matrices are subject to the same estimation problems of covariance matrices: there are too many parameters to be estimated. However, one is interested in estimating the rank of the matrices and not the entire matrix. I can summarize the conclusions of my discussion as follows.

If the sample has a structure of eigenvalues that supports identification of static factors with principal components, then dynamic factors can be identified and estimated with one of the current procedures from static principal components. However, if the sample has a structure of eigenvalues that does not support the identification of static factors with principal components, then model estimation is more difficult.

In this latter case, one could first determine the number of dynamic factors using the methods in Peña and Box (1987). Observe that this methodology determines the number of common dynamic factors. However the residuals might still have a factor structure. The factor structure of the residuals might be predominant in explaining cross correlations, both in terms of the number of factors and the strength of correlations. In other words, a sample of time series might admit a representation where there are purely static factors that explain most of the covariance between time series plus a number of dynamic factors that are responsible for both additional cross correlation plus autocorrelations.

In practical applications there are two different objectives: 1) to determine the exposure to risk factors and 2) to determine factors that might allow to forecast. Factors required for the two objectives are not necessarily the same. If the principal components model were applicable and learnable, then one would get a complete picture of factor exposure and factor forecasting. However, in financial markets the principal components model cannot really be applied to returns without being required to estimate a number of components that are ultimately not learnable.

Let's revisit criteria for model selection. I suggested above that mutual information (or communications channel capacity) might be employed as a model selection criterion. If applied to the static factor representation, this criterion does not distinguish between static and dynamic factors. Current methods to estimate dynamic factors from static factors ─ for example Forni, Lippi, and Reichlin (2005) and Breitung and Kretschmer (2005) ─ work in the infinite model case where factors are unique. These criteria assume that all static factors are formed by copies of the dynamic factors at different lags.

However in finite models the above assumption is too restrictive. In practice, in finite models where factors cannot be represented as principal components, one finds a number of factors that are essentially static and a number of factors that might appear at different lags. In practical applications in asset management, one often considers all factors at different lags. However, given the need of a parsimonious representation, it is important to select those factors that have maximum forecasting power. Therefore, given a factor model, for example a sector factor model, this Dissertation contributes criteria to separate a reduced number of factors with forecasting power from a larger number of essentially static factors.

### *1.4.3 Dynamic factor models of integrated processes (prices)*

I will now discuss factor models of prices. Observe that from a purely theoretical point of view, asset pricing models are primarily models of prices. For example, Merton's Intertemporal

Capital Asset Pricing Model (Merton, 1973) is established in terms of prices. This is reasonable in that prices are the long memory of returns and reflect accumulated wealth.

Assuming that prices are integrated processes, I(1), the representation of prices with factor models hinges on the presence of cointegration among price processes. A fundamental result in Stock and Watson (1988) establishes that if there are *K* cointegrating relationships in *N* integrated time series, then there are exactly *N-K* common trends which are integrated processes. Each of the *N* time series can be represented as a linear combination of the common trends plus a stationary process. Common trends can be established with PCA applied to price processes. Peña and Poncela (2004a,b) establish a general theory of factor models where factors can be any mix of integrated and stationary processes. Estimation of the number of factors and of factors themselves hinges on the estimation of generalized autocovariances.

Observe that the theory of factor models for integrated processes is a finite model theory. The distinction between factors and non factors is based on a clear-cut distinction between integrated processes and stationary processes. From the population point of view, there is no identification issue. Therefore, besides estimation problems, there is no ambiguity in deciding the number of factors and there is no need to consider asymptotic results with an infinite number of time series.

If we can represent asset prices with a factor model, then the covariance matrix of returns shows a slow decay of the eigenvalues. This is due to the cointegrating relationships that bind prices to factors. *This Dissertation seeks to demonstrate that if asset prices can be represented with a factor model with K integrated factors, then returns cannot be represented with principal components because cointegration will introduce an additional factor structure of returns that will result in a covariance matrix with a slow decay of eigenvalues. This is the major technical contribution of this Dissertation. The practical implication for asset management is that any factor model of returns can be mapped to the long-term factor model of returns implied by the factor model of prices. Deviations from this core model of returns are due to the mean-reverting behaviour of prices that tend to revert to factors. This fact creates instability in return factors*. Table 1.1 summarizes the key facts that I will attempt to prove.

|  | Static factor models of stationary processes | Dynamic factor models of stationary processes | Dynamic factor models of integrated processes |
|---|---|---|---|
| The signal-to-noise ratio is very high for the same number of components chosen by AIC or BIC. PCA has good generalization capabilities | Approximate factor models apply to returns; factor models are similar to principal components | Standard methods to disentangle dynamic from static factors apply | Dynamic factor models of prices do not apply |
| Eigenvalues decay smoothly; the signal-to-noise ratio remains close to 1 | Approximate factor models do not apply to returns; factor models are not similar to principal components; different factor models coexist | Coexistence of static and dynamic factors | Dynamic factor models of prices apply |

*Tab 1.1 Summary of key facts of factor models.*

## 2. Survey of the literature

In this section I survey the literature on the topics that I discuss in this Dissertation. I will do so by topics. I first survey the results from random matrix theory (RMT) that are relevant for factor models. RMT offers tools to discriminate, at various levels, meaningful information from random noise. Second, I discuss static factor models and, third, I survey estimation methods for static factor models including RMT and Information Theory.

I identify three stages in the literature on static factor models. In this Dissertation I call the number of observations *T* and the number of time series *N.* The first stage is the definition of the classical, strict factor model; the second stage includes its generalization to "large *T*, large *N*" strict and approximate factor models in the a-temporal setting of independent samples; the third stage includes the study of "large *T*, large *N*" static approximate factor models. The latter are compatible with a time series dynamics though they do not explicitly include such dynamics. For the sake of clarity, the literature on dynamic factor models is surveyed as a separate topic as dynamic factor models were developed in parallel with static factor models. I also survey separately the literature on the application of Information Theory to the analysis of data sets and factor models, and the literature on sparse principal components analysis, a technique that constrains the number of non-zero coefficients in principal components.

### *2.1 Random Matrix Theory*

Random matrices are matrix-variate random variables. Random matrix theory (RMT) was originally developed in the 1920s in biometrics and general multivariate statistics, to respond to specific application needs. RMT became a key tool in quantum physics in the 1950s and is now applied to many fields of science, from quantum mechanics, statistical physics, and wireless communications to number theory and financial econometrics. RMT milestones with a bearing on financial econometrics can be summarized as follows.

- In the late 1920s, John Wishart introduced the concept of a matrix-variate random variable and discovered the Wishart distribution as a generalization of the Chi-square distributions to matrix-variate random variables (see Wishart, 1928).
- In the 1950s, Eugene Wigner, co-recipient of the 1963 Nobel Prize in Physics, developed the theory of square matrices with random entries, known today as Wigner matrices. Wigner also discovered the first asymptotic limit law for the distribution of eigenvalues and the spacing between eigenvalues of a random matrix.

- In the 1960s, V. A. Marčenko and L. A. Pastur discovered the limit distribution of the eigenvalues of a Wishart matrix (see Marčenko and Pastur, 1967).
- From the 1980s onward, a series of papers fully characterized the behaviour of the hedges of the distribution of the eigenvalues of a random matrix, both in the null hypothesis of no correlation and in a number of cases beyond the null. In addition, the distribution of eigenvalues of asymmetric matrices was determined.

I will briefly sketch RMT before surveying recent results with a bearing on factor models.[1] As already observed, random matrices are matrix-variate random variables. More precisely, a random matrix model (RMM) is a probability space $(\Omega, P, F)$ where the sample space is a set of matrices. Tracy and Widom (2008) survey the principal classic RMMs and the distributions associated with them.

I will consider three types of matrices: Gaussian matrices, Wigner matrices, and Wishart matrices. Based on Tulino and Verdù (2004), I will describe three important distributions related to random matrices (if applicable): the pdf of a random matrix as a probabilistic object, i.e., the pdf related to the probability *P* of the RMM; the joint distribution of its eigenvalues; and the marginal distribution of the eigenvalues. The asymptotic distribution of eigenvalues ─ a central object of RMT ─ will be treated separately.

### *2.1.1. Gaussian matrices*

A standard real/complex Gaussian matrix is an $m \times n$ matrix $H$ whose entries are independent and identically distributed (i.i.d.) normal variables $h_{ij}$ with the same variance $\sigma^2 = {}^1\!/_m$: The pdf $p(H)$ of a matrix $H$ is the probability that the matrix elements are in the infinitesimal volume $\prod dh_{ij}$. The pdf of a standard complex Gaussian matrix is given by the following expression: $p(H) = \left(\pi\sigma^2\right)^{-mn} \exp\left[-\frac{trace(HH^*)}{\sigma^2}\right] = \left(\pi\sigma^2\right)^{-mn} \exp\left[-\frac{\|H\|_F^2}{\sigma^2}\right]$

where $\|H\|_F^2$ is the square of the Frobenius norm of the matrix *H***.** The Frobenius norm of a matrix $A$ is defined as: $\|A\|_F = \sqrt{trace(A^*A)} = \sqrt{trace(AA^*)}$. Because *H* is a rectangular matrix, in

[1] For a full account of the theory, see Mehta (1991) and Tulino and Verdù (2004) and the survey papers by Edelman and Rao (2005) and Johnstone (2006).

the general case its eigenvalues do not exist. The exponential of the trace of a matrix **A** is sometimes called *ert*: $ert(A) = \exp(tr(A))$.

### 2.1.2. *Wigner matrices*

A complex Wigner matrix $W$ is a Hermitian square $n \times n$ complex matrix whose upper triangular elements are independent zero-mean variables with the same variance. If the variance is $\sigma^2 = 1/n$, the matrix is called a standard Wigner matrix. If entries are independent and identically distributed (i.i.d.), zero-mean, unit variance (standard) normal variables, the matrix pdf is the following:

$$p(W) = 2^{-n^2/2} \pi - n^2/2 \exp\left[ - \frac{trace(W^2)}{2} \right].$$

In the general case of non-Gaussian Wigner matrices, there is no closed formula for the joint distribution of eigenvalues. For complex Gaussian Wigner matrices, the joint distribution of the ordered eigenvalues $\lambda_1 > \cdots > \lambda_n$ is given by the following expression:

$$p_\Lambda(\lambda_1, \ldots, \lambda_n) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n} \lambda_i^2} \left( \prod_{i=1}^{n-1} \frac{1}{i!} \prod_{i<j}^{n} (\lambda_i - \lambda_j)^2 \right).$$

and the marginal distribution of the unordered eigenvalues is the following:

$$p_\Lambda(\lambda) = n^{-1} \sum_{i=0}^{n-1} \frac{1}{2^i i! \sqrt{2\pi}} \left( e^{-\frac{\lambda^2}{4}} H_i(\lambda) \right)$$

where $H_i(\lambda)$ is the *i*th Hermite polynomial.

### 2.1.3. *Wishart matrices*

Given the $m \times n$ matrix $H$ whose columns are independent real/complex zero-mean Gaussian vectors with covariance matrix $\mathbf{\Sigma}$, the matrix $A = HH^*$ is called a central Wishart matrix $W_m(n, \mathbf{\Sigma})$ with *n* degrees of freedom and covariance $\mathbf{\Sigma}$. If the entries of $H$ are not zero-mean, the Wishart matrix is non-central. The pdf of a central Wishart matrix with *n*>*m* has the following form:

$$p_W(A) = \frac{\pi^{-\frac{m(m-1)1}{2}}}{(\det \Sigma)^n \prod_{i=1}^{m} (n-i)!} \exp\left[- trace(\Sigma^{-1} A)\right] \det A^{n-m}.$$

The joint pdf of the ordered strictly positive eigenvalues $\lambda_1 > \cdots > \lambda_n$ of a Wishart matrix is given by the following expression:

$$f_\Lambda(\lambda_1,\ldots,\lambda_n) = e^{-\sum_{i=1}^{t}\lambda_i}\left(\prod_{i=1}^{t}\frac{\lambda_i^{r-t}}{(t-i)!(r-i)!}\prod_{i<j}^{t}(\lambda_i-\lambda_j)^2\right),\ t = \min(m,n), r = \max(m,n).$$

The pdf of the marginal distribution of the unordered eigenvalues is the following:

$$g(\lambda) = \frac{1}{t}\sum_{k=0}^{t-1}\frac{k!}{(k+r-t)!}\left(L_k^{r-t}(\lambda)\right)^2\lambda^{r-t}e^{-\lambda},\ t = \min(m,n), r = \max(m,n).$$

where $L_k^{r-t}(\lambda)$ are the Laguerre polynomials.

### *2.2. Asymptotic distribution of the bulk of eigenvalues*

I am interested in applying RMT for estimating a covariance matrix and for determining the number of factors. The basic methodologies proposed in the literature consist either in estimating the theoretical distribution of the eigenvalues from the empirical distribution of the eigenvalues or in constructing tests for specific asymptotic distributions of eigenvalues. Hence I survey the literature on the asymptotic distributions of eigenvalues. The RMT distinguishes between the distribution of the *bulk* of the distribution of eigenvalues and the distribution of the *edges*. I will first discuss the bulk of the eigenvalue distribution.

Results for the bulk of the distribution of eigenvalues can be summarized as follows. Anderson (1963) proved that the empirical distribution of the eigenvalues of a square *N*x*N* matrix tends to the distribution of the eigenvalues of the true covariance matrix when the number of samples tends to infinity. However, if both the number of samples and the number of entries of the covariance matrix tend to infinity, then the empirical eigenvalues are not consistent estimators of the true eigenvalues.

The first asymptotic limit law was discovered by Wigner in the 1950s. Wigner (1955, 1956, 1958, 1959) proved the famous *semicircle law* which states that the limit distribution of the eigenvalues of a Wigner matrix assuming that 1) $n \to \infty$ and 2) the fourth moments of the entries of the matrix are of order $O\left(\frac{1}{N^2}\right)$, is the deterministic semicircle law:

$$w(\lambda) = \begin{cases} 0 \quad \text{for} \quad |\lambda| > 2 \\ \dfrac{1}{2\pi}\sqrt{4-\lambda^2} \quad \text{for}\ |\lambda| \le 2 \end{cases}.$$

Considering a Gaussian matrix instead of a Wigner matrix. Girko (1984) proved that the (complex) eigenvalues have an asymptotic uniform distribution on the unit circle. This result is called the *Girko's full circle law*.

The semicircle and the full circle laws are related to square matrices. The next fundamental asymptotic result was proved in Marčenko and Pastur (1967) for rectangular matrices. I will first state the Marčenko and Pastur law. Consider a $T \times N$ matrix $\mathbf{H}$ whose entries are i.i.d real or complex zero mean variables with variance $1/T$ and fourth moments of order $O\left(\frac{1}{T^2}\right)$. Marčenko and Pastur (1967) proved that the asymptotic distribution of the eigenvalues of the matrix $A = H^*H$ when $T, N \to \infty, \frac{N}{T} \to \gamma$ has the following density:

$$f_\gamma(x) = \left(1 - \frac{1}{\gamma}\right)^+ \delta(x) + \frac{\sqrt{(x-a)(b-x)}}{2\pi\gamma x}, a = \left(1-\sqrt{\gamma}\right)^2 \le x \le b = \left(1+\sqrt{\gamma}\right)^2 .$$
$$f_\beta(x) = 0, x < a, x > b$$

where $(z)^+ = \max(0,z)$. Under the same assumptions, the asymptotic distribution of the eigenvalues of the matrix $HH^*$ when $T, N \to \infty, \frac{N}{T} \to \gamma$ has the following density:

$$\tilde{f}_\gamma(x) = (1-\gamma)^+ \delta(x) + \frac{\sqrt{(x-a)(b-x)}}{2\pi x}, a = \left(1-\sqrt{\gamma}\right)^2 \le x \le b = \left(1+\sqrt{\gamma}\right)^2$$
$$f_\beta(x) = 0, x < a, x > b$$

If $\gamma = 1$, the distribution of singular values, which are the square roots of the corresponding eigenvalues, is the *quarter circle law*:

$$q(x) = \frac{\sqrt{4-x^2}}{\pi}, 0 \le x \le 2,$$
$$q(x) = 0, x < 0, x > 2$$

Dirac's delta at the origin reflects the fact that a fraction $\frac{N-T}{N}$ of the eigenvalues are zero if $\gamma \ge 1$.

Actually Marčenko and Pastur (1967) proved a more general result. They were able to determine the limit eigenvalue distribution of a matrix of the form $A = W_0 + HTH^*$ where $W_0$ is a Hermitian matrix whose distribution of the empirical eigenvalues converges to a non random limit $L$ and $T$ is a diagonal real matrix whose distribution of the empirical eigenvalues converges to a non random limit *H*. Call *F* the limit distribution of the matrix $A$. The Stieltjes transform $S(z)$ of a distribution *F* is the integral: $S_F(z) = \int_{-\infty}^{+\infty} \frac{dF(y)}{y - z}$. Marčenko and Pastur (1967) proved that the following relationship holds:

$$S_F(z) = L\left[z - \beta \int \frac{ydH(y)}{1 + yS_F}\right].$$

Solving this equation and inverting the Stieltjes transform, one can determine the limit distribution *F*.

This result has been extended and refined in many different ways. Silverstein (1995) proved an extension of Marčenko-Pastur (1967) dropping the condition that the fourth moments exist and that *T* is diagonal. Suppose the entries of the $T \times N$ matrix *H* are i.i.d. real or complex variables with zero mean, unit variance, and finite fourth moments. Let $T_N$ be a fixed $N \times N$ Hermitian (unitary if real) matrix. Assume the sample vector is $T_N^{\frac{1}{2}} H$. This implies that $T_N$ is the population covariance matrix. Consider the sample covariance matrix: $B_N = \frac{1}{N} T_N^{\frac{1}{2}} HH' T_N^{\frac{1}{2}}$. Silverstein (1995) proved that if the distribution of the eigenvalues of the matrices $T_N$ tends to a non-random distribution, then the empirical covariance matrices $B_N$ also tend to a non-random distribution and the following equation still holds

$$S_F(z) = L\left[z - \gamma \int \frac{ydH(y)}{1 + yS_F}\right].$$

Burda, Görlich, A. Jarosz and Jurkiewicz (2004) proved the Marčenko-Pastur law using the method of the resolvent and diagrammatic techniques from quantum mechanics. Burda, Jurkiewicz, and Waclaw (2005) extended the Marčenko-Pastur law to samples that are both correlated and autocorrelated. Burda, Goerlich, and Waclaw (2006) determined explicit formulas in the case of Student-*t* distributions up to integrals. Similar results were obtained by Sengupta and Mitra (1999).

### *2.3. Asymptotic distribution of the largest eigenvalues*

I will now review the literature that deals with the size and asymptotic distribution of the largest and the smallest eigenvalues. In RMT we distinguish between the bulk of the distribution of the eigenvalues and the *edges*, that is, the smallest and largest eigenvalues. The Marčenko-Pastur law has finite support and therefore the bulk of the eigenvalues remains confined in a finite segment of the real line. However, the Marčenko-Pastur law is compatible with the existence of a few stray eigenvalues that are at the right (left) of its rightmost (leftmost) edge. Geman (1980) and Silverstein (1985) demonstrated that this is not the case. The largest eigenvalue $\lambda_1$ of the covariance matrix of of a $T \times N$ i.i.d matrix $H$ when $T, N \to \infty, \frac{N}{T} \to \gamma$ converges a.s. to the value $b = \left(1 + \sqrt{\gamma}\right)^2$ and the eigenvalue $\lambda_k, k = \min(T, N)$ converges to the value $a = \left(1 - \sqrt{\gamma}\right)^2$ with $\lambda_{k+1} = \lambda_N = 0$ if $T < N$. Equivalently the scaled eigenvalues $\lambda_1, \lambda_k$ of the matrix $A = H^*H$ converge to the same limits: $n^{-1}\lambda_1 \underset{a.s.}{\to} \left(1 + \sqrt{\gamma}\right)^2, n^{-1}\lambda_k \underset{a.s.}{\to} \left(1 - \sqrt{\gamma}\right)$.

Yin, Bai, and Krishnaiah (1988) demonstrated that a necessary and sufficient condition for $n^{-1}\lambda_1 \underset{a.s.}{\to} \left(1 + \sqrt{\gamma}\right)^2$ to hold is the existence of finite fourth moments of the distribution of the matrix entries.

The latter result does not say anything about the asymptotic distribution of the largest eigenvalue. Tracy and Widom (1996) determined the asymptotic distribution of the largest eigenvalue of a Gaussian real or complex symmetric matrix. The two Tracy-Widom distributions are called, respectively, $F_{GOE}$ and $F_{GUE}$. None of these distributions has a closed-formula expression but they are expressed as an integral of the solution of the Painlévé equations of type II.

Forrester (1993), Johansson (2000), and Johnstone (2001) generalized the above result to rectangular matrices. Overall, this result can be stated as follows. Consider a *TxN* matrix *H* with i.i.d real or complex standard Gaussian entries. Consider the eigenvalues $l_1 \geq l_2 \geq \cdots \geq l_N$ of the matrix *H*H.* Then there are centering and scaling constants $\mu_n, \sigma_n$ such that, if $m, n \to \infty$, then the distribution of the centered and rescaled largest eigenvalue $\frac{l_1 - \mu_n}{\sigma_n}$ tends to the Widom-Tracy limit law of order 1 or 2, as in the Gaussian square case.

However, there are differences regarding the constants and how the limits are taken. Forrester (1993) considered complex matrices and assumed $T - N = const., T \to \infty$. Johansson (2000) considered complex matrices and assumed $T = \frac{1}{\gamma}N + O\left(N^{\frac{1}{3}}\right), \gamma \leq 1, N \to \infty$. Johnstone

(2001) considered real matrices, assumed $T = \frac{1}{\lambda} N, \gamma \leq 1, T \to \infty$ , and considered the following centring and scaling constants:

$$\mu_T = \left(\sqrt{T-1} + N\right)^2,$$
$$\sigma_T = \left(\sqrt{T-1} + \sqrt{N}\right)\left(\frac{1}{\sqrt{T-1}} + \frac{1}{N}\right)^{\frac{1}{3}}.$$

These results have been further extended and refined in different ways. Soshnikov (2002) extended Johnstone (2001) to the first *k* eigenvalues and not only to the largest eigenvalue. Using slightly different centering and scaling constants, El Karoui (2003) proved that the rate of convergence to the Tracy-Widom law is of the order $N^{\frac{2}{3}}$. El Karoui (2006) dropped the assumption $\gamma \leq 1$ and allowed $N/T \to \gamma$ where $\gamma \in (0, \infty)$ or $\gamma = 0$ or $\gamma = \infty$ . Soshnikov (2002) proved that the largest eigenvalue still converges to a Tracy-Widom law if the entries of the matrix *H* are not Gaussian but have sufficiently light tails and *n-p* is of order $p^{-\frac{1}{3}}$.

The behaviour of the largest eigenvalue changes completely if the matrix *H* is heavy-tailed. Soshnikov and Fyodorov (2005) and Soshnikov (2006) proved that the distribution of the largest eigenvalue exhibits a weak convergence to a Poisson process. Biroli, Bouchaud, and Potters (2005) showed that the largest eigenvalue of a square random matrix whose entries have distributions with power law tails exhibits a phase transition for the Tracy-Widom law to a Frechet distribution with tail index 4.[2]

Thus far I have discussed the behaviour of the largest eigenvalue(s) under the null hypothesis of i.i.d entries of the matrix *H*. In terms of factor models, this null hypothesis is very restrictive as it is the null of the residuals of a strict scalar factor model. I will now survey results for the distribution of the largest eigenvalue(s) under the assumption that the matrix *H* is formed by independent observations of correlated vectors.

Bai and Silverstein (1998) proved that there is no eigenvalue outside of the support of the asymptotic distributions of the eigenvalues of the matrix *H*H* under the assumption that observations are $H = T_N^{\frac{1}{2}} Z$ where $T_N^{\frac{1}{2}}$ is the square root of an Hermitian matrix whose eigenvalues converge to a proper probability distribution and **Z** has i.i.d standard complex entries. Bai and Silverstein (1999) proved an asymptotic exact separation theorem which states that, for any interval that separates true eigenvalues, there is a corresponding interval that separates corresponding empirical eigenvalues.

[2] If the tails decay as a power law the tail index is the exponent of that power law

Johnstone (2001) introduced the "spiked" covariance model where the population covariance matrix is diagonal with *N-r* eigenvalues equal to 1 while the first *r* largest eigenvalues are larger than 1:

$$\left(\underbrace{l_1,\ldots,l_r}_{r}\underbrace{1,\ldots,1}_{N-r}\right).$$

Recall that the eigenvalues are invariant under an orthogonal transformation and that I call $l_i$ the *i*-th eigenvalue of the population covariance matrix and $\lambda_i$ the *i*-th eigenvalue of the empirical covariance matrix.

Consider a complex *T*x*N* matrix *H***.** Assume that the eigenvalues of the covariance matrix $S = \frac{1}{N} H * H$ are $\left(\underbrace{l_1,\ldots,l_r}_{r}\underbrace{1,\ldots,1}_{N-r}\right)$.

Péchè (2003) proved that if $l_i \leq 2, i = 1,\ldots,r, l_i = 1, i > r$ then:

$$P\left|(\lambda_1 - 4)2^{-\frac{4}{3}}T^{\frac{2}{3}} \leq x\right| \underset{T,N\to\infty}{\rightarrow} F_{GUE}(x).$$

Baik, Ben Arous, and Péché (2005) generalized this result and proved a phase transition law. In fact, they proved the following. Consider a complex *T*x*N* matrix *H* whose rows are independent samples extracted from a multivariate distribution such that the eigenvalues of the covariance matrix $S = \frac{1}{N} H * H$ are

$$\left(\underbrace{l_1,\ldots,l_r}_{r}\underbrace{1,\ldots,1}_{N-r}\right).$$

Assume $N, T \to \infty, \frac{N}{T} \to \gamma < 1$. Consider a $k, 0 \leq k \leq r$. Then, if

$$\begin{array}{l} l_i = \left(1 + \sqrt{\gamma}\right), i = 1,\ldots,k, \\ l_i < \left(1 + \sqrt{\gamma}\right), k < i \leq r \end{array}, \text{ then}$$

$$P\left[\left(\lambda_1 - \left(1 + \sqrt{\gamma}\right)^2\right)\left(1 + \sqrt{\gamma}\right)^{-\frac{4}{3}}T^{\frac{2}{3}} \leq x\right] \underset{T,N\to\infty}{\rightarrow} F_K(x).$$

If $l_i > \left(1 + \sqrt{\gamma}\right), i = 1,\ldots,k, l_i \leq l_1, k \leq i \leq r$, then

$$P\left[\left(\lambda_1 - \left(l_1 + \frac{l_1\gamma}{l_1 - 1}\right)^2\right)\frac{\sqrt{T}}{\sqrt{\left(l_1^2 - {l_1^2\gamma}\big/{(l_1 - 1)^2}\right)}} \leq x\right] \underset{T,N\to\infty}{\to} G_K(x)$$

where $F_K$ and $G_K$ are special functions defined in Baik, Ben Arous, and Péché (2005). As $F_0 = F_{GUE}$, I again find the particular result obtained in Péché (2003). Onatski (2007) generalized this result to the case of singular Wishart matrices, that is, $\gamma \geq 1$.

The limit behaviour of the largest eigenvalues was generalized in Baik and Silverstein (2006) to include non-one eigenvalues, complex and real data, and distribution assumptions different from the Gaussian. The main result in Baik and Silverstein (2006) can be stated as follows. Suppose the entries of the $N \times T$ matrix **Z** are i.i.d. real or complex variables with zero mean, unit variance, and finite fourth moments. Let $T_N$ be a fixed $N \times N$ Hermitian (unitary if real) matrix. Assume the sample vector is $T_N^{\frac{1}{2}}Z$. This implies that $T_N$ is the population covariance matrix. Consider the sample covariance matrix: $B_N = \frac{1}{N}T_N^{\frac{1}{2}}ZZ'T_N^{\frac{1}{2}}$. For some unitary matrix

$$U_B B_N U_B^{-1} = \begin{bmatrix} s_1^N & 0 & 0 \\ 0 & \ddots & \\ 0 & 0 & s_N^N \end{bmatrix}$$

where the $s$ are the sample eigenvalues in decreasing order. Suppose the eigenvalues have the following structure:

$$s = \left(s_1^N, \ldots, s_N^N\right) = \left(\underbrace{\alpha_1 \ldots \alpha_1}_{k_1}, \ldots, \underbrace{\alpha_M, \ldots, \alpha_M}_{k_M}, \underbrace{1, \ldots, 1}_{N-r}\right),$$

where $M$ is a non negative integer, $k_1 + \cdots + k_M = r$ and the $\alpha_j$ are fixed real numbers.

Assume that $N = N(T), \frac{N}{T} \to \gamma, T \to \infty$. Then the following holds:

*Case 1.* Suppose $0 < \gamma < 1$. Let $M_0$ be the integer such that $\alpha_j > 1 + \sqrt{\gamma}$ and let $M - M_1$ be the integer such that $\alpha_j < 1 - \sqrt{\gamma}$;

For $1\le j\le M_0, s_{k_1+\cdots k_{j-1}+i}\to \alpha_j+\frac{\gamma\alpha_j}{\alpha_j-1}, 1\le i\le k_j$, a.s.

$s_{k_1+\cdots k_{M_0}+1}\to \left(1+\sqrt{\gamma}\right)^2$, a.s.

$s_{T-r+k_1+\cdots k_{M_1}}\to \left(1-\sqrt{\gamma}\right)^2$, a.s.

$M_1+1\le j\le M, s_{T-r+k_1+\cdots k_{j-1}+i}\to \alpha_j+\frac{\gamma\alpha_j}{\alpha_j-1}, 1\le i\le k_j$

*Case 2.* Suppose $\gamma>1$. Let $M_0$ be the integer such that $\alpha_j>1+\sqrt{\gamma}$;

For $1\le j\le M_0, s_{k_1+\cdots k_{j-1}+i}\to \alpha_j+\frac{\gamma\alpha_j}{\alpha_j-1}, 1\le i\le k_j$, a.s.

$s_{k_1+\cdots k_{M_0}+1}\to \left(1+\sqrt{\gamma}\right)^2$, a.s.

$s_T\to \left(1-\sqrt{\gamma}\right)^2$, a.s.

$s_{T+1}=\cdots=s_N=0$

*Case 3.* Suppose $\gamma=1$. Let $M_0$ be the integer such that $\alpha_j>2$;

For $1\le j\le M_0, s_{k_1+\cdots k_{j-1}+i}\to \alpha_j+\frac{\gamma\alpha_j}{\alpha_j-1}, 1\le i\le k_j$, a.s.

$s_{k_1+\cdots k_{M_0}+1}\to 4$, a.s.

$s_{\min(T,N)}=0$

### *2.4 Asymmetric matrices*

Thus far I have reviewed the literature on symmetric covariance matrices. As it will be reviewed in Section 2.10, there is extensive literature on the application of the results from RMT to the analysis of static factor models. However, when one needs to forecast prices or returns one has to consider auto-cross correlations, that is, correlations between return *i* at time *t* and return *j* at time *t−k.* The study of the asymptotic distribution of eigenvalues of asymmetric matrices allows one to test the null of absence of significant non-zero auto-cross correlations and to determine the number of factors that might be used in forecasting.

The theory of asymmetric random matrices is less developed than the theory of symmetric random matrices. Ginibre (1965) proposed a random matrix model, now called the GinOE, which stands for Ginibre Orthogonal Ensemble. The GinOE is formed by pairs of independent matrices *X,Y* whose entries are i.i.d zero-mean Gaussian variables: $X = \{X_{t,i}\}, Y = \{Y_{t,j}\}$ $t = 1,\ldots,T, i,j = 1,\cdots,N$. The covariance matrix $C = \frac{1}{T}X'Y$ is a square matrix in general non symmetric. Therefore, the eigenvalues and the eigenvectors $Cv_k = \lambda_k v_k$ of the matrix *C* are in general complex.

The pdf of the matrix *C* is given by $P_{GinOE} = (2\pi)^{-\frac{N^2}{2}} \exp\left(-Tr\left(\frac{CC'}{2}\right)\right)$. The limit distribution of the eigenvalues of $C$ when both *N* and *T* tend to infinity is an ellipse in the complex plane:

$$p(\lambda) = \begin{cases} (\pi ab)^{-1}, \left(\frac{\operatorname{Re} z}{a}\right)^2 + \left(\frac{\operatorname{Im} z}{b}\right)^2 \leq 1 \\ 0, \left(\frac{\operatorname{Re} z}{a}\right)^2 + \left(\frac{\operatorname{Im} z}{b}\right)^2 > 1 \end{cases}$$

where $a = 1 + s, b = 1 - s$ and $s$ is a degree of matrix symmetry: $s = 1$ symmetric, real eigenvalues, $s = -1$ antisymmetric, imaginary eigenvalues, $s = 0$ asymmetric, complex eigenvalues. Forrester and Nagao (2007) provide a full set of asymptotic statistics for the eigenvalues.

Bouchaud *et al.* (2005) studied a related problem. They considered a *N*×*T* matrix *X* and a *M*×*T* matrix *Y*. Diagonalize the matrices *X* and *Y* . Consider the diagonalized matrices $\hat{X}, \hat{Y}$ and form the matrix $G = \hat{X}\hat{Y}'$. The *N*x*M* matrix *G* is the covariance matrix between *X* and *Y* . The singular values of *G* are the canonical correlations between the two diagonalized matrices. Bouchaud *et al.* (2005) demonstrated that the asymptotic distribution of the singular values, when *T,N,M* tend to infinity with *n* = *N*/*T*, *m* = *M*/*T*, is given by the following expression:

$$\rho(s) = \max(1 - m, 1 - n)\delta(s) + \max(m + n - 1)\delta(s - 1) + \frac{\sqrt{(s^2 - \theta_-)(\theta_+ - s^2)}}{2s(1 - s^2)}.$$

$$\theta_\pm = n + m - 2mn \pm \sqrt{mn(1 - m)(1 - n)}$$

### *2.5. Static Factor Models*

I will first fix the notation that I will use throughout the dissertation. Consider a multivariate time series of returns:

$r_t = (r_{it}),\ i = 1,2,\ldots,N,\ t = 1,2,\ldots,T$

If returns are stationary, I assume that the constant means are subtracted so that $\mathbf{r}_t$ is an $N$-vector of possibly correlated and autocorrelated zero-mean variables. A static linear factor model of returns is a model of the following type:

$$r_t = \beta f_t + \varepsilon_t$$

where

$$\beta = \begin{array}{ccc} \beta_{11} & \cdots & \beta_{1Q} \\ \vdots & \ddots & \vdots \\ \beta_{N1} & \cdots & \beta_{NQ} \end{array}$$

is the $N \times Q$ matrix of factor loadings, $f_t$ is a $Q$-vector of factors, and $\boldsymbol{\varepsilon}_t$ is a $N$-vector of residuals. The term $\beta f_t$ is called the common component while the terms $\boldsymbol{\varepsilon}_t$ are the idiosyncratic components.

The above is a static factor model; a dynamic factor model would explicitly include lagged factors: $r_t = \beta_0 f_t + \cdots \beta_P f_{t-P} + \varepsilon_t$. Adding factors and defining $F_t = \left(f_t, \ldots, f_{t-P}\right)$, a dynamic factor model can always be cast in a static form. A static factor model can be compactly written in terms of a sample of observed data. Suppose one has $T$ observations of the multivariate vector $\mathbf{r}_t = \left(r_{it}\right), i = 1,2,\ldots,N, t = 1,2,\ldots,T$. Define the matrix of observations, factors and residuals as follows:

$$R = \begin{bmatrix} r_{11} & \cdots & r_{N1} \\ \vdots & \ddots & \vdots \\ r_{1t} & \cdots & r_{Nt} \\ \vdots & \ddots & \vdots \\ r_{1T} & \cdots & r_{NT} \end{bmatrix}, \quad F = \begin{bmatrix} f_{11} & \cdots & f_{Q1} \\ \vdots & \ddots & \vdots \\ f_{1t} & \cdots & f_{Qt} \\ \vdots & \ddots & \vdots \\ f_{1T} & \cdots & f_{QT} \end{bmatrix}, \quad E = \begin{bmatrix} \varepsilon_{11} & \cdots & \varepsilon_{N1} \\ \vdots & \ddots & \vdots \\ \varepsilon_{1t} & \cdots & \varepsilon_{Nt} \\ \vdots & \ddots & \vdots \\ \varepsilon_{1T} & \cdots & \varepsilon_{NT} \end{bmatrix}$$

where each row is an observation of $N$ variables, $Q$ factors and $N$ residuals. I can compactly write the factor models as:

$$R = F\beta\beta + E$$

I will first survey the static factor models defined in the literature, starting with the classical factor model. Classical factor models have a long history which goes back to the formalization of psychometric models. Spearman (1904) introduced a one-factor model of mental abilities, Thurstone (1938, 1947) introduced the first multifactor model, and Hotelling (1933) described principal components analysis. Classical factor models as described, for example, in Anderson (2003), are strict factor models with a finite number of variables. In a strict factor model, residuals are mutually uncorrelated and uncorrelated with factors. This implies that all correlations are due to factors. A strict factor model is called a scalar strict factor model if all residuals have the same variance.

Without additional assumptions, strict factor models are not identifiable. In fact, one obtains observationally equivalent models if one multiplies the matrix of factors by any non singular matrix and the matrix of loadings by its inverse. In order to identify a strict factor model, additional assumptions are needed. For example, factors can be assumed to be orthogonal, unit variance variables.

The setting of classical strict static factor models is one of independent samples extracted from a population with a multivariate Gaussian distribution. Though a factor model might describe a multivariate time series, no dynamics is allowed. In the setting of strict factor models, as observations are i.i.d. vectors, there is no dynamics even if samples are taken at different times. Strict factor models are "fixed $N$ large $T$" models as the number of variables is kept fixed while the number of samples is allowed to grow. This assumption works well in the original empirical setting of psychometric studies where the number of individuals largely exceeds the number of variables.

The first factor model used in financial econometric was the Capital Asset Pricing Model (CAPM) of Sharpe-Lintner-Mossin. The CAPM is a single-factor theoretical model based on General Equilibrium principles. The first multifactor model in financial econometrics was proposed by Ross (1976) in his asset pricing theory (APT). The APT model is a strict multifactor model.

Connor and Korajczyk (1986, 1988) proposed a "fixed $T$ large $N$" model. This model is suggested by empirical settings where the number of series exceeds the number of samples. They allow residuals to be cross-sectionally correlated. Connor and Korajczyk (1988) also allow heteroscedasticity in the time dimension. Because of these assumptions, their model cannot be considered a classical model of independent samples.

The extension from "fixed $N$, large $T$" to doubly infinite models was motivated by several considerations. First, in practice, in macroeconomic and financial applications, the number $N$ of empirical time series of returns is of the same order of magnitude or even exceeds the number of observations $T$. There are also theoretical considerations as the no-arbitrage arguments of Ross are obtained in the limit of an infinite market (i.e., when both $N$ and $T$ diverge).

When both $T$ and $N$ are large, the asymptotic results developed in the case of fixed $N$, $T \to \infty$ are not applicable. Therefore, it seems reasonable to develop a factor model with infinite $T$ and infinite $N$ and to determine asymptotic results when both $T$ and $N$ tend to infinity. However, it was observed that, if one allows the number of stock returns to grow indefinitely, it is unlikely that residuals are uncorrelated if one retains only a finite (small) number of factors.

To address the problem, Chamberlain and Rothschild introduced the notion of approximate linear factor models (Chamberlain, 1983; Chamberlain and Rothschild, 1983). An approximate factor model[3] as defined in Chamberlain and Rothshild (1983) allows residuals to be mutually correlated but requires that the correlations of residuals be only local while the correlations of returns be global, driven by factors. The setting of the approximate factor model as defined in Chamberlain and Rothshild is still one of independent samples extracted from a distribution, though the number of series grows with the size of the sample. This setting does not allow any time dynamics. Assume that there are $T$ i.i.d. observations of a multivariate correlated vector $r_t = (r_{it}), i = 1,2,\ldots,N, t = 1,2,\ldots,T$. It is assumed that both $T,N \to \infty$ but there is no time dynamics as observations are independent.

While in a strict factor model data are correlated and residuals are uncorrelated, in the Chamberlain and Rothschild approximate factor model both data and residuals are correlated but not autocorrelated. However the essential feature of an approximate factor model is the requirement that residuals have only local correlations (and therefore no factor structure) while the data have a correlation structure due to common causes, i.e., common factors.

Chamberlain and Rothschild (1983) expressed the condition of local correlations of residuals in the limit of an infinite market as follows. Given $N$ return time series, the covariance matrix of an approximate factor model with $Q$ factors where factors and residuals are mutually uncorrelated can be written as:

$$\Sigma_N = \beta_N' \beta_N + R_N$$

where all $\beta_N$ are $NxQ$ matrices. Call $\lambda_{iN}$ the $i$-th eigenvalue of matrix $R_N$. The condition that correlations are local is imposed by requiring that the sequence of covariance matrices of residuals $R_N$ for $N \to \infty, T \to \infty$ has uniformly bounded eigenvalues, that is, there is a $M>0$ such that $\lambda_{iN} \leq M, \forall i, \forall N$. As the matrix $\beta_N' \beta_N$ has rank $Q$, $N\text{-}Q$ eigenvalues of the sequence $\mathbf{\Sigma}_N$ are uniformly bounded while $Q$ eigenvalues are unbounded. The size of $M$ is arbitrary.

The condition of serially independent observations is too strong both for financial econometrics and for macroeconomics, where factor models were widely applied. The third stage in the study of static factor models was the definition of models able to describe time series with a dynamics. Stock and Watson (1998, 2002) defined an approximate static factor model of stationary

[3] The terminology "approximate factor models" and "approximate factor structure" is slightly misleading insofar as a linear approximate factor model should be a perfectly well-specified model. There is no approximation implied by an approximate factor model.

but possibly autocorrelated processes in the asymptotic limit of $T, N \to \infty$. The Stock and Watson model assumes a constant covariance matrix of factors and makes assumptions about the covariances between returns. It is not necessarily a Gaussian model but conditions on the second and fourth moments of the distributions impose constraints as to the type of admissible non-normality. In the same paper, Stock and Watson also considered the possible non stationarity of the factor model allowing factor loadings to be time dependent.

Bai (2003) and Bai and Ng (2002) generalized the definition of an approximate factor model to include the possibility of serial autocorrelations of returns and to allow for heteroscedasticity. To make these generalizations, Bai (2003) and Bai and Ng (2002) imposed conditions on covariances similar to those in Stock and Watson (1998, 2002). The assumptions in Stock and Watson (1998, 2002), Bai (2003) and Bai and Ng (2002) still imply that the first $Q$ eigenvalues of the covariance matrix of returns diverge while the remaining $N$-$Q$ are uniformly bounded. If residuals are neither autocorrelated nor heteroscedastic, these assumptions coincide with the assumptions made by Chamberlain and Rothschild (1983).

Though the largest $Q$ eigenvalues are assumed to grow without bounds with $N$, the rate of growth is not fixed. Stock and Watson (2002) assumed that $N - Q$ eigenvalues of the sequence $\mathbf{\Sigma}_N$ are $O(N)$ while the remaining $Q$ eigenvalues are $o(1)$.

### *2.6. Estimation of static factor models*

A sound asymptotic estimation theory of classical static factor models is available. Anderson (1963) proved that, for Gaussian distributions, the empirical covariance matrix of X tends to the population covariance matrix. He also proved that the eigenvalues of the sample covariance matrix tend to the eigenvalues of the population matrix and have an asymptotic normal distribution. If one assumes that variables are normally distributed, factor loadings can be estimated with maximum likelihood estimation principles. Estimated factor loadings are normally distributed and the rate of convergence is the square root of $T$.

However, as proved in Anderson (2003), factors cannot be estimated with maximum likelihood. If all the residuals have the same variance, factors can be estimated with principal components even if factors are not normally distributed. If residuals have different variances, in general principal components are not guaranteed to be consistent estimators of factors (Breitung, Jorg and Uta Kretschmer, 2005).

However, when one moves to the large $N$, large $T$ setting, that is, when one considers asymptotic results for $N, T \to \infty$, estimation of factors and factor loadings simplify as factors can indeed be consistently estimated with principal components. Chamberlain and Rothschild (1983) demonstrated that, assuming the covariance matrix is known, asymptotic principal components span the factor space. Their result is still a population property as Chamberlain and Rothschild did not consider sample variability and asymptotic consistency.

When one allows $N,T \to \infty$, one has to consider how one takes the limits as different paths $N = N(T)$ can be specified. Assuming that the number of factors is known, Connor and Koraczik (1986, 1988) showed that factors can be consistently estimated with principal components when $T$ is fixed and $N \to \infty$. This result is obtained in Connor and Korajczyk (1986), where they introduced asymptotic principal components. Connor and Korajczyk (1986) showed that the limit for $N$ that tends to infinity of the eigenvectors of the $T \times T$ cross product matrix of returns span the factor space. Connor and Korajczyk (1988) extended the procedure to account for cross-sectional heteroskedasticity and Jones (2001) extended the procedure to account for time-series heteroskedasticity.

Stock and Watson (1998) showed that, in the $N,T \to \infty$ limit, factors can be consistently estimated with principal components. Stock and Watson (2002) subsequently extended the theory to allow both long time series and large cross-sectional samples as well as time-varying factor betas. In addition, they provided a quasi-maximum likelihood interpretation of the technique. Bai (2003) analyzed the large-sample distributions of the factor returns and factor beta matrix estimates in a generalized version of this approach. Bai (2003) and Bai and Ng (2002) showed that principal components estimated from data span the factor space. Bai (2003) and Bai and Ng (2002) determined the asymptotic distribution and convergence rates of estimated factors under the assumptions that $N,T \to \infty$ and without any restriction as to how we take the limit.

As regards the number of factors, Chamberlain and Rothschild (1983) assumed that the number of factors is known. Connor and Korajczyk (1993) introduced a formal test of the number of factors. Bai and Ng (2003) introduced three equivalent tests for determining the asymptotic number of factors. These tests are model selection tests based on Information Theory. Bai and Ng (2002) first showed that the usual Bayesian information criterion does not work if both $T$ and $N$ diverge and then introduced their own test which is a generalization of the Bayesian information criterion to include both $T$ and $N$.

I will now review the application of RMT on static factor models.

### *2.7. Application of RMT to static factor models*

Results from RMT have been applied to factor models to determine the number of factors, to understand the nature of the residuals and, in particular, to determine if residuals are correlated. The paper by Galluccio, Bouchaud, and Potters (1998) is the first paper where results from RMT are applied to a financial optimization problem. The paper by Plerou *et al*. (2002) is the first application of RMT to financial econometrics. The latter paper includes a thorough analysis of the empirical spectrum of eigenvalues of a large universe of returns. Plerou *et al.* analyzed the distribution of the eigenvalues of the covariance matrix of returns in two large datasets. The first dataset includes 30-minute returns of the largest 1,000 stocks traded on the New York Stock Exchange (NYSE), NASDAQ, and American Stock Exchange (AMEX) in the two-year period 1994-1995. The second dataset includes daily returns of all stocks in the CRSP files that survived in the 35-year period 1962-1996.

Plerou *et al*. (2002) showed that the bulk of the distribution of empirical eigenvalues for both datasets is in good agreement with the theoretical Marčenko-Pastur law for uncorrelated variables. However, in both cases, large eigenvalues appear. The authors introduced a number of methodological considerations. They first observed that the agreement between the bulk of empirical and theoretical distributions of eigenvalues is not sufficient to conclude that the eigenvalues in the bulk of the distribution correspond to zero correlation. Plerou *et al*. therefore introduced three additional tests based on universal properties of random matrices: the distribution of the nearest-neighbor eigenvalue spacing, the distribution of the next-nearest-neighbor eigenvalue spacing, and a special statistics called "number variance" to gauge long-range correlations between eigenvalues. Empirical data passed the three tests. The authors also considered the composition of eigenvectors and measured the concentration of eigenvectors using the inverse participation ratio, defined as the sum of the inverse of the fourth power of eigenvector coefficients. The paper identified the first ten eigenvectors with the entire market and with specific industrial sectors.

It should be observed that Plerou *et al*. (2002) did not assume any factor model. Only at the end of the paper did they suggest common factors as a possible explanation of the empirical results. I will now review several papers that did assume a factor structure and that used RMT to estimate the number of factors and to understand the nature of factors.

Kapetanios (2004) suggested the following method for determining the number of factors based on RMT under the assumption that $N/T \to \gamma$ . First determine an a priori parameter $d$ and compute $b = \left(1 + \sqrt{\gamma}\right)^2 + d$ . Recall that $\left(1 + \sqrt{\gamma}\right)^2$ is the a.s. limit of the largest eigenvalue of a white Wishart matrix. Normalize the data so that they have unit variance and compute the eigenvalues

$\lambda_1 \geq \cdots \geq \lambda_N$ of the covariance matrix of normalized data. Ensure that the data support a factor structure, that is, that there is at least a pervasive factor, by checking that $\lambda_1 > b$. If this is the case, compute the first principal component, regress the data on the first principal component and repeat the procedure on residuals. If the test fails after *r* + 1 steps, then there are *r* factors. The parameter *d* is not determined formally though the author suggests to set *d* equal to the average eigenvalue of the covariance matrix of normalized data. If data are normalized, *d* = 1. In section 4 of Kapetanios (2004), this methodology is extended to data whose covariance matrix is non-diagonal provided that the average of covariances is zero.

Onatsky (2005, 2006 a,b) proposed the following family of estimators of the number of factors of an approximate factor model:

$$\hat{r}_\delta = \#\{i \leq n : \lambda_i > (1+\delta)\hat{u}\}$$

where $\delta$ is a positive fixed real number and $\hat{u} = w\lambda_{r_{\max}+1} + (1-w)\lambda_{2r_{\max}+1}$, and $w = 2^{\frac{2}{3}}\left(2^{\frac{2}{3}} - 1\right)$ and $r_{\max} = \min\left(N^\alpha, T^\alpha\right), 0 < \alpha < 1$. In other words, the number of factors is the number of eigenvalues that are larger than $(1+\delta)\hat{u}$. This estimator is based on the fact - established in the same paper - that if the right edge of the support of the distribution of eigenvalues is $\hat{u}$, then for any sequence of eigenvalues $\lambda_{j(T)}$ such that $\frac{j(T)}{T} \to 0$ then $\lambda_{j(T)} \to \hat{u}$. This property suggests that $\lambda_{r_{\max}+1}$ is a consistent estimator of $\hat{u}$; the particular form of the estimator suggested in Onatski (2006b) is intended to improve small-sample performance.

Onatski (2008) proposed a test of the null that the number of factors is $k_0$ against the alternative hypothesis that it is larger than $k_0$ but less than $k_1 > k_0$. The test statistics is:

$$\max_{\lambda_{k_0} < \lambda_k < \lambda_{k_1}} \frac{\lambda_k - \lambda_{k+1}}{\lambda_{k+1} - \lambda_{k+2}}.$$

Onatski computes the critical values of the test statistics based on the joint distribution of the last *k* eigenvalues established in El Karoui (2006) and extended to singular Wishart matrices in Onatski (2007 b).

Onatski (2007) studied "weak" factors. The paper defines weak factors as factors associated with bounded eigenvalues. It showed that principal components are inconsistent estimators of weak factors and quantifies the amount of inconsistency.

Harding (2008 a) explained the "single factor bias" described in the literature. Many authors, in particular Brown (1989), observed that there is a bias towards identifying a single market factor, while weaker factors cannot be disentangled from noise. Assuming "weak" factors, Harding (2008a) uses results from the distribution of eigenvalues in spiked models to justify this effect.

Harding (2008b) proposed a methodology for identifying the number of factors based on the distribution of the moments of the distribution of eigenvalues. He described an identification strategy based on the fact, proved in Bai and Silverstein (2004), that the moments of the distribution of eigenvalues satisfy a central limit theorem.

## *2.8. Dynamic Factor Models*

Dynamic factor models are models that allow to specify a dynamics for factors and for the processes themselves. Dynamic factor models now have important applications outside the area of financial econometrics, for example in ecological studies (see, for example, Zuur, Tuck and Bailey, 2003). The development of dynamic factor models is recent in comparison with static factor models. While modern static multi-factor models were proposed by Thurstone and Hotelling in the 1930s, the first dynamic factor models were proposed in econometrics only in 1977 by Geweke (1977) and by Sargent and Sims (1977). The subsequent development of dynamic factor models followed three lines: 1) dynamic factor models of stationary processes in the "finite *N*, large (infinite) *T*" case, 2) dynamic factor models of stationary processes in the "large (infinite) *N*, large (infinite) *T*" case, and 3) dynamic factor models of integrated processes. The literature on dynamic factor models of integrated processes overlaps with the large literature on cointegration.

Dynamics enter factor models in three different ways: 1) specifying a dynamics for the factors, 2) specifying a dynamics for the residuals, and 3) allowing regression on lagged factors. Dynamics is typically specified as an autoregressive process.

I will start to review the literature on dynamic factor models with models involving a small number of variables and a number of observations that tends to infinity. Dynamic models of this type are instances of state-space models (see Lutkepohl, 1991). Estimation of these models is achieved either with maximum likelihood and the Kalman filter or in the frequency domain.

Sargent and Sims (1977) and Geweke (1977) proposed a dynamic factor model of the type:

$$r_t = \sum_{i=0}^{\infty} \beta_i f_{t-i} + \varepsilon_t$$

where returns are an $N \times 1$ vector, the $\beta_i$ are $N \times Q$ matrices, $f_t$ is a $K \times 1$ vector for each $t$ and $\boldsymbol{\varepsilon}_t$ is a $N \times 1$ vector. It is assumed that *N* is finite, *K*<<*N* and $T$ tend to infinity. It is also assumed that factors and residuals are uncorrelated and that residuals are mutually uncorrelated though possibly autocorrelated. This model is the dynamic equivalent of the strict factor model.

Estimation is performed with maximum likelihood in the frequency domain. The number of factors is determined with a likelihood ratio test.

Engle and Watson (1981), Sargent (1989), and Stock and Watson (1991) proposed similar models of a small number of variables. In these papers, estimation of the model parameters is performed in the time domain with maximum likelihood and factors are recovered with the Kalman filter. Quah and Sargent (1993) studied larger models ($N$ up to 60) using the Expectation Maximization algorithm.

Peña and Box (1987) studied the following more general model:

$$\begin{aligned} r_t &= \beta f_t + \varepsilon_t \\ \Phi(L) f_t &= \Theta(L)\eta_t \\ \Phi(L) &= I - \Phi_1 L - \cdots - \Phi_p L^p \\ \Theta(L) &= I - \Theta_1 L - \cdots - \Theta_q L^q \end{aligned}$$

where factors are stationary processes, $L$ is the lag operator, $\boldsymbol{\varepsilon}_t$ is white noise with a full covariance matrix but is serially uncorrelated, $\eta_t$ has a full-rank covariance matrix and is serially uncorrelated and $\boldsymbol{\varepsilon}_t$ and $\eta_t$ are mutually uncorrelated at all lags. That is, the common dynamic structure comes only from the common factors while the idiosyncratic components can be correlated but no autocorrelation is allowed.

Peña and Box (1987) proposed the following methodology for determining the number of factors and estimating the factors. Assume that factors are normalized through the identification conditions $\beta'\beta = I$. Consider the covariance matrices $\Gamma_r(k) = E(r_t r_{t-k}), k = 0,1,2,\ldots$ and $\Gamma_f(k) = E(f_t f_{t-k}), k = 0,1,2,\ldots$. The following relationships hold:

$$\begin{aligned} \Gamma_r(0) &= \beta\Gamma_f(0)\beta' + \Sigma_\varepsilon, k = 0 \\ \Gamma_r(k) &= \beta\Gamma_f(k)\beta', k \geq 1 \end{aligned}$$

Compute the eigenvalues and eigenvectors of $\Gamma_r(k) \geq 1$. The number of factors is the common rank $Q$ of the matrices $\Gamma_r(k) \geq 1$. Use the non-zero eigenvectors of $\Gamma_r(k) \geq 1$ to estimate the loading matrix $\beta$. Use the loading matrix to recover factors.

The setting of dynamic models discussed thus far is that of classical statistics: a fixed number of time series and a number of samples that tends to infinity. In a series of papers, Stock and Watson discuss the problem of forecasting a time series using a large number of predictors. This methodology is referred to as creating diffusion indexes from a large number of predictors.

The motivation for suggesting this procedure is the large number of variables available to macroeconomists. Stock and Watson observed that the availability of large number of observed time series ─ in the range of hundreds of series ─ makes it impossible to use the classical Vector Autoregressive (VAR) models used by macroeconomists to model a carefully selected number of variables. They advocated a different procedure based on constructing a number of "diffusion indexes" from a large number of observed series.

Stock and Watson (1998) introduced a static factor model with an infinite $N$ and an infinite $T$. The authors observed that this model is compatible with a dynamic factor model with a finite number of lags, but not with an infinite number of lags. As discussed in Section 2.7, Stock and Watson (1998) demonstrated that in the limit $N,T \to \infty$, factors can be estimated with principal components. Therefore, any dynamic factor model with a finite number of lags can be put in a static form and estimated with principal components.

Principal components do not disentangle factors from their lagged copies. Stock and Watson (1998) suggested estimating the number of factors with information criteria. The model is used to forecast one variable that is regressed on lagged factors, hence there is no need to forecast factors. The paper demonstrated that "feasible forecasts", that is, forecasts based on factors estimated with principal components, asymptotically coincide with the "unfeasible forecasts" performed using the unknown true factors.

Forni, Hallin, Lippi, and Reichlin (2000), introduced the *generalized dynamic factor model*, which is a model with $N,T \to \infty$ and a finite number $Q$ of factors but allowing an infinite number of lags. Factors are assumed to be orthonormal white noise and factor loadings are assumed to be constant in time. The idiosyncratic components are possibly correlated and autocorrelated but uncorrelated with factors at every lag. The major difference with respect to the model described in Stock and Watson (1998) is the allowance of an infinite number of lags and the imposition of constant factor loadings.

Consider the spectral density matrix of the returns and of the idiosyncratic components. Call dynamic eigenvalues the eigenvalues of the spectral density at each frequency. Forni, Hallin, Lippi and Reichlin (2000) assumed that the first $Q$ dynamic eigenvalues diverge while the first dynamic eigenvalue of the idiosyncratic components is uniformly bounded. These conditions are the dynamic equivalent of the conditions on the eigenvalues of an approximate factor model. The authors estimated the model computing principal components in the frequency domain. Forni, Hallin, Lippi, and Reichlin (2004) determined the rates of convergence in function of the convergence path $N = N(T), T \to \infty$.

Thus far I have discussed two major methodologies for estimating dynamic factor models: maximum likelihood in the classical small $N$ and $T \to \infty$ factor model applied either in the frequency domain in Geweke (1977), Sargent and Sims (1977), and Peña and Box (1987) or in the time domain in Engle and Watson (1981), Sargent (1989), Stock and Watson (1991) and Quah and Sargent (1993) and principal components in the $N, T \to \infty$ case applied in the time domain in Stock and Watson (1989) and in the frequency domain in Forni, Hallin, Lippi, and Reichlin (2000, 2004).

Doz, Giannone, and Reichlin (2006) reconciled these two approaches. Their paper demonstrated that, under the same assumptions as in Stock and Watson (2002 a,b), a dynamic factor model can be estimated with quasi-maximum likelihood. The basic idea in Doz, Giannone, and Reichlin (2006) was to estimate a dynamic factor model with maximum likelihood and the Kalman filter as a misspecified exact factor model and to show that the error vanishes asymptotically.

Heaton and Solo (2003 and 2006) reconciled the small *N* and the large *N* approaches by introducing the signal-to-noise ratio. The setting of the paper is the same as in Stock and Watson (1998), that is, forecasting a variable using a small number of diffusion indexes. They assumed a fixed *N* and determined the bounds on the forecasting error in function of the signal-to-noise ratio when factors are approximated with principal components.

### *2.9 Dynamic factor models of integrated processes*

I will now review the literature on dynamic factor models of integrated processes. The notion of a factor model of integrated processes is rooted in the concept of cointegration. Following Granger and Engle, who were jointly awarded the 2003 Nobel Memorial Prize in Economic Sciences for the discovery of cointegration and autoregressive conditional heteroskedasticity (ARCH) behaviour, two or more integrated time series are cointegrated if there is a linear combination $\sum_{i=1}^{N} \alpha_i x_{it}$ of the series that is stationary. The linear combinations $\sum_{i=1}^{N} \alpha_i x_{it}$ that are stationary are called cointegrating relationships.

As observed in Galeano and Peña (2000), the idea that two or more time series can be individually integrated but that a linear combination of the series is stationary had already been put forward by Box and Tiao (1977) in introducing canonical correlation analysis. There is a vast literature on cointegration and on determining the number of cointegrating relationships. The state-of-the-art cointegration test is the Johansen test. Johansen (2000) and Hendry and Juselius (2000) offer a concise presentation of cointegration.[4]

[4] See Hendry (1995**)**.

The first link between cointegration and dynamic factor models appeared in Stock and Watson (1988). This landmark paper demonstrated that if a set of $N$ time series is cointegrated with $K$ cointegrating relationships, then there are $Q=N-K$ integrated common trends and the $N$ series can be described as regressions on the common trends. The common trends are obtained performing a generalized principal components analysis, that is, the $Q$ common trends are determined by the eigenvectors corresponding to the $Q$ largest eigenvalues of the generalized covariance matrix $\Omega = \frac{1}{T}\left(X - \overline{X}\right)'\left(X - \overline{X}\right)$. Escribano and Peña (1994) established that common trends are equivalent to common dynamic factors in the sense that the statement that there are $K$ cointegrating relationships is equivalent to the statement that data can be represented by $N$-$K$ dynamic factors.

Peña and Poncela (2004a) generalized the methodology put forward in Peña and Box (1987). They introduced a generalized covariance matrix for integrated processes and showed that a procedure similar to the analysis in the frequency domain holds also for integrated processes. The paper proposed a test for the number of common factors based on analyzing the eigenvalues of the generalized covariance matrices. Factors are estimated with maximum likelihood. Peña and Poncela (2004b) analyzed the forecasting performance of dynamic factor models with possibly integrated factors. Pesaran and Shin (1997) presented a theory of *autoregressive distributed lag* (ARDL) models where the regressors are time series that are integrated but not cointegrated.

### *2.10. Application of RMT to dynamic factor models*

The application of RMT to dynamic factor models requires the consideration of asymmetric matrices. The literature on this topic is much less developed that the literature on applications of RMT to static factor models. Kwapie, Drożdz, Górski, and Oświęcimka (2006) apply RMT to the problem of understanding the auto cross correlations between the returns of stocks in the Dow Jones Industrial Average (DJIA) in the United States and the returns of stocks in the Deutscher Aktien IndeX 30 (DAX30) in Germany. The paper considers returns computed over very short time intervals, from 3 to 120 seconds, and considers lags between 0 to 5 minutes. The empirical finding is that all eigenvalues but one stay confined in the Ginibre Orthogonal Ensemble (GinOE) theoretical distribution for time lags less than 5 minutes. After 5 minutes, all eigenvalues are in the GinOE theoretical distribution, which signals that after a 5-minute lag, the two universes, the DJIA and the DAX30, have no mutual influence.

Bouchaud *et al.* (2005) propose a different analysis. In their paper, they compute the asymptotic distribution of the singular values of the covariance matrix between two sets of $M$

random processes *X* and *N* processes *Y* when *M, N* and the number of observations *T* tends to infinity. In order to avoid mixing correlations with cross autocorrelations, both series *X,Y* are orthogonalized and normalized before computing the singular values. It is well known that the singular values measure the strength of correlation of the corresponding canonical correlations. Therefore, the authors essentially find the asymptotic distribution of the canonical correlation coefficients between uncorrelated processes. Because canonical correlation coefficients are obtained by optimizing two portfolios formed with the *X* and the *Y*, this methodology exhibits one Dirac's delta at zero and one Dirac's delta at one if the ratio *M/T, N/T, (M+N)/T* is less than one. This is because, in this case, a fraction of correlations can be set algebraically to 1 or to zero. Therefore to be viable, this method requires a large number of samples in comparison with the number of time series.

### *2.11. Application of information theory to factor models*

One could reasonably suppose that information theory would have a strong impact on the development of finance theory: after all, information is a compact, coherent measure on the amount of uncertainty carried by a probability distribution. However this is not the case and financial applications of information theory are still relatively rare.[5] Part of the problem is due to the fact that the concept of information can be consistently defined only for discrete distributions.

Given a discrete probability distribution $p_i, i = 1,\ldots,N$, Shannon (1948a, 1948b) defined the amount of information carried by the distribution as $I = \sum_{i=1}^{N} p_i \log(p_i)$. The quantity *I* can be interpreted as an amount of information. It reaches a maximum at zero when one probability is 1 and it has a minimum at $I = \log\left(\frac{1}{N}\right)$ when all probabilities are equal. There is no lower bound to information, as $\log\left(\frac{1}{N}\right) \to \infty$ when $N \to \infty$.

Consider now two distributions, $p_i, q_i, i = 1,\ldots,N$. The Kullback-Leibler divergence is defined as follows:

$$I = \sum_{i=1}^{N} p_i \log\left(\frac{p_i}{q_i}\right).$$

[5] See Dionísio, Menezes, and Mendes (2005) for a discussion on the use of entropy as a measure of uncertainty in finance.

However, the extension of the quantity $I$ to a continuous setting is not straightforward. The obvious extension of the definition of information to a continuous probability distribution would be to replace summation with integral and to define $I = \int p \log(p) dp$ . But this is not a sound definition as the integral might diverge. However, the Kullback-Leibler divergence can be extended to the continuous case defining

$$I(f,g) = \int p(x) \log\left( \frac{p(x)}{q(x)} \right) dx .$$

Otter and Jacobs (April 2006) and Otter and Jacobs (July 2006) use the Kullback-Leibler divergence to determine the number of factors and, more in general, to determine the amount of information in a set of data. The starting point of their analysis is the information contained in a covariance matrix, which is proportional to the logarithm of the matrix determinant. Hence, the information contained in a matrix is proportional to the sum of eigenvalues. Otter and Jacobs (April 2006) and Otter and Jacobs (July 2006) compute the Kullback-Leibler divergence between the distribution of eigenvalues of two matrices. Based on this basic computation, these papers derive criteria for determining the number of factors by computing the Kullback-Leibler divergence between the empirical distribution of eigenvalues and a diagonal matrix.

### *2.12. Sparse Principal Components Analysis*

The survey of the literature on factor models has shown the importance of PCA in factor analysis. Summarizing the conclusions from the literature, principal components are consistent estimators of factors in scalar strict factor models, in infinite strict factor models, in infinite approximate factor models, and in finite dynamic factor models. In addition, principal components accurately approximate factors in finite factor models if the signal-to-noise ratio is sufficiently large.

From the point of view of financial applications, principal components are portfolios formed with all the assets present in the market. This is an inconvenience as many if not most assets will only marginally contribute to principal components. The basic idea of *sparse principal components analysis* (SPCA) is to constrain principal components so that they are formed by only a small number of assets.

However, while PCA is easy to compute numerically, SPCA is a difficult combinatorial problem, as shown by Moghaddam, Weiss, and Avidan (2006) using results from Natarajan (1995). Systematic approaches to the problem of computing SPCA are based on nonconvex algorithms.[6]

[6] SCoTLASS by Jolliffe (2003) and SLRA by Zhang *et al.* (2002). Zou 2006 allows the application of the LASSO algorithm proposed by Tibshirani (1996). The LASSO algorithm is a penalization technique. Johnstone and Yu Lu (2004) proposes a simple algorithm for nonfinancial applications. D'Aspremont, Bach and El Ghaoui (2007) propose an efficient greedy algorithm for SPCA.

# 3. Mathematical Methods

In this chapter I will review the mathematical methods used in Random Matrix Theory (RMT) and in Dynamic Factor Analysis (DFA). I will first review the classical methods of factor analysis based on maximum likelihood and then introduce different versions of principal components analysis and independent components analysis. Next I will introduce the methods used in RMT where new, elegant, and simpler methods have recently been introduced. After briefly discussing methods used in cointegration analysis, I will conclude with a discussion on the application of the above methods to DFA.

As discussed in the literature review in Chapter 2, there are different strains of literature and research in factor analysis, with paths from static to dynamic factors, from small *N* to large *N* models, and from stationary to integrated variables. In moving from small *N* to large *N* models, there is a significant theoretical simplification insofar as robust methods such as PCA and singular value decomposition can be used. However, this simplification comes at a price: when the number of time series approaches the number of observations, estimation becomes more problematic as there is an explosion of the number of parameters to estimate per observation. The literature that I surveyed in Chapter 2 found many important asymptotic results that solve the problem of separating meaningful estimations from random noise, for example, identifying true factors.

However when asymptotic results are applied to finite samples, it is difficult to separate information from randomness because there might be a large amount of apparent randomness generated by the finite size of the sample. One of the objectives of this dissertation is to identify additional criteria that allow one to use asymptotic results in finite samples. To this end, I need to show what conditions are critical for obtaining the sample results. Hence I need to discuss the key methods used to obtain asymptotic results.

## *3.1 Classical methods of factor analysis*

Classical factor analysis applies to static strict factor models under the assumption that the number of variables *N* (time series of returns) is fixed, while the number of observations *T* is allowed to grow to infinity. Classical factor analysis is based on maximum likelihood estimates and on distance minimization, that is, generalized least squares as described, for example, in Anderson (2003). There are two types of identification issues related to classical strict factor models: identification of the model's parameters and identification of factors. Let's first address the question of the identification and the estimation of the model's parameters.

Consider a strict factor model: $R = F\Lambda + U$ where $R$ and $U$ are $T \times N$ matrices, $\Lambda$ is a $N \times Q$ matrix of factor loadings and $F$ is a $T \times Q$ matrix. Each row of the matrix $R$ contains an observation of $N$ returns. We assume that the number of factors is known. To make the model identifiable, assume that factors are orthonormal (i.e., uncorrelated with unit variance) and that factors and residuals are mutually uncorrelated, so that the covariance matrix of returns can be written as $\Sigma = \Lambda\Lambda' + \Psi$ where $\Psi$ is a diagonal matrix. The model is still not completely identified under these assumptions as factors can be rotated, that is, multiplied by an orthogonal matrix.

Let's assume returns are normalized by subtracting the mean, and that returns, factors and residuals are normally distributed. The likelihood function can then be written as follows:

$$L = (2\pi)^{-\frac{1}{2}TN} |\Sigma|^{-\frac{1}{2}N} \exp\left( -\frac{1}{2}\sum_{t=1}^{T} r_t' \Sigma^{-1} r_t \right)$$

This function is called the *concentrated likelihood* because unobserved factors and residuals have been *concentrated out* and thus do not appear. The log-likelihood function is:

$$\log L = -\tfrac{1}{2}TN \log(2\pi) - \tfrac{1}{2}T \log|\Sigma| - \frac{1}{2} trace\left(A\Sigma^{-1}\right)$$

where $A = R'R$ is $T$ times the sample covariance matrix. In order to estimate $\Lambda, \Psi$ the concentrated log likelihood has to be maximized with respect to $\Lambda, \Psi$, imposing the relationship $\Sigma = \Lambda\Lambda' + \Psi$. Maximization cannot be achieved explicitly as the log likelihood is a highly non-linear function. Numerical methods are therefore called for. The Expectation-Maximization (EM) algorithm maximizes the log likelihood looking at factors such as missing data.

Maximum likelihood estimation allows one to estimate the model's parameters. Anderson (2003) proves that maximum likelihood cannot be used to estimate factors. If $\Lambda, \Psi$ are known, factors can be estimated using the equation $R = F\Lambda + U$ as a cross sectional generalized regression. However, regression is not unique and different factors are compatible with the same parameters $\Lambda, \Psi$.

Thus far it has been assumed that the number of factors is known. In practice, however, the number of factors is not known *a priori*, but needs to be determined. One way to determine the number of factors in classical factor analysis is to perform a likelihood ratio test on a growing number of factors. If the test is passed for the first time for a number $Q$ of factors that is reasonably small, then it is assumed that data have a factor structure and that $Q$ is the true number of factors. Otherwise it is possible to conclude that the data do not have a non-trivial linear factor structure.

The likelihood test is written as the ratio between the unconstrained likelihood and the likelihood under the assumption of $Q$ factors:

$$\frac{\max_{\Lambda,\Psi} L(\Lambda\Lambda' + \Psi)}{\max_{\Sigma} L(\Sigma)} = \frac{|C|^{\frac{1}{2}T}}{\left|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}\right|^{\frac{1}{2}T}}$$

Note that the perspective of classical factor analysis is to continue adding factors until the model fits under the assumption of uncorrelated residuals.

Anderson (2003) proves that if the empirical covariance matrix **C** is a consistent estimator of the true covariance matrix $\Lambda\Lambda' + \Psi$, then the empirical likelihood, normalized dividing by $N$, converges to the true limit normalized likelihood. It is also proved that, under the same hypotheses, the maximum likelihood estimators $\hat{\Lambda}, \hat{\Psi}$ converge to $\Lambda, \Psi$ and that $\sqrt{N(\hat{\Lambda} - \Lambda)}$ and $\sqrt{N(\hat{\Psi} - \Psi)}$ have a limiting normal distribution.

The maximum likelihood method requires that the distribution function be known. Methods based on distance minimization do not require specific assumptions as regards the distribution. One method consists in minimizing the distance between the empirical covariance matrix and the matrix $\Lambda\Lambda' + \Psi$. This is a generalized least squares problem as observations are not independent.

### *3.2 Factor uniqueness*

Looking at the literature, there is a proliferation of different factor models of stock returns determined with different techniques and including a different number of factors. While the original work of Ross (1976) and then of Chamberlain and Rothschild (1983) and Rothschild (1983) used strict and approximate factor models in the sense defined above, many other academic works proposed different methodologies. To cite a few, consider the fundamental models introduced by Rosenberg (1974). In Rosenberg's fundamental factor model, factor loadings are predetermined and factors are recovered through cross sectional regressions. Among the fundamental models, perhaps the best known is the Fama-French three-factor model (Fama and French, 1993). Carhart (1997) and Jagadeesh and Titman (1993, 2001) added a momentum factor to the Fama-French factor model. A variety of country and industry sector models have also been proposed, with papers showing that sector factor models have outperformed country factor models in recent years. A variety of commercial models are also available.

How should we interpret this variety of linear factor models for describing returns? Are they all equivalent models, each model implementing a different estimation process? Or are there genuine differences among these models such that they are not equivalent? If returns can be described by a true approximate linear factor model, then different factor models, if correctly specified, are equivalent among themselves and they are all equivalent to a PCA-based model up to rotations and within the fluctuations implicit in finite samples.

If, on the other hand, no true approximate linear factor model describes returns, for example because of non-linearities or due to dynamic effects, then linear factor models are approximations —not in the sense of approximate factor models but in the sense that factor models are mis-specified— and might be genuinely different. Otherwise formulated, is there a true factor model that represents returns or are there many different approximate regressions of returns on different factors?

In classical factor models, there is a fundamental indeterminacy in factor models. Steiger (1979 and 1996) and Steiger and Schonemann (1978) provide an historical account as well as a statement of the problem of factor indeterminacy in factor analysis. As described in these papers, there have been periods of interest and debate on the problem of factor indeterminacy. As mentioned above, factor models were originally introduced by Spearman in psychometrics. Spearman (1904) introduced a single factor model where the factor, called g, is identified as the general intelligence. Spearman believed that all human abilities are determined primarily by the intelligence factor g. Attributing much importance to his discovery, Spearman believed that the measurement of intelligence and human abilities opened a new era where Governments would be able to measure the intelligence of every child and select the optimal career path suited to each individual. Factor indeterminacy was a blow to these claims and therefore produced a heated debate.

Essentially the problem can be stated as follows. If there are a sufficient number of factors, classical factor models are identifiable in the sense that the parameters of the model, ultimately the factor loadings and the covariance matrix of residuals, are uniquely determined. In fact, consider the covariance matrix decomposition: $\varOmega = BB' + \varSigma$, where $\varOmega$ is the covariance matrix of the observed variables, $\mathbf{\Sigma}$ is the diagonal covariance matrix of the residuals and $B$ is the matrix of factor loadings. If there are $m$ observed variables and $p$ factors, there are $\frac{1}{2}m(m+1)$ equations plus $p^2$ equations due to orthonormality constraints while we have to determine $m \times k$ factor loadings plus $m$ residual variances. The model is identified if the number of conditions exceeds the number of parameters. Therefore, if the model is identifiable, we can determine the model parameters by maximum likelihood or by generalized least squares (Anderson 2003).

However, it is impossible to write down a maximum likelihood estimate of the model parameters plus the factors because the loglikelihood does not have a maximum (Anderson 2003). Even if the $B$ and $\Sigma$ are known, the factors are not uniquely determined. Factors are not, in general, linear combinations of the observed variables, that is, in financial terms factors are not, in general, portfolios. Guttman (1955) demonstrated that factors can be decomposed in the sum of two terms: the first term is a linear combination of observed variables, the second term is not a linear combination of variables and is arbitrary. Therefore, there is an essential arbitrariness in factor determination.

Schönemann [1971] derived a simplified formula for the minimum average correlation between equivalent orthogonal factors: $r = \frac{m - p}{m + p}$. In financial applications where $m$ is of the order of hundreds and $p$ is less than 10, the correlation between equivalent factors is very close to 1. It was observed that if the number of variables tends to infinity, equivalent factors become perfectly correlated. This is an early statement of the fact that factor models become perfectly identified in infinite markets.

### *3.3 Spectral analysis*

Let’s recall a few facts related to the analysis of time series in the frequency domain following Priestly (1983) and Cox and Miller (1977). The basis for spectral analysis is the Fourier series and the Fourier transforms. A periodic function $x(t)$ with period *2τ* can be represented as a Fourier series formed with a denumerably-infinite number of sine and cosine functions:

$$x(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\left( a_n \cos\left(\frac{\pi}{\tau}t\right) + b_n \sin\left(\frac{\pi}{\tau}t\right)\right).$$

This series can be inverted in the sense that the coefficients can be recovered as integrals through the following formulas:

$$\begin{aligned} a_n &= \frac{1}{\tau}\int_{-\tau}^{+\tau} x(t)\cos\left(\frac{\pi}{\tau}t\right)dt, \\ b_n &= \frac{1}{\tau}\int_{-\tau}^{+\tau} x(t)\sin\left(\frac{\pi}{\tau}t\right)dt \end{aligned}.$$

If the function $x(t)$ is square integrable, it can be represented as a Fourier integral:

$$x(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{i\omega t} F(\omega)d\omega$$

where the function $F(\omega)$ is called the Fourier transform of $x(t)$:

$$F(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-i\omega t} x(t)dt .$$

In both cases, periodic and nonperiodic, Parseval's Theorem holds:

$$\int_{-\infty}^{+\infty} x^2(t)dt = 2\pi\sum_{n=1}^{\infty} c_n^2, c_0 = \tfrac{1}{2}a_0, c_n = \sqrt{\tfrac{1}{2}\left(a_n^2 + b_n^2\right)}$$

$$\int_{-\infty}^{+\infty} x^2(t)dt = \int_{-\infty}^{+\infty} |F(\omega)|^2 d\omega$$

In engineering terms, Parseval's Theorem states that, if a signal has a finite energy $\int_{-\infty}^{+\infty} x^2 dt < \infty$ , then the energy can be expressed in terms of the frequency spectrum. Therefore, a periodic signal or a non periodic square integrable signal can be represented either directly as a function of time (time domain) or through its spectral representation given by its Fourier transforms. The knowledge of the spectrum allows one to completely recover the time signal and vice versa.

The above Fourier analysis applies to a deterministic function $x(t)$. Suppose now that $x(t)$ is a univariate stationary stochastic process in continuous time $x(t)$. A stochastic process is a set of paths. As the process is infinite and stationary, its paths are not periodic, they do not decay to zero when time goes to infinity, and they cannot be square integrable as functions of time. Therefore the paths of the series cannot be expressed either as Fourier series or as Fourier integrals. However it is still possible to recover a spectral representation of a stationary stochastic process as stochastic integrals. In fact, the *Cramer representation* of a stationary process represents a stationary process as a stochastic integral. Consider first continuous time processes. Given a zero-mean, (stochastically continuous) stationary process $x(t), -\infty < t < +\infty$ , then there is an orthogonal process $Z(\omega), -\infty < \omega < +\infty$ such that, for all $t$, the following representation holds:

$$x(t) = \int_{-\infty}^{+\infty} e^{it\omega} dZ(\omega)$$

where the integral is defined in the mean-square sense (stochastic integral). The process $Z(\omega)$ is a stochastic process with the following properties:

$$E[dZ(\omega)] = 0, \forall \omega$$
$$E\left[|dZ(\omega)|^2\right] = dH(\omega), \forall \omega$$
$$\operatorname{cov}[dZ(\omega)dZ(\omega')] = E\left[dZ(\omega)^* dZ(\omega')\right] = 0, \forall \omega, \omega', \omega \neq \omega'$$

Note that $\int_{-\infty}^{+\infty} e^{it\omega} dZ(\omega)$ is not a Fourier integral and therefore the relationship $x(t) = \int_{-\infty}^{+\infty} e^{it\omega} dZ(\omega)$ cannot be inverted as a Fourier integral. However it can be proved that the following holds:

$$Z(\omega_2) - Z(\omega_1) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\frac{e^{-it\omega_2} - e^{-it\omega_1}}{-it}\right)x(t)dt$$

The spectral representation carries over to discrete series with only minor modifications. It requires changing the integration limits from $-\infty, +\infty$ to $-\pi, +\pi$ and dropping the condition that the process be stochastically continuous, a condition which would be meaningless in the discrete case. We can therefore establish that, given a zero-mean stationary process $x(t), t = 0, \pm 1, \pm 2, \ldots$, there is an orthogonal process $Z(\omega)$ such that, for all $t = 0, \pm 1, \pm 2, \ldots t$, the following representation holds in the interval $-\pi, +\pi$ :

$$x(t) = \int_{-\pi}^{+\pi} e^{it\omega} dZ(\omega)$$

I will now consider the power of the signal and the power spectra. Given a stationary series (signal), its energy (i.e., the integral of its square) is infinite, but the power of the series (i.e., its energy divided by time) might tend to a finite limit. Suppose that the functions $x(t)$ are truncated at $\pm S$ and define the function $x_S(t)$ such that $x_S(t) = x(t)$ for $-S \leq t \leq +S$ and zero elsewhere. Call $F_S(\omega)$ the Fourier transform of $x_S(t)$. The functions $\frac{|F_S(\omega)|^2}{2S}$ are different for each path of the process. Let's average and take the limit for *S* that tends to infinity, defining the function:

$$h(\omega) = \lim_{S \to \infty}\left[E\left(\frac{|F_S(\omega)|^2}{2S}\right)\right].$$

The function $h(\omega)$ is called the *non-normalized power spectral function of* $x(t)$. The integral $H(\omega) = \int_{-\infty}^{\omega} h(\omega)d\omega$ is called the *integrated spectrum.*

I now consider the autocovariance function of the process $x(t)$ defined as:

$$R(\tau) = E[x(t)x(t-\tau)].$$

It can be demonstrated (see Priestly, 1983) that the function $h(\omega)$ is the Fourier transform of the autocovariance function: $h(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\omega\tau} R(\tau)d\tau$ and, therefore, $R(\tau) = \int_{-\infty}^{+\infty} e^{i\omega\tau} h(\omega)d\omega$ . (Note that in this definition I adopt the convention of apportioning the factor $\frac{1}{2\pi}$ only to one integral.) The demonstration hinges on the fact that the Fourier transform of a convolution of two functions is the product of the respective Fourier transforms.

Define $\sigma_x^2 = R(0) = \int_{-\infty}^{+\infty} h(\omega)d\omega$ . From this definition it is clear that $\sigma_x^2$ is the total power of the process but it is also the variance of the process. Define the normalized power spectral density function as the power spectral density divided by the variance of the process:

$$f(\omega) = {h(\omega)}/{\sigma_x^2}$$

From these definitions, I obtain that the normalized power spectral density function is the Fourier transform of the autocorrelation function of the process defined as:

$$\rho(\tau) = {R(\tau)}/{\sigma_x^2},$$

that is, I obtain that:

$$f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\omega\tau}\rho(\tau)d\tau \text{ and } \rho(\tau) = \int_{-\infty}^{+\infty} e^{i\omega\tau} f(\omega)d\omega .$$

The integral $F(\omega) = \int_{-\infty}^{\omega} f(\omega)d\omega$ is called the *normalized integrated spectrum.*

The integrated spectra are more general than the spectral density function insofar as the integrated spectra might exist even if the density does not. It is possible to generalize the previous definitions using the Stieltjes integrals and the Fourier-Stieltjes transforms:

$$\rho(\tau) = \int_{-\infty}^{+\infty} e^{i\omega\tau} dF(\omega).$$

The Wiener-Kintchine theorem states that a necessary and sufficient condition that $\rho(\tau)$ be the autocorrelation function of some continuous-time stationary stochastic process is that there is a non decreasing function $F(\omega)$ such that $F(-\infty)=0$ and $F(\infty)=1$ and such that:

$$\rho(\tau)=\int_{-\infty}^{+\infty} e^{i\omega\tau}\,dF(\omega).$$

The same property can be established in terms of $R(\tau), H(\omega)$:

$$R(\tau)=\int_{-\infty}^{+\infty} e^{i\omega\tau}\,dH(\omega).$$

Note that the inverse formulas do not hold in general as the functions *h* and *f* do not necessarily exist. However, if *f, h* do exist, then $dH(\omega)=h(\omega)d\omega$ and $dF(\omega)=f(\omega)d\omega$ , the usual Riemann integrals can be used, and the previous invertible Fourier transforms hold.

Thus far I have outlined the spectral theory for continuous-time processes. There is a parallel theory for discrete-time series. Consider a stationary time series $x(t), t=0,\pm 1,\pm 2,\ldots$ For discrete processes of this type, the autocovariance and autocorrelation functions are defined only for integer values; integration has to be replaced with summation and integrals must be taken in the interval $[-\pi,+\pi]$. The latter condition is related to the Nyqist-Shannon Sampling Theorem which states that discrete sampling can reveal a spectrum whose size is half the sampling frequency.

For a stationary discrete-time series, the Wiener-Kintchine theorem is replaced by Wold's theorem. Wold's theorem states that a necessary and sufficient condition for the sequence $\rho(t), t=0,\pm 1,\pm 2,\ldots$ to be the autocorrelation function of a discrete stationary process $x(t), t=0,\pm 1,\pm 2,\ldots$ is that there is a non decreasing function $F(\omega)$ such that $F(-\pi)=0$ and $F(+\pi)=1$ and such that:

$$\rho(r)=\int_{-\pi}^{+\pi} e^{i\omega r}\,dF(\omega).$$

Assuming that the function $F(\omega)$ is differentiable and $\frac{dF(\omega)}{d\omega}=f(\omega)$, I can write:

$$\rho(r)=\int_{-\pi}^{+\pi} e^{i\omega r} f(\omega)\,d\omega .$$

This relationship can be inverted in terms of a Fourier series:

$$f(\omega)=\frac{1}{2\pi}\sum_{r=-\infty}^{r=+\infty}\rho(r)e^{-i\omega r}$$

If the series $x(t), t=0,\pm 1,\pm 2,\ldots$ is real-valued, then $\rho(r)$ is an even sequence and I can write:

$$f(\omega)=\frac{1}{2\pi}\sum_{r=-\infty}^{r=+\infty}\rho(r)\cos r\omega=\frac{1}{2\pi}+\frac{1}{\pi}\sum_{r=1}^{r=+\infty}\rho(r)\cos r\omega .$$

Similar relationships can be established for covariances. In particular:

$$R(r) = \int_{-\pi}^{+\pi} e^{i\omega r} dH(\omega)$$

and if there is a density $\frac{dH(\omega)}{d\omega} = h(\omega)$, the previous formula becomes:

$$R(r) = \int_{-\pi}^{+\pi} e^{i\omega r} h(\omega) d\omega$$

This expression can be inverted:

$$h(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{r=+\infty} R(r) e^{-i\omega r}$$

If the time series is real-valued:

$$h(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{r=+\infty} R(r) \cos r\omega = \frac{\sigma_x}{2\pi} + \frac{1}{\pi} \sum_{r=1}^{r=+\infty} R(r) \cos r\omega \ .$$

I can now establish the connection between the Cramer representation and the integrated spectrum. In fact, it can be demonstrated that the following relationship holds:

$$E\left[|dZ(\omega)|^2\right] = dH(\omega), -\pi \le \omega \le +\pi$$

Thus far I have discussed univariate processes. I will now discuss multivariate processes. Consider a complex-valued, vector-valued, jointly stationary, discrete-time series:

$$(X_{1,t}, \ldots, X_{N,t})', t = 0, \pm 1, \pm 2, \ldots$$

As the process *X* is assumed to be stationary, I can assume that each $X_{i,t}$ has zero-mean. The process is characterized by a covariance matrix at lag *s* defined as follows:

$$\mathbf{R}(s) = \{R_{i,j}(s)\}, i = 1, \ldots, N, j = 1, \ldots, N$$
$$R_{ij}(s) = E\left[X^*_{j,t} X_{i,t+s}\right]$$

As the process is assumed to be jointly stationary, $R_{ij}(s)$ depends only on *s* and not on *t.* If *i=j*, $R_{ii}(s)$ is the autocovariance function of each process $X_{i,t}$. I can therefore write the Cramer spectral representation of $X_{i,t}$ as a univariate process:

$$X_{i,t} = \int_{-\pi}^{+\pi} e^{it\omega} dZ_i(\omega)$$

where the $dZ_i(\omega)$ are orthogonal and cross orthogonal, that is, $E[dZ_i^*(\omega)dZ_i(\omega')] = 0$ for $\omega \neq \omega'$ and $E[dZ_i^*(\omega)dZ_j(\omega')] = 0$ for $\omega \neq \omega', i \neq j$. The spectral representation of the covariance matrix is:

$$R_{i,j}(s) = \int_{-\pi}^{+\pi} e^{is\omega} dH_{i,j}(\omega)$$
$$dH_{i,j}(\omega) = E[dZ_i(\omega)dZ_i^*(\omega)]$$

The matrix $H_{ij}(\omega)$ is called the *integrated spectral matrix* or *spectral distribution matrix*. If the $H_{i,j}(\omega)$ are differentiable, then $\frac{dH_{i,j}(\omega)}{d\omega} = h_{i,j}(\omega)$ holds. The matrix-valued function $h_{i,j}(\omega)$ is called the *spectral density matrix* and the following relationship holds:

$$R_{i,j}(s) = \int_{-\pi}^{+\pi} e^{is\omega} h_{i,j}(\omega)d\omega$$

The spectral density matrix can be inverted to yield:

$$h_{i,j}(\omega) = \frac{1}{2\pi}\sum_{s=-\infty}^{s=+\infty} R_{i,j}(s)e^{is\omega} .$$

The above formula establishes the relationship between the autocovariance matrix function $R_{i,j}(s)$ of the process *X* and the spectral density function. The latter can be thought of as the covariance function of the process *Z*. In summary, I can establish the correspondences between the representations in the time domain and those in the frequency domain as shown in Table 3.1.

| | **Time domain** | **Frequency domain** |
|---|---|---|
| Deterministic functions<br>Fourier analysis | $x_{i,t} = \int_{-\infty}^{+\infty} e^{it\omega} dF(\omega)$ | $= F(\omega) = \int_{-\infty}^{+\infty} e^{it\omega} x(t)dt$ |
| Stochastic processes | $X_{i,t}$, zero-mean, jointly stationary | $Z_i(\omega)$, orthogonal and cross orthogonal |
| Cramer representation | | $X_{i,t} = \int_{-\pi}^{+\pi} e^{it\omega} dZ_i(\omega)$ |
| | $R_{i,j}(s) = \int_{-\pi}^{+\pi} e^{is\omega} dH_{i,j}(\omega)$ | |
| $\frac{dH_{i,j}(\omega)}{d\omega} = h_{i,j}(\omega)$ | $R_{i,j}(s) = \int_{-\pi}^{+\pi} e^{is\omega} h_{i,j}(\omega)d\omega$ | $h_{i,j}(\omega) = \frac{1}{2\pi}\sum_{s=-\infty}^{s=+\infty} R_{i,j}(s)e^{is\omega}$ |
| | | $dH_{i,j}(\omega) = E[dZ_i(\omega)dZ_i^*(\omega)]$<br>$dh_{i,j}(\omega)d\omega = E[dZ_i(\omega)dZ_i^*(\omega)]$ |

*Table 3.1: Analysis of time series in the time and frequency domains.*

Cox and Miller (1977) offer some intuition on the analysis of time series in the time or frequency domain. The analysis in the time domain seeks a representation of a time series in terms of an infinite sequence of white noise $e_t$. In particular, the analysis in the time domain can be thought of as a regression with an infinite number of regressors. The analysis in the frequency domain seeks a representation in terms of an infinite sequence of terms $Ze^{i\omega t}$ where $Z$ is a random variable. In particular, a process $X$ is represented as a continuous sum of terms of the type $e^{i\omega t}dZ_i(\omega)$, that is, as a stochastic integral: $X_{i,t} = \int_{-\pi}^{+\pi} e^{it\omega}\, dZ_i(\omega)$.

*3.4 Principal components*

Consider the same setting as in Section 3.1. PCA solves the problem of finding linear combinations of the variables (returns) $\beta_i r_t, i = 1,\ldots,N$ that have maximum variance and that satisfy the condition: $\beta_i'\beta_i = 1, \beta_i'\beta_j = 0, j = 1,\ldots,i-1$, that is, portfolios that have maximum variance, are mutually uncorrelated, and such that the vector of loadings have length 1.

Let $\Sigma$ be the covariance matrix of returns. The PCA problem can be written as follows:

$$\beta_i' = \arg\max[\beta_i'\Sigma\Sigma_i]$$
$$s.t. \beta_i'\beta_i = 1, \beta_i'\beta_j = 0, j \neq i$$

This problem can be solved analytically as a constrained maximization problem. It can be demonstrated that the $\beta_i$ must satisfy the conditions $\Sigma\beta_i = \lambda_i\beta_i$ with the normalization $\beta_i'\beta_i = 1$, that is, the $\beta_i$ are the normalized eigenvectors of the covariance matrix up to a factor $\pm 1$. The corresponding eigenvalues $\lambda_i$ are the variances of the relative linear combinations. Collecting the $\beta_i$ as columns of a matrix *B***,** the following relationships holds: $B'B = BB' = I$ and $\Sigma B = \Lambda B$ or, equivalently, $B'\Sigma' = \Lambda$. The matrix $\Lambda$ is a diagonal matrix on whose diagonal the eigenvalues are in decreasing order.

The linear combinations $\beta_i r_t$ are called principal components: $\eta_i = \beta_i r_t$ or, in matrix form $P = B'r_t$ where the columns of *P* are the principal components. The following relationship holds:

$$r_t = Ir_t = BB'r_t = BP.$$

Multiplying by $\mathbf{\Lambda}$, principal components are normalized to have unitary variance: $r_t = (B_1\Lambda)^{\frac{1}{2}}\left(\Lambda^{-\frac{1}{2}}P\right)$, that is, variables can be completely recovered as linear combinations of normalized principal components. However, the interest for PCA resides not in this identity but in the possibility of using only a small number of principal components $Q<<N$. I partition the matrix $B$ and $P$ respectively in two matrices $B=[B_1 \quad B_2]$ and $P=[P_1 \quad P_2]$ where the $N \times Q$ matrix $B_1$ contains the first $Q$ eigenvectors and the $N \times (N-Q)$ matrix $B_2$ contains the remaining *N-Q* eigenvectors and the $N \times Q$ matrix $P_1$ contains the first $Q$ principal components and the $N \times (N-Q)$ matrix $P_2$ contains the remaining *N-Q* principal components. I then write the following relationship: $r_t = B_1P_1 + B_2P_2$. This relationship can be rewritten as the sum of two orthogonal components: $r_t = B_1P_1 + w$.

Consider the diagonal matrix $\Lambda_1$ having on the main diagonal the first $Q$ eigenvalues in decreasing order. The following relationship holds: $\Sigma B_1 = B_1\Lambda_1$ and $B_1\Sigma B_1 = \Lambda_1$. Multiplying the matrix of principal components by $\Lambda_1^{-\frac{1}{2}}$, the first $Q$ principal components are normalized to have variance one. It is therefore possible to represent variables in terms of the first $Q$ normalized principal components as follows: $r_t = \left(B_1\Lambda_1^{\frac{1}{2}}\right)\left(\Lambda_1^{-\frac{1}{2}}P_1\right) + w$. This relationship is formally similar to the relationship that defines a factor model.

Thus far principal components have been defined as linear combinations of the data variables (returns) with loadings given by the eigenvectors of the data covariance matrix. Principal components can be estimated computing the eigenvectors and eigenvalues of the empirical covariance matrix: $\frac{1}{T}R'R$. Dividing by the respective eigenvalues, principal components can be normalized to have variance one. I can now estimate principal components using singular value decomposition (SVD) working directly on the data matrix $R$. The SVD of $R$ is the following decomposition: $R = USV$ where $U$ and $V$ are the eigenvectors of the matrices $RR'$ and $R'R$ respectively and $S$ is the matrix of singular values which are the square roots of the eigenvalues of the covariance matrix of $R$. Writing $U = SV'R$ shows that $\Lambda^{-\frac{1}{2}}P = \frac{1}{T}U$.

After the introduction of PCA by Hoteling (1933), it was conjectured that principal components and factors are essentially the same thing. However, Anderson (2003) and Jolliffe (2002) observe that factor models and PCA are conceptually different for two main reasons. First, factor models assume a theoretical model for the data while principal components do not assume any model. Second, principal components are linear combinations of the data variables while factors might be non linear combinations of data. Schneeweiss and Mathes (1995) established that if the covariance matrix of residuals is $\Psi = \sigma^2 I$, then estimation with maximum likelihood and with principal components analysis yields the same result. This is true for the model parameters as factors are not uniquely determined.

### *3.5 Dynamic principal components*

Time series can be analyzed in the time domain as well as in the frequency domain through their spectral representations. Principal components analysis can be applied directly to the time series, as seen in section 3.3, or it can be applied to their spectral representation. In the latter case, principal components are called *dynamic principal components* (see Brillinger 1981).

Consider a complex-valued, vector-valued, zero-mean, jointly stationary, discrete time series: $(X_{1,t},\ldots,X_{N,t})', t = 0,\pm 1,\pm 2,\ldots$ and consider its spectral density $h_{i,j}(\omega)$. The spectral density is a matrix-valued function of $\omega$ . $h(\omega) = \{h_{i,j}(\omega)\}$. For each value of $\omega$ , the spectral density matrix has a set of *N* eigenvalues $\lambda_i(\omega), i = 1,\ldots,N$, and a set of *N* row eigenvectors $p_i(\omega), i = 1,\ldots,N$, which are $1\times N$ vectors such that: $p_i(\omega)h(\omega) = \lambda_i(\omega)p_i(\omega)$. The $\lambda_i(\omega)$ are referred to as *dynamic eigenvalues* and the $p_i(\omega)$ are called *dynamic eigenvectors*.

Suppose that for each $(\omega)$ the eigenvalues are all distinct. The eigenvectors *p* can be expanded in Fourier series. The expansion yields:

$$p_j(\omega) = \sum_{k=-\infty}^{k=+\infty}\left[\frac{1}{2\pi}\int_{-\pi}^{-\pi} p_j(s)e^{iws}ds\right]e^{-ik\omega}$$

Consider now the filter $p_j(L) = \sum_{k=-\infty}^{k=+\infty}\left[\frac{1}{2\pi}\int_{-\pi}^{-\pi} p_j(s)e^{iws}ds\right]L^k$ where *L* is the lag operator. The filter $p_j(L)$ is square summable. The scalar process $DPC_{j,t} = p_j(L)X_{j,t}, t = 0,\pm 1,\pm 2,\ldots$ is called the *j*-th dynamic principal component of the process *X*. The spectral density of the process $DPC_{j,t}$ is $\lambda_i(\omega)$. The processes $DPC_{j,t}$ and $DPC_{k,t}$ are orthogonal at any lead and at any lag. Brillinger (1964, 1981) introduced the notion of dynamic principal components which are principal components in the frequency domain.

### *3.6 Sparse principal components*

Principal components are linear combinations of the original variables. In general, principal components involve all original variables, that is, loadings are all non-zero numbers. In many applications this fact is undesirable. In financial applications, for example, it would be more desirable if principal components reflected “sectors”; it would thereby have many zero loadings. SPCA (or S-PCA) constrains a number of loadings to be zero. The SPCA problem can be formulated as follows:

$$\begin{aligned}&\beta_i' = \arg\max\left[\beta_i'\Sigma\Sigma_i\right] - \rho Card\left(\beta_i\right)\\&s.t.\beta_i'\beta_i = 1\end{aligned}$$

where $Card\left(\beta_i\right)$ is the number of non-zero elements of the vector $\beta_i$. A similar formulation of the SPCA problem is the following:

$$\begin{aligned}&\beta_i' = \arg\max\left[\beta_i'\Sigma\Sigma_i\right]\\&s.t.\beta_i'\beta_i = 1, Card\left(\beta_i\right) \leq k\end{aligned}$$

Both $k$ or $\rho$ are tuning parameters that need to be determined as a function of the application.

While PCA involves a “simple” maximization problem, SPCA is a difficult maximization problem; special algorithms are required to circumvent the problem and to find an approximate solution. A number of computer programs are now available to perform SPCA.

### *3.7 Independent Components*

Suppose *N* separate signals $x_{i,t}, i = 1,\ldots,N$ (prices, returns) are observed, and suppose that the $x_{i,t}$ are the weighted sum of *N* hidden, statistically independent, signals $s_{i,t}, i = 1,\ldots,N$. Call *A* the matrix of weights so that I can write: $x_{i,t} = As_{i,t}$. The signals $s_{i,t}$ are called *independent components*. Independent Components Analysis (ICA) seeks to recover $s_{i,t}$ from the observed $x_{i,t}$ without any *a priori* knowledge of the matrix *A*. ICA is also called *Blind Signal Separation* (BSS) as it seeks to separate the hidden $s_{i,t}$ “blindly”, that is, without any other information. An illustration of ICA (or BSS) is the cocktail-party problem, which is the problem of recovering each individual conversation in a room filled with people, each with a microphone picking up their conversation.

There are similarities and differences between ICA, PCA, and factor analysis. Both ICA and factor analysis assume a statistical model for the data while PCA can be applied without any statistical assumption on the data. However ICA is not, *per se*, a dimensionality reduction technique. The distinctive characteristic of ICA is the independence of components. The level of the independent components is not determined by the model: if I multiply all components by $\alpha$ and the matrix $A$ by $\frac{1}{\alpha}$, I obtain an observationally equivalent model.

ICA can be performed only if the independent components $s_{i,t}$ are not Gaussian variables. If independent components $s_{i,t}$ exist and are not Gaussian, ICA seeks the linear transformation of the observed variables that are maximally non-Gaussian. That is, given that data are represented by a model $x_{i,t} = As_{i,t}$, where the $s_{i,t}$ are independent and not Gaussian, ICA seeks the matrix $W$ such that $s_i = Wx_i$ are maximally non-Gaussian. The key idea behind ICA is that the sum of non-Gaussian variables is less non-Gaussian than the individual variables. For example, if we are standing outside of a cocktail party, though individual voices can be highly non-normal, what we hear is essentially white noise.

The various algorithms proposed to perform ICA differ in how the non-normality of $f(s)$ is measured and estimated. To date, three approaches have been proposed. The first approach measures non-normality through a function of higher moments of $f(s)$. In a normal distribution, all moments except the first two are zero. The presence of non zero higher moments signals a non-Gaussian distribution. A second approach is based on the negentropy, defined as the difference between the entropy of the distribution $f(s)$ and the entropy of the normal distribution. A third approach is based on the Kullback-Leibler divergence between $f(s)$ and a normal distribution.

ICA is a relatively new technique; it was proposed for the first time as BSS in Herault and Jutten (1986). Though relatively new, ICA has attracted a significant amount of research and algorithms coded in major languages such as Matlab or Mathematica are now available. Most applications are in the area of communications and signal recognition. Back and Weigend (1995) were the first to use ICA in finance.

### *3.8 Random Matrix Models (RMM)*

In this section I will outline the mathematical concepts behind Random Matrix Models (RMMs). A RMM is a probability space $(\Omega, P, F)$ where the set $\Omega$ is formed by matrices. Random matrices are matrix-variate random variables whose entries are random variables and eigenvalues (in case of square matrices) are real/complex valued random vectors. A RMM is characterized by its joint density as a matrix-variate random variable. A number of RMMs have received special attention and are relevant for this dissertation.

The following conventions have been adopted in this field. In RMT, models relative to real-valued matrices are denoted with $\beta$ = 1, models relative to complex-valued matrices are denoted with $\beta$ = 2, and models of matrices whose entries are quaternions are denoted with $\beta$ = 4. In this dissertation I will limit my discussion to models with $\beta$ = 1,2 as quaternions have not (yet) found applications in financial econometrics.

The mean of a complex-valued random variable is a complex number defined as follows: $E(x+iy)=E(x)+iE(y)$. The variance of a complex-valued random variable is a real number defined as follows: $\mathrm{var}(x+iy)=\mathrm{var}(x)+\mathrm{var}(y)=E(x-E(x))^2+E(y-E(y))^2$. For a zero mean variable $\mathrm{var}(x+iy)=E(x^2)+E(y^2)$. A standard real-valued normal variable is normally distributed with mean 0 and variance 1, that is, it is distributed according to: $N(0,1)$. A standard, complex-valued, normal variable is formed by a pair of independent normal variables distributed according to $N\left(0,\frac{1}{\sqrt{2}}\right)$, so that it has mean 0 and variance 1.

Real-valued Gaussian matrices, denoted by $G_1(m,n)$ are *mxn* matrices whose entries are i.i.d. standard normal variables. Complex-valued Gaussian matrices, denoted by $G_2(m,n)$ are *mxn* matrices whose entries are i.i.d. standard complex variables. The density of a Gaussian matrix **A** is

$(2\pi)^{-\beta\frac{mn}{2}}\exp\left(-\frac{trace(AA^*)}{2}\right)dA=(2\pi)^{-\beta\frac{mn}{2}}\exp\left(-\frac{\|A\|_F^2}{2}\right)dA$, where $\|A\|_F=\sqrt{\sum_{i,j}a_{i,j}^2}$ is the Frobenius matrix norm and $dA=\prod_{i,j}dA_{ij}$ for $G_1(m,n)$, $dA=\prod_{i,j}d\,\mathrm{Re}\,A_{ij}d\,\mathrm{Im}\,A_{ij}$ for $G_2(m,n)$. This expression is derived as the product of the normal standard distributions of the entries. A fundamental property of Gaussian matrices is their invariance under orthogonal transformations. This invariance is a consequence of the fact that the trace, or equivalently the Frobenius norm, is invariant under orthogonal transformations.

For example, for $3\times 3$ real-valued matrices, this expression means that the probability that $A_{ij}<a_{ij}<A_{ij}+dA_{ij}$, where $\{a_{ij}\}$ are the entries of the matrix A, is:

$$(2\pi)^{-\frac{9}{2}}\exp\left(-\frac{1}{2}\left(A_{11}^2+A_{22}^2+A_{33}^2+A_{13}^2+A_{12}^2+A_{23}^2+A_{31}^2+A_{21}^2+A_{32}^2\right)\right)dA_{11}dA_{12}dA_{13}dA_{22}dA_{23}dA_{33}dA_{31}dA_{21}dA_{32}$$

The Gaussian Orthogonal Ensemble, written as (GOE, $\beta$ = 1), is formed by square, symmetric $N\times N$ matrices with independent, zero mean, real-valued Gaussian entries. It is assumed that the off-diagonal terms have variance equal to 1 while the diagonal terms have variance equal to ½. The density of the GOE model is formally the same as the $G_1(m,n)$

$$\left(\pi\sigma^2\right)^{-\frac{N}{\sigma^2}}\exp\left(-\frac{trace\left(A^2\right)}{2}\right)dA=\left(\pi\sigma^2\right)^{-\frac{N}{\sigma^2}}\exp\left(-\frac{\|A\|_F^2}{2}\right),$$

but, in this case $dA=\prod_i dA_{ii}\prod_{i<j} dA_{ij}$. For example, for $3\times 3$ matrices and $\sigma=1$, this expression means that the probability that $A_{ij}<a_{ij}<A_{ij}+dA_{ij}$, where $\{a_{ij}\}$ are the entries of the matrix A, is:

$$(\pi)^{-3}\exp\left(-\left(A_{11}^2+A_{22}^2+A_{33}^2+A_{13}^2+A_{12}^2+A_{23}^2\right)\right)dA_{11}dA_{12}dA_{13}dA_{22}dA_{23}dA_{33}.$$

As the trace is invariant with respect to an orthogonal transformation, the density is invariant with respect to orthogonal transformations, hence the name Gaussian Orthogonal Ensemble.

The Gaussian Unitary Ensemble, written as (GUE, *β*=2) is formed by square, Hermitian $N\times N$ matrices with independent, zero-mean, complex-valued Gaussian entries. It is assumed that the off-diagonal terms have variance equal to 1 while the diagonal terms have variance equal to ½. The density of the model is

$$\left(\pi\sigma^2\right)^{-\frac{N}{\sigma^2}}\exp\left(-\frac{trace\left(A^2\right)}{2}\right)dA=\left(\pi\sigma^2\right)^{-\frac{N}{\sigma^2}}\exp\left(-\frac{\|A\|_2^2}{2}\right),$$

where $dA=\prod_i dA_{ii}\prod_{i<j} d\operatorname{Re}A_{ij}\,d\operatorname{Im}A_{ij}$.

As the trace is invariant with respect to a unitary transformation, the density is invariant with respect to orthogonal transformations, hence the name Gaussian Unitary Ensemble.

The Hermite ensemble, also called Wigner matrices, are square $n\times n$ Hermitian matrices with i.i.d. entries with variance 1/*n* off-diagonal and variance sqrt(2)/*n* on the diagonal. Wigner matrices are not necessarily Gaussian and therefore their density depends on the distribution of their entries.

The Wishart Ensemble is defined as follows. Consider *N* real or complex random variables $(X_1,\ldots,X_N)$ that follow a normal distribution with mean 0 and covariance matrix $\Sigma_N$ $N(0,\Sigma_N)$ and with probability density function:

$$f(X)=\left|\sqrt{2\pi}\,\Sigma\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}X'\Sigma^{-1}X\right).$$

Consider $T$ independent samples extracted from the distribution $N(0, \Sigma_N)$ and organized in a $T \times N$ *matrix* $H$ where each row is an observation. Suppose that $T > N$. The $N \times N$ matrix $A = H^* H$ is said to have a $N$-variate Wishart distribution on $T$ degrees of freedom. The Wishart distribution has the following density:

$$f_W(A) = c_N |\Sigma|^{-\frac{N}{2}} |A|^{\frac{n-p-1}{2}} \exp\left( -\frac{1}{2} trace\left(\Sigma^{-1} A\right) \right)$$

where $c_N$ is a normalizing constant:

$$c_N = \left( \frac{1}{2^{\frac{NT}{2}} \Gamma_N\left(\frac{T}{2}\right)} \right).$$

The Ginebre Orthogonal Ensamble is defined as the Gaussian Orthogonal Ensamble without the symmetry assumption. Therefore, the Ginebre Orthogonal Ensamble, denoted as GinOE, is the ensemble of square $N \times N$ matrices with independent, zero mean, real-valued Gaussian entries. The density of the GinOE model is formally the same as the $G_1(N, N)$

$$\left(\pi \sigma^2\right)^{-\frac{N^2}{2}} \exp\left( -\frac{trace\left(A^2\right)}{2} \right) dA = \left(\pi \sigma^2\right)^{-\frac{N}{\sigma^2}} \exp\left( -\frac{\|A\|_F^2}{2} \right), \text{ with } dA = \prod_{i,j} dA_{ij}$$

The Ginibre Unitary Ensemble is defined as the Gaussian Unitary Ensemble without the Hermitian assumption. Therefore, the Ginibre Orthogonal Ensemble, denoted as GinUE, is the ensemble of square $N \times N$ matrices with independent, zero mean, complex-valued Gaussian entries. The density of the GinOE model is formally the same as the $G_2(N, N)$

$$\left(\pi \sigma^2\right)^{-\frac{N^2}{2}} \exp\left( -\frac{trace\left(AA^*\right)}{2} \right) dA = \left(\pi \sigma^2\right)^{-\frac{N}{\sigma^2}} \exp\left( -\frac{\|A\|_F^2}{2} \right),$$

with $dA = \prod_{i,j} d \operatorname{Re} A_{ij} d \operatorname{Im} A_{ij}$

### *3.9 Transforms*

In this section I introduce the main transforms used in RMT. Transforms play a key role in proving many theorems on the distributions of eigenvalues, in particular the Marčenko-Pastur law. I will describe the Stieltjes transform, the $\eta$-transform, the R and S transforms.

### *3.9.1 The Stieltjes transform*

Consider a random variable *X* and its distribution $F_X$. The Stieltjes transform is a complex-valued function defined as follows:

$$S_X(z) = E\left(\frac{1}{X - z}\right) = \int_{-\infty}^{+\infty} \frac{1}{\lambda - z} dF_X(\lambda), z \in C$$

where *z* is a complex number. Usually the Stieltjes transform is defined for values of *z* with positive imaginary part. If the distribution $F_X$ admits a density $f_X$, then the Stieltjes transform becomes

$$S_X(z) = E\left(\frac{1}{X - z}\right) = \int_{-\infty}^{+\infty} \frac{1}{\lambda - z} f_X(\lambda)d\lambda\ , z \in C$$

In this form, the Stieltjes transform can be inverted and the inversion formula is:

$$f_X = \lim_{\omega \to 0^+} \frac{1}{\pi} \mathrm{Im}[S_X(x + i\omega)]$$

Assuming $F_X$ has compact support, we can expand the Stieltjes transform in a Laurent series and, exchanging integration and summation, I can write:

$$S_X(z) = -\frac{1}{z} \sum_{k=1}^{\infty} \frac{E(X^k)}{z^k}$$

### *3.9.2 The η-transform*

The *η*-transform of a non-negative random variable is a real-valued function defined as follows:

$$\eta_X(\gamma) = E\left[\frac{1}{1 + \gamma X}\right],$$

where $\gamma$ is a non negative real number. There is a simple relationship between the *η*-transform and the Stieltjes transform:

$$\eta_X(\gamma) = \frac{S_X\left(-\frac{1}{\lambda}\right)}{\gamma}$$

and therefore:

$$\eta_X(\gamma) = \sum_{k=0}^{\infty} \gamma^k E(X^k).$$

*3.9.3 The R and S-transforms*

The R-transform plays a key role in the application of free probability concepts to RMT. The R-transform is defined as follows. Consider the functional inverse of the Stieltjes transform, that is, consider the function of complex variable $S_X^{-1}(z)$ such that $z = S_X^{-1}(S_X(z))$. The R-transform is defined as:

$$R_X(z) = S_X^{-1}(-z) - \frac{1}{z}.$$

For any *a>0*, the following relationship holds: $R_{aX}(z) = aR_X(z)$. If the random variable $X$ has compact support, it is possible to represent the R-transform as a series:

$$R_X(z) = \sum_{k=1}^{\infty} c_k z^{k-1}.$$

The coefficients $c_k$ are called free cumulants and can be expressed in terms of the moments of the variable $X$ as follows:

$$E[X^m] = \sum_{k=1}^{m} c_k \sum_{m1+\cdots mk=m} E[X^{m1-1}]\cdots E[X^{mk-1}].$$

The S-transform is defined as follows:

$$\Sigma_X(x) = .-\frac{x+1}{x}\eta_X^{-1}(1+x),$$

*3.10 Free Probability*

Free Probability is a mathematical theory originally developed by Dan-Virgil Voiculescu (1986, 1987) in relation to operator algebras. I will begin by putting free probability in the context of probability theory in general.

Kolmogorov (1933) gave a rigorous, abstract mathematical formulation of classical probability theory based on the notion of events and measures. The fundamental object in the theory of Kolmogorov is a triple $(\Omega, \Im, P)$ formed by a set of objects $\Omega$, a sigma-algebra of events which are subsets of $\Im$ and a probability measure *P*. Random variables are real-valued functions defined on $\Omega$. A complex-valued random variable *Z* is defined as follows: *Z=X+iY* where *X,Y* are real-valued random variables.

Roughly contemporaneous with Kolmogorov's giving an axiomatic foundation to probability theory, John von Neumann (1932, reprinted 1996) gave an axiomatic foundation to quantum mechanics based on the theory of Hilbert spaces. The work of von Neumann made it clear that, due to the nature of quantum objects, quantum mechanics would require a type of probability theory different from the classical probability theory of Kolmogorov. This new type of probability theory was to be called *quantum probability*. Quantum probability is a non-commutative, algebraic theory of probability. It was later realized that it is possible to define an algebraic theory of probability that encompasses classical probability, quantum probability and free probability. I will briefly sketch the concept of algebraic probability theory.

### *3.10.1 Algebraic probability theory*

I first recall that in classical probability theory, complex-valued random variables are assumed to have the following properties:

1. Complex constants are random variables.
2. The sum of two random variables is a random variable.
3. The product of two random variables is a random variable.
4. Addition and multiplication of random variables are both commutative
5. There is a notion of conjugation of random variables, satisfying (ab)* = b* a* and a** = a for all random variables a, b, which coincides with complex conjugation if a is a constant. If a = a*, the random variable a is called "real".

The above means that random variables form a complex commutative C-*-algebra. In fact, I recall the definition of *-algebra, in particular C-*-algebra on the field C. A C-*-algebra *A* over the field of complex numbers C is a vector space endowed with a C-bilinear operation called product or multiplication $A \to A$ (where the image of $(x,y)$ is written as $xy$) with the following properties:

- $(xy)\, z = x\, (y\, z)$ for all $x$, $y$ and $z$ in $A$.
- $(x + y)\, z = x\, z + y\, z$ for all $x$, $y$, $z$ in $A$,
- $x\, (y + z) = x\, y + x\, z$ for all $x$, $y$, $z$ in $A$,
- $a\, (x\, y) = (a\, x)\, y = x\, (a\, y)$ for all $x$, $y$ in $A$ and $a$ in $K$.

If $A$ contains an identity element, i.e., an element 1 such that $1x = x1 = x$ for all $x$ in $A$, then $A$ is called an *associative algebra with one* or a unital (or unitary) associative algebra.

An expectation E on an algebra A of random variables is a normalized, positive linear functional, that is, the function E: $A \rightarrow$ C has the following properties:

- *E(k) = k* where *k* is a constant;
- *E(a* a)* ≥ 0 for all random variables *a*;
- *E(a + b) = E(a) + E(b)* for all random variables *a* and *b*; and
- *E(za) = zE(a)* if *z* is a constant.

The above properties of random variables and expectations that hold in classical probability theory become the basis for algebraic probability theory. While classical probability theory is centered on the concept of event, the algebraic theory of probability is centered on the concept of random variables and measures, and related concepts are derived through the theory of representation. The algebraic theory of probability is formalized as follows.

An algebraic probability space is a pair $(A,\phi)$ where *A* is a C-*-algebra on the field C and $\phi$ are positive, normalized linear functionals called *states in quantum probability* (*expectations* in classical probability and *traces* in free probability). The functionals $\phi$ have the following properties:

1. $\phi(1) = 1$;
2. $\phi(a*a) \geq 0, \forall a \in A$;
3. $\phi(a+b) = 1\phi(a) + \phi(b), \forall a,b \in A$;
4. $\phi(ka) = k\phi(a), \forall a \in A, k \in C$.

The distribution (in a generalized sense) of a random variable $a \in A$ is defined through the (infinite) set of moments: $\alpha_k = \phi(a^k)$. Joint moments can be defined analogously:

$$\alpha_{k_1 \ldots k_n} = \phi\left(a_1^{k_1} \cdots a_n^{k_n}\right).$$

If the algebra *A* is commutative, then algebraic probability theory is equivalent to Kolmogorov's classical probability theory. If the algebra *A* is non-commutative, then it is possible to specify alternative probability theories such as quantum probability. I will now focus on free probability.

Free probability is a concept that applies within the domain of non-commutative algebraic probability theories. Consider a non-commutative algebraic probability space $(A,\phi)$ where *A* is a C-*-algebra on the field C and $\phi$ are positive, normalized linear functionals called *traces*.

The concept of freeness can be introduced as follows. Let $a_1, a_2 \in A$ be two random variables. Consider the sub-algebras $A_1$ and $A_2$ formed by the polynomials of $a_1$ and $a_2$ respectively (that is, the subalgebras generated by $a_1,1$ and by $a_2$). The random variables $a_1$ and $a_2$ are said to be free if $\phi(Z_1,\ldots,Z_m) = 0$ whenever $\phi(Z_k) = 0$ for $k = 1,\ldots,m$ and $Z_k \in A_{i(k)}, i(k) = 1,2$ and consecutive indices are distinct: $i(k) \neq i(k+1)$.

### *3.10.2 Freeness*

Freeness is the free probability equivalent of independence in classical probability theory. Observe the following property of free random variables. From the definition of freeness, if *a* and *b* are free, one can write: $\phi(a - \phi(a)1) = 0$ and therefore $[\phi(a - \phi(a)1)][\phi(b - \phi(b)1)] = 0$ hence $\phi(ab) = \phi(a)\phi(b)$.

The concepts introduced thus far are summarized in Table 3.2.

| | **Classical probability theory** | **Algebraic formulation of classical probability theory** | **Algebraic formulation of non-classical probability theory** |
|---|---|---|---|
| | Probability space based on events: $(\Omega, \Im, P)$ | Algebraic probability space based on random variables: $(A,\phi)$ | Algebraic probability space based on random variables: $(A,\phi)$ |
| | | *A* commutative C-*-algebra,<br>$\phi$ expectation | *A* non-commutative C-*-algebra<br>$\phi$ different interpretations, e.g., matrix trace |
| Free random variables | NA | NA | $\phi(Z_k) = 0 \Rightarrow$<br>$\phi(Z_1,\ldots,Z_m) = 0$ |
| Independence vs freeness | Independence:<br>$P(A \cap B) = P(A)P(B)$ | | Freeness:<br>$\phi(ab) = \phi(a)\phi(b)$ |

*Table 3.2: Classical and Algebraic Probability Theory.*

### *3.10.3 Asymptotic freeness of random matrices*

Thus far I have defined the algebraic probability spaces and the concept of freeness. I will now make the connection between these concepts and random matrices. My exposition draws on a number of sources including Nica and Speicher (2006) and Tulino and Verdù (2005). Recall that random matrix models are classical probability models, that is, an $N \times N$, complex, self-adjoint (Hermitian) random matrix is a classical probability object. Consider now an infinite sequence of complex-valued, Hermitian random matrices $A_N, N \to \infty$. It is impossible to define the limit $\lim_{N\to\infty} A_N$ as an infinite matrix. However, it is possible to define the following quantity:

$$a_k = \lim_{N\to\infty} \frac{1}{N} E\left[trace\left(A_N^k\right)\right].$$

Consider now an abstract non-commutative algebraic probability space $(A,\phi)$ and consider the random variable $a \in A$ such that $\phi\left(a^k\right) = a_k, \forall k = 1,2,\ldots$. It is said that the random variable *a* is the limit of the sequence $A_N$: $a = \lim_{N\to\infty} A_N$. Note explicitly that random matrices $A_N$ are non-commutative objects because matrix product is not commutative but a random matrix model is a classical probability space. It is the space of limit random variables that is a non-commutative algebraic probability space.

It can be demonstrated that the moments of the limit distribution of the eigenvalues of $A_N, N \to \infty$ are the $a_k$. Two sequences $A_N, N \to \infty$ and $B_N, N \to \infty$ are said to be asymptotically free if the limits $a = \lim_{N\to\infty} A_N$ and $b = \lim_{N\to\infty} B_N$ are free.

Asymptotically free random matrices have a number of important properties.

- Consider two asymptotically free random matrices A and B. The R-transform of the asymptotic spectrum of the sum A+B is the sum of the R-transforms of of the asymptotic spectra of A and of B: $R_{A+B}(z) = R_A(z) + R_B(z)$.
- Consider two asymptotically free random matrices A and B. The S-transform of the asymptotic spectrum of the product AB is the product of the S-transforms of the asymptotic spectra of A and of B: $S_{AB}(z) = S_A(z) S_B(z)$.

### *3.11 The method of the Resolvent*

I will now outline the proofs of the Marčenko-Pastur law. I will sketch a new proof proposed by Burda, Goerlich, Jarosz, and Jurkiewicz (2004) and subsequently extended in Burda, Jurkiewicz, and Waclav (2005). First I recall a property of the moments of the distribution of eigenvalues. Consider an $N \times N$ Hermitian random matrix *A* and denote $\{\lambda_1 \leq \cdots \leq \lambda_k \leq \cdots \leq \lambda_N\}$. Consider the distribution of the eigenvalues of *A*, indicated as ESD (acronym of Empirical Spectral Distribution).

The ESD can be written as follows: $F^A(x) = \frac{1}{N} \neq \{k \leq N, \lambda_k \leq x\}$ where $\neq$ denotes the number of elements in the set indicated. That is, $F^A(x)$ represents the proportion of eigenvalues of the matrix A that are $\leq x$. The following property of the moments of $F^A(x)$ holds:

$$m_k(A) = \int x^k F^A(x)dx = \frac{1}{N} trace(A^k).$$

Note also that the Stieltjes transform of the ESD of a matrix A is :

$$S_{F^A}(z) = \frac{1}{N} trace((A - I_N z)^{-1})$$

I start by proving the Marčenko-Pastur law for correlated variable and uncorrelated samples. Consider a statistical model with *N* zero-mean variables (degrees of freedom) distributed according to the following probability distribution: $p(x_1, \ldots, x_N)\prod_{i=1}^{N} dx_i$. The covariance matrix of the model is:

$$C_{ij} = \int p(x_1, \ldots, x_N)\prod_{n=1}^{N} dx_n .$$

Assume the system belongs to the Gaussian universality class, which implies that the probability distribution can be approximated by a normal probability distribution:

$$p(x_1, \ldots, x_N)\prod_{i=1}^{N} dx_i = [(2\pi)^N \det|C|]\exp\left(-\frac{1}{2}\sum_{i,j} x_i C_{ij} x_j\right)\prod_{i=1}^{N} dx_i$$

Now consider *T* independent observations arranged in a $N \times T$ matrix *X* and consider the empirical covariance matrix $c = \frac{1}{T} XX'$. Call $\Lambda_n, n = 1, \ldots, N$ the eigenvalues of the true covariance matrix matrix $C$ and $\lambda_n, n = 1,\ldots,N$ the eigenvalues of the empirical covariance matrix **c**. It is convenient to define the spectral density, that is, the density of the eigenvalues, as follows:

$$\rho_0(\Lambda) = \frac{1}{N}\sum_{n=1}^{N} \delta(\Lambda - \Lambda_n)$$

where $\delta$ is Dirac's delta function. Note that in integrating this density we obtain the distribution $F^C(x)$ as defined above. We can compute the moments of the distribution of the true eigenvalues as follows:

$$M_k = \frac{1}{N} trace\left(C^k\right) = \frac{1}{N}\sum_{n=1}^{N} \Lambda_n^k = \int \Lambda^k \rho_0(\Lambda) d\Lambda$$

Consider now the empirical covariance matrix and its eigenvalues and form the expectation of the distribution:

$$\rho(\lambda) = \frac{1}{N} E\left[\sum_{n=1}^{N} \delta(\lambda - \lambda_n)\right]$$

and the expectation of the moments:

$$m_k = \frac{1}{N} E\left[trace\left(c^k\right)\right] = \int \lambda^k \rho(\lambda) d\lambda \ .$$

Let's now introduce the following resolvents:

$$G(Z) = \left(ZI_N - C\right)^{-1}$$
$$g(z) = E\left[\left(zI_N - c\right)^{-1}\right].$$

Formally expanding the resolvents in powers of $\frac{1}{Z}$ and $\frac{1}{z}$, we see that the resolvents can be interpreted as the generating functions of the moments:

$$M(Z) = \frac{1}{N}\left[trace(ZG(Z))\right] - 1 = \sum_{k=1}^{\infty} \frac{1}{Z^k} M_k$$
$$m(z) = \frac{1}{N}\left[trace(zg(z))\right] - 1 = \sum_{k=1}^{\infty} \frac{1}{z^k} m_k$$

Using planar diagrammatic techniques from quantum mechanics, Burda, Goerlich, Jarosz, and Jurkiewicz (2004) demonstrate that when *N,T* tend to infinity with a constant ratio *r* if

$$Z = \frac{z}{1 + rm(z)}$$

then $M(Z) = m(z)$.

From this last expression we can recover the true eigenvalue density from the empirical covariance matrix in the limit of infinite *T,N.* The eigenvalue distribution is given by the formula:

$$\rho(\lambda) = \frac{1}{\pi} \operatorname{Im} g\left(\lambda + i0^{+}\right).$$

Burda, Jurkiewicz, and Waclav (2005) extend the above proof to the case where samples are not independent. Burda, Jurkiewicz, and Waclav assume that cross correlations and autocorrelations can be factorized as CA where C is a cross correlation matrix and A is an autocorrelation matrix.

# 4. Forecasting with factors: New results

## *4.1 Key findings*

I will restate the problems posed by the application of factor analysis to portfolio management, problems that motivated the choice of subject matter for this dissertation. Classical factor analysis cannot be correctly applied to large panels of data such as econometric time series or financial returns for two reasons. First, it is very onerous to carry the maximum likelihood estimation for large samples and second, the assumption of uncorrelated residuals is not empirically tenable. However, abandoning the assumption of uncorrelated residuals in finite models proved to be challenging. In fact, it is very difficult to establish theoretically sound and empirically meaningful criteria to separate common factors from correlated residuals.

The literature on factor models has taken a different path, proposing the paradigm of approximate factor models which are infinite in both the number *N* of observed variables and the number of observations. In approximate factor models, the separation between common factors and residuals is achieved in a natural way assuming that the eigenvalues of the covariance matrix corresponding to the common factors diverge while the remaining eigenvalues remain bounded. Under these assumptions, factors are unique (up to a rotation) and can be estimated with principal components.

In the theoretical literature, it is generally assumed that large samples ─ including financial returns and macroeconomic variables ─ can be analyzed with approximate factor models. The practice of asset management, however, has taken a different path, proposing a number of factor models constructed with criteria based on identifying fundamental parameters, sectors, countries, or exogenous variables.

The problem I want to discuss can be stated simply as follows: If the approximate factor model paradigm is applicable, then practical efforts to find unique models are futile, factor models are unique and factors can be determined with principal components. If the approximate factor model paragidm is not applicable, we need 1) robust criteria to determine when the paradigm of approximate factor models is applicable and 2) criteria to select factor models created with methods different from both factor analysis and principal components analysis.

The theoretical and practical solution I propose in this dissertation, which I will discuss in this section, can be summarized in the following five points:

1. The theoretical paradigm of approximate factor models cannot be applied blindly to large samples. There must be a neat separation between “large” and “small” eigenvalues of the covariance matrix of data.
2. Criteria to separate “large” and “small” eigenvalues are arbitrary. I propose to sidestep the problem by defining directly criteria that determine the ability to approximately identify factors and principal components, also taking into account the ability to “learn” the model from the data.
3. When factors cannot be approximately identified with principal components, I propose to look at factor models as multiple communication channels and to use the channel capacity (i.e., the average mutual information) as criterion for choosing among the different models.
4. Financial returns cannot be faithfully represented with unique approximate static factor models because the eigenvalues decay smoothly.
5. I try to prove that the inability to find unique static factor models of returns might be due to the presence of dynamic factors and of cointegration in models of asset returns. I try to prove that both dynamic factor models and cointegration-based models are likely to produce a smooth decay of eigenvalues of the covariance matrix of static factors.

The rest of this chapter follows the above scheme. First, I discuss just what conditions are responsible for the fact that the approximate factor model paradigm is not applicable. Second, I introduce the conditions of finite samples under which factors and principal components are very similar; similarity of factors and principal components allows factors to be effectively estimated with principal components. Then I introduce criteria based on mutual information, suggesting the analogy of channel capacity. I show empirically that financial returns do not qualify for approximate factor models. As a consequence, factors are not uniquely identified and different factor models of returns offer different partial explanations of returns. Next, I discuss the applicability of dynamic factor models to returns. Finally, I show that dynamic factor models of prices can be estimated in a natural way, I discuss model uniqueness and I show that, if prices follow a dynamic factor model, then it is unlikely that factor models of returns are unique. This is ultimately the reason why there are different competing factor models of returns.

### *4.2 Conditions for applying the paradigm of approximate factor models*

In this section, I use Monte Carlo simulation to demonstrate that approximate factor models are not applicable if the signal-to-noise ratio is too low. The existing literature discusses the problem of estimation of factor models in the "large *T*, large *N*" assumption, where "large" is meant to be synonymous with infinite. Equating "large" and "infinite" is a strategy typical of classical statistics, supported by asymptotic results such as the law of large numbers or the central limit theorem and the asymptotic distribution of estimators. However, the statistics of "large" factor models is not classical because the number of parameters grows with the size of the model. Therefore "large" might be very different from "infinite".

Empirically, all samples have a maximum size. In statistical physics, the maximum size of samples is often a truly enormous number. Take for example the Avogadro constant, which states that the number of atoms in 12 grams of carbon-12 is:

.02214179x100.000.000.000.000.000.000.000.[7]

However, in financial econometrics, the maximum size of samples available today is in the range of thousands. In practice, the largest available universes of stock returns are of the order of 10,000-15,000 stocks. The maximum available time depth for daily data is again in the range of 10,000 days.

High-frequency data might multiply the number of available data points by a factor of 100 or more. However, it is not clear if and how it is possible to use high-frequency data to make forecasts at a time horizon pertinent to asset management. While using high-frequency data to capture phenomena at very short time horizons might be useful for strategies based on derivative products, the transaction costs associated with exploiting high-frequency data in equity asset management renders the exercise of dubious value.

The numbers of assets in today's markets are large but not infinite by any practical sense of the term. If we segment by countries and by sectors, for example using the GICS codes, we rapidly form hundreds of different groupings, each containing only hundreds of stocks. What meaning can we assign to the fact that the number of returns tend to infinity? A growing number of hypothetical countries? Or a growing number of hypothetical sectors? Or a growing number of stocks in each country/sector segment? And if we interpret factors as economic causes, do we assume that, in an infinite market, there are still a finite number of common causes or do we assume that the number of causes also grows to infinity?

Clearly any assumption that we make is somewhat arbitrary. There is no "reasonable" assumption as regards how a universe of stock returns might tend to infinity. What we need are asymptotic results that are close to the result that we would obtain if we were able to perform true small-sample estimations. For these reasons, when we discuss factor models in the "large *T*, large

[7] Named after the Italian chemist Amedeo Avogadro, the Avogadro principle (1811) states that the number of atoms in a mass proportional to the atomic number of a substance is a constant.

*N*” assumption, we need to ensure that a finite sample has all the essential characteristics that we use to obtain the asymptotic results.

### *4.2.1 Pitfalls in determining the number of factors*

I will start by discussing criteria to determine the number of factors of static factor models. I will show that none of the criteria proposed thus far is effective in determining the number of factors unless “sample qualification criteria” are added.

In the literature, three basic solutions to the problem of determining the number of factors have been proposed: criteria based on information theory (Bai and Ng, 2002), criteria based on the rank of matrices (Peña and Box, 1987), and criteria based on random matrix theory (Kapetanios, 2004, Onatski, 2005 and 2006). Criteria based on information theory introduce a penalization function that increases the mean square error in function of the number of parameters to be estimated but decreases with the number of samples. Criteria based on the rank of matrices, typically used in dynamic factor models, assume a finite $T$ or $N$. Criteria based on random matrix theory compare the empirical distribution of eigenvalues with the asymptotic distribution of eigenvalues of random matrices. I will start with a discussion of criteria based on information theory as proposed by Bai and Ng (2002).

The Bai and Ng criteria apply to approximate factor models of the type described in Bai and Ng (2002, 2003). In particular, these criteria require that a finite number of eigenvalues of the empirical covariance matrix diverge while all other eigenvalues remain bounded. The Bai and Ng criteria based on information theory are an optimal model selection theory and are therefore based on some optimal trade-off. However, Bai and Ng’s criteria differ from classical model selection criteria such as the AIC or BIC or the Minimum Description Length principle due the asymptotic nature of the factor models.

I will briefly digress on model selection criteria in order to discuss the difference between classical model selection criteria and the Bai and Ng criteria. Model selection criteria such as AIC or BCI apply to models of finite complexity estimated on a variable sample. AIC, BIC, model selection criteria introduce a penalty function that grows with the number of parameters so that the sum of the loglikelihood minus the penalty function attains a non trivial maximum. The AIC/BIC model selection criteria work in similar ways but are based on different principles. The AIC is based on information theory. It adds the number of parameters to the loglikelihood so that the best model maximizes $L(T)-k$. The BIC of Schwarz is based on Bayesian criteria and uses the penalty function $k\times\log(T)/2$ so that the best model maximizes $L(T)-k\times\log(T)/2$.

The Minimum Description Length (MDL) principle of Jorma Rissanen is another model selection criteria based on the idea of finding the code with the shortest length to describe the data and the model. The MDL principle is grounded in the theory of the complexity of models and distributions. Early formulations of the MDL principle were similar to the BIC criterion but the MDL principle admits (in practice) different (not always easy to implement) formulations.

Vapnik and Chervonenkis developed a complete theory of statistical learning presented, for example, in Vapnik (1998). In principle the Vapnik-Chervonenkis theory of statistical learning can also be applied to factor model selection, but its implementation is complicated and to my knowledge it has not been used to determine the number of factors.

The AIC and BIC work as optimal model selection criteria in finite samples. They implement, or tend to implement, the best trade-off between model complexity and size of the model. If only small samples are given, typically a model simpler than the true model will be chosen. In general, the optimal number of parameters grows with sample size and reaches asymptotically the true number of parameters. Note, however, that the BIC converges asymptotically while the AIC does not (see, for example, Gourieraux and Monfort, 1995).

However, approximate factor models as those described in Bai and Ng (2002, 2003) are models that, asymptotically, include an infinite number of parameters. In fact, as there are an infinite number of time series, the number of loadings is infinite. More importantly, approximate factor models are well defined only in the limit of an infinite market; it does not make sense to estimate the number of factors in a finite model. Criteria for determining the number of factors are inherently asymptotic criteria. Bai and Ng (2002) introduce general criteria for model selection (i.e., criteria for selecting the number of factors) formulated through the following theorem.

Suppose factors are estimated with principal components. Call $V_k$ the residuals normalized by $1/NT$ using the largest $k$ eigenvalues. Consider now a function $g(k,N,T)$ and suppose that:

$$\lim_{N,T\to\infty} g(k,N,T)=0,\ \lim_{N,T\to\infty} C^2_{N,T}\, g(k,N,T)=\infty\ , C_{N,T}=\min\left(\sqrt{T},\sqrt{N}\right).$$

Form the difference ($V$-$g$). Bai and Ng (2002) demonstrate that

$$\arg\min_k \left(V_k - g(k,N,T)\right)$$

converges to the true number of factors if both $N$ and $T$ tend to infinity. Bai and Ng show that neither the BIC nor the AIC satisfy the assumption of the above theorem. They then introduce three new criteria, that is, three explicit expressions for the function $g$ that do satisfy the assumptions of the theorem. Bai and Ng do not demonstrate that their criteria work in a finite sample. Actually there is no theory to support the claim that the Bai and Ng criteria work in finite samples. Bai and Ng show that their criteria work asymptotically and then illustrate, through simulation, that their criteria have good finite sample performance.

However, the Bai and Ng criteria are asymptotic criteria which might fail in finite samples. It is easy to construct counterexamples to make the Bai and Ng criteria fail in the sense that they grossly underestimate or overestimate the number of factors in finite samples. For example, consider an infinite population with one single common factor and many weak factors, that is, factors that influence only a finite number of return series. If the weak factors correspond to large eigenvalues, that is, if finite sets of series are strongly correlated while the common factor has only weak loadings, the Bai and Ng criteria will find that there are many factors in a finite sample and converge to only one factor with the growth of the sample.

Criteria based on estimating the rank of a matrix are typically used in estimating the number of factors in a dynamic factor model where all autocorrelation matrices have the same rank, which corresponds to the number of factors. Hence, the number of factors is determined estimating the rank of the autocorrelation matrices. An alternative equivalent procedure consists in estimating the rank of the spectral density matrix (Camba-Mendez and Kapetanios, 2004). All these criteria assume a finite $N$ or a finite $T$. There are many criteria for estimating the rank of a matrix.

I will now discuss criteria based on RMT. Criteria based on RMT are inherently asymptotic insofar as the asymptotic distribution of the eigenvalues of random matrices is compared with the empirical distributions of eigenvalues. The basic idea underlying the application of RMT is that the eigenvalues of the empirical covariance matrix of uncorrelated data do not converge to the true eigenvalues, not even in the asymptotic limits. This occurs because the ratio between the number of observations and the number of parameters is constant. It is therefore assumed that, given an empirical eigenvalue distribution, those eigenvalues that are within the limit of the theoretical distribution of purely random eigenvalues do not carry information.

Observe explicitly that random matrix theory assumes a specific distribution of population eigenvalues. For example, the null of zero correlation assumes that all eigenvalues are equal to 1. The asymptotic distribution of eigenvalues has been determined only for a small number of models that are, in general, perturbations of the null of zero correlations. As we have seen in Chapter 3, more complex population distributions can be studied but results are difficult to determine numerically.

There are two problems associated with applying criteria based on random matrices. First, as discussed in Chapter 2, the asymptotic distribution of the eigenvalues relative to uncorrelated series is constrained between

$$a = \left(1 - \sqrt{\gamma}\right)^2 \text{ and } b = \left(1 + \sqrt{\gamma}\right)^2,$$

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

If we allow correlations that originate bounded eigenvalues as in the spiked model, a number of eigenvalues will fall to the right of the interval (a,b) in positions whose expected values are described in Chapter 2. The magnitude of these "large" eigenvalues is not constrained: they can assume any value. One has therefore to select arbitrarily the threshold that divides "spiked" residuals from common components.

Second, all results obtained in Chapter 2 are asymptotic results. In finite samples, eigenvalues can assume values different from the asymptotic limits. There is no simple way to evaluate the small sample deviations. The result is that criteria based on random matrix theory depend on arbitrary decisions as regards the thresholds.

The conclusion of the above remarks is that any factor model based on the assumption that both *N,T* go to infinity is based on some assumption that cannot be justified within factor model theory itself. What are called for are criteria for model selection that are able to suggest just what asymptotic results can be reasonably applied to finite samples.

### *4.2.2 Pitfalls in estimating factors*

I will now discuss the estimation of factors. In Chapter 3, I explained that in classical factor models the parameters of the model can be consistently estimated but that, in general, factors cannot be consistently estimated with maximum likelihood. Factors have to be estimated, in general, as cross sectional regressions. However in classical factor analysis there is a serious problem of factor indeterminacy; this problem was discussed in Section 3.1.1. In general, principal components are not consistent estimators of factors. Only in the case of scalar factor models (i.e., factor models where residuals are i.i.d. variables with the same variance) can factors be identified with principal components.

However in approximate factor models, if both the number of series and the number of observations tend to infinity, then factors can be consistently estimated with principal components up to a rotation. Stated differently, principal components span the factor space. Estimating factors with principal components is one of the key features of modern static factor theory for large models.

Maximum likelihood methods are not applicable when the number of series is too large while principal components can be determined with robust techniques, in particular Singular Value Decomposition (SVD). Therefore, the theory of infinite approximate factor models is welcome from the theoretical as well as the practical point of view.

In practice, however, the application of the theory of approximate factor models as described in Chamberlain and Rothshild (1983), Bai and Ng (2002, 2003), and Stock and Watson (2002 a,b) is very difficult. The major difficulty lies in determining whether the approximate factor model paradigm applies to a finite sample and, if so, how many factors actually exist. I observed above that criteria to determine the number of factors are asymptotic criteria and might not work at all in finite samples. In short, given a finite sample, it is impossible to distinguish between eigenvalues that will diverge and eigenvalues that are large but will remain bounded.

In order to address this problem, Onatski (2007) and, implicitly, Harding (2008 b), proposed to replace "strong" factors that influence an infinite number of returns with "weak" factors that influence only a finite number of returns. Under this assumption, all large eigenvalues are considered as factors. However, if we drop the assumption that a finite number of eigenvalues tend to infinity while the others remain bounded, there is no possibility of consistent estimation of factors via principal components. This fact has been proved in the context of random matrix theory by determining the asymptotic distribution of eigenvalues and the distribution of the largest eigenvalues. Onatski (2007) demonstrates that weak factors, that is factors whose corresponding eigenvalues stay bounded, cannot be consistently estimated with principal components. Harding (2008 a) used results from spiked models, which are equivalent to factor models with a finite number of weak factors, to prove the well-known fact that factor models tend to overestimate the largest factor.

### *4.2.3 Monte Carlo simulation of approximate factor models*

I first observe that in the area of financial returns, the paradigm of an infinite number of infinitely long return series is subject to serious limitations. One reason why this is the case is that financial markets are finite open systems where new return processes are generated and old return processes cease to exist. If we want to include long series in our samples, we must reduce the number of series because a smaller number of longer series is available. Conversely, if we want to increase the number of series, then we have to reduce the time window. This fact introduces significant survivorship biases and makes it difficult to support the assumption that both $N,T$ tend to infinity.

The above is not without importance as biases can be significant. Consider, for example, the Russell 1000 universe. Suppose that we consider samples formed by all of the Russell 1000 stocks that exist in 200-week-long moving windows. Our samples will include only a fraction of the constituents of the Russell 1000 at the beginning, end, or any given time in the time window because many public firms are created or cease to exist during the window and therefore have to be excluded from the sample. For example, we find that in the period January 2002-November 2006, a 200-week-long moving window includes an average of less than 800 return series. If we consider a 500-week moving window, the average number of available series drops to less than 400. Table 4.2 illustrates the number of return series of the Russell 1000 that can be defined for time windows of 200 and 500 weeks respectively in the period January 2002-December 2006. Time windows are taken at 16-week intervals and end at the following dates.

| Date<br>Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | 2002 | 2002 | 2002 | 2002 | 2003 | 2003 | 2003 | 2004 | 2004 | 2004 | 2005 | 2005 | 2005 | 2005 | 2006 | 2006 | 2006 |
| Month | 01 | 04 | 08 | 12 | 03 | 07 | 11 | 02 | 06 | 10 | 01 | 05 | 09 | 12 | 04 | 08 | 11 |
| Day | 03 | 25 | 15 | 05 | 27 | 17 | 06 | 26 | 17 | 07 | 27 | 19 | 08 | 29 | 20 | 10 | 30 |

*Table 4.1 Beginning dates of the time windows.*

| | Evolution of the number of return series in the Russell 1000 universe in time windows of 200/500 weeks in the period: | | | | | | | | | | | | | | | | | Av. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Window<br>Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | |
| Year | 2002 | 2002 | 2002 | 2002 | 2003 | 2003 | 2003 | 2004 | 2004 | 2004 | 2005 | 2005 | 2005 | 2005 | 2006 | 2006 | 2006 | |
| Month | 01 | 04 | 08 | 12 | 03 | 07 | 11 | 02 | 06 | 10 | 01 | 05 | 09 | 12 | 04 | 08 | 11 | |
| Day | 03 | 25 | 15 | 05 | 27 | 17 | 06 | 26 | 17 | 07 | 27 | 19 | 08 | 29 | 20 | 10 | 30 | |
| 200 w | 676 | 673 | 690 | 694 | 691 | 741 | 750 | 753 | 809 | 799 | 797 | 859 | 846 | 833 | 814 | 860 | 846 | 732 |
| 500 w | 283 | 300 | 306 | 313 | 333 | 341 | 353 | 383 | 395 | 406 | 416 | 462 | 467 | 473 | 505 | 518 | 521 | 367 |

*Table 4.2 - Evolution of the number of returns series in 17 consecutive windows of length 200/500 weeks at a distance of 16 weeks for the period January 2002-November 2006.*

In practice, we have a sample of returns of finite size. Based on this sample, we need to determine 1) if the paradigm of approximate factor models applies and 2) the confidence bands of our estimates. I will now show how the asymptotic results obtained in Bai (2003) do not always apply to finite samples of size comparable to that of financial markets. Bai (2003) demonstrates that if we estimate factors with PCA:

$$\underset{N,T\to\infty}{p\lim}\frac{\tilde{F}'F}{T} = Q$$

where Q is invertible, $\mathbf{F}$ are the true factors, and $\tilde{\mathbf{F}}$ their principal components estimates. This means that there is a well-defined, invertible asymptotic covariance matrix between estimated and true factors. Bai (2003) also demonstrates that the asymptotic distribution of the estimated common component is the following:

$$\frac{\left(\widetilde{C}_{it} - C_{it}\right)}{\left(\frac{V_{it}}{N} + \frac{W_{it}}{T}\right)^{\frac{1}{2}}} \xrightarrow{d} N(0,1) \quad .$$

where $C_{it}$ is the true common component, $\widetilde{C}_{it}$ is the estimated common component, and $V, W$ are estimated by the following quantities:

$$\widetilde{V}_{it} = \widetilde{\lambda}_i'\left(\widetilde{\Lambda}'\widetilde{\Lambda}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} \widetilde{e}_{it}^{2}\widetilde{\lambda}_i'\widetilde{\lambda}_i'\right)\left(\widetilde{\Lambda}'\widetilde{\Lambda}\right)^{-1}\widetilde{\lambda}_i$$

$$\widetilde{W}_{it} = \widetilde{F}_t'\widetilde{\Theta}_i\widetilde{F}_t'$$

where $\widetilde{\Theta}_i$ is the Heteroscedasticity and Autocorrelation Consistent (HAC) estimator of Newey and West (1987) constructed with the series $\widetilde{F}_t e_{it}$. See Bai (2003), who also demonstrates that the asymptotic distribution of the factors $\sqrt{N}\left(\widetilde{F}_t - KF_t\right)$ is normal and computes the asymptotic variances (Avar).

I will clarify the meaning of this expression. It says that for each return and for each moment, the difference in the limit of infinite $T,N$ between the estimated and the true component is normally distributed with mean 0 and variance $\frac{V_{it}}{N} + \frac{W_{it}}{T}$. This is an asymptotic distribution, valid only in the large $N,T$ limit. If one assumes that our model is close to the asymptotic limit, it is possible to determine confidence bands for the common components using the asymptotic distribution.

The above formula can be interpreted as follows. Given a sample and given a PCA-based estimator of factors and factor loadings, and assuming that the sample is sufficiently large for the asymptotic distribution to apply, the true components will lie in a band that we can determine for each confidence interval. For example, if one chooses a 95% confidence interval, the true components will lie in the interval

$$\left(\widetilde{\mathbf{C}}_{it} - 1.96\left(\frac{V_{it}}{N} + \frac{W_{it}}{T}\right)^{\frac{1}{2}} N^{\frac{1}{2}}, \widetilde{\mathbf{C}}_t + 1.96\left(\frac{V_{it}}{N} + \frac{W_{it}}{T}\right)^{\frac{1}{2}} N^{\frac{1}{2}}\right) \quad t = 1,2,\ldots,T \, .$$

Alternatively, if one knows the true factors, the same formula tells us that the estimated PCA-based components will lie in the same interval. This observation cannot be applied to real factor models but it is useful in analyzing how well one can estimate approximate factor models with PCA using Monte Carlo simulations.

### *4.2.4 Simulation results for confidence bands*

Bai (2003) finds that the results of Monte Carlo simulation performed with one factor and with uncorrelated residuals agree with the theoretical prediction. This is a simplified context. I performed a more realistic test, estimating with PCA a simulated 5-factor model. The objective is to understand what parameters other than $N,T$ might influence the adequacy of the asymptotic results.

Thus I conducted Monte Carlo simulations with five factors and with a number of series and data points very close to our empirical data based on the Russell 1000. In the Russell 1000 data sets, if we choose a time window of 200 weeks, our samples will include on average nearly 800 return series, while if we choose a time window of 500 weeks, our samples will include on average 400 return series. I therefore performed Monte Carlo simulations in the two cases of 400 series and 500 data points and 800 series and 200 data points.

I simulated factor models as in Bai (2003), creating random, zero-mean, unit-variance, i.i.d. factors and a normal random matrix of factor loadings. That is, I simulated different models $R = F\beta\beta + \varepsilon$ where:

- $F$ are 200×5 or 500×5 matrices that simulate factors formed by random numbers after subtracting the column means;
- $\beta'$ are 5×800 or 5×400 matrices that respect the relative factor loadings formed by random numbers;
- $\varepsilon$ are 200×800 or 500×400 matrices formed by random numbers after subtracting the column means and multiplied by a common error variance.

I estimated factors with the first five principal components. The entire set of Monte Carlo simulations was repeated with errors of different size. Because in this experiment the true factors are known, for each series and for each moment, I created the 95% confidence intervals for each component and for each factor.

I tested the adequacy of the asymptotic approximation counting the percentage of factors and of estimated common components that fall outside the theoretical confidence band. Given the large number of data points involved, this numerical simulation is highly reliable. If the distribution of common components and of factors follows the theoretical distribution, the percentage of points that exceed the 95% confidence band should be close to 0.05. If the percentage of points that exceed the 95% confidence band is significantly different from 0.05, we can reject the assumption that a model is similar to its asymptotic limit.

Table 4.3 shows the percentage of points of factors and components that fall outside the theoretical band for different lengths of the estimation periods, different numbers of returns series, and different magnitudes of the residual terms.

| *Test* | Res std= 0.1, 200 weeks, 800 series | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Components | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Factors | 0.064 | 0.183 | 0.305 | 0.305 | 0.247 | 0.179 | 0.369 | 0.225 | 0.312 | 0.092 | 0.224 | 0.204 | 0.262 | 0.252 | 0.291 | 0.239 | 0.321 |
| | Res std= 1, 200 weeks, 800 series | | | | | | | | | | | | | | | | |
| Components | 0.044 | 0.043 | 0.038 | 0.043 | 0.040 | 0.050 | 0.043 | 0.046 | 0.044 | 0.044 | 0.046 | 0.044 | 0.046 | 0.042 | 0.042 | 0.047 | 0.045 |
| Factors | 0.164 | 0.344 | 0.386 | 0.232 | 0.279 | 0.324 | 0.273 | 0.233 | 0.242 | 0.175 | 0.373 | 0.257 | 0.310 | 0.309 | 0.278 | *0.148* | *0.300* |
| | Res std= 5, 200 weeks, 800 series | | | | | | | | | | | | | | | | |
| Components | 0.700 | 0.701 | 0.697 | 0.708 | 0.704 | 0.705 | 0.695 | 0.705 | 0.704 | 0.700 | 0.695 | 0.706 | 0.705 | 0.703 | 0.704 | 0.705 | 0.703 |
| | 0.716 | 0.695 | 0.730 | 0.722 | 0.733 | 0.724 | 0.700 | 0.719 | 0.762 | 0.685 | | 0.740 | 0.720 | 0.717 | 0.726 | 0.705 | 0.734 |
| | Res std= 0.1, 500 weeks, 800 series | | | | | | | | | | | | | | | | |
| Components | *0* | *0* | *0* | *0* | *0* | *0* | *0* | *0* | *0* | *0* | *0* | *0* | *0* | *0* | *0* | *0* | *0* |
| Factors | 0.050 | 0.084 | 0.056 | 0.025 | 0.024 | 0.038 | 0.036 | 0.100 | 0 | 0.028 | 0.084 | 0.105 | 0.020 | 0.070 | 0.042 | 0.039 | 0.007 |
| | Res std= 1, 500 weeks, 400 series | | | | | | | | | | | | | | | | |
| Components | 0.044 | 0.044 | 0.044 | 0.048 | 0.046 | 0.045 | 0.044 | 0.045 | 0.045 | 0.043 | 0.045 | 0.043 | 0.043 | 0.045 | 0.045 | 0.043 | 0.051 |
| Factors | 0.204 | 0.107 | 0.131 | 0.154 | 0.104 | 0.122 | 0.140 | 0.154 | 0.106 | 0.164 | 0.142 | 0.088 | 0.124 | 0.108 | 0.102 | 0.081 | 0.108 |
| | Res std= 5, 500 weeks, 400 series | | | | | | | | | | | | | | | | |
| Components | 0.707 | 0.704 | 0.707 | 0.710 | 0.701 | 0.707 | 0.715 | 0.703 | 0.709 | 0.707 | 0.694 | 0.711 | 0.703 | 0.712 | 0.707 | 0.712 | 0.707 |
| Factors | 0.703 | 0.692 | 0.696 | 0.706 | 0.705 | 0.705 | 0.704 | 0.684 | 0.697 | 0.715 | 0.694 | 0.695 | 0.712 | 0.701 | 0.716 | 0.704 | 0.710 |

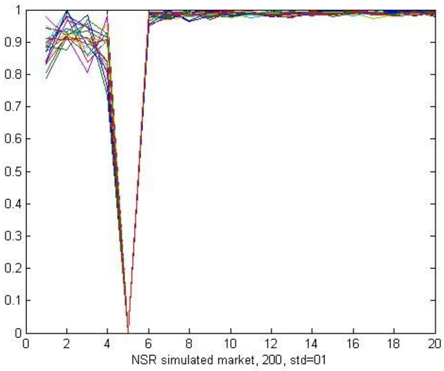
*Table 4.3 - Percentages of factors / common components that fall outside the confidence band in a simulated factor model with different parameters.*

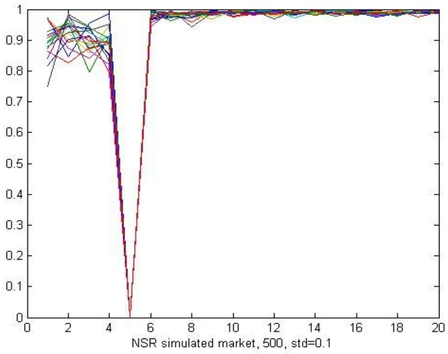
As shown in the Table 4.3, given the sample size, the behaviour of the common components and the factors depends on the magnitude of the residual term in the simulations. Results can be summarized as follows. For a time window of 200 weeks and 800 series (upper half of the table):

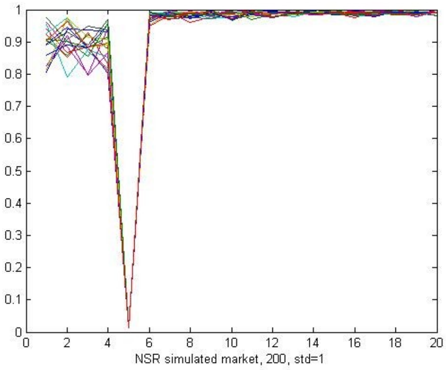
- If the magnitude of the residuals' variance is 0.1, an average of 0% of Components values and 3% of Factors values exceed the confidence band.
- If the magnitude of the residuals' variance is 1, an average of 4% of Components values and 30% of Factors values exceed the confidence band.
- If the magnitude of the residuals' variance is 5, an average of 70% of Components values and 70% of Factors values exceed the confidence band.
- 
- For a time window of 500 weeks and 400 series (lower half of the table):
- If the magnitude of the residuals' variance is 0.1, an average of 0% of Components values and 5% of Factors values exceed the confidence band.
- If the magnitude of the residuals' variance is 1, an average of 4% of Components values and 15% of Factors values exceed the confidence band.
- If the magnitude of the residuals' variance is 5, an average of 70% of Components values and 70% of Factors values exceed the confidence band.

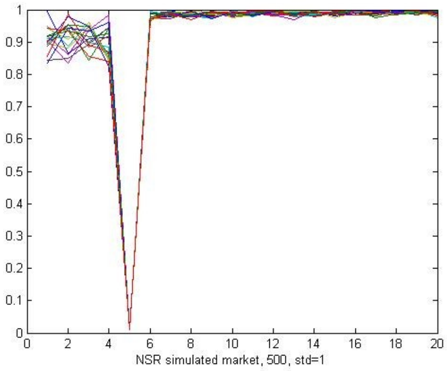
These results suggest that the assumptions of large factor models are or are not satisfied not only in function of the size of *T,N* but also in function of the magnitude of the residuals. In my Monte Carlo simulations, if *T,N* are large and the residuals are also large (in a sense to be made precise), then the distribution of factors and components does not follow the theoretical distribution. Therefore we cannot confidently assume that our finite model is close to the asymptotic limits.

Let's reconsider the assumption of approximate factor models. An approximate factor model requires that the covariance matrix of the noise processes have bounded eigenvalues while the first *r* eigenvalues of the covariance matrix of both factors and returns diverge with *N,T.* This notion can be defined precisely only in the limit of infinite *N,T.* In a finite context, it would seem reasonable to require that, in addition to large *N,T* , we require a sudden, large change in the magnitude of the eigenvalues of the returns covariance matrix so that the smallest eigenvalue of the returns covariance matrix are large with respect to the largest eigenvalue of the residuals covariance matrix. In addition, if we believe that our data have a well-defined factor structure, we expect a change in magnitude of the eigenvalues at a well-defined *k.* If we were to find more than one large change, we might expect a nested factor structure.









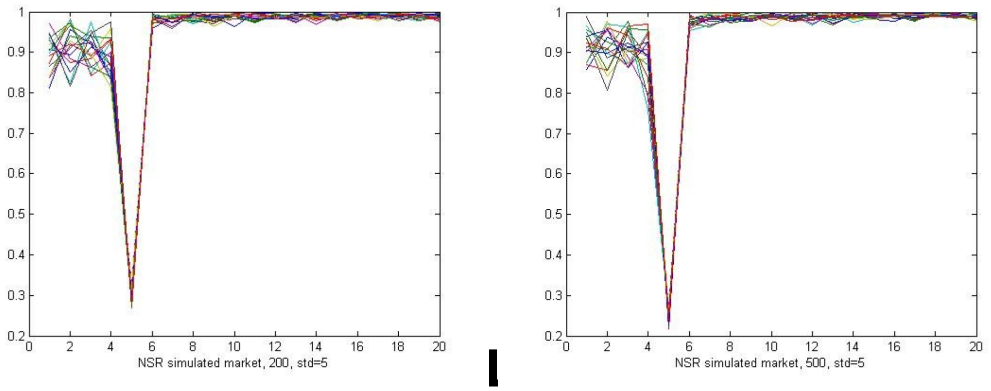
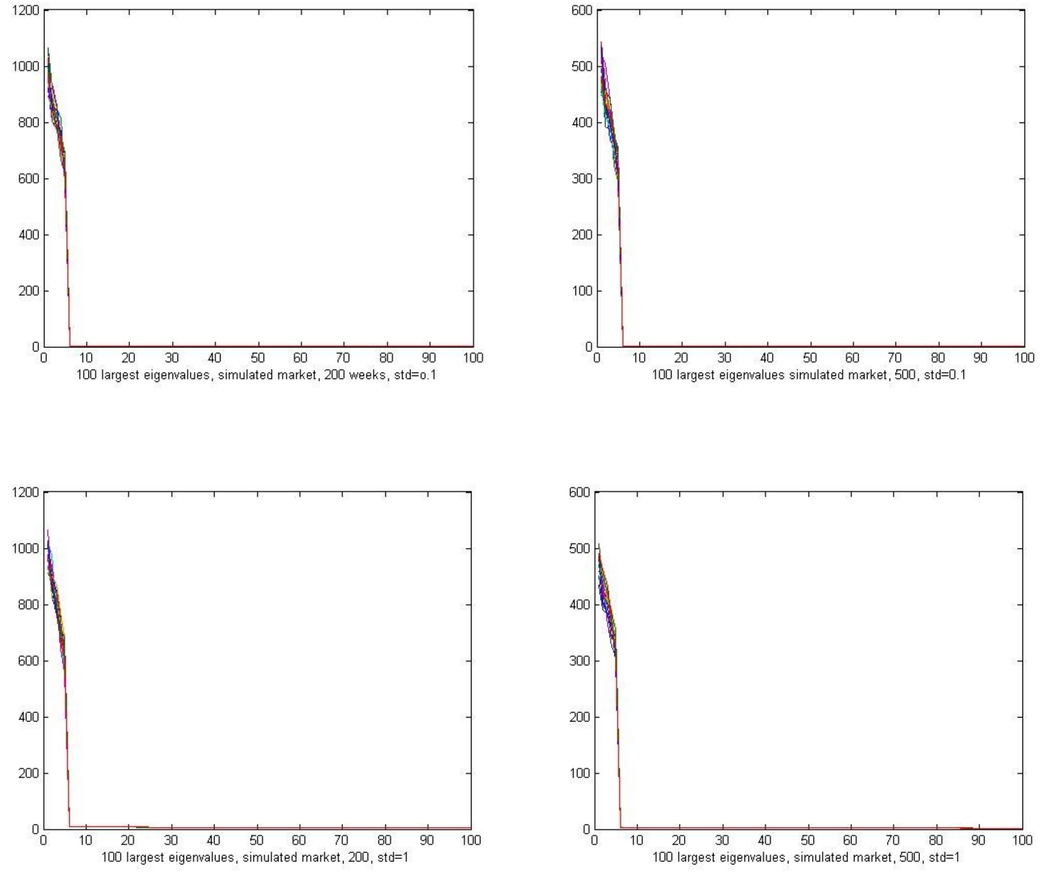


*Figure 4.1. Plot of the NSR of the simulated market in function of the number of factors and for different values of the standard deviation of residuals. The plot is made superposing the plots of the NSR in different time windows. The figure on the left is relative to a 200-week time window, the figure on the right is relative to a 500-week time window.*



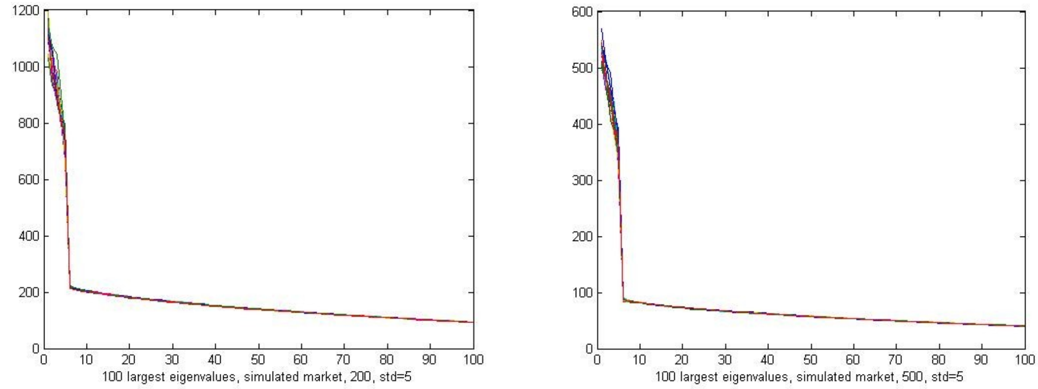


*Figure 4.2. Plot of the 100 largest eigenvalues of the covariance matrix of the simulated market for different values of the standard deviation of residuals. The figures on the left are relative to a 200-week time window, the figures on the right to a 500-week time window. The plots are made superposing the plots of the 100 largest eigenvalues in different time windows.*

### *4.3 When can factors be estimated with principal components?*

The Monte Carlo simulations show that principal components estimated on finite samples do not always correctly estimate the population factor model. As described in Chapter 1, I propose to identify the conditions that allow factors to be estimated with principal components.

Under what conditions can factors be estimated with principal components? In order to solve this problem, I first need to introduce concepts of distance between factors. As observed in Chapters 2 and 3, there is a fundamental indeterminacy in factor models given that factors can be multiplied by an invertible matrix and generate an observationally equivalent model. Note that this is not the factor indeterminacy discussed in Chapter 3: the latter is indeterminacy between factors and residuals and does not affect factor loadings, while rotational indeterminacy affects only factors and their loadings. Therefore factor distance must be insensitive to factor rotation even after imposing the condition that factors are orthonormal variables.

Schneeweiss and Mathes (1995) define the closeness of two random vectors using the canonical correlation coefficients between the respective factors. More precisely, suppose that $\xi_1,\xi_2$ are two random $q$-vectors defined on the same probability space. Consider the canonical correlations $\rho_i = \rho_i(\xi_1,\xi_2), i = 1,\ldots,q$ in decreasing order.

Recall that, given two random $q$-vectors $\xi_1,\xi_2$, in canonical correlation analysis one looks for linear combinations $x_{1i},x_{2i}$, $i = 1,\ldots,q$ of the vectors $\xi_1,\xi_2$ such that the variables $\xi'_1 x_{1i},\xi\xi_2 x_{2i}$ are maximally mutually correlated and uncorrelated for different values of *i*. The variables $\xi'_1 x_{1i},\xi\xi_2 x_{2i}$ are called *canonical variates* and the relative correlation coefficients are called *canonical* *correlations*. Consider now the covariance matrices $\Sigma_{11}, \Sigma_{22}, \Sigma_{12}, \Sigma_{21}$ of each vector and between vectors $\xi_1, \xi_2$. In canonical correlation analysis, it can be demonstrated that the squares of the $\rho_i$ are the eigenvalues of the matrix $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$.

Schneeweiss and Mathes (1995) define the closeness between two models with factors $f_1,f_2$ not necessarily orthonormal as the sum of the squares of the canonical correlation coefficients between the factors:

$$r = \sum_{i=1}^{q} \rho_i^2 = trace\left(\mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21}\right).$$

The quantity *r* assumes values $0 \leq r \leq q$. It attains the maximum value *q* iff there is a non-singular matrix $K$ such that: $f_1 = Kf_2$. Given that canonical correlations are invariant after orthogonal transformations, the quantity *r* is clearly invariant after an orthogonal transformation. If $f_1$ are orthonormal factors, that is, $\Sigma_{11} = I$, and if $\Sigma_{22} = \Lambda = diag(\lambda_1,\ldots,\lambda_q)$ then the measure *r* becomes: $r = \sum_{i=1}^{q} \rho_i^2 = trace\left(\mathbf{\Sigma}_{12}\, \mathbf{\Lambda}^{-1}\mathbf{\Sigma}_{21}\right)$. If we divide each principal component by the square root of the corresponding eigenvalue, we obtain the normalized principal components $f_2\Lambda^{-\frac{1}{2}}$ and the measure *r* becomes $r = \sum_{i=1}^{q} \rho_i^2 = trace\left(\mathbf{\Sigma}^2\right) = \sum_{i=1}^{q} \eta_i^2$ where $\mathbf{\Sigma} = \mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{21}'$ is the covariance matrix between the orthonormal factors and the normalized principal components and $\eta_i$ are its eigenvalues.

I define another measure of closeness between factor models using the concept of Procrustes analysis (See Gower and Dijksterhuis, 2004). Given two $m \times n$ matrices $A,B$, Procrustes analysis finds an $n \times n$ orthogonal matrix **C** that minimizes the Frobenius norm of the matrix $A - BC$:

Procrustes problem: $\min_C \|A - BC\|_F$ subject to $C'C = I$

To solve the Procrustes problem, consider first the SVD of the matrix $B'A$: $B'A = USV',\ S = diag[\sigma_i]$. It can be demonstrated (see Golub and Van Loan, 1996) that the solution of the Procrustes problem is the following matrix: $C = UV'$ and that the following holds:

$$\min_C \|A - BC\|_F^2 = \|A\|_F^2 + \|B\|_F^2 - 2\sum_{i=1}^{n} \sigma_i$$

Consider the same setting as before where $f_1, f_2$ are the factors of two factor models. Suppose that $f_1, f_2$ are orthonormal random $q$-vectors defined on the same probability space. I define the distance $d$ between the two factor models as follows:

$$\frac{1}{2}\min_C \|f_1 - f_2 C\|_F^2 \text{ subject to } C'C = I$$

From the above discussion of Procrustes analysis, given that the factors $f_1, f_2$ are orthonormal $q$-vectors, the distance between factors is:

$$d = \frac{1}{2}\|f_1 - f_2 UV'\|_F \text{ where } f_2' f_1 = USV'$$

which yields:

$$d(f_1, f_1) = q - \sum_{i=1}^{q} \sigma_i = q - \sum_{i=1}^{q} \sqrt{\eta_i}$$

where $\sigma_i$ are the singular values of the matrix $f_1 f'_2$, that is, the square roots of the eigenvalues of the covariance matrix between the factors. If there is an orthogonal matrix $C$ such that $f_1 = f_2 C$, then $d$=0. Conversely if $d$=0, $f_1 - f_2 UV' = 0$ because the Frobenius norm of a matrix is zero iff the matrix is zero. Therefore $f_1 = f_2 C, C = UV'$. Therefore $d = 0 \Leftrightarrow r = q$. More in general, the following relationship holds: $d \leq q - r$

Let's now go back to the problem of model identifiability and estimation. In the literature, the following conditions for identification and estimation have been determined. Scalar factor models (i.e., strict factor models with uniform variance of residuals) are uniquely identified and factors can be estimated with principal components. Finite strict factor models with heterogeneous variances admit identification and estimation of the model parameters but factors are subject to indeterminacy and can be estimated with generalized least squares (GLS) but not with principal components. Infinite factor models with a finite number of common factors with *infinite eigenvalues* can be identified and factors are estimated with asymptotic principal components. Infinite factor models with a finite number of common factors with *finite eigenvalues* cannot be identified and factors are estimated with asymptotic principal components. We can summarize the facts relative to factor model identification and factor estimation in the following Table 4.4

| **Scalar factor model** | **Strict factor model** | **Approximate factor model** | **Bai and Ng factor model** | **Weak factors model** |
|---|---|---|---|---|
| Finite N, infinite T | Finite N, infinite T | Infinite T, infinite N | Infinite N, infinite T | Infinite N, infinite T |
| Model identified and estimated with ML | Model identified and estimated with ML | Model identified and estimated with asymptotic PCA or as a limit of strict factor models | Model identified and estimated with asymptotic PCA | Model not identified |
| Factors can be estimated with principal components | Factors cannot be estimated with ML. Factors not identified and estimated with GLS or principal components | Factors identified with principal components or as limit of strict factor models | Factors identified with principal components | Factors estimated with principal components |

*Table 4.4 Factor model identification and estimation.*

Finite models with possibly correlated and autocorrelated residuals do not fall into any of the categories in Table 4.4. Still these models are the most important models in practice. In the area of econometrics, financial returns as well as macroeconomic variables form finite models. Current practice is to use asymptotic results from infinite models. However, simulations in section 4.2.4 demonstrate that asymptotic results might not be applicable to finite samples. In addition, from the economic point of view there is no natural way to allow the model to become infinite. Therefore it is necessary to develop a theory of how finite samples can be represented with possibly many different factor models and determine criteria for choosing the best model(s).

Let's first discuss factor models from the population point of view. Consider a finite sample with $T$ samples of an $m$-vector $X$. Consider a finite factor model: $X = FB' + E$ with $m$ observed variables and $p$ factors. Assume that factors are orthonormal and that factors and residuals are mutually uncorrelated. Then, the covariance matrix $\mathbf{\Omega}$ of the observed variables can be written as $\Omega = BB' + \Sigma$ where $\mathbf{\Sigma}$ is the covariance matrix of the residuals.

If no restriction is placed on the matrices $B,\Sigma$ clearly the model is not identifiable because there are more parameters than conditions. For finite models, the typical restriction imposed in the literature is that $\mathbf{\Sigma}$ be diagonal. This restriction characterizes strict factor models. Strict factor models are identifiable because the number of parameters of the matrix $\mathbf{\Sigma}$ is reduced from $\frac{1}{2}m(m+1)$ to $m$. In the case of scalar factor model, the number declines to 1.

It is possible to construct progressive "perturbations" of the strict factor model that are still identifiable and that could be estimated with ML. For example, an obvious candidate is a near-scalar model where the residual covariance matrix has homogeneous variances and a constant covariance, thereby reducing the number of parameters of the residual covariance matrix to 2. The next step would be to accept a small number of heterogeneous variances and covariances.

This exercise is, however, practically without much interest for three reasons. The first reason is that restrictions on the covariance matrix of the residuals are arbitrary and any specific restriction is unlikely to meet the empirical test. The second reason is that the computation of ML estimates of the parameters would become rapidly impossible when the number of observed variables is in the range of hundreds, as is typical of financial econometrics and, presently, also of macroeconomics. The final reason is that learning theory constraints have to be taken into account. The objective of factor models is not to describe sample data but to generalize to out-of-sample data. If too many independent parameters are added, the generalization ability of the model decays rapidly.

Therefore, in this Dissertation I propose to focus on those criteria that allow one to 1) establish that data can be represented by a factor model and 2) estimate factors with principal components and loadings with the corresponding eigenvectors. In practice this is equivalent to establishing under what conditions current “large factor model theories” can be applied to finite samples.

Let’s first consider the distance *r* proposed by Schneeweiss and Mathes (1995), which we can write:

$$r = \sum_{i=1}^{q} \rho_i^2 = = trace\left(\left(E\left(f'_1 f_1\right)\right)^{-1} E\left(f'_1 f_2\right)\left(E\left(f'_2 f_2\right)\right)^{-1} E\left(f'_2 f_1\right)\right) = trace\left(\Sigma_{12}\, \Lambda^{-1} \Sigma_{21}\right)$$

As discussed earlier in this section. Schneeweiss and Mathes (1995) prove that the distance between principal components and the factors of the data set is: $r = q - trace\left(\mathbf{\Lambda}^{-1}\mathbf{\widetilde{B}'\Sigma\widetilde{B}}\right)$. In fact:

$$E\left(f'_1 f_1\right) = I_q,$$
$$E\left(f_1 f'_2\right) = E\left(f_1 x\widetilde{B}\right) = E\left(f_1\left(f'_1 B' + w\right)\widetilde{B}\right) = E\left(f_1 f'_1 B'\widetilde{B}\right) = B'\widetilde{B}$$
$$E\left(f_2 f'_2\right) = \widetilde{B}xx'\widetilde{B}' = \widetilde{B}\Omega\widetilde{B}' = \Lambda^{-1}$$
$$r = \sum_{i=1}^{q} \rho_i^2 = = trace\left(\left(E\left(f'_1 f_1\right)\right)^{-1} E\left(f_1 f'_2\right)\left(E\left(f_2 f'_2\right)\right)^{-1} E\left(f_2 f'_1\right)\right) =$$
$$trace\left(B'\widetilde{B}\Lambda^{-1}\widetilde{B}'B\right)$$

The trace has the property that $trace\left(AB\right) = trace\left(BA\right)$ (but the trace is not commutative for 3 or more matrices). Therefore:

$$
\begin{aligned}
r &= trace\left(B'\tilde{B}\Lambda^{-1}\tilde{B}'B\right) = trace\left(\tilde{B}'BB'\tilde{B}\Lambda^{-1}\right) = trace\left(\tilde{B}'(\Omega - \Sigma)\tilde{B}\Lambda^{-1}\right) = \\
&= q - trace\left(\tilde{B}'\Sigma\tilde{B}\Lambda^{-1}\right)
\end{aligned}
$$

This formula represents the distance between factors and principal components in terms of the loadings of the principal components (which are the first *q* eigenvectors of the covariance matrix of the data), the first *q* eigenvalues of the covariance matrix of the data, and the covariance matrix of the residuals.

Now suppose that the population can be described by a factor model with *q* factors. Let $\sigma^2$ be equal to the largest eigenvalue of $\mathbf{\Sigma}$ (the covariance matrix of residuals) and *b* be equal to the smallest of the largest eigenvalues of $\mathbf{\Omega}$. If the eigenvalues of the covariance matrix of the data are ordered in decreasing order, $b = \lambda_q$. Consider now the $p \times p$ matrix $B'B$ and call its eigenvalues $(d_1, \ldots, d_p)$ in decreasing order. The rank of the matrix $B'B$ is *q* and therefore only *q* of its eigenvalues are greater than zero: $d_1 \geq \cdots d_q > 0, d_{q+1} = \cdots = d_p = 0$. On the other hand, in general the matrix $\mathbf{\Sigma}$ has full rank *p* and all its eigenvalues are positive. The following property holds (Golub and Van Loan, 1996, pp 396):

$$
\begin{aligned}
& d_i \leq \lambda_i \leq d_i + \sigma^2, i = 1, \ldots, q \\
& \lambda_i \leq \sigma^2, i = q + 1, \ldots, p
\end{aligned}
.
$$

Call SNR(*q*) (signal-to-noise ratio) the ratio ${}^{\lambda_q}\!/_{\sigma^2}$. Observe that SNR(*q*) is defined in terms of a factor model whose residuals have a covariance matrix $\mathbf{\Sigma}$ with maximum eigenvalue $\sigma^2$. It is not defined in terms of the principal components. The corresponding ratio for the principal components is ${}^{\lambda_q}\!/_{\lambda_{q+1}} \geq {}^{\lambda_q}\!/_{\sigma^2}$. This implies that the principal components have the highest possible SNR(*q*) for every *q*.

Suppose data are generated by a factor model with *q* factors and with a given SNR(*q*). Building on Schneeweiss and Mathes (1995), we can make the following derivation:

$$
\begin{aligned}
& d \leq = q - r = trace\left(\tilde{B}'\Sigma\tilde{B}\Lambda^{-1}\right) \\
& d \leq trace\left(\tilde{B}'\Sigma\tilde{B}\Lambda^{-1}\right) \leq \sigma^2 trace\left(\Lambda^{-1}\right) \leq q\frac{\sigma^2}{b}
\end{aligned}
$$

To prove the last result, let's first introduce the notation $A \geq B\ (A > B)$, which means that the matrix $A - B$ is positive semidefinite (definite). A matrix $A$ is positive semidefinite (definite) if $x'Ax \geq 0\ (> 0)$ for every vector $\mathbf{x}$. In a positive semidefinite (definite), the largest entries are on the diagonal and the diagonal entries are all positive (non negative).

We can now state that $\Lambda \geq bI$ and $\widetilde{B}'\Sigma\widetilde{B} \leq \sigma^2 I$. The first follows immediately from the properties of the eigenvalues established above. The second statement can be proved as follows. Consider the eigenvalues $\sigma_i$ of the matrix $\mathbf{\Sigma}$ and form the matrix $diag(\sigma_i)$ so that $\Sigma = Q'diag(\sigma_i)Q$. The following property holds because of the definition of $\sigma^2$: $diag(\sigma_i) \leq \sigma^2 I$. Hence: $diag(\sigma_i) - \sigma^2 I$ is negative semidefinite and therefore $Q'(diag(\sigma_i) - \sigma^2 I)Q$ is also negative semidefinite and therefore $\widetilde{B}'Q'(diag(\sigma_i) - \sigma^2 I)Q\widetilde{B}$ is negative semidefinite which implies $\widetilde{B}'Q'(diag(\sigma_i))Q\widetilde{B} \leq \sigma^2 I$.

It has therefore been established the result that the distance from the factors of any factor model with SNR(*q*) and the first *q* principal components is $d \leq q - r \leq q\frac{\sigma^2}{b}$. At first sight this result might look counterintutive: in a measure of the distance between factors and principal components, one would expect to find both ratios $\frac{\sigma^2}{b}$ and $\frac{\lambda_{k+1}}{b}$. However, consider that the formula $r \geq q - q\frac{\sigma^2}{b}$ is a lower bound estimate based on the ratio between the smallest of the first *q* largest eigenvalues and the largest eigenvalue of the residuals of the factor model. The actual *r* can be larger. For example, if the factor model is formed by normalized principal components, then NSR(*q*)=$\frac{\lambda_{k+1}}{b}$ but *r=q* because in this case $\widetilde{B}'\Sigma$=0. Note also that if the factor model is a scalar factor model, we still have $r \geq q - q\frac{\sigma^2}{b}$. However, in the case of a scalar model, the estimate of *r* could be refined. In fact, if $\Sigma = \sigma^2 I$, then $trace(\widetilde{B}'\Sigma\widetilde{B}\Lambda^{-1}) = trace(\sigma^2 I\Lambda^{-1}) = trace(\sigma^2\Lambda^{-1})$ and $r = trace(\sigma^2\Lambda^{-1}) \geq q\frac{\sigma^2}{b}$

The above analysis determines the distance between the factors of a factor model and the principal components of the same population. It tells us that if a factor model has a high SNR($q$), then its factors can be approximated with principal components with a level of precision given by the relative $r$ and $d$. Under this assumption, factors are almost uniquely determined and there is not much room for determining different models.

These considerations apply from the population point of view. In practice, however, we are given a set of empirical data $X$ and we have to decide if and how the data can be represented with a factor model. From the above considerations, a sensible way to understand if the data $X$ admit a factor model representation is to analyze the vector $\frac{\lambda_{k+1}}{\lambda_k}, k = 1,\ldots,p-1$. Suppose that the ratio $\frac{\lambda_{k+1}}{\lambda_k}$ is very small only for $k = q$. Any factor model with NSR($q$) close to $\frac{\lambda_{q+1}}{\lambda_q}$ will have factors that can be well mimicked by the first $q$ principal components and, for any other value of $k$, no factor model will be close to principal components.

This latter fact calls for a comment. For $k = q$, any factor model with NSR($q$) close to $\frac{\lambda_{q+1}}{\lambda_q}$ will have residuals whose magnitude is close to the magnitude of the residual principal components. Any such model will, from the point of view of in-sample residuals, be quite similar to principal components. Consider now $k = q+1$. Suppose the ratio $\frac{\lambda_{q+2}}{\lambda_{q+1}}$ is large, say $\frac{\lambda_{q+2}}{\lambda_{q+1}} \approx 1$. For any factor model with $q+1$ factors $\frac{\sigma^2}{b} \geq \frac{\lambda_{q+2}}{\lambda_{q+1}}$ because, by construction, principal components exhibit the lowest possible NSR($q$).

Because of this fact, the factors of any model with $q+1$ factors can be very different from the first $q+1$ principal components even if in-sample residuals remain small. That is, for $k = q$ it is possible to say that models with similar NSR($q$) will be very similar and therefore establish a criterion of near identification of factors. For any other number of factors, there can be very different factor models with residuals of approximately the same magnitude. Therefore, I have thus far established the following:

If a factor model with $q$ factors exhibits a very small NSR, then the factors of that model are very similar to the first principal components. Similarity can be measured either by the coefficient $r$ or by the Procrustes distance $d$. Factor models with $q$ factors are nearly unique.

Let's now discuss the question of determining the number of factors. As observed, empirically we have a fixed number of returns time series and a possibly variable but certainly bounded number of samples. From the economic point of view, there is no compelling reason to assume that the empirical finite sample is a finite sample of an infinite market with an infinite number of both returns and observations. Still, from the economic point of view, there is no compelling reason to choose any specific path to infinity. Such assumptions are arbitrary and are warranted only if the sample behaves approximately like its asymptotic limit.

Empirically one can only make an asymptotic assumption plausible. In the theory of approximate factor models, the NSR tends to zero and asymptotically factors are unique up to a rotation. Therefore, we can establish the first criterion:

*Criterion 1: If NSR(q) is very small, it is plausible to assume that the sample is representative of an approximate factor model. For any number of factors where NSR(q) is large (close to 1) the assumption that the sample is representative of an approximate factor model is not tenable.*

Let's now analyse how to determine the number of factors. As observed in Chapter 3, there are two main techniques for determining the number of factors: information criteria and random matrix theory. Consider first information criteria. The Bai and Ng (2003) criteria are inherently asymptotic. The authors propose criteria based on modifying the BIC and AIC to take into account the fact that the number of parameters grows with both $T$ and $N$. Though the need to consider both dimensions is obvious asymptotically, in a finite sample, however, it should be sufficient to count the number of parameters correctly. Given the proliferation of criteria, in a finite sample it is difficult to decide whether or not the number of factors that eventually correspond to a drop in the NSR can also be chosen by the information-based criteria.

Let's now discuss criteria based on random matrix theory. Let's assume that the empirical distribution of eigenvalues exhibit a bulk distribution in reasonable agreement with the fundamental Marčenko-Pastur law plus a number of well identified isolated eigenvalues. As discussed in Chapter 3, the behaviour of the largest eigenvalues depends on both small sample effects and the presence of correlations in residuals. In addition, in "spiked" models, which represent local correlations, eigenvalues follow an asymptotic distribution. In principle, in a random matrix model, there is no certainty that the ordering of the largest empirical eigenvalues effectively corresponds to the largest true eigenvalues. In addition, deciding the threshold between "spiked" local correlations and correlations due to common factors is arbitrary. As a result, random matrix theory is compatible with choosing a number of factors equal to the number $q$ where there is a large drop in the NSR.

*Criterion 2: If NSR(q) is very small it is plausible, on the basis of results from random matrix theory, to assume that the number of factors is q. If the NSR(q) is large (close to 1) for every,q, then random matrix theory cannot offer any guidance.*

Let's comment. In random matrix theory we have different asymptotic models in function of different assumptions as regards correlations, namely, in particular the null of no correlation and spiked factor models. Any empirical distribution of eigenvalues is compatible with either a spiked model or with a null model for residuals plus diverging eigenvalues or with a spiked model for residuals plus diverging eigenvalues.

The choice of which asymptotic results, if any, to adopt, can only be made considering additional features of the sample and choosing those features that are mostly in agreement with all tests. In particular, one situation that lends itself to the adoption of a model is the presence of a small NSR for a small number of factors. This situation is compatible with:

- Asymptotic principal components to estimate unique factors
- Random matrix theory to estimate the number of factors.

However, a sample where the empirical distribution of eigenvalues drops slowly and there is no truly small NSR cannot be considered a sample extracted from an approximate factor model population. Onatski (2007) proposes the use of spiked models when the empirical distribution of eigenvalues drops slowly. This model makes the assumption of Gaussian i.i.d. residuals and is exposed to the same criticism of strict factor models.

### *4.4 Factor models as noisy multiple communication channels*

I introduce the notion that factor models can be viewed as noisy communication channels where factors are the emitters and the observed variables are the receivers. I then introduce criteria based on the efficiency and capacity of this idealized communication channel. The motivation for adopting this analogy is twofold: 1) to use results from the vast literature on communication channels and 2) to use basic information theory concepts in understanding and evaluating factor models. The fundamental idea underlying this analogy is that a "good" factor model can be likened to a "good" communication channel with a large capacity. In other words, in practice, finite samples can only approximately be represented with factor models. We need criteria to choose the optimal assumptions.

I propose criteria based on how efficiently information is transferred from factors to the observed time series. Actually the interest in using a factor model is effectively related to how efficiently it transmits information from factors to returns (or to any other series). For example, an asset manager is interested in a factor model of returns only if returns can be "explained" (for risk management) or "predicted" (for forecasting purposes) from the factors. This implies that information from the factors must be efficiently transmitted to returns.

This view might seem in contrast with the idea that factors are "diffusion indexes", that is, a summary of information dispersed in a large number of series. In fact, one of the motivations for using factor models is effectively dimensionality reduction. Factor models allow one to efficiently capture information that is dispersed in a large number of mutual relationships and that cannot be handled directly. However, there is no real contradiction between the two points of view. A "communication theory" point of view can be used to evaluate factor models after estimating factors.

### *4.4.1 Channel Capacity*

Let's now write down the first result of looking at factor models as communication channels. In communications theory, the *capacity* of a noisy communication channel is defined as the superior of the mutual information between receiver and transmitter. It has been demonstrated (see Tulino and Verdù 2005) that the capacity of a multiple communication channel described by a linear equation of the type:

$$y = Hx + n$$

where **x** is the $K\times 1$ input vector, $n$ is Gaussian noise, **y** is the $N\times 1$ output vector and $H$ is a $N\times K$ matrix of transfer coefficients is given by the following expression:

$$C = E\left[I\left(y,x\middle|H\right)\right] = \frac{1}{N}\log\det\left(I + SNRHH^*\right) =$$
$$= \frac{1}{N}\sum_{i=1}^{N}\log\left(1+\lambda_i\left(HH^*\right)\right) = \int_0^\infty \left(1+SNRx\right)dF_{HH^*}$$

where SNR is the source signal-to-noise ratio. If we interpret the inputs as factors, the transfer coefficients as factor loadings and the output as the observed variables, we have a theoretical measure of factor model efficiency. But the previous caveat applies: the number C represents the in-sample factor model efficiency obtained through a process that implies an optimization process. In the next section I will discuss how to apply this concept to determine the optimal factor model.

### *4.4.2 Criteria for model choice*

When the empirical distribution of eigenvalues drops slowly, it is reasonable to assume that the data cannot be described by any model that implies uniqueness of factors. In this case, we have to choose between different models, none of which can be considered to be the true factor representation of data. In this situation, criteria for choosing the optimal model would be useful. This is a particularly pressing problem for equity portfolio managers that are trying to improve their factor models by adding new factors.

When factors cannot be surely identified with principal components, they might be identified through non linear methods such as clustering. In addition, factors are not necessarily estimated by portfolios - a situation which opens the possibility of using factors that are exogenous, for example macroeconomic factors. In practice, factor models become regression models.

Model selection criteria, such as the Bai and Ng criteria, are based on a trade-off between a reduction of in-sample residuals and a penalty for the number of parameters. Model selection criteria based on random matrix theory try to separate directly noise from information based on a specific model for the covariance matrix. However, the assumption of a smooth decay of the empirical distribution of the eigenvalues makes it difficult to apply these criteria.

I propose another criterion which is based on the transfer of information from factors to returns (observed variables). In Section 4.4.1. the capacity of a multiple linear information channel was defined as the mutual information input-output which has the following expression:

$$C = \frac{1}{N}\sum_{i=1}^{N} \log\left(1 + SNR\lambda_i\left(HH^*\right)\right)$$

Let's look at a factor model as a multiple communication channel where factors are the inputs, returns (observed variables) are the output, and residuals are transmission noise. I propose to use the channel capacity (i.e., the mutual information) as a criterion for selecting factor models. Recall that, in this context, the SNR is the ratio between the average "power" of the factors and the average "power" of the residuals:

$$SNR = \frac{N\|F\|^2}{K\|n\|^2}$$

Mutual information as a measure of the strength of mutual dependence has been used in many instances in biomathematics. In our case, the matrix $H$ is the matrix of factor loadings assuming factors are orthonormal. Therefore I introduce the third criterion:

*Criterion 3: If NSR(q) is large (close to 1) for every q, then model selection criteria look at the mutual information between factors and observed variables given by the expression:* $C = \frac{1}{N}\sum_{i=1}^{N} \log\left(1+ SNR\lambda_i\left(HH^*\right)\right)$. *This criterion chooses those models that have more weight in the eigenvalues of BB'.*

In summary, in this section I proved that if a factor model has a small noise-to-signal ratio then its factors are very close to the first principal components and can be estimated with principal components. Factors are approximately unique and we can use the asymptotics results of approximate factor models.

If, on the other hand, the noise-to-signal ratio is never small, then factor models are not unique and factors are not close to principal components. In this case, we can choose between different factor models using the criterion of maximizing the mutual information between factors and observed variables.

### *4. 5. Empirical results relative to the Russell 1000*

In this section we apply our analysis to the stocks in the Russell 1000 universe. Let's first compare the NSR of the Russell 1000 with that of our simulated market. Given a universe of time series such as the Russell 1000, we can estimate the NSR for different time windows and for different numbers of factors. To do so, we compute the eigenvalues of the covariance matrix of returns in each time window and the ratios $\rho_k = {}^{\lambda_k}\!/\!_{\lambda_{k+}}$ for different values of *k.* Tables 4.5 a and b show the NSR of the Russell 1000 for the first *k*=1,…,20 principal components, using all available series in time windows of 200 and 500 weeks in the period January 2002-December 2006. As shown in Table 4.5, time windows end at the 17 dates included in this period with a spacing of 16 weeks.

| Date | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | 2002 | 2002 | 2002 | 2002 | 2003 | 2003 | 2003 | 2004 | 2004 | 2004 | 2005 | 2005 | 2005 | 2005 | 2006 | 2006 | 2006 |
| Month | 01 | 04 | 08 | 12 | 03 | 07 | 11 | 02 | 06 | 10 | 01 | 05 | 09 | 12 | 04 | 08 | 11 |
| Day | 03 | 25 | 15 | 05 | 27 | 17 | 06 | 26 | 17 | 07 | 27 | 19 | 08 | 29 | 20 | 10 | 30 |

*Table 4.5 – The end date of the 17 time windows.*

Table 4.6a is relative to 200-week-long time windows. It is formed by 17 columns and 20 rows of data. Each column represents the 20 NSR($k$) relative to the first 20 largest eigenvalues for each time window. Each column is relative to a time window that ends at the date indicated in the top cell of the table.

Table 4.6b is relative to 500-week-long time windows. It is formed by 17 columns and 20 rows of data. Each column represents the 20 NSR($k$) relative to the first 20 largest eigenvalues for each time window. Each column is relative to a time window that ends at the date indicated in the top cell of the table. In both tables (i.e., Tables 4.6a and 4.6b) the NSR shows a smooth growth in each column. The NSR starts at approximately 0.2-0.3 for the first principal component and arrives at values close to 1 after 10 to 12 principal components.

| W.N. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *NSR(*1) | 0.37 | 0.36 | 0.42 | 0.38 | 0.37 | 0.33 | 0.31 | 0.23 | 0.18 | 0.18 | 0.15 | 0.16 | 0.16 | 0.17 | 0.18 | 0.17 | 0.19 |
| *NSR(*2) | 0.48 | 0.45 | 0.41 | 0.40 | 0.39 | 0.37 | 0.38 | 0.63 | 0.56 | 0.59 | 0.75 | 0.90 | 0.94 | 0.90 | 0.80 | 0.65 | 0.66 |
| *NSR(*3) | 0.64 | 0.68 | 0.79 | 0.79 | 0.80 | 0.83 | 0.82 | 0.66 | 0.84 | 0.76 | 0.57 | 0.49 | 0.53 | 0.54 | 0.61 | 0.85 | 0.61 |
| *NSR(*4) | 0.91 | 0.81 | 0.84 | 0.93 | 0.90 | 0.91 | 0.88 | 0.86 | 0.87 | 0.82 | 0.88 | 0.95 | 0.92 | 0.97 | 0.93 | 0.92 | 0.86 |
| *NSR(*5) | 0.88 | 0.85 | 0.79 | 0.71 | 0.74 | 0.84 | 0.84 | 0.85 | 0.76 | 0.81 | 0.96 | 0.94 | 0.89 | 0.95 | 0.96 | 0.75 | 0.85 |
| *NSR(*6) | 0.87 | 0.91 | 0.93 | 0.91 | 0.88 | 0.84 | 0.85 | 0.86 | 0.92 | 0.98 | 0.91 | 0.83 | 0.91 | 0.85 | 0.88 | 0.90 | 0.97 |
| *NSR(*7) | 0.94 | 0.95 | 0.86 | 0.94 | 0.91 | 0.97 | 0.98 | 0.97 | 0.88 | 0.87 | 0.95 | 0.95 | 0.92 | 0.93 | 0.89 | 0.89 | 0.96 |
| *NSR(*8) | 0.92 | 0.94 | 0.98 | 0.97 | 0.98 | 0.95 | 0.93 | 0.91 | 0.95 | 0.93 | 0.89 | 0.99 | 0.90 | 0.90 | 0.90 | 0.98 | 0.94 |
| *NSR(*9) | 0.97 | 0.98 | 0.95 | 0.89 | 0.93 | 0.94 | 0.96 | 0.95 | 0.96 | 0.97 | 0.94 | 0.90 | 0.94 | 0.93 | 0.95 | 0.91 | 0.92 |
| *NSR(*10) | 0.93 | 0.93 | 0.95 | 0.98 | 0.98 | 0.93 | 0.96 | 0.97 | 0.94 | 0.99 | 0.98 | 0.94 | 0.94 | 0.97 | 0.91 | 0.98 | 0.99 |
| *NSR(*11) | 0.97 | 0.94 | 0.98 | 0.96 | 0.96 | 0.95 | 0.92 | 0.96 | 0.94 | 0.91 | 0.95 | 0.98 | 0.93 | 0.90 | 0.96 | 0.94 | 0.95 |
| *NSR(*12) | 0.97 | 0.95 | 0.93 | 0.99 | 0.94 | 0.99 | 0.98 | 0.94 | 0.99 | 0.97 | 0.91 | 0.94 | 0.95 | 0.97 | 0.95 | 0.96 | 0.98 |
| *NSR(*13) | 0.91 | 0.94 | 0.97 | 0.95 | 0.97 | 0.99 | 0.98 | 0.96 | 0.96 | 0.95 | 0.98 | 0.92 | 0.96 | 0.94 | 0.96 | 0.97 | 0.97 |
| *NSR(*14) | 0.98 | 0.96 | 0.95 | 0.93 | 0.93 | 0.95 | 0.92 | 0.92 | 0.94 | 0.99 | 0.96 | 0.95 | 0.97 | 1.00 | 0.96 | 0.94 | 0.99 |
| *NSR(*15) | 0.99 | 0.99 | 0.95 | 0.94 | 0.98 | 0.93 | 0.96 | 0.93 | 0.97 | 0.96 | 0.97 | 0.96 | 0.94 | 0.93 | 0.94 | 0.95 | 0.96 |
| *NSR(*16) | 0.96 | 0.97 | 0.96 | 0.97 | 0.94 | 0.97 | 0.98 | 0.99 | 0.98 | 0.98 | 0.96 | 0.95 | 0.99 | 0.95 | 0.92 | 0.98 | 0.97 |
| *NSR(*17) | 0.96 | 0.95 | 0.98 | 0.98 | 0.98 | 0.98 | 0.96 | 0.98 | 0.96 | 0.97 | 0.93 | 0.99 | 0.96 | 0.99 | 0.96 | 1.00 | 0.96 |
| *NSR(*18) | 0.97 | 0.98 | 0.98 | 0.99 | 0.99 | 0.96 | 0.97 | 0.94 | 0.96 | 0.93 | 0.98 | 0.97 | 0.94 | 0.91 | 0.98 | 0.99 | 0.96 |
| *NSR(*19) | 0.98 | 0.99 | 0.97 | 0.99 | 0.98 | 0.99 | 0.98 | 0.98 | 0.97 | 0.96 | 0.99 | 0.96 | 0.97 | 0.98 | 0.96 | 0.97 | 0.98 |
| *NSR(*20) | 0.98 | 0.97 | 0.97 | 0.96 | 0.98 | 0.98 | 0.98 | 0.97 | 0.99 | 0.98 | 0.99 | 1.00 | 0.95 | 0.96 | 0.98 | 0.97 | 0.98 |

*a. 200-week period January 2002- November 2006.*

| W.N. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *NSR(*1) | 0.29 | 0.31 | 0.31 | 0.29 | 0.30 | 0.30 | 0.30 | 0.31 | 0.29 | 0.29 | 0.29 | 0.30 | 0.29 | 0.29 | 0.27 | 0.27 | 0.26 |
| *NSR(*2) | 0.52 | 0.45 | 0.44 | 0.43 | 0.41 | 0.40 | 0.40 | 0.36 | 0.37 | 0.38 | 0.38 | 0.37 | 0.39 | 0.41 | 0.50 | 0.51 | 0.53 |
| *NSR(*3) | 0.68 | 0.69 | 0.95 | 0.97 | 0.90 | 0.95 | 0.95 | 0.90 | 0.89 | 0.87 | 0.86 | 0.83 | 0.83 | 0.80 | 0.71 | 0.72 | 0.70 |
| *NSR(*4) | 0.86 | 0.88 | 0.77 | 0.73 | 0.86 | 0.84 | 0.84 | 0.80 | 0.80 | 0.77 | 0.77 | 0.79 | 0.78 | 0.77 | 0.89 | 0.86 | 0.85 |
| *NSR(*5) | 0.92 | 0.96 | 0.87 | 0.90 | 0.81 | 0.80 | 0.80 | 0.85 | 0.84 | 0.88 | 0.90 | 0.84 | 0.83 | 0.83 | 0.92 | 0.94 | 0.94 |
| *NSR(*6) | 0.95 | 0.97 | 0.96 | 0.96 | 0.95 | 0.92 | 0.93 | 0.92 | 0.92 | 0.92 | 0.91 | 0.98 | 0.95 | 0.93 | 0.76 | 0.72 | 0.72 |
| *NSR(*7) | 0.96 | 0.90 | 0.88 | 0.85 | 0.83 | 0.88 | 0.88 | 0.96 | 0.95 | 0.92 | 0.91 | 0.88 | 0.92 | 0.93 | 0.93 | 0.98 | 0.99 |
| *NSR(*8) | 0.97 | 0.97 | 0.98 | 0.98 | 0.97 | 0.96 | 0.94 | 0.95 | 0.96 | 0.94 | 0.93 | 0.93 | 0.94 | 0.95 | 0.93 | 0.93 | 0.93 |
| *NSR(*9) | 0.92 | 0.94 | 0.93 | 0.93 | 0.98 | 0.95 | 0.99 | 0.90 | 0.89 | 0.93 | 0.94 | 0.94 | 0.93 | 0.92 | 0.92 | 0.91 | 0.92 |
| *NSR(*10) | 0.96 | 0.97 | 0.97 | 0.97 | 0.91 | 0.94 | 0.90 | 0.99 | 0.99 | 0.98 | 0.98 | 0.98 | 0.97 | 0.93 | 0.98 | 0.98 | 0.97 |
| *NSR(*11) | 0.94 | 0.96 | 0.97 | 0.97 | 0.98 | 0.97 | 0.99 | 0.92 | 0.92 | 0.94 | 0.94 | 0.94 | 0.94 | 0.95 | 0.96 | 0.98 | 0.96 |
| *NSR(*12) | 0.96 | 0.96 | 0.93 | 0.98 | 0.92 | 0.93 | 0.91 | 0.97 | 0.98 | 0.97 | 0.97 | 0.96 | 0.97 | 0.98 | 0.99 | 0.92 | 0.95 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *NSR(*13) | 0.89 | 0.95 | 0.97 | 0.93 | 0.97 | 0.98 | 0.99 | 0.92 | 0.93 | 0.94 | 0.92 | 0.94 | 0.95 | 0.95 | 0.94 | 0.98 | 0.97 |
| *NSR(*14) | 0.97 | 0.98 | 0.95 | 0.95 | 0.99 | 0.98 | 0.97 | 0.97 | 0.97 | 0.96 | 0.98 | 0.96 | 0.94 | 0.96 | 0.99 | 0.93 | 0.90 |
| *NSR(*15) | 0.98 | 0.96 | 0.96 | 0.99 | 0.94 | 0.95 | 0.96 | 0.95 | 0.92 | 0.96 | 0.95 | 0.97 | 0.98 | 0.97 | 0.90 | 0.96 | 0.97 |
| *NSR(*16) | 0.95 | 0.95 | 0.96 | 0.95 | 0.97 | 0.99 | 0.98 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.96 | 1.00 | 0.98 | 0.97 | 0.99 |
| *NSR(*17) | 0.97 | 0.97 | 0.97 | 0.98 | 0.96 | 0.96 | 0.98 | 0.97 | 1.00 | 0.96 | 0.95 | 0.99 | 0.99 | 0.94 | 0.96 | 0.96 | 0.95 |
| *NSR(*18) | 0.98 | 0.95 | 0.95 | 0.94 | 0.99 | 0.97 | 0.97 | 1.00 | 0.98 | 0.99 | 1.00 | 0.94 | 0.93 | 0.97 | 0.98 | 0.99 | 0.97 |
| *NSR(*19) | 0.95 | 0.97 | 0.99 | 0.98 | 0.96 | 0.96 | 0.95 | 0.98 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.97 | 0.98 | 0.99 |
| *NSR(*20) | 0.97 | 0.99 | 0.95 | 0.98 | 0.97 | 0.98 | 0.99 | 0.99 | 0.98 | 0.97 | 0.98 | 0.98 | 0.98 | 0.96 | 0.99 | 0.97 | 0.99 |

*b. 500-week period January 2002- December 2006.*

*Table 4.6 - Noise-to-signal ratio of the Russell 1000 for two time windows.*

Perhaps a still more intuitive representation is given by Figure 4.3 which graphically illustrates how the NSR changes in function of the number of factors in different time windows in the cases of 200- and 500-week-long time windows, and by Figure 4.4 which shows the plot of the magnitude of the first 100 eigenvalues for each time window in the two cases of 200- and 500-week-long time window. The plot is smooth without any sudden jump and becomes almost a straight line after 10-12 principal components. In other words: The eigenvalues of the covariance matrix of the Russell 1000 universe for time windows of 200 and 500 weeks show a smooth pattern.

The NSR of principal components of the same universe and in the same time windows also exhibit a smooth behaviour without ever reaching levels below 20%.

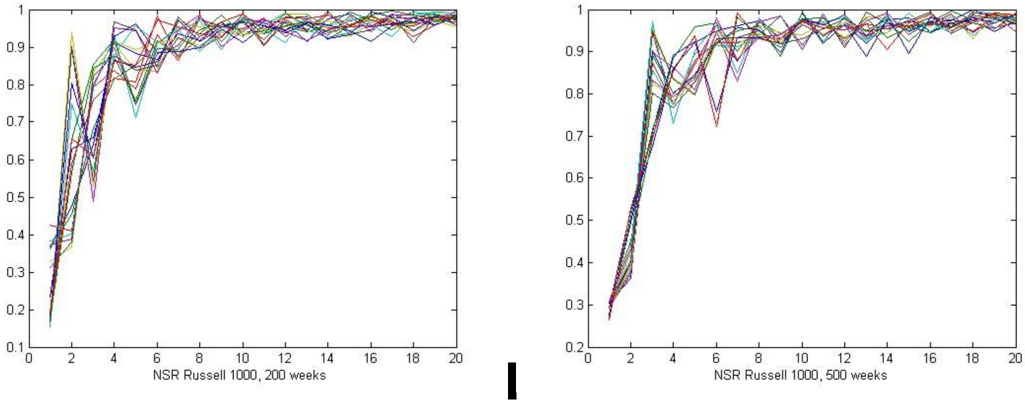


*Fig 4.3. Plot of the NSR of the Russell 1000 in function of the number of factors. The plot is made superposing the plots of the NSR in different time windows. The figure on the left is relative to a 200-week time window, the figure on the right is relative to a 500-week time window.*

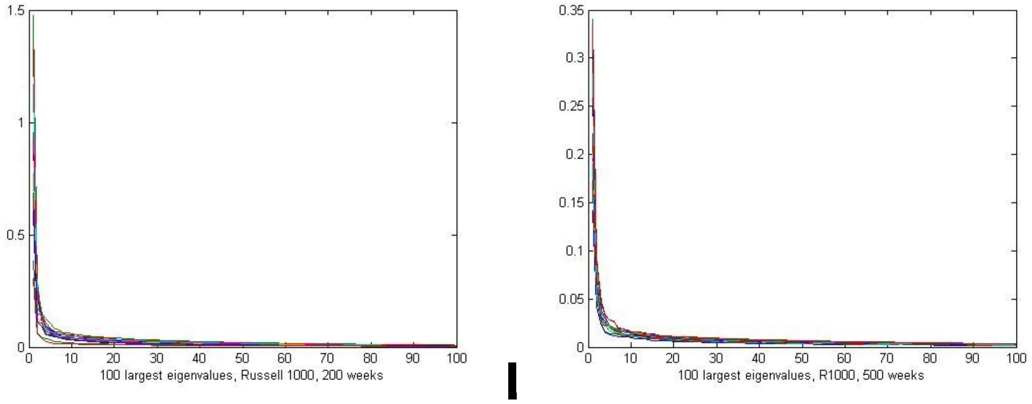


*Figure 4.4. Plot of the 100 largest eigenvalues of the covariance matrix of the Russell 1000. The figure on the left is relative to a 200-week time window, the figure on the right to a 500-week time window. The plot is made superposing the plots of the 100 largest eigenvalues in different time windows.*

I compared these empirical data with equivalent data computed on our simulated markets for time windows and number of returns 200/800 and 500/400, respectively. These numbers of time points and returns series are close to the corresponding numbers for the Russell 1000. Results are shown in the six panels of Table 4.7 which are constructed in a way analogous to those of Table 4.6.

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.90 | 0.84 | 0.73 | 0.91 | 0.77 | 0.83 | 0.90 | 0.91 | 0.89 | 0.97 | 0.97 | 0.89 | 0.97 | 0.93 | 0.81 | 0.94 | 0.87 |
| 0.93 | 0.94 | 0.98 | 0.87 | 0.97 | 0.95 | 0.86 | 0.91 | 0.84 | 0.90 | 0.89 | 0.90 | 0.89 | 0.88 | 0.83 | 0.91 | 0.96 |
| 0.98 | 0.95 | 0.84 | 0.88 | 0.90 | 0.95 | 0.81 | 0.96 | 0.85 | 0.90 | 0.91 | 0.97 | 0.95 | 0.92 | 0.96 | 0.79 | 0.78 |
| 0.90 | 0.89 | 0.92 | 0.92 | 0.77 | 0.85 | 0.91 | 0.85 | 0.92 | 0.93 | 0.88 | 0.90 | 0.75 | 0.80 | 0.72 | 0.96 | 0.86 |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.97 | 0.94 | 1.00 | 0.96 | 0.95 | 0.97 | 0.97 | 0.98 | 1.00 | 0.95 | 0.97 | 0.98 | 0.97 | 0.98 | 0.96 | 0.97 | 0.98 |
| 0.99 | 0.99 | 0.99 | 0.95 | 1.00 | 0.99 | 0.98 | 0.99 | 0.98 | 1.00 | 0.99 | 0.96 | 0.97 | 0.99 | 0.99 | 0.99 | 0.98 |
| 0.99 | 0.99 | 0.97 | 0.99 | 0.97 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.98 | 0.98 | 1.00 | 0.99 | 0.99 | 0.99 |
| 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 |
| 0.99 | 1.00 | 0.99 | 0.98 | 0.99 | 0.97 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.97 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 |
| 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.97 | 0.99 |
| 0.98 | 0.98 | 0.99 | 0.98 | 0.99 | 1.00 | 0.99 | 0.98 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.97 | 0.99 | 0.99 | 0.99 |
| 1.00 | 0.97 | 0.98 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.97 | 0.98 | 0.99 | 0.99 | 0.99 | 0.98 | 1.00 | 0.98 |
| 0.99 | 0.99 | 1.00 | 0.99 | 0.98 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.99 | 1.00 | 0.99 | 0.99 | 1.00 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 | 0.99 |
| 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.97 |
| 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.98 | 0.99 | 1.00 | 1.00 | 0.99 |
| 0.98 | 0.99 | 1.00 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.98 | 0.98 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 |
| 0.99 | 1.00 | 1.00 | 0.99 | 0.99 | 1.00 | 0.98 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 0.99 |
| 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.97 | 0.99 | 0.99 | 0.97 | 0.99 | 0.99 |

*a. Time windows 200 weeks 800 returns noise std = 0.1, NSR almost zero.*

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.97 | 0.88 | 0.86 | 0.95 | 0.78 | 0.92 | 0.97 | 0.98 | 0.87 | 0.86 | 0.98 | 0.92 | 0.87 | 0.83 | 0.94 | 0.90 | 0.92 |
| 0.93 | 0.83 | 0.82 | 0.90 | 0.94 | 0.83 | 0.92 | 0.97 | 0.94 | 0.93 | 0.75 | 0.93 | 0.88 | 0.85 | 0.88 | 0.94 | 0.78 |
| 0.98 | 0.93 | 0.87 | 0.86 | 0.89 | 0.94 | 0.87 | 0.89 | 0.96 | 0.95 | 0.89 | 0.92 | 0.90 | 0.95 | 0.91 | 0.95 | 0.96 |
| 0.88 | 0.93 | 0.88 | 0.94 | 0.86 | 0.94 | 0.87 | 0.88 | 0.82 | 0.90 | 0.98 | 0.95 | 0.93 | 0.81 | 0.87 | 0.94 | 0.84 |
| 0.01 | 0.01 | 0.02 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 |
| 0.97 | 0.99 | 0.98 | 0.99 | 0.99 | 0.96 | 0.98 | 0.99 | 0.98 | 0.98 | 0.97 | 0.98 | 0.98 | 0.96 | 0.99 | 0.98 | 0.98 |
| 0.98 | 0.99 | 0.97 | 1.00 | 0.97 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.97 | 0.99 | 0.98 | 0.98 | 0.99 | 0.98 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.00 | 0.99 | 0.98 | 0.97 | 0.98 | 0.99 | 0.99 | 0.99 | 0.98 | 0.97 | 0.99 | 0.99 | 0.97 | 0.97 | 0.98 | 0.98 | 0.98 |
| 0.98 | 0.98 | 0.98 | 0.99 | 0.98 | 0.98 | 0.97 | 0.98 | 0.98 | 1.00 | 0.98 | 0.99 | 0.99 | 0.98 | 0.97 | 1.00 | 0.99 |
| 0.98 | 0.98 | 0.99 | 0.97 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 1.00 | 0.98 | 0.98 | 0.99 | 0.99 | 0.98 | 0.98 |
| 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.98 | 1.00 | 0.99 | 0.99 | 0.97 | 0.99 | 0.98 | 0.99 | 1.00 | 0.98 | 1.00 |
| 0.99 | 0.98 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.98 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 1.00 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.98 | 0.99 | 1.00 | 0.98 |
| 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.97 | 0.98 | 1.00 | 0.99 | 1.00 | 0.99 |
| 0.98 | 0.99 | 0.98 | 0.99 | 1.00 | 0.98 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 |
| 0.99 | 1.00 | 1.00 | 0.98 | 1.00 | 1.00 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.98 | 0.98 | 0.99 | 0.98 |
| 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.98 | 0.99 |
| 0.99 | 0.98 | 0.99 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |

*b. Time windows 200 weeks 800 returns noise std = 1, NSR approximately 0.01.*

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.96 | 0.78 | 0.89 | 0.92 | 0.90 | 0.91 | 0.96 | 0.94 | 0.94 | 0.96 | 0.93 | 0.93 | 0.93 | 0.89 | 0.94 | 0.96 | 0.91 |
| 0.86 | 0.96 | 0.96 | 0.97 | 0.95 | 0.89 | 0.81 | 0.91 | 0.81 | 0.86 | 0.83 | 0.97 | 0.90 | 0.90 | 0.88 | 0.82 | 0.91 |
| 0.92 | 0.99 | 0.93 | 0.95 | 0.79 | 0.92 | 0.94 | 0.90 | 0.96 | 0.93 | 0.88 | 0.87 | 0.94 | 0.91 | 0.83 | 0.88 | 0.89 |
| 0.99 | 0.73 | 0.92 | 0.94 | 0.86 | 0.95 | 0.98 | 0.88 | 0.82 | 0.92 | 0.93 | 0.89 | 0.85 | 0.90 | 0.88 | 0.89 | 0.86 |
| 0.29 | 0.33 | 0.28 | 0.27 | 0.33 | 0.28 | 0.29 | 0.29 | 0.30 | 0.31 | 0.30 | 0.29 | 0.29 | 0.29 | 0.31 | 0.31 | 0.31 |
| 0.95 | 0.98 | 0.98 | 0.98 | 0.98 | 0.96 | 0.96 | 0.99 | 0.99 | 0.95 | 0.97 | 0.98 | 0.98 | 0.97 | 0.99 | 0.97 | 0.98 |
| 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.98 | 0.98 | 1.00 | 0.98 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 | 0.98 |
| 0.98 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.98 | 1.00 | 0.97 | 0.99 | 0.99 | 0.95 | 0.98 | 0.98 | 0.98 | 0.98 |
| 0.99 | 0.98 | 1.00 | 0.99 | 0.99 | 0.99 | 0.98 | 1.00 | 0.98 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 |
| 0.99 | 0.99 | 0.98 | 0.98 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.97 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.99 | 0.98 | 1.00 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 1.00 | 0.99 | 0.98 | 0.98 | 0.99 | 0.98 | 0.98 |
| 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 |
| 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.97 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 |
| 0.99 | 1.00 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 |
| 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.97 | 0.99 | 0.98 | 0.98 | 1.00 | 0.98 | 0.99 |
| 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.98 | 0.99 | 0.98 | 1.00 | 1.00 | 1.00 | 0.99 | 0.99 | 0.98 | 1.00 | 1.00 |
| 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 | 0.98 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.99 | 1.00 | 1.00 | 0.98 | 0.99 | 1.00 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 |
| 0.99 | 0.98 | 0.98 | 1.00 | 0.99 | 0.98 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 |

*c. Time windows 200 weeks 800 returns noise std = 5, NSR approximately 0.3.*

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.84 | 0.83 | 0.95 | 0.91 | 0.87 | 0.89 | 0.90 | 0.93 | 0.93 | 0.84 | 0.95 | 0.90 | 0.94 | 0.78 | 0.90 | 0.88 | 0.84 |
| 0.91 | 0.91 | 0.88 | 0.88 | 0.87 | 0.89 | 0.91 | 0.91 | 0.93 | 0.87 | 0.81 | 0.88 | 0.93 | 0.93 | 0.95 | 0.84 | 0.95 |
| 0.95 | 0.93 | 0.91 | 0.92 | 0.86 | 0.92 | 0.89 | 0.88 | 0.82 | 0.95 | 0.89 | 0.90 | 0.94 | 0.85 | 0.88 | 0.94 | 0.95 |
| 0.87 | 0.87 | 0.80 | 0.80 | 0.94 | 0.90 | 0.99 | 0.84 | 0.90 | 0.85 | 0.96 | 0.94 | 0.84 | 0.95 | 0.86 | 0.89 | 0.87 |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.97 | 0.99 | 0.99 | 0.97 | 0.98 | 0.98 | 0.98 | 0.99 | 0.98 | 0.97 | 0.96 | 1.00 | 0.96 | 0.98 | 0.97 | 0.98 | 0.99 |
| 0.99 | 0.99 | 0.98 | 0.96 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 1.00 | 0.99 | 0.98 | 0.98 | 0.97 | 1.00 | 0.99 |
| 0.99 | 0.99 | 0.99 | 0.99 | 0.97 | 0.98 | 0.98 | 0.97 | 0.97 | 0.97 | 0.99 | 0.98 | 0.99 | 0.98 | 0.98 | 0.98 | 0.97 |
| 0.98 | 0.98 | 1.00 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 1.00 | 0.98 | 0.99 |
| 1.00 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.97 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 | 1.00 |
| 0.99 | 0.98 | 0.98 | 0.99 | 1.00 | 0.99 | 1.00 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.98 | 1.00 | 0.99 | 0.97 |
| 0.99 | 1.00 | 0.99 | 0.99 | 0.98 | 0.98 | 1.00 | 0.96 | 0.98 | 0.99 | 0.99 | 0.98 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 |
| 0.99 | 1.00 | 1.00 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.98 | 0.98 | 0.99 | 1.00 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 | 0.98 | 0.98 | 0.99 | 0.99 |
| 0.99 | 0.99 | 1.00 | 0.98 | 1.00 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 |
| 0.99 | 0.99 | 0.98 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 |
| 0.98 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 1.00 | 0.99 | 0.98 |
| 0.99 | 0.98 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.98 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 | 1.00 | 0.98 | 1.00 | 1.00 | 0.99 | 0.99 |
| 1.00 | 0.98 | 0.98 | 0.97 | 0.98 | 0.98 | 1.00 | 0.99 | 0.99 | 1.00 | 1.00 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 |

*d. Time windows 500 weeks 400 returns noise std = 0.1, NSR almost zero.*

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.86 | 0.95 | 0.90 | 0.83 | 0.96 | 0.87 | 0.84 | 0.92 | 0.94 | 0.84 | 0.93 | 0.93 | 0.94 | 0.89 | 0.97 | 0.98 | 0.97 |
| 0.88 | 0.92 | 0.86 | 0.90 | 0.91 | 0.94 | 1.00 | 0.90 | 0.99 | 0.95 | 0.84 | 0.95 | 0.92 | 0.89 | 0.89 | 0.90 | 0.89 |
| 0.78 | 0.84 | 0.95 | 0.94 | 0.89 | 0.95 | 0.92 | 0.78 | 0.84 | 0.87 | 0.95 | 0.85 | 0.89 | 0.91 | 0.91 | 0.95 | 0.93 |
| 0.91 | 0.89 | 0.83 | 0.91 | 0.86 | 0.97 | 0.87 | 0.92 | 0.91 | 0.85 | 0.90 | 0.87 | 0.94 | 0.92 | 0.85 | 0.83 | 0.82 |
| 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| 0.99 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.98 | 0.98 | 0.97 | 0.97 | 0.97 | 0.99 | 0.96 | 0.98 | 0.99 | 0.99 | 0.99 |
| 0.97 | 0.98 | 0.98 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 1.00 | 0.99 | 0.97 | 0.98 | 0.98 | 0.99 |
| 1.00 | 0.98 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.97 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.97 | 0.98 | 0.96 | 0.98 |
| 0.98 | 0.99 | 0.99 | 0.99 | 0.98 | 0.97 | 0.97 | 0.99 | 0.98 | 0.98 | 0.98 | 0.98 | 1.00 | 1.00 | 0.98 | 0.99 | 0.97 |
| 0.98 | 0.98 | 0.99 | 0.98 | 0.99 | 1.00 | 0.98 | 0.99 | 1.00 | 1.00 | 0.99 | 0.99 | 0.97 | 0.99 | 0.98 | 0.98 | 0.99 |
| 0.99 | 0.99 | 1.00 | 0.98 | 0.99 | 0.98 | 1.00 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 1.00 | 0.98 | 0.99 | 0.99 | 0.99 |
| 0.98 | 1.00 | 0.98 | 0.99 | 0.98 | 0.99 | 1.00 | 0.99 | 0.97 | 0.98 | 0.99 | 0.98 | 0.98 | 0.99 | 1.00 | 1.00 | 0.99 |
| 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 |
| 0.99 | 1.00 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 1.00 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 |
| 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 |
| 0.99 | 0.99 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 |
| 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 1.00 | 0.98 | 0.98 | 0.98 | 0.98 | 0.99 | 0.99 |
| 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.98 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 | 1.00 | 0.99 | 0.99 | 1.00 | 0.99 |

*e. Time windows 500 weeks 400 returns noise std = 1, NSR approximately 0.01.*

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.94 | 0.93 | 0.91 | 0.94 | 0.92 | 0.82 | 0.92 | 0.73 | 0.93 | 0.99 | 0.86 | 0.91 | 0.92 | 0.93 | 0.93 | 0.92 | 0.87 |
| 0.84 | 0.90 | 0.96 | 0.86 | 0.96 | 0.98 | 0.94 | 0.98 | 0.97 | 0.87 | 0.94 | 0.89 | 0.90 | 0.96 | 0.82 | 0.94 | 0.98 |
| 0.90 | 0.95 | 0.91 | 0.92 | 0.85 | 0.96 | 0.91 | 0.86 | 0.95 | 0.97 | 0.88 | 0.94 | 0.92 | 0.91 | 0.92 | 0.91 | 0.95 |
| 0.92 | 0.98 | 0.96 | 0.93 | 0.97 | 0.83 | 0.89 | 0.85 | 0.92 | 0.88 | 0.90 | 0.95 | 0.90 | 0.93 | 0.90 | 0.82 | 0.93 |
| 0.26 | 0.22 | 0.24 | 0.23 | 0.23 | 0.24 | 0.23 | 0.27 | 0.23 | 0.23 | 0.24 | 0.22 | 0.24 | 0.21 | 0.25 | 0.25 | 0.24 |
| 0.96 | 1.00 | 0.96 | 0.99 | 0.98 | 0.97 | 0.97 | 0.96 | 0.98 | 0.99 | 0.97 | 0.97 | 0.98 | 0.98 | 0.99 | 1.00 | 0.96 |
| 0.99 | 0.97 | 0.99 | 0.99 | 0.97 | 0.99 | 0.98 | 0.99 | 0.98 | 0.97 | 0.99 | 0.98 | 0.98 | 0.99 | 0.98 | 0.98 | 0.99 |
| 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 0.98 | 0.99 | 0.99 | 1.00 | 0.98 | 0.97 |
| 0.98 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.97 | 0.99 | 0.98 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 1.00 | 0.98 |
| 1.00 | 0.99 | 0.99 | 0.97 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 |
| 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.97 | 0.98 | 0.97 | 0.99 | 0.98 | 1.00 | 1.00 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 |
| 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 0.99 | 0.99 | 0.98 | 0.99 |
| 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 1.00 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 |
| 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.98 |
| 1.00 | 0.99 | 0.98 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 |
| 0.98 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 |
| 1.00 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 |
| 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 |
| 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.98 | 1.00 | 0.99 | 0.98 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 |
| 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 0.98 | 0.98 | 0.99 | 1.00 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |

*f. Time windows 500 weeks 400 returns noise std = 5, NSR approximately 0.25.*

*Table 4.7. - Noise-to-signal ratio of a simulated market, for 200-and 500-week time windows, 800 and 400 returns respectively, various noise standard deviations, and various NSRs.*

The behaviour of the eigenvalues and of the NSR of our simulated market is different with respect to that of the Russell 1000. In the simulated market, the NSR is high for all numbers of factors except for the case where $k = 5$, the true number of factors, where it drops to near 0 when the standard deviation of the residual terms are 0.1 and grows to 0.25-0.3 when the standard deviation of the noise term reaches 5. In other words: The eigenvalues of the covariance matrix of our simulated market for time windows of 200 and 500 weeks and 800 and 400 stocks respectively show a sudden drop for $k = 5$ and lie approximately on a straight line elsewhere.

The NSR of principal components of the same universe and in the same time windows is close to 1 except for $k = 5$, where it assumes values that depend on the magnitude of the standard deviation of the noise term. When the standard deviation is equal to 5, the NSR is 0.2-0.3.

We can now reinterpret the simulation results of Table 4.7 in terms of the NSR. The results in Table 4.7 show that if we generate data with a simulated factor model that produces an NSR in the range of 0.3, the estimation of factors and components with PCA is subject to errors much larger than the errors predicted by the theory of approximate factor models. The magnitude of errors shows that the number of data points that fall outside the 95% confidence band is more than ten times the number predicted by the theory. Therefore, if we work with return series such as the Russell 1000 where the NSR assumes similar values in the range 0.2-0.3, the NSR is too large to allow to confidently use the conclusions of the theory of approximate factor models.

### *4. 6. Dynamic factor models of returns*

Thus far I have offered reasons to believe that approximate factor models cannot be correctly applied to returns, that there is no reason to believe that PCA of returns reveals the true factors, and that possibly different factor models can explain the same returns processes. The discussion focused on static factor models of returns as dynamic factor models with a finite number of lags can be cast in a static form. In this section we discuss specifically dynamic factor models. I focus the discussion on returns but the considerations I will make are valid for stationary processes in general.

Recall that Burda, Jurkiewicz, and Waclaw (2005) determine the distribution of eigenvalues when samples are correlated. Expressions are very complex and do not lend themselves easily to interpretations and generalizations. However, in the case of exponential autocorrelations, Burda, Jurkiewicz, and Waclaw are able to find closed-form expressions. This paper essentially establishes that the distribution of the bulk of eigenvalues spreads out.

In other words, given a sample of data, if we compute the empirical distribution of eigenvalues of the covariance matrix we find that the distribution is also influenced by correlations between samples, that is, by autocorrelations. A possible explanation for the slow decay of eigenvalues of the covariance matrix is that it is due to autocorrelations: autocorrelations blur the boundary between the bulk of eigenvalues and the largest eigenvalues.

Not all factors that drive financial returns have the same forecasting power. Stated differently, not all factors are dynamic. There are purely static factors that affect only same-time relationships. One of the important objectives of dynamic factor models is to separate, if possible, the purely dynamic from the purely static structure.

There are two forms of the theory of dynamic factor models of stationary processes. The first is described in Peña and Box (1987) and is relative to dynamic factor models with a finite number of observed variables. The second form is the dynamic version of approximate factor models. In Chapter 2, I surveyed the literature on these processes. In dynamic factor models with infinite *N,T,* static factors are determined with PCA and dynamic factors are disentangled with different procedures which ultimately rely on separating a finite number of diverging eigenvalues from an infinite number of bounded eigenvalues. Dynamic factor models with infinite *N,T* are based on some generalized approximate factor model of static factors and are therefore subject to the same considerations that were made in the previous sections. Therefore it is not possible that a generalized dynamic factor model applies when no generalized static approximate factor model applies.

The dynamic factor model in Peña and Box (1987) does not suffer from the above problem. Recall that this model was described in Section 2.8.

$$\begin{aligned}
&r_t = \beta f_t + \varepsilon_t \\
&\Phi(L) f_t = \Theta(L)\eta_t \\
&\Phi(L) = I - \Phi_1 L - \cdots - \Phi_p L^p \\
&\Theta(L) = I - \Theta_1 L - \cdots - \Theta_q L^q
\end{aligned}$$

where factors are stationary processes, $\varepsilon_t$ is white noise with a full covariance matrix but it is serially uncorrelated, $\eta_t$ has a full-rank covariance matrix and it is serially uncorrelated, and $\boldsymbol{\varepsilon}_t$ and $\eta_t$ are mutually uncorrelated at all lags. Consider the covariance matrices $\Gamma_r(k) = E(r_t r_{t-k}), k = 0,1,2,\ldots$ and $\Gamma_f(k) = E(f_t f_{t-k}), k = 0,1,2,\ldots$. The number of factors is the common rank $Q$ of the matrices $\Gamma_r(k) \geq 1$.

The model has a significant advantage: it allows to identify dynamic factors under the mild assumption that residuals are not autocorrelated. This is a reasonable assumption when applied to returns because returns are not individually autocorrelated and it is therefore reasonable to assume that autocorrelation (when it occurs) is due to common factors.

The dynamic factor model in Peña and Box (1987) is one way to assess the global autocorrelation structure of returns. The number of factors is equal to the common rank of the autocovariance matrices. The methodology is therefore based on estimating the rank of a large matrix. Otter and Jacobs (2006 a,b) propose a method based on information theory which ultimately requires to determine the rank of autocovariance matrices; the Otter-Jacobs method uses the same information as is used in Peña and Box method.

Bouchaud, Laloux, Miceli and Potters (2005) offer a different methodology based on random matrix theory applied to autocorrelation matrices. The method in Bouchaud Laloux, Miceli and Potters (2005) determines the distribution of the canonical correlation coefficients of the observed variables under the null hypothesis of zero autocorrelation. This methodology requires a number of observations much larger than the number of time series, otherwise the continuous part of the spectrum occupies the entire range of correlation. The reason is that canonical correlations depend on two sets of parameters, one for each variable, where each set includes as many parameters as there are variables.

### *4. 7. Dynamic factor models of prices*

The theory of dynamic factor models of prices is described in Escribano and Peña (1994) and in Peña and Poncela (2004 a,b). These papers builds on and generalizes Stock and Watson (1988). The latter paper established that if a multivariate process with $N$ variables integrated of order (1) exhibits $q$ cointegrating relationships, then each process can be represented as a linear combination of $N$-$q$ common integrated trends plus a possibly autocorrelated stationary process. The term trend indicates here a stochastic trend. Bossaerts (1988) observed that common trends are the most predictable portfolios and conjectured that common trends can be determined with canonical correlation analysis.

Escribano and Peña (1994) generalize this idea and establish the equivalence between common trends, common factors and cointegration. In particular, Escribano and Peña (1994) establish the following. First the paper introduces the general concept of factor models for integrated processes. Building on Peña and Box (1987), Escribano and Peña write a general factor model as follows:

$$X_t = Af_t + u_t$$
$$\phi(B)f_t = \vartheta(B)a_t$$

where $X_t$ is the *n*-vector of observed variables, $f_t$ is the *k*-vector of common factors, *A* is the *nxk* factor loading matrix and *u,a* are uncorrelated white noise with covariance matrices $\Sigma_u, \Sigma_a$ respectively, $\Sigma_a$ diagonal. Integrated and stationary factors can coexist in this model. The number of factors *k* can be less than the number of observed variables, $k<n$ , but it can also be equal to the number of observed variables *k=n*. In the latter case, the model is identifiable if $\Sigma_u = 0$, that is, *u* is a vector of constants. The condition *k=n* does not make the model trivial because integrated and stationary factors can coexist.

Recall that *n* integrated variables $X_t$ are said to be cointegrated of order 1 with rank *r* if there are *r* linearly independent linear combination $\beta' X_t$ which are stationary. Escribano and Peña (1994) established that the following three conditions are equivalent:

1. $X_t$ are cointegrated of order 1 with rank *r*.
2. $X_t$ are generated by *n-r* common trends.
3. $X_t$ have *n-r* common integrated factors and *r* common stationary factors.

If the $X_t$ are generated by *n-r* common trends, they can be written as:

$$X_t = \beta_\perp \tau_t + e_t$$
$$\tau_{tt} = \tau_{t-1} + v_t$$

where $\beta_\perp$ is the orthogonal complement of $\beta$ , *e* is a stationary process and *v* is white noise.

### *4.7.1 Empirical evidence of cointegration in equity prices*

Though there are positive indications, for example Kanas and Kouretas (2005), a definitive empirical test as to whether equity prices exhibit cointegration and can therefore be represented as factor models with integrated factors is very difficult to obtain. The key reason is the fact that markets are open systems where stocks are continuously created and destroyed. In order to perform a test of the hypothesis that markets can be represented with integrated factor models, one needs to form a sample with long time series. As shown in Section 4.5, this condition in practice limits the number of time series to those that have survived in the entire period. The percentage of survivors is not very close to 1 if periods of the orders of five to ten years are considered. This fact makes it impossible to run definitive tests of the assumption of cointegration among prices. In the literature several tests have been proposed. The result of these tests is that there is evidence of cointegration among stock prices. However, it is very difficult to establish the number of common trends or factors because of the abovementioned phenomenon of time series that exist only for a partial fraction of the time windows considered.

Stated differently, cointegration implies mean reversion. It is reasonable to assume that in the long run all prices are mean reverting to just one common dynamic factor. However, to prove this statement empirically one would need to consider very long time series, thus exposing the test to severe survivorship biases. It is also reasonable to assume that, over shorter periods of time, equity prices would probably exhibit multiple integrated trends. However, determining the number of factors active in any time window is a delicate estimation problem.

These considerations apply to every factor model. More in general, any model that requires long time series and that does not explicitly consider that markets are open systems is subject to the same problem: samples will be biased. Any test can only be partial and cannot be definitive. There are many ways to circumvent this problem. Perhaps the most obvious is to consider the quality of forecasts that can be obtained with different models. Forecasts can be made truly out-of-sample, in the sense that models can be estimated on moving windows and forecasts made on a small numbers of periods ahead. One can compare results obtained, making forecasts based on different procedures. This type of result is basically unbiased.

In this dissertation I do not perform a detailed comparative analysis of the performance of different types of factor models. However, I will now explore the following question: Suppose that asset prices are cointegrated and therefore can be described by a factor model with integrated factors. Suppose that a factor model of returns is estimated on the same market. Would price cointegration imply any particular bias when estimating factor models of returns?

Recall that empirical results from Section 4.5 seem to indicate that factor models of returns are not uniquely identifiable because the decay of the eigenvalues of the covariance matrix is too smooth. My objective is to explore whether this fact could be partially explained by the cointegrating relationships of prices.

### *4.7.2 Cointegration and correlation*

Cointegration and correlation are related but different phenomena. They are related insofar as both deal with the fact that two series stay close together. In fact, if prices are cointegrated we can say that prices move together, while if returns are correlated returns move together. Consider two time series of prices $x_t, y_t$. If prices are cointegrated one can regress *y* over *x* and obtain:

$$y_t = \alpha x_t + \nu_t$$

while if returns are correlated (that is, returns are jointly stationary) one can write:

$$\Delta y_t = \beta \Delta x_t + \eta_t$$

where both $\nu$ and $\eta$ are stationary processes possibly autocorrelated. It is immediate to see that, from the population point of view, the correlation coefficient between two cointegrated processes is not constrained to assume any specific value. For example, suppose that $y_t = x_t + \nu_t$ and compute the correlation coefficient between the relative returns:

$$\rho(\Delta x_t, \Delta y_t) = \frac{E[\Delta x_t(\Delta x_t + \Delta \nu_t)]}{\sqrt{E[\Delta x_t^2] E[(\Delta x_t + \Delta \nu_t)^2]}}$$

As $\Delta \nu_t$ is an arbitrary process, if $\Delta x_t = \Delta \nu_t$ the correlation coefficient assumes the value 1, if $\Delta x_t = -\Delta \nu_t$ the correlation coefficient assumes the value zero and if $\Delta x_t = -2\Delta \nu_t$ the correlation coefficient assumes the value -1 with all possible intermediate levels. That is, two processes can be cointegrated but the relative returns can be perfectly correlated, perfectly anticorrelated, uncorrelated or exhibit any intermediate correlation value.

The global framework of analysis is the following. Suppose *prices* are described by a factor model of the type:

$$X_t = Af_t + u_t$$
$$\phi(B) f_t = \vartheta(B) a_t$$

where *n-r* factors $f_t$ are I(1) and the *r* factors $u_t$ are I(0). Suppose a *factor model of returns* is estimated on the observed variables. How many factors would be detected?

In order to tackle this problem, let's consider a simplified situation with only one factor, that is, consider a model of the type: $X_t = f_t + u_t$. The correlations between returns can be written as follows:

$$E[\Delta x_{it} \Delta x_{jt}] = E[\Delta(f_t + u_{it})]\Delta(f_t + u_{jt}) =$$
$$= E\left[\Delta(f_t)^2\right] + E\left[\Delta(u_{it})\Delta(u_{jt})\right] + E[\Delta(u_{it})]\Delta(f_t) + E[\Delta(f_t)]\Delta(u_{jt})$$

Now, the number of factors in the factor model of prices is not influenced by the correlations and autocorrelations between the *u*.

Assuming that the *u* are all mutually uncorrelated and uncorrelated with the factor, the population would exhibit a single large eigenvalue and therefore one factor for returns would be detected. However, this implies the unrealistic assumption that all returns have exactly the same mutual correlation. If we assume that the *u* can be mutually correlated and autocorrelated then the covariance matrix of returns is not an identity matrix.

The distribution of eigenvalues of the covariance matrix of correlated samples of correlated variables was studied in Burda, Jurkiewicz, and Waclaw (2005) who found that the presence of autocorrelations has the effect of "spreading" the distribution of eigenvalues as if the *N/T* ratio were larger. Clearly no conclusion can be reached without knowing the distributions of eigenvalues of the covariance matrix of the residuals *u.* In the simplest case, the distribution of eigenvalues of the covariance matrix of the *u* will not add any factor and the factor analysis of returns will find one factor as the factor analysis of prices. However, in the most general case, the distribution of the residuals *u* will have a factor structure of its own and a factor analysis of returns will detect multiple factors.

The previous analysis was conducted in the case of one integrated factor. However, considering multiple factors complicates the algebra but reaches the same conclusions. If there are multiple integrated factors, and if the *u* have no factor structure, an eventual factor analysis of returns will detect only the common integrated factors. However if the residuals after the factor analysis of prices still have a factor structure, this factor structure will be superimposed to the factor structure generated by integrated factors.

Therefore, it is possible to conclude that if one runs a factor analysis of returns, or more fundamentally if one computes the eigenvalues of the covariance matrix of returns, one finds a number of factors that is the sum of two factor structures: one generated by cointegration with its common trends and one generated by the factor structure of the residuals of the common trends.

This analysis might explain why equity returns do not have a unique factor structure. Actually, the cointegration relationships add a factor structure of their own which results in a proliferation of factors and the consequent inability to estimate all factors given the size of available samples.

# 5. Summary and Future Direction of Research

This dissertation has discussed the problem of performing factor analysis of large universes of financial and economic variables, in particular financial returns. Presently, hundreds of financial and economic time series are available. These large universes cannot be correctly analyzed with classical factor models because maximum likelihood estimation is practically not feasible with large universes and because any reasonably small number of factors leaves residuals correlated.

The state-of-the-art theoretical approach is that of generalized approximate factor models infinite in both the number of observations and the number of time series. Generalized approximate factor models offer a theoretically impeccable solution to the problem of factor models and present nice features such as the ability to estimate factors with principal components.

The identification of factors and principal components has been conjectured since Hotelling (1932) proposed principal components analysis. Approximate factor models prove that in infinite markets, under appropriate assumptions, principal components and factors coincide. This result is mathematically important and is profoundly related to results in Random Matrix Theory. However, approximate factor models do not solve the problem of determining factors in finite sample, albeit large.

This dissertation proposes a novel framework for factor analysis in finite samples. First, it proposes criteria to determine upfront in finite samples when factors and principal components can be identified. These criteria require a large Signal-to-Noise Ratio under the additional condition that all factors/principal components are learnable given the size of the sample. Next, when the identification of factors and principal components is not feasible, this dissertation proposes to analyze factor models as communication channels and to choose those models that offer the best communication capacity between factors and observed variables.

The dissertation then makes the claim that, in practice, financial returns do not lend themselves to being analyzed in terms of principal components. The dissertation suggests that this might be due to cointegration effects among prices. It attempts to prove that if prices can be analyzed with dynamic factor models, which implies that prices are cointegrated, then the covariance matrix of returns exhibits a slow decay of the eigenvalues, a condition that implies the impossibility of identifying factors and principal components of returns.

There are several areas for future research. First, the empirical analysis of cointegration of prices. Cointegration of prices is not a simple phenomena. In fact, if we observe prices, we see that there are many different time horizons of mean reversion. It is arguable that, in the long run, all prices revert to a single common factor. Therefore, a first future direction for research is the empirical analysis of factor models of prices. To my knowledge this is an area largely unexplored. Anecdotal evidence claims that practical applications such as pair trading are based on cointegration. However, extensive academic studies of price cointegration are still missing.

A second area of research is tackling models of markets as open systems. Empirical studies of factor models and/or cointegration are fundamentally biased in that samples are formed by time series that exist in the entire time window. These biases are ineliminable unless models become models of open systems. Perhaps it is possible to borrow ideas from the statistical mechanics of open systems. For example, cointegrating prices could perhaps be considered as mechanical statistical systems formed by particles that flow in ducts with partially permeable walls.

The key consideration in applying these analogies is that simple models are needed in financial econometrics: Only relatively simple models can be effectively learned given the size of the present financial systems, which is many orders of magnitude smaller that the size of physical particle systems.

In summary, this dissertation attempts to prove that dynamic effects, in particular cointegration effects, are responsible for making factor models of returns "fuzzy" and indetermined. In practice, modeling cointegration is challenging. Future research should analyze the multiple time horizons of mean reversion, and therefore of cointegration, and ultimately propose a framework for analyzing markets as open systems.

# Notation

In this Dissertation I tried to maintain uniformity of notation as much as possible, given the large number of subjects. Follows some remarks on notation

Returns are denoted with the letter **r**. I denote factors of factor models with the letter **f** and the corresponding factor loadings with the Greek letter *β*. I denote with *R* the matrix that contains all the observations of returns, with **F** the matrix with the time series of factors and with **E** the matrix with the time series of residuals.

I tend to write general real or complex-valued matrixes with the letter *H*. The adjoint of a complex-valued matrix *H* is denoted *H******. If the matrix *H* is real-valued its transpose is denoted with *H***'.** If *H* is real, conjugate and the transpose coincide: *H*=H'***.** The ratio *N/T* between the number of series and the number of observations is called the aspect ratio of a matrix and is denoted with the letter *γ*. The lag operator L shifts a process by one period: $L(x_t) = x_{t-1}$.

Given a random variable $x$ a tilde denotes its estimation: $\tilde{x}$.

# References

Amengual, Dante and Mark W. Watson, 2006, “Consistent Estimation of The Number Of Dynamic Factors In A Large N And T Panel”, July 3, 2006

Anderson, T. W. (1963). “Asymptotic theory for principal component analysis”. Annals of Mathematical Statistic 34, 122–148.

Anderson, T. W. ,2003, *An introduction to multivariate statistical analysis*, Third ed. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.

D’Aspremont, Alexandre, Francis Bach, and Laurent El Ghaoui, 2007, “Optimal Solutions for Sparse Principal Component Analysis”, arXiv:0707.0705v4 [cs.AI] 9 Nov 2007

Back, Andrew D. and Andreas S. Weigend, 1997, A First Application of Independent Component Analysis to Extracting Structure from Stock Returns, International Journal of Neural Systems, Vol. 8, No.5 (October, 1997).

Bai, Jushan, 2003, “Inferential theory for factor models of large dimensions”, Econometrica, 71, 135–171, 2003.

Bai, Jushan and Serena Ng, 2002, “Determining the number of factors in approximate factor models.” Econometrica, 70, 191–221. (Errata, Determining the number of factors in approximate factor models, by Jushan Bai and Serena Ng.)

Bai. Z. D. and Jack W. Silverstein, 1998, “No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices”, Annals of Probability, 26, 316-345.

Bai. Z. D. and Jack W. Silverstein, 1999, “Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices”, Annals of Probability, Volume 27, Number 3 (1999), 1536-1555.

Bai, Z. D. and J. W. Silverstein, 2004, “CLT for linear spectral statistics of large dimensional sample covariance matrix.”, Annals of Probability, 32, 553-605.

Baik, Jinho, Gérard Ben Arous and Sandrine Péché, 2005, “Phase Transition Of The Largest Eigenvalue For Nonnull Complex Sample Covariance Matrices,” Annals of Probability, 2005, Vol. 33, No. 5, 1643–1697

Baik, Jinho and Jack W. Silverstein, 2006, Eigenvalues of large sample covariance matrices of spiked population models., Journal of Multivariate Analysis 97(6) (2006), pp. 1382-1408.

Biroli, Giulio, Jean-Philippe Bouchaud, Marc Potters, 2006, “On the top eigenvalue of heavy-tailed random matrices”, DSM/SPhT-T06/216 http://www-spht.cea.fr/articles/T06/216/

Bossaerts, Peter, 1988, “Common Nonstationary Components Of Asset Prices”, Journal of Economic Dynamics and Contrcl 12 (1988) 347-364. North-Holland

Box, G. and Tiao, G., 1977, “a canonical analysis of multiple time series”, Biometrika, 64, 355-65

Bouchaud, Jean-Philippe, Laurent Laloux, M. Augusta Miceli, and Marc Potters, 2005. “Large dimension forecasting models and random singular value spectra”, Science & Finance (CFM) working paper archive 500066, Science & Finance, Capital Fund Management.

Breitung, Jorg and Uta Kretschmer, 2005, “Dynamic factor models”, Deutsche Bundesbank Discussion Paper Series 1: Economic Studies, No 38/2005

Brillinger, David R., 1964, A frequency approach to the technique of principal components, factor analysis and canonical variates in the case of stationary time series, Royal Statistical Society Conference, October 1964

Brillinger, David R., 1981, Time Series: Data Analysis and Theory, SIAM, 1981

Brown, Stephen, 1989, "The number of factors in security returns", Journal of Finance, 64, 1247-1262.

Burda, Zdzislaw, Andrzej T. Görlich, A. Jarosz and Jerzy Jurkiewicz, 2004, "Signal and Noise in Financial Correlation Matrices", arXiv:cond-mat/0312496v2 [cond-mat.stat-mech] 3 Feb 2004

Burda, Zdzislaw, Andrzej T. Görlich, and Bartlomiej Waclaw, 2006, "Spectral properties of empirical covariance matrices for data with power-law tails", arXiv:physics/0603186v2 [physics.data-an] 20 Apr 2006

Burda, Zdzislaw, Jerzy Jurkiewicz, and Bartlomiej Waclaw, 2005, "Eigenvalue density of empirical covariance matrix for correlated samples", arXiv:cond-mat/0508451v1 [cond-mat.stat-mech] 19 Aug 2005

Camba-Mendez, Gonzalo and George Kapetanios, 2004, "Estimating The Rank Of The Spectral Density Matrix", European Central Bank, Working Paper Series No. 3 4 9 / April 2004

Carhart, Mark, 1997, "On Persistence in Mutual Fund Performance.", Journal of Finance 52:1, 57-82.

Chamberlain, Gary and Michael Rothschild, 1983, "Arbitrage, factor structure and mean-variance analysis in large asset markets", Econometrica 51, 1305–1324.

Chamberlain, Gary, 1983, "Funds, factors, and diversification in arbitrage pricing models", Econometrica, 51, 1281–1304.

Connor, Gregory and Robert Korajczyk, 1986, Performance measurement with the arbitrage pricing theory: a new framework for analysis, Journal of Financial Economics, 15, 373-394

Connor, Gregory and Robert Korajczyk, 1988, "Risk and Return in an Equilibrium APT: Application of a New Test Methodology", Journal of Financial Economics, 15, 373-394

Connor, Gregory and Robert Korajczyk, 1993, "A Test for the Number of Factors in an Approximate Factor Model", Journal of Finance 48 (September 1993): 1263-1291.

Cox, D.R. and H.D. Miller, *The Theory of Stochastic Processes*, Chapman & Hall/CRC , 1977

Dionísio, Andreia, Rui Menezes and Diana A. Mendes, 2005, "Uncertainty analysis in financial markets: can entropy be a solution?" August, 2005

Doz, Catherine, Domenico Giannone and Lucrezia Reichlin, 2006, "A Quasy Maximum Likelihood Approach For Large Approximate Dynamic Factor Models", ECB Working Paper Series No 674, September 2006

Edelman, Alan and N. Raj Rao, 2005, "Random Matrix Theory", Acta Numerica pp 1-65

El Karoui, Noureddine, 2003, "On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity", 2003 arXiv:math.ST/0309355.

El Karoui, Noureddine, 2006, “A Rate Of Convergence Result For The Largest Eigenvalue Of Complex White Wishart Matrices”, The Annals of Probability, 2006, Vol. 34, No. 6, 2077–2117

Escribano, A. and D. Peña, 1994. “Cointegration and common factors.”, Journal Time Series Analysis 15, 577–586.

Fabozzi, Frank J., Sergio M. Focardi, and Petter N. Kolm, 2006 *Financial Modeling of the Equity Market: from CAPM to Cointegration*, 2006, Wiley, Hoboken, New Jersey

Fabozzi, Frank J., Sergio Focardi, and Caroline Jonas, 2008. *Challenges in Quantitative Equity Management* (CFA Institute Research Foundation, Charlottesville, VA).

Fama, Eugene F. and Kenneth R. French, 1993, “Common risk factors in the returns on stocks and bonds.” Journal of Financial Economics 33, 3-56.

Forni, Mario Marc Hallin, Marco Lippi, and Lucrezia Reichlin, 2000, “The generalized dynamic factor model: Identification and estimation”, Review of Economics and Statistics, 82, 540–554.

Forni, Mario, Marc Hallin, Marco Lippi and Lucrezia Reichlin, 2004, “The generalized dynamic factor model consistency and rates”, Journal of Econometrics 119 (2004) 231 – 255

Forni, Mario, Marco Lippi, and Lucrezia Reichlin, 2005, “Opening the black box: Structural factor models versus structural VARs”, CEPR Discussion Paper 4133.

Forrester, P. J. (1993). “The spectrum edge of random matrix ensembles”. Nuclear Physics B 402 709–728. MR1236195

Forrester, Peter J. and Taro Nagao, “Eigenvalue statistics of the real Ginibre ensemble”, arXiv:0706.2020v1 [cond-mat.stat-mech] 14 Jun 2007

Galeano, Pedro and Daniel Peña, 2000, “Multivariate Analysis In vector Time Series”, Universidad Carlos III de Madrid, Working Paper 01-24, Statistics and Econometrics Series 15, March 2000

Galluccio, Stefano, Jean-Philippe Bouchaud, and Marc Potters, 1998, “Rational decisions, random matrices and spin glasses”, Physica A 259 (1998) 449{456

Geman, S. 1980, “A limit theorem for the norm of random matrices.” The Annals of Probability 8 252–261.

Geweke, J. (1977) The dynamic factor analysis of economic time series. In D.J. Aigner and A.S. Goldberger, Eds., Latent Variables in Socio-Economic Models, North Holland, Amsterdam.

Ginibre, J., 1965, “Statistical ensembles of complex, quaternion, and real matrices”, Journal Mathematical Physics, 19, 133

Girko, V. L., “Circular law,” Theory Prob. Appl., vol. 29, pp. 694–706, 1984.

Golub, Gene H. and Charles F. Van Loan 1996. *Matrix Computations*, Third Edition, The Johns Hopkins Studies in Mathematical Sciences, Baltimore, MD.

Gourieroux, Christian and Alain Monfort, 1995, *Statistics and Econometric Models: Volume 1 and 2*, Cambridge University Press, 1995

Gower, John C. and Garmt B. Dijksterhuis, 2004, *Procrustes Problems*, Oxford University Press, New York.

Guttman, “The Determinacy of Factor Score Matrices With Implications for Five Other Basic Problems of Common-Factor Theory.”, British Journal of Statistical Psychology, 1955,8, 65-81.

Harding, Matthew C., 2008a, “Explaining the Single Factor Bias of Arbitrage Pricing Models in Finite Samples”, Economics Letters, 99(1), 85-88, 2008

Harding, Matthew C., 2008b, “Structural estimation of high-dimensional factor models”, MIT, Department of Economics, Job Market Paper

Heaton, Chris and Victor Solo, 2003, “Asymptotic Principal Components Estimation of Large Factor Models”

Heaton, Chris and Victor Solo, 2006, “Estimation Of Approximate Factor Models: Is It Important To Have A Large Number Of Variables?” Presented at the North American Summer Meeting of the Econometric Society at the University of Minnesota in June 2006

Hendry, David F. 1995, *Dynamic Econometrics*. New York: Oxford University Press.

Hendry, David F. and Katarina Juselius, 2000, Explaining Cointegration Analysis: Part II, November 12, 2000

Herault J. and C. Jutten, 1986, “Space or time Adaptive Signal Processing by Neural Networks Model”, International Conference on Neural Networks for computing, Snowbird (Utah, USA), AIP conference proceedings n° 151, J. Denker (Ed.), pp. 206-211, 1986

Hotelling, Harold, 1933, “Analysis of a complex of statistical variables into principal components.” Journal of Educational Psychology, 24, 417-441.

Jegadeesh, N., Titman, S., 1993, “Returns to buying winners and selling losers: Implications for stock market efficiency.” The Journal of Finance 48, 65-91.

Jegadeesh, N., Titman, S., 2001, “Profitability of momentum strategies: an evaluation of alternative explanations.” The Journal of Finance 56, 699-720.

Johansen, Søren, 2000, “Modelling of cointegration in the vector autoregressive model”, Economic Modelling 17, 2000, 359-373.

Johansen, Søren and David Lando, 1996, “Multi-period models as cointegration models.” Discussion paper. University of Copenhagen.

Johansson, Kurt, 2000, “Shape fluctuations and random matrices.” Communications in Mathematical Physics, 209:437–476, 2000.

Jones, S., Christopher, 2001, “Extracting factors from heteroskedastic asset returns”, Journal of Financial Economics 62 (2001) 293–325

Johnstone, M. Iain, 2001, “On the distribution of the largest eigenvalue in principal components analysis”, The Annals of Statistics 2001, Vol. 29, No. 2, 295–327

Johnstone, M. Iain and Arthur Yu Lu, 2004, “Sparse Principal Components Analysis”, Technical Report, Stanford University, January 1, 2004

Johnstone, M. Iain, 2006, “High Dimensional Statistical Inference and Random Matrices, Proceedings International Congress of Mathematicians, arXiv:math/0611589v1 [math.ST] 19 Nov 2006

Jolliffe, Iain T., *Principal Components Analysis*, Springer; 2nd edition, October 1, 2002

Jolliffe, I. (2003). “A modified principal component technique based on the LASSO.” Journal of Computational and Graphical Statistics, 12, 531-547.

Kanas, Angelos and Georgios P Kouretas, 2005, “A cointegration approach to the lead-lag effect among sizesorted equity portfolios”, International Review of Economics and Finance, 2005, 14, 181-201

Kapetanios, George, 2004, “A new method for determining the number of factors in factor models with large datasets”, Queen Mary University of London, Working Paper No. 525, October 2004

Kolmogorov, A. N., *Foundations of the Theory of Probability*, By A.N. Kolmogorov, Chelsea Publishing Company, New Yori, 1956

Kwapie, J., S. Drożdz, A.Z. Górski and P. Oświęcimka, 2006, “Asymmetric Matrices In An Analysis Of Financial Correlations”, Acta Physica Polonica B , Vol. 37 (2006), No 11

Lütkepohl, H. L. (1991). *Introduction to multiple time series analysis*. Berlin, Springer.

Marčenko, V. A. and L. A. Pastur, “Distributions of eigenvalues for some sets of random matrices,” Math. USSR-Sbornik, vol. 1, pp. 457–483, 1967.

Mehta, M. L. *Random Matrices and the Statistical Theory of Energy Levels*. New York, Academic Press, 1967. 1991, Second Edition, Academic Press, Boston.

Merton, C. Robert, 1973, An Intertemporal Capital Asset Pricing Model , Econometrica, Vol. 41, No. 5. (Sep., 1973), pp. 867-887.

Moghaddam, B., Y. Weiss, and S. Avidan, 2006, “Generalized spectral bounds for sparse LDA.” In Proceedings ICML, 2006.

Natarajan, B. K., 1995, “Sparse approximate solutions to linear systems.”, SIAM Journal of Computing, 24(2): 227–234, 1995.

Nica, Alexandru and Roland Speicher, 2006, Lectures on the Combinatorics of Free Probability (London Mathematical Society Lecture Note Series) Cambridge University Press, 2006

Newey, Whitney K and Kenneth D. West, 1987. “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix”, Econometrica, 55, 703–708.

Onatski, Alexei, 2005, “Determining the number of factors from empirical distribution of eigenvalues”, Discussion Papers Columbia University Year 2005,

Onatski, Alexei, 2006a, “Determining the Number of Factors from Empirical Distribution of Eigenvalues”, Discussion Papers, Columbia University, May 23, 2006

Onatski, Alexei, 2006b, “A formal statistical test for the number of factors in the approximate factor models”, Economics Department, Columbia University, September 26, 2006

Onatski, Alexei, 2007, “Asymptotics of the Principal Components Estimator of Large Factor Models with Weak Factors and i.i.d. Gaussian Noise”, manuscript, Columbia University.

Onatski, Alexei, 2008, “Testing hypotheses about the number of factors in large factor models.”, January 25, 2008

Otter, W., Pieter and Jan P.A.M. Jacobs, 2006a, “Determining the number of factors and lag order in dynamic factor models: A minimum entropy approach” Working Paper, University of Groningen, April 2006

Otter, W., Pieter and Jan P.A.M. Jacobs, 2006b, On information in static and dynamic factor models, CCSO Centre for Economic Research University of Groningen, Working paper July 2006/05,

Péché, S., 2003, "Universality of local eigenvalue statistics for random sample covariance matrices." Ph.D. Thesis, Ecole Polytechnique Fédérale de Lausanne.

Peña Daniel and George E. P. Box, 1987, Identifying a Simplifying Structure in Time Series Journal of the American Statistical Association, Vol. 82, No. 399 (Sep., 1987), pp. 836- 843

Peña, Daniel and Pilar Poncela, 2004a, "Nonstationary dynamic factor analysis", Journal of Statistical Planning and Inference, 2004

Peña Daniel and Pilar Poncela, 2004b, "Forecasting with nonstationary dynamic factor models", Journal of Econometrics 119 (2004) 291 – 321

Pesaran, M. Hashem and Yongcheol Shin, 1997, "An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis", Working Paper, Trinity College, Cambridge, England January, 1997

Phillips, P. C. B. and S. Ouliaris, 1988, "Testing for Cointegration Using principal Components Methods." Journal of Economic Dynamics and Control 12, 1988, 205-230.

Plerou, Vasiliki , Parameswaran Gopikrishnan, Bernd Rosenow, Luıs A. Nunes Amaral, Thomas Guhr and H. Eugene Stanley, 2002, "Random matrix approach to cross correlations in financial data", Physical Review E, Volume 65, 066126

Priestley, M. B., *Spectral Analysis and Time Series*. 1983, Academic Press

Quah, Danny and Thomas J. Sargent, 1993. "A Dynamic Index Model for Large Cross Sections", CEP Discussion Papers 0132, Centre for Economic Performance, LSE.

Rosenberg, Barr, 1974. "Extra-market components of covariance in security returns", Journal of Financial and Quantitative Analysis 9, 263-274.

Ross, Stephen A., 1976. "The arbitrage theory of capital asset pricing", Journal of Economic Theory 13, 341-360.

Sargent, Thomas J., Christopher Sims, 1977, "Business Cycle Modelling Without Pretending to Have Too Much A Priori Economic Theory", Working Paper 55, Federal Reserve Bank of Minneapolis

Scherrer, Wolfgang and Christiaan Heij, 1998, "Estimation of factor models by realization-based and approximation methods", Econometric Institute Report 9831

Schneeweiss, Hans, 1997, "Factors and principal components in the near spherical case", Multivariate Behavioral Research, 32, 375-401.

Schneeweiss, Hans and Hans Mathes, 1995, "Factor analysis and principal components.", Journal of Multivariate Analysis, 55,105-124.

Schonemann, P. H. "The Minimum Average Correlation Between Equivalent Sets of Uncorrelated Factors.", Psychometrika, 197 1, 36.2 1-30.

Sengupta A.M. and P.P. Mitra, 1999, "Distributions of singular values for some random matrices", Physical Review E, Vol. 60, No. 3, September 1999.

Sentana, Enrique, “Factor representing portfolios in large asset markets”, Journal of Econometrics 119 (2004) 257 – 289

Shannon, Claude E. 1948a, “A mathematical theory of communication”. Bell System Technical Journal 27 (July), 379-423

Shannon, Claude E., 1948b, “A mathematical theory of communication”. Bell System Technical Journal 27 (October), 623-656.

Silverstein, Jack W., 1985, “The smallest eigenvalue of a large dimensional Wishart matrix.”, Annals of Probability, 13(4):1364{1368, 1985.

Silverstein, Jack W., 1986, “Eigenvalues and Eigenvectors Of Large Dimensional Sample Covariance Matrices”, Contemporary Mathematics, Volume 50, 1986

Silverstein, Jack W., 1995, “Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices.”, Journal of Multivariate Analysis 55(2) (1995), pp. 331-339.

Silverstein, Jack W. and Z.D. Bai, “On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices”, Journal of Multivariate Analysis 54(2) (1995), pp. 175-192.

Silverstein, J. W. and S. Choi, 1995, “Analysis of the limiting spectral distribution of large dimensional random matrices.”, Journal of Multivariate Analysis 54, 295-309.

Silverstein, Jack W. and Debashis Paul, 2008, “No Eigenvalues Outside the Support of the Limiting Empirical Spectral Distribution of a Separable Covariance Matrix.”

Silverstein, Jack W. and Antonia M. Tulino, 2006, “Theory of Large Dimensional Random Matrices for Engineers”

Soshnikov, Alexander, 2002 “A note on universality of the distribution of the largest eigenvalues in certain classes of sample covariance matrices.”, Journal of Statistical Physics, 108:1033–1056, 2002.

Soshnikov, Alexander, 2006, “Poisson statistics for the largest eigenvalues in random matrix ensembles.”, In Mathematical physics of quantum mechanics, volume 690 of Lecture Notes in Phys., pages 351–364. Springer, Berlin, 2006.

Soshnikov, Alexander and Yan V. Fyodorov, 2005, “On the largest singular values of random matrices with independent Cauchy entries.”, Journal of Mathematical Physics, 46(3): 033302, 15, 2005.

Spearman, C. (1904), “General intelligence, objectively determined and measured.”, American Journal of Psychology, 15, 201-293.

Speicher, Roland, “Free Probability Theory And Random Matrices”

Steiger, H., James , 1979, “Factor Indeterminacy In The 1930’s And The 1970’s Some Interesting Parallels”, Psychometrika, Vol 44. , No. 1. June, 1979

Steiger, James H., 1996, “Coming Full Circle in the History of Factor Indeterminacy”, Multivariate Behavioral Research, 31 (4), 617-630

Steiger, James H. and Peter H. Schonemann, 1978, “A History of Factor Indeterminacy”, in S. Shye Editor, Theory construction and data analysis in the behavioural sciences, 1978, Jossey-Bass, San Francisco.

Stock, James H. and Mark W. Watson, 1988, “Testing for common trends”, Journal of the American Statistical Society, Vol. 83 No. 404, December 1988

Stock, James H. and Mark W. Watson, 1998, “Diffusion Indexes”, NBER Working Paper 6702, august 1998

Stock, James H. and Mark W. Watson, 2002a, “Forecasting using principal components from a large number of predictors”, Journal of the American Statistical Association, 97, 1167-79

Stock, James H. and Mark W. Watson, 2002b, “Macroeconomic forecasting using diffusion indexes”, Journal of Business and Economics Statistics, 20:147–162, 2002b.

Stock, James, H. and Mark W. Watson, 2004, “Forecasting With Many Predictors”, August 2004, Handbook of Economic Forecasting

Stock, James, H. and Mark W. Watson, 2005, “Implications Of Dynamic Factor Models For VAR Analysis”, June 2005

Stock, James, H. and Mark W. Watson, 2007, “Forecasting In Dynamic Factor Models Subject To Structural Instability”, August 2007

Thurstone, L. L., 1938, Primary mental abilities, Chicago, University of Chicago Press.

Thurstone, L. L., 1947, *Multiple-Factor Analysis*. Chicago, University of Chicago Press.

Tibshirani, R., “Regression shrinkage and selection via the LASSO.” Journal of the Royal Statistical Society, series B, 58(1):267–288, 1996.

Tracy, C.A., Widom, H. (1996) “On orthogonal and symplectic matrix ensambles” Communications Mathematical Physics 177, 727-754.

Tracy, Craig A. and Harold Widom, 2008, “The Distributions of Random Matrix Theory and their Applications.”

Tulino, Antonia M. and Sergio Verdù, 2004, *Random Matrix Theory and Wireless Communications*., Now Publishers Inc.

Vapnik, Vladimir N., 1998, *Statistical Learning Theory*, 1998, Wiley-Interscience.

Voiculescu, Dan-Virgil, “Addition of certain non-commuting random variables,” Journal Functional Analysis, vol. 66, pp. 323–346, 1986.

Voiculescu, Dan-Virgil,“Multiplication of certain non-commuting random variables,” Journal Operator Theory, vol. 18, pp. 223–235, 1987.

von Neumann, John, 1996 *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, 1996

Wachter, K. W., 1978, “The strong limits of random matrix spectra for sample matrices of independent elements.”, Annals of Probability 6, 1, 1–18.

Watson, Mark W., 2000, “Macroeconomic Forecasting Using Many Predictors”, July 2000

Weenink, David,2003, “Canonical Correlation Analysis”, Institute of Phonetic Sciences, University of Amsterdam, Proceedings 25 (2003), 81–99.

Wigner, E. , “Characteristic vectors of bordered matrices with infinite dimensions,” Annals of Mathematics, vol. 62, pp. 546–564, 1955.

Wigner, E. , “Results and theory of resonance absorption,” in Conference on Neutron Physics by Time-of-Flight, Nov. 1-2 1956. Oak Ridge National Lab. Report ORNL-2309.

Wigner, E. , “On the distribution of roots of certain symmetric matrices,” Annals of Mathematics, vol. 67, pp. 325–327, 1958.

Wigner, E. , “Statistical properties of real symmetric matrices with many dimensions,” Proc. 4th Canadian Math. Congress, pp. 174–176, 1959.

Wigner, E. , “Distribution laws for the roots of a random Hermitian matrix,” in Statistical Theories of Spectra: Fluctuations, (C. E. Porter, ed.), New York: Academic, 1965.

Wigner, “Random matrices in physics,” SIAM Review, vol. 9, pp. 1–123, 1967.

Wishart, J., “The generalized product moment distribution in samples from a normal multivariate population,” Biometrika, vol. 20 A, pp. 32–52, 1928.

Yin, Y.Q., Z.D. Bai, and P.R. Krishnaiah, 1988, “On the Limit of the Largest Eigenvalue of the Large Dimensional Sample Covariance Matrix”, Probability Theory and Related Fields 78, 509-521 (1988)

Zhang, Z., H. Zha, and H. Simon, 2002, “Low rank approximations with sparse factors I: basic algorithms and error analysis.” SIAM journal on matrix analysis and its applications, 23 (3):706–727, 2002.

Zou, Hui, Trevor Hastie, Robert Tibshirani, 2004, “Sparse Principal Component Analysis”, April 26, 2004

Zuur, A., F., I.D. Tuck, and N. Bailey, 2003, “Dynamic factor analysis to estimate common trends in fisheries time series”, Can. J. Fish. Aquat. Sci. 60: 542–552 (2003)

## STATUTORY DECLARATION

Herewith I affirm that this dissertation was written independently without any help of third parties, that any other aids were completely and precisely stated, and that all content, whether taken altered or unaltered from my own publications or the works of others, was cited respecting current academic rules.

Karlsruhe, 26 June, 2009