Review of Linear Factor Models

Surprising Common Principles, the Systematic-plus-Idiosyncratic Myth, and the Misread Relationship with Financial Theory

Attilio Meucci[1]
attilio_meucci@symmys.com

this draft (please do not circulate): July 30 2010
latest version available at http://ssrn.com/abstract=1635495

Abstract

This is work in progress. Please check back soon for the final version

JEL Classification: C1, G11

Keywords: generalized r-square, dimension reduction fundamental factor models, macroeconomic factor models, factor analysis, regression, random matrix theory, GICS industry classification, cross-sectional models, time-series models, statistical models

[1]The author is very grateful to Eva Chan, Yingjin Gan, Sridhar Gollamudi, and Muting Ren

Contents

1 Introduction

Linear factor models represent a powerful dimension reduction technique to estimate the distribution of a large dimensional market. A few approaches to linear factor modeling, such as "time series", "cross-section", "statistical", etc., have become standard in the industry.

In this article we present the general framework to factor model construction from which all the standard approaches can be derived as special cases by suitably imposing restrictions and making assumptions. This framework is the constrained maximization of the generalized multivariate r-square. The purpose of the general framework is threefold.

First, the general framework provides guidance to learn and implement any factor model. Indeed, without knowledge of how a given approach fits in the bigger picture, it is easy to become overwhelmed by the intricacies and nuances of that specific approach. Therefore, the general framework is a powerful didactical tool that allows us to review systematically and in depth all the approaches to factor modeling.

Second, the general framework helps detecting common issues at a glance. For instance, as we shall see, using the general framework we prove mathematically that it is *impossible* to separate the randomness in the market into a systematic component and an idiosyncratic component. As a result, *none* of the standard approaches to factor modeling is of systematic-plus-idiosyncratic type and thus the standard risk computations provided by most vendors, which are based on the systematic-plus-idiosyncratic assumption, are incorrect.

Third, only by understanding the model in full generality and realizing why certain restrictions and assumptions are imposed can we take advantage of all the flexibility at our disposal to tailor and enhance any factor model.

This review is organized as follows. In Section 2 we introduce the general framework for factor model construction based on a dominant-plus-residual decomposition of the market. Then we present a summary of all the models commonly used in the industry and how they fit into the general framework.

In the subsequent Sections 3-6 we start from the general framework to derive and discuss in depth all the models commonly used in the industry. For each model we provide real-life case studies, we analyze the key properties, we delve into the consequences of incorrectly assuming a systematic-plus-idiosyncratic decomposition, and we discuss potential flexible generalizations. In particular, in Section 3 we cover the "pure residual" models, used in fixed-income risk management. In Section 4 we discuss the "time series" approach, used in equity return analysis. In Section 5 we discuss the "cross-section" approach, used in equity and corporate bond portfolio management. In Section 6 we analyze the "statistical" approach, used across asset classes, which includes "principal component analysis", closely related to "random matrix theory", "factor analysis", and hybrid approaches.

In Section 7 we discuss the often misunderstood relationships between linear factor models, the pillars of financial theory such as CAPM and APT, the theory of risk estimation, and risk attribution. The implementation of a platform for risk managers, portfolio managers and traders that connects modularly and flexibly these different aspects is detailed in the companion article on "Factors on Demand" Meucci (2010b).

In the appendix we prove some technical results. Fully commented MATLAB code for all the case studies and the figures is available for download, please refer to Meucci (2009a).

2 The general framework

We denote by $\mathbf{X} \equiv (X_1, \ldots, X_N)'$ the risk drivers of a market, i.e. a set of N yet to be realized random variables that fully determine the randomness in that market. The entries of $\mathbf{X}$ can be the returns of each stock in an equity market, or the credit spreads of corporate bonds, or the entries of implied volatility surfaces for options, etc. We will refer to the risk drivers as "the market".

2.1 Linear factor models

A linear factor model is a decomposition of the market as follows

$$\mathbf{X} \equiv \mathbf{a} + \mathbf{B}\mathbf{F} + \mathbf{U}. \tag{1}$$

In this expression $\mathbf{a} \equiv (a_1, \ldots, a_N)'$ are N constants; $\mathbf{F} \equiv (F_1, \ldots, F_K)'$ are K factors, i.e. yet to be realized random variables that are correlated with the market $\mathbf{X}$; $\mathbf{B}$ is a $N \times K$ matrix of coefficients that transfers the randomness of the factors into the randomness of the risk drivers; and $\mathbf{U} \equiv (U_1, \ldots, U_N)'$ are defined as the N residuals that make (1) an identity. For any given market $\mathbf{X}$, different specifications of $\mathbf{a}$, $\mathbf{B}$ and $\mathbf{F}$ in the linear decomposition (1) give rise to different residuals $\mathbf{U}$ and thus different factor models. Since typically $K \ll N$, factor models are dimension reduction techniques used for estimation. However, factor models can also be used to interpret the risk in a portfolio, which is a $N \equiv 1$-dimensional market, as discussed in the companion article Meucci (2010b) on "Factors on Demand".

We stress that in (1) we model *random variables* with a fully arbitrary distribution, instead of modeling *data*. In (1) there is no notion of time-series, historical observations, the time dimension, and the like. Instead, all the variables are random and yet-to-be realized in the future.

2.2 Dominant-plus-residual factor models

A *dominant-plus-residual* factor model is a decomposition of the market $\mathbf{X}$

$$\mathbf{X} \equiv \mathbf{a}^* + \mathbf{B}^*\mathbf{F}^* + \mathbf{U}. \tag{2}$$

where the term $\mathbf{a}^* + \mathbf{B}^*\mathbf{F}^*$ is optimized to explain the largest portion of the randomness in the market $\mathbf{X}$ under a potential set of constraints $\mathcal{C}$ on the decision variables, according to a general measure of fitness $\mathcal{F}\{\cdot,\cdot\}$:

$$(\mathbf{a}^*, \mathbf{B}^*, \mathbf{F}^*) \equiv \underset{(\mathbf{a},\mathbf{B},\mathbf{F})\in\mathcal{C}}{\operatorname{argmax}} \mathcal{F}\{\mathbf{a}+\mathbf{BF}, \mathbf{X}\}. \tag{3}$$

The fitness $\mathcal{F}\{\cdot,\cdot\}$ can be defined in terms of quantiles, conditional value at risk, skewness, kurtosis, etc. We refer the reader to the companion paper Meucci (2010b) and to Meucci (2010a) for a fully general discussion with applications.

In this article, we will focus on the fitness provided by the multivariate weighted r-square, see also Meucci (2005)

$$\mathcal{F}\{\mathbf{Y}, \mathbf{X}\} \equiv R^2_{\mathbf{W}}\{\mathbf{Y}, \mathbf{X}\} \equiv 1 - \frac{\operatorname{tr}\left(\operatorname{Cov}\{\mathbf{W}(\mathbf{Y}-\mathbf{X})\}\right)}{\operatorname{tr}\left(\operatorname{Cov}\{\mathbf{WX}\}\right)}. \tag{4}$$

In this expression "tr" denotes the trace of a matrix, i.e. the sum of its diagonal elements; $\operatorname{Cov}\{\mathbf{Y}\}$ is the matrix of the covariances among the entries of the generic random vector $\mathbf{Y}$; and $\mathbf{W}$ is an full-rank square matrix of weights to rescale and potentially rotate the variables. Since $\operatorname{tr}(\operatorname{Cov}\{\mathbf{Y}\}) \propto \frac{1}{N}\sum_{n=1}^{N}\operatorname{Var}\{Y_n\}$, the average variance of $\mathbf{Y}$, the numerator in (4) is always positive, and only zero if the vector $\mathbf{Y}$ replicates $\mathbf{X}$ exactly; and the denominator is a normalization term, meant to make the overall r-square scale-independent. The closer the r-square is to 1, the more $\mathbf{Y}$ is responsible for the randomness in $\mathbf{X}$.

Substituting the r-square definition of fitness (4) in the general definition (3), in this article a *dominant-plus-residual* linear factor model is a decomposition of the market (2) where

$$(\mathbf{B}^*, \mathbf{F}^*) \equiv \underset{(\mathbf{B},\mathbf{F})\in\mathcal{C}}{\operatorname{argmax}} R^2_{\mathbf{W}}\{\mathbf{BF}, \mathbf{X}\}. \tag{5}$$

Notice that the constant vector $\mathbf{a}$ does not play a role in the maximization (5) that defines a dominant-plus-residual model, due to the shift invariance of the covariances in (4).

2.3 Systematic-plus-idiosyncratic factor models

A *systematic-plus-idiosyncratic* factor model is a selection of $\mathbf{a}$, $\mathbf{B}$, and $\mathbf{F}$ in the linear decomposition of the market $\mathbf{X} \equiv \mathbf{a}+\mathbf{BF}+\mathbf{U}$ such that the residuals are uncorrelated with each other and they are uncorrelated with the factors:

$$\begin{aligned} \operatorname{Cor}\{U_n, F_k\} &= 0, \quad n = 1,\ldots,N, k = 1,\ldots,K \\ \operatorname{Cor}\{U_n, U_m\} &= 0, \quad n \neq m = 1,\ldots,N. \end{aligned}$$

Notice again that in reality the choice of the constant vector $\mathbf{a}$ does not play a role in the conditions (6)-(6) that define a systematic-plus-idiosyncratic model, due to the shift invariance of the correlation.

A systematic-plus-idiosyncratic factor model displays convenient aggregation properties. Assume that the market are the linear returns on a set of securities

$\mathbf{R} \equiv \mathbf{a}+\mathbf{BF}+\mathbf{U}$ and that we need to compute the risk of a portfolio with weights $\mathbf{w}$, where risk is represented by the variance of the portfolio return $R_{\mathbf{w}} \equiv \mathbf{R}'\mathbf{w}$.

For a general factor model the variance of the portfolio return is the sum of three terms, one stemming from the K^2 covariances $\operatorname{Cov}\{F_k, F_j\}$ among the factors; one from the KN covariances $\operatorname{Cov}\{F_k, U_n\}$ between factors and residuals; one from the N^2 covariances $\operatorname{Cov}\{U_n, U_m\}$ among the residuals x

$$\begin{aligned} \operatorname{Var}\{R_{\mathbf{w}}\} \quad = \quad & \underbrace{\sum_{m,n=1}^{N} \sum_{j,k=1}^{K} w_n B_{n,k} \operatorname{Cov}\{F_k, F_j\} B_{m,j} w_{m'}}_{\text{factors}} \\ & \underbrace{+2 \sum_{m,n=1}^{N} \sum_{k=1}^{K} w_m B_{m,k} \operatorname{Cov}\{F_k, U_n\} w_{n'}}_{\text{cross-term}} \\ & \underbrace{+ \sum_{m,n=1}^{N} w_n w_m \operatorname{Cov}\{U_n, U_m\}.}_{\text{residual}} \end{aligned} \tag{6}$$

On the other hand, for a systematic-plus-idiosyncratic model the the variance of the portfolio return is the sum of two simpler terms. Indeed, from the definition of idiosyncratic residual (6) instead of N^2 covariances among the residuals we only need to account for the N variances $\operatorname{Var}\{U_n\} = \operatorname{Cov}\{U_n, U_n\}$; and from the definition of systematic factors (6), the cross-terms $\operatorname{Cov}\{F_k, U_n\}$ disappear. Therefore for a systematic-plus-idiosyncratic model the variance of the portfolio return reads

$$\operatorname{Var}\{R_{\mathbf{w}}\} = \underbrace{\sum_{m,n=1}^{N} \sum_{j,k=1}^{K} w_n B_{n,k} \operatorname{Cov}\{F_k, F_j\} B_{m,j} w_{m'}}_{\text{systematic}} + \underbrace{\sum_{n=1}^{N} w_n^2 \operatorname{Var}\{U_n\}}_{\text{idiosyncratic}}. \tag{7}$$

The amount of time and memory required by computers to calculate (7) is much lower than for (6). As a result, portfolio analytics vendors routinely rely on (7) for risk computations.

2.4 Summary of most common factor models

As summarized in Figure 1, *all* the approaches to factor modeling are instances of the optimization (5) under different constraints, and thus they represent dominant-plus-residual models.

In particular, the "pure residual" model of Section 3 corresponds to a full specification of both $\mathbf{B}^*$ and $\mathbf{F}^*$ in the dominant-plus-residual optimization (5). The "time series" approach of Section 4 corresponds to a full specification of $\mathbf{F}^*$ in the dominant-plus-residual optimization (5). When in addition $\mathbf{B}^*$ follows

Factor models $\mathbf{X} \equiv \mathbf{a} + \mathbf{BF} + \mathbf{U}$			$\mathrm{Cor}\{U_n, U_m\} = 0$	$\mathrm{Cor}\{U_n, F_k\} = 0$
"Pure residual" ($\mathbf{B}$ and $\mathbf{F}$ fully constrained)			X	X
"Time-series" ($\mathbf{F}$ fully constrained)	**"OLS"**	$\mathbf{B}$: fully unconstrained; fit: R^2	X	✓
	Generalized	$\mathbf{B}$: constrained; fit: R^2, CVaR,...	X	X
"Cross-section" ($\mathbf{B}$ fully constrained)	**"w-OLS"**	$\mathbf{F}$: fully unconstrained; fit: R^2	X	X
	Generalized	$\mathbf{F}$: constrained; fit: R^2, CVaR ,...	X	X
"Statistical"	**"PCA"**	$\mathbf{B}$, $\mathbf{F}$: fully unconstrained; fit: R^2	X	✓
	"FA"	$\mathbf{B}$, $\mathbf{F}$: overconstrained	X	✓
	Hybrid	$\mathbf{B}$, $\mathbf{F}$ constrained; fit: R^2, CVaR ,...	X	X

Figure 1: Factor models do not satisfy systematic-plus-idiosyncratic conditions

from an unconstrained optimization, we obtain the standard ordinary-least-square (OLS) approach. The "cross section" approach of Section 5 corresponds to a full specification of $\mathbf{B}^*$ in the dominant-plus-residual optimization (5). When in addition $\mathbf{F}^*$ follows from an unconstrained optimization, we obtain the standard weighted ordinary-least-square (w-OLS) approach. The statistical approach of section 6 corresponds to the case when neither $\mathbf{B}^*$ nor $\mathbf{F}^*$ are specified exogenously in the dominant-plus-residual optimization (5). When the optimization is fully unconstrained, we obtain principal component analysis (PCA); when the maximization is over-constrained, we obtain factor analysis (FA).

The above approaches can be extended to more powerful hybrid models by tailoring the constraints and the fitness target in the more general dominant-plus-residual optimization (3). This is the essence of the "Factors on Demand" methodology discussed in the companion paper Meucci (2010b).

On the other hand, as summarized in Figure 1, *none* of the approaches to factor modeling satisfies both the idiosyncratic condition (6) and the systematic condition (6) and therefore *no* factor model is of systematic-plus-idiosyncratic type. This follows because the idiosyncratic and systematic conditions form a set of $N(N-1)/2 + NK$ constraints, which far exceeds the number of model parameters. From this observation we derive many caveats while implementing factor models. In particular, it is incorrect to rely on the simplified two-term formula (7) to compute the risk of a portfolio.

3 "Pure residual" factor models

In this section we discuss the "pure residual" approach. We consider a market $\mathbf{X}$ and a linear factor model as (1)

$$\mathbf{X} \equiv \mathbf{a} + \mathbf{BF} + \mathbf{U}, \tag{8}$$

where the factors $\mathbf{F}$, the loadings $\mathbf{B}$ and the constant $\mathbf{a}$ are all imposed exogenously.

For example, consider the government bond market. In this context $\mathbf{X}$ are the yet-to-be realized returns on N bonds from the current date to the investment horizon of, say, one week; $\mathbf{a}$ are the deterministic components of the returns, namely the "carry", or "roll-down"; $\mathbf{F}$ are the yet-to-be realized changes from the current date to the investment horizon of K key points of the government interest rates curve; and the entries of $\mathbf{B}$ are the $N \times K$ key rate durations of the bonds, given the current market conditions.

Assume that the joint distribution of the market and the factors $f_{\mathbf{X},\mathbf{F}}$ is known. Then the residuals and their distribution are fully determined by the factor model (2)

$$\mathbf{U} \equiv \mathbf{X} - \mathbf{a} - \mathbf{BF}. \tag{9}$$

These residuals are *not* uncorrelated with the factors and they are *not* idiosyncratic, i.e. they do not satisfy (6)-(6). Therefore the "pure residual" factor models are not of systematic-plus-idiosyncratic type.

Correlations rate changes/residuals

	Δ 6m	Δ 1y	Δ 2y	Δ 3y	Δ 5y	Δ 7y	Δ 10y	Δ 20y	Δ 30y
Bond 1	23%	15%	32%	36%	37%	33%	40%	44%	47%
Bond 2	-25%	-25%	-9%	-11%	-29%	-32%	-31%	-23%	-12%
Bond 3	-31%	-34%	3%	4%	-4%	-7%	-5%	1%	10%

Correlations residuals/residuals

	Bond 1	Bond 2	Bond 3
Bond 1	100%	28%	43%
Bond 2	28%	100%	47%
Bond 3	43%	47%	100%

Figure 2: Pure residual model typical in fixed-income is not-systematic plus-idiosyncratic

Consider our example of the government bond market. We model the joint distribution of the yet-to-be realized bond returns $\mathbf{X}$ and key rate changes $\mathbf{F}$ by means of empirical distribution

$$f_{\mathbf{X},\mathbf{F}} \equiv \frac{1}{T}\sum_{t=1}^{T} \delta^{(\mathbf{x}_t,\mathbf{f}_t)}. \tag{10}$$

In this expression $\{\mathbf{x}_t, \mathbf{f}_t\}_{t=1,\ldots,T}$ denote the time series of the joint historical realizations of the bond returns and the key rate changes; and $\delta^{(\mathbf{y})}$ denotes the Dirac-delta, which concentrates a unit probability mass on the generic point $\mathbf{y}$, see Appendix A.1 for more details and for a visual interpretation. Note that in general in the bond market the returns are not invariants, because the presence of the maturity date makes close-to-maturity returns much different from far-from-maturity returns, see Meucci (2005). However, the invariance assumption holds true in approximation for far-from-maturity bonds over a limited time span.

Using (9) we can compute the joint distribution of the factors, i.e. the key rate changes, and the residuals

$$f_{\mathbf{F},\mathbf{U}} \equiv \frac{1}{T}\sum_{t=1}^{T} \delta^{(\mathbf{f}_t,\mathbf{u}_t)}, \tag{11}$$

where $\mathbf{u}_t \equiv \mathbf{x}_t - \mathbf{a} - \mathbf{B}\mathbf{f}_t$ are the empirical realizations of the residual. From this distribution, we can compute the cross-correlations.

In Figure 2 we report the correlations of the residuals with the key rate changes and with each other in a real example, please refer to Meucci (2009a) for more details and for the MATLAB code. As we can see in the figure, these correlations are not zero and therefore the model is not of systematic-plus-idiosyncratic type.

Although the "pure residual" approach is not an instance of systematic-plus-idiosyncratic factor modeling, it is an instance of the dominant-plus-residual approach. Indeed, we can interpret this approach as a constrained optimization (5), where all the parameters are exogenously restricted to pre-specified values. In practical applications, such values are gauged to give the ensuing factor model the highest explanatory power.

4 "Time series", or "macroeconomic" factor models

Here we discuss the "time series" class of factor models. In Section 4.1 we characterize this class of models. In Section 4.2 we construct the best model in this class according to the r-square maximization framework of Section 2, which turns out to be familiar ordinary least square model. In Section 4.3 we use again the r-square maximization framework to generalize the discussion to constrained "time series" factor models.

4.1 Definition of "time series" models

For a given market $\mathbf{X}$, the class of "time series" linear models refers to decompositions such as (1), which we report here for convenience

$$\mathbf{X} \equiv \mathbf{a} + \mathbf{B}\mathbf{F} + \mathbf{U}, \tag{12}$$

where $\mathbf{a} \equiv (a_1, \ldots, a_N)'$ are N constants; $\mathbf{F} \equiv (F_1, \ldots, F_K)'$ are K factors, i.e. yet to be realized random variables that are correlated with the market $\mathbf{X}$; $\mathbf{B}$ is a $N \times K$ matrix of coefficients that transfers the randomness of the factors into the randomness of the risk drivers; and $\mathbf{U} \equiv (U_1, \ldots, U_N)'$ are defined as the N residuals that make (12) an identity.

In a "time series" model the factors $\mathbf{F}$ are specified exogenously. Therefore the starting point in "time series" models is the joint distribution of the market $\mathbf{X}$ and the exogenous factors $\mathbf{F}$, represented by the fully general pdf $f_{\mathbf{X},\mathbf{F}}$.

To illustrate, we consider the US stock market. In this context $\mathbf{X}$ are the yet-to-be realized returns on $N \approx 500$ stocks in the S&P 500 from the current date to the investment horizon of, say, one week. Then, consider as factors $\mathbf{F}$ the yet-to-be realized returns of K industry indices from the current date to the investment horizon. In particular, we choose the $K \equiv 10$ MSCI US sector indices.

We model the joint distribution of $\mathbf{X}$ and $\mathbf{F}$ by means of the empirical distribution

$$f_{\mathbf{X},\mathbf{F}} \equiv \frac{1}{T} \sum_{t=1}^{T} \delta^{(\mathbf{x}_t, \mathbf{f}_t)}, \tag{13}$$

where $\delta^{(\mathbf{y})}$ denotes the Dirac-delta, which concentrates a unit probability mass on the generic point $\mathbf{y}$, see Appendix A.1 for more details and for a visual interpretation.

Given the joint distribution of the market $\mathbf{X}$ and the factors $\mathbf{F}$, each choice of the loadings $\mathbf{B}$ and the constant $\mathbf{a}$ in (12) gives rise to a different "time series" model with different residuals $\mathbf{U} \equiv \mathbf{X} - \mathbf{a} - \mathbf{B}\mathbf{F}$. To complete the model, we must specify $\mathbf{B}$ and $\mathbf{a}$.

These models are known as "time series" factor models because the distribution of $\mathbf{X}$ and $\mathbf{F}$, and thus the specification of $\mathbf{B}$ and $\mathbf{a}$, are determined by time series analysis, see e.g. Straumann and Garidi (2007) for an application in the equity market. However, we emphasize that "time series" is a misnomer, because also all the other approaches, including for instance the "cross-section" approach of Section 5, ultimately rely on time series analysis to estimate the joint distribution of the market, the factors, and the residuals.

These models are also known as "macroeconomic" factor models, because, when applied to the financial markets, the factors can be specified as macroeconomic variables.

4.2 OLS, or unconstrained models

To determine the loadings $\mathbf{B}$, we follow our general principle (5) to maximize the r-square of the model. Since in the present case the factors are fully specified we obtain

$$\mathbf{B}^* \equiv \underset{\mathbf{B}}{\operatorname{argmax}}\, R^2_{\mathbf{W}} \{\mathbf{BF}, \mathbf{X}\}. \tag{14}$$

As we show in Meucci (2005), in the absence of constraints the dominant-plus-residual optimization (14) can be solved explicitly

$$\mathbf{B}^* \equiv \operatorname{Cov}\{\mathbf{X}, \mathbf{F}\} \operatorname{Cov}\{\mathbf{F}\}^{-1}. \tag{15}$$

Notice that the result does not depend on the weights $\mathbf{W}$. The maximum unconstrained r-square provided by the solution (15) reads

$$R^2_{\mathbf{W}} \left\{\operatorname{Cov}\{\mathbf{X}, \mathbf{F}\} \operatorname{Cov}\{\mathbf{F}\}^{-1} \mathbf{F}, \mathbf{X}\right\} = ???, \tag{16}$$

see Meucci (2005).

Then we can set the constant vector $\mathbf{a}$ in such a way that the expectation of the residuals be null, $\mathrm{E}\{\mathbf{U}\} \equiv \mathbf{0}$, which implies

$$\mathbf{a}^* \equiv \mathrm{E}\{\mathbf{X}\} - \mathbf{B}^* \mathrm{E}\{\mathbf{F}\}. \tag{17}$$

With the optimal loadings $\mathbf{B}^*$ and the optimal constant vector $\mathbf{a}^*$ we can compute the residuals $\mathbf{U} \equiv \mathbf{X} - \mathbf{a}^* - \mathbf{B}^*\mathbf{F}$. The residuals are not uncorrelated among each other, i.e. $\operatorname{Cor}\{U_n, U_m\} \neq 0$ and thus the residuals are not idiosyncratic.

However, as we show in Meucci (2005) the residuals are uncorrelated with the factors, i.e. $\operatorname{Cor}\{U_n, F_k\} = 0$. It is remarkable that all these NK equalities are satisfied ex-post without being imposed ex-ante in the r-square maximization (14). Equivalently, it is remarkable that the NK coefficients $\mathbf{B}^*$ determined by the NK equations $\operatorname{Cor}\{U_n, F_k\} \equiv 0$ happen to maximize the r-square of the model.

We continue with our stock market case study. The optimal loadings (15) stemming from the empirical distribution (13) are the standard ordinary least squares (OLS) coefficients

$$\mathbf{B}^* \equiv [\frac{1}{T}\sum_{t=1}^{T} (\mathbf{x}_t - \widehat{\boldsymbol{\mu}}_X)(\mathbf{f}_t - \widehat{\boldsymbol{\mu}}_F)'][\frac{1}{T}\sum_{t=1}^{T} (\mathbf{f}_t - \widehat{\boldsymbol{\mu}}_F)(\mathbf{f}_t - \widehat{\boldsymbol{\mu}}_F)']^{-1}, \tag{18}$$

where $\widehat{\boldsymbol{\mu}}_Y \equiv \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}_t$ denotes the average. The optimal constant vector (17) that follows from the empirical distribution (13) reads

$$\mathbf{a}^* \equiv \widehat{\boldsymbol{\mu}}_X - \mathbf{B}^*\widehat{\boldsymbol{\mu}}_F. \tag{19}$$

With the specification of the loadings (18) and the constant (19) we can compute the realized residuals $\mathbf{u}_t \equiv \mathbf{a}^* - \mathbf{x}_t - \mathbf{B}^* \mathbf{f}_t$. Then we can compute the joint distribution of the factors, i.e. the returns of the indices, and the residuals

$$f_{\mathbf{F},\mathbf{U}} \equiv \frac{1}{T} \sum_{t=1}^{T} \delta^{(\mathbf{f}_t, \mathbf{u}_t)}. \tag{20}$$

From this distribution we can compute all the cross-correlations. In Figure ??? we report the correlations of the residual with the factors and among each other. Please refer to Meucci (2009a) for more details and for the MATLAB code.

4.3 Models with general constraints

More in general, we can determine the loadings $\mathbf{B}$ in a "time series" approach by imposing general constraints $\mathcal{C}$ on $\mathbf{B}$ and then applying our general principle (5) to maximize the r-square. Then (14) becomes

$$\mathbf{B}^* \equiv \underset{\mathbf{B} \in \mathcal{C}}{\operatorname{argmax}} R^2_{\mathbf{W}} \{\mathbf{BF}, \mathbf{X}\}. \tag{21}$$

As in (17) we set the constant vector in such a way that the residuals have null expectation

$$\mathbf{a}^* \equiv \mathrm{E}\{\mathbf{X}\} - \mathbf{B}^* \mathrm{E}\{\mathbf{F}\}. \tag{22}$$

Then with the optimal loadings $\mathbf{B}^*$ and the optimal vector $\mathbf{a}^*$ we can compute the residuals $\mathbf{U} \equiv \mathbf{X} - \mathbf{a}^* - \mathbf{B}^* \mathbf{F}$.

As discussed at the end of Section 2.4, "time series" models cannot be of systematic-plus-idiosyncratic type because the conditions $\mathrm{Cor}\{U_n, F_k\} = 0$ and $\mathrm{Cor}\{U_n, U_m\} = 0$ for all the factors and the residuals correspond to $NK + N(N-1)/2$ equations, whereas the unknown loadings are only NK. Notice that the presence of the constraints in (21) breaks the remarkable result $\mathrm{Cor}\{U_n, F_k\} = 0$ which applies in the unconstrained case (14).

We continue with our stock market case study. We impose the constraint that all the loadings be bound from below by $\underline{B} \equiv 0.8$ and from above by $\overline{B} \equiv 1.2$ and that the market-capitalization weighted sum of the loadings be one

$$0.8 \leq B_{n,k} \leq \overline{B}, \quad n = 1, \ldots, N, k = 1, \ldots, K \tag{23}$$

$$\sum_{n=1}^{N} M_n B_{n,k} \equiv 1. \tag{24}$$

We compute the optimal "time series" loadings $\mathbf{B}^*$ as in (21). This is an instance of quadratic programming, which can be easily solved numerically.

The optimal constant vector (22) that follows from the empirical distribution (13) is computed as (19), where $\mathbf{B}^*$ is the newly defined constrained matrix of loadings $\mathbf{a}^* \equiv \overline{\mathbf{x}}. - \mathbf{B}^* \overline{\mathbf{f}}.$.

As in the unconstrained case, with the loadings $\mathbf{B}^*$ and the constant $\mathbf{a}^*$ we compute the realized residuals $\mathbf{u}_t \equiv \mathbf{a}^* - \mathbf{x}_t - \mathbf{B}^* \mathbf{f}_t$. The joint distribution of the factors, i.e. the returns of the indices, and the residuals has the same functional expression as in the unconstrained case (20) but the residuals $\mathbf{u}_t$ take on different values

$$f_{\mathbf{F},\mathbf{U}} \equiv \frac{1}{T} \sum_{t=1}^{T} \delta^{(\mathbf{f}_t, \mathbf{u}_t)}. \tag{25}$$

From this distribution we can compute all the cross-correlations. In Figure ??? we report the correlations of the residual with the factors and among each other. Please refer to Meucci (2009a) for more details and for the MATLAB code.

5 "Cross section", or "fundamental" factor models

Here we discuss the "cross section" class of factor models. In Section 5.1 we characterize this class of models. In Section 5.2 we construct the best model in this class according to the r-square maximization framework of Section 2, which turns out to be familiar weighted ordinary least square cross-sectional model. In Section 5.3 we use again the r-square maximization framework to generalize the discussion to constrained "cross section" factor models.

5.1 Definition of "cross section" models

For a given market $\mathbf{X}$, the class of "cross section" linear models refers to decompositions such as (1), which we report here for convenience

$$\mathbf{X} \equiv \mathbf{a} + \mathbf{B}\mathbf{F} + \mathbf{U}, \tag{26}$$

where $\mathbf{a} \equiv (a_1, \ldots, a_N)'$ are N constants; $\mathbf{F} \equiv (F_1, \ldots, F_K)'$ are K factors, i.e. yet to be realized random variables that are correlated with the market $\mathbf{X}$; $\mathbf{B}$ is a $N \times K$ matrix of coefficients that transfers the randomness of the factors into the randomness of the risk drivers; and $\mathbf{U} \equiv (U_1, \ldots, U_N)'$ are defined as the N residuals that make (26) an identity.

In a "cross section" model the loadings $\mathbf{B}$ are specified exogenously. The rationale of this approach is to impose loadings $\mathbf{B}$ that accurately characterize the market $\mathbf{X}$ and then extract from the market the factors $F_k \equiv g_k(\mathbf{X})$, where $k = 1, \ldots, K$, that best explain those characteristics, by means of suitable functions g_k. For simplicity, and for consistence with the principles of linear factor modeling, we impose that the factors be extracted from the market by means of affine functions

$$\mathbf{F} \equiv \mathbf{d} + \mathbf{G}\mathbf{X}, \tag{27}$$

where $\mathbf{d}$ is a K-dimensional constant vector and $\mathbf{G}$ is a $K \times N$ matrix of coefficients.

Then, given the market $\mathbf{X}$ with distribution $f_{\mathbf{X}}$ and the loadings $\mathbf{B}$, each choice of the factor extraction coefficients $\mathbf{d}$ and $\mathbf{G}$ gives rise to a different "cross section" model (26)-(27), with residuals

$$\mathbf{U} \equiv (\mathbf{I} - \mathbf{BG})\,\mathbf{X} - (\mathbf{a} + \mathbf{Bd})\,, \tag{28}$$

whose distribution is determined by the market distribution $f_{\mathbf{X}}$. To complete the model, we must specify factor-extraction coefficients $\mathbf{G}$ and $\mathbf{d}$ in (27).

This approach is known as "cross section" because, as we shall see below, in special cases the factors extraction coefficients are the coefficients of a weighed cross sectional regression. This approach is also known as "fundamental", because, when applied to the equity market, the loadings are often defined in terms of fundamental book variables, see e.g. Fama and French (1993), or Menchero, Morozov, and Shepard (2008).

5.2 Weighted OLS, or unconstrained models

The starting point in "cross section" models is the market $\mathbf{X}$, with a fully general distribution represented by the pdf $f_{\mathbf{X}}$.

For example, we consider again the US stock market discussed in the case study in Section 4. In this context, $\mathbf{X}$ are the yet-to-be realized returns on $N \equiv 500$ stocks from the current date to the investment horizon of one week and we model the distribution of $\mathbf{X}$ by means of the empirical distribution

$$f_{\mathbf{X}} \equiv \frac{1}{T} \sum_{t=1}^{T} \delta^{(\mathbf{x}_t)}, \tag{29}$$

where $\{\mathbf{x}_t\}_{t=1,\ldots,T}$ is the time series of the historical realizations of the stock returns, and $\delta^{(\mathbf{y})}$ denotes the Dirac-delta, which concentrates a unit probability mass on the generic point $\mathbf{y}$, see Appendix A.1 for more details and for a visual interpretation.

To determine the factor-extraction coefficients, we follow our general principle (5) to maximize the r-square of the model. In the present case, the loadings $\mathbf{B}$ are fully specified and the constant $\mathbf{d}$ plays no role in the computation of the r-square (4). Therefore (5) becomes

$$\mathbf{G}^* \equiv \underset{\mathbf{G}}{\operatorname{argmax}}\, R^2_{\mathbf{W}} \{\mathbf{BGX}, \mathbf{X}\}\,, \tag{30}$$

As we show in Appendix A.2, the maximization (30) can be solved explicitly. Assuming that $\mathbf{B}$ has full rank and defining $\mathbf{\Phi} \equiv \mathbf{W}'\mathbf{W}$ we obtain

$$\mathbf{G}^* = (\mathbf{B}'\mathbf{\Phi}\mathbf{B})^{-1}\,\mathbf{B}'\mathbf{\Phi}. \tag{31}$$

Then the maximum r-square provided by this solution reads

$$R^2_{\mathbf{W}} \left\{\mathbf{B}\,(\mathbf{B}'\mathbf{\Phi}\mathbf{B})^{-1}\,\mathbf{B}'\mathbf{\Phi}\mathbf{X}, \mathbf{X}\right\} = ???, \tag{32}$$

see Appendix A.2. Notice that (31) are the coefficients of a weighted OLS cross-sectional regression. However, we emphasize that at no point did we run a regression on data. Instead, in (30) we maximized the r-square of the *distribution* of the factor model.

Using the convenient choice $\mathbf{d}^* \equiv \mathbf{0}$ for the constant vector in the factor-extraction function (27) the factors are defined as

$$\mathbf{F} \equiv (\mathbf{B}'\mathbf{\Phi}\mathbf{B})^{-1}\,\mathbf{B}'\mathbf{\Phi}\mathbf{X}. \tag{33}$$

To illustrate, we continue with our US equity example (29). We set the loadings as dummy variables relative to the $K \equiv 10$ exhaustive and mutually exclusive industry sectors of the MSCI classification

$$B_{nk} \equiv \begin{cases} 1 & \text{if stock } n \text{ is in sector } k \\ 0 & \text{if stock } n \text{ is not in sector } k. \end{cases} \tag{34}$$

Then we set the weights matrix $\mathbf{W}$ as a diagonal whose entries are the inverse of the volatility of the yet-to-be-determined residual, which we proxy with the standard deviation of the stocks. Therefore

$$\mathbf{\Phi} \equiv \mathbf{W}'\mathbf{W} = [\operatorname{diag}(\operatorname{Cov}\{\mathbf{X}\})]^{-1}, \tag{35}$$

where follows from the empirical distribution (29) and results in the sample covariance.

From (33) we can extract the factors

$$\mathbf{f}_t \equiv (\mathbf{B}'\mathbf{\Phi}\mathbf{B})^{-1}\,\mathbf{B}'\mathbf{\Phi}\mathbf{x}_t, \quad t = 1, \ldots, T. \tag{36}$$

This is a set of of weighted OLS cross-sectional regressions of the market returns $\mathbf{x}_t$, performed for each time in the observation period.

Notice that these factors can be interpreted as the joint returns at time t of K portfolios of our $N \equiv 500$ stocks, where the weights of the generic k-th portfolio is represented by the k-th row in the matrix $(\mathbf{B}'\mathbf{\Phi}\mathbf{B})^{-1}\,\mathbf{B}'\mathbf{\Phi}$. Given the structure of the loadings (34), we expect these factors to resemble the returns of the explicit indices that appear in the "time series" approach case study (13). In Figure ??? we display the overall correlations ???

Then we set the constant vector $\mathbf{a}$ in (26) in such a way that the residuals expectations be null, $\operatorname{E}\{\mathbf{U}\} \equiv \mathbf{0}$, which implies

$$\mathbf{a}^* \equiv \left(\mathbf{I} - \mathbf{B}\,(\mathbf{B}'\mathbf{\Phi}\mathbf{B})^{-1}\,\mathbf{B}'\mathbf{\Phi}\right)\operatorname{E}\{\mathbf{X}\}. \tag{37}$$

With the optimal factor-extraction coefficients $\mathbf{G}^*$ defined in (30) and the optimal constant vector $\mathbf{a}^*$ defined in (37) we can compute the residuals (28), which read

$$\mathbf{U} \equiv \left(\mathbf{I} - \mathbf{B}\,(\mathbf{B}'\mathbf{\Phi}\mathbf{B})^{-1}\,\mathbf{B}'\mathbf{\Phi}\right)(\mathbf{X} - \operatorname{E}\{\mathbf{X}\}). \tag{38}$$

The residuals (38) are *not* uncorrelated with the factors, i.e. $\mathrm{Cor}\{U_n, F_k\} \neq 0$ and thus the factors are not systematic. Furthermore, the residuals (38) are *not* uncorrelated among each other, i.e. $\mathrm{Cor}\{U_n, U_m\} \neq 0$ and thus they are *not* idiosyncratic. As discussed at the end of Section 2.4, this happens because the $NK + N(N-1)/2$ restrictions $\mathrm{Cor}\{U_n, F_k\} \equiv 0$ and $\mathrm{Cor}\{U_n, U_m\} \equiv 0$ are too many to be satisfied with only NK decision variables in the r-square maximization (30).

In our stock example we use (38) and (31) to extract a residual for each observation $\mathbf{x}_t$ of the market returns

$$\mathbf{u}_t \equiv \left(\mathbf{I} - \mathbf{B}\left(\mathbf{B}'\boldsymbol{\Phi}\mathbf{B}\right)^{-1}\mathbf{B}'\boldsymbol{\Phi}\right)\left(\mathbf{x}_t - \widehat{\boldsymbol{\mu}}_X\right), \quad t = 1, \ldots, T \tag{39}$$

Then from the market distribution (29) and the extracted factors and residuals (36)-(39) we derive the the joint distribution of the factors and the residuals

$$f_{\mathbf{F},\mathbf{U}} = \frac{1}{T}\sum_{t=1}^{T} \delta^{(\mathbf{f}_t, \mathbf{u}_t)}. \tag{40}$$

From this distribution we can compute all the cross-correlations. In Figure ??? we report the correlations of the residual with the factors and among each other. Please refer to Meucci (2009a) for more details and for the MATLAB code.

5.3 Models with general constraints

More in general, we can determine the factors in a "cross section" approach by imposing constraints $\mathcal{C}$ on the factor-extraction affine transformation (27) and applying our general principle (5) to maximize the r-square. Then (30) becomes

$$\mathbf{G}^* \equiv \underset{\mathbf{G}\in\mathcal{C}}{\mathrm{argmax}}\, R^2_{\mathbf{W}}\{\mathbf{BGX}, \mathbf{X}\}. \tag{41}$$

Using again the convenient choice $\mathbf{d}^* \equiv \mathbf{0}$ for the constant vector in the factor-extraction function (27) the factors are defined in general as

$$\mathbf{F} \equiv \mathbf{G}^*\mathbf{X}. \tag{42}$$

Setting the constant vector $\mathbf{a}$ in (26) in such a way that the residuals expectations be null, $\mathrm{E}\{\mathbf{U}\} \equiv \mathbf{0}$, we obtain

$$\mathbf{a}^* \equiv (\mathbf{I} - \mathbf{BG}^*\boldsymbol{\Phi})\,\mathrm{E}\{\mathbf{X}\}. \tag{43}$$

Then the residuals (28) read

$$\mathbf{U} \equiv (\mathbf{I} - \mathbf{BG}^*)(\mathbf{X} - \mathrm{E}\{\mathbf{X}\}). \tag{44}$$

These residual cannot give rise to a systematic-plus-idiosyncratic "cross section" factor model because the $NK + N(N-1)/2$ restrictions $\mathrm{Cor}\{U_n, F_k\} \equiv 0$ and

$\text{Cor}\{U_n, U_m\} \equiv 0$ and the additional constraints $\mathcal{C}$ in the r-square maximization (41) are too many to be satisfied with only NK decision variables.

To illustrate ???

6 "Statistical" factor models

Here we discuss the "statistical" class of factor models. In Section 6.1 we characterize this class of models. In Section 6.2 we construct the best model in this class according to the r-square maximization framework of Section 2, which turns out to be familiar principal component analysis approach. In Section 6.3 we use again the r-square maximization framework to generalize the discussion to arbitrarily constrained hybrid factor models. In Section 6.4 we over-constrain the model, trying but failing to achieve a systematic-plus-idiosyncratic decomposition. As a result we obtain the familiar factor analysis approach.

6.1 Definition of "statistical" models

For a given market $\mathbf{X}$, the class of "statistical" linear models refers to decompositions such as (1), which we report here for convenience

$$\mathbf{X} \equiv \mathbf{a} + \mathbf{BF} + \mathbf{U}. \tag{45}$$

where $\mathbf{a} \equiv (a_1, \ldots, a_N)'$ are N constants; $\mathbf{F} \equiv (F_1, \ldots, F_K)'$ are K factors, i.e. yet to be realized random variables that are correlated with the market $\mathbf{X}$; $\mathbf{B}$ is a $N \times K$ matrix of coefficients that transfers the randomness of the factors into the randomness of the risk drivers; and $\mathbf{U} \equiv (U_1, \ldots, U_N)'$ are defined as the N residuals that make (45) an identity.

In "statistical" linear models neither the coefficients $\mathbf{a}$ and $\mathbf{B}$ nor the factors $\mathbf{F}$ are specified exogenously. Therefore, in this class of models coefficients and factors must be extracted from the market $\mathbf{X}$ and its distribution $f_{\mathbf{X}}$ using statistical techniques. Notice that in this approach also the number K of the factors is a decision variable.

For simplicity, and for consistence with the principles of linear factor modeling, as in the cross-sectional approach discussed in Section 5 we impose that the factors be extracted from the market by means of affine functions

$$\mathbf{F} \equiv \mathbf{d} + \mathbf{GX}, \tag{46}$$

where $\mathbf{d}$ is a K-dimensional constant vector and $\mathbf{G}$ is a $K \times N$ matrix of coefficients. Then the randomness captured by the factors is mapped back into the market space by means of the yet to be specified $N \times K$ matrix of loadings $\mathbf{B}$ and the constant vector $\mathbf{a}$. Each combination $(\mathbf{a}, \mathbf{B}, \mathbf{d}, \mathbf{G})$ gives rise to a different "statistical" model (45)-(46), with residuals

$$\mathbf{U} \equiv (\mathbf{I} - \mathbf{BG})\,\mathbf{X} - (\mathbf{a} + \mathbf{Bd}) \tag{47}$$

whose distribution is determined by the market distribution $f_{\mathbf{X}}$. To complete the model, we must specify all the coefficients $(\mathbf{a}, \mathbf{B}, \mathbf{d}, \mathbf{G})$.

6.2 Principal component analysis, or unconstrained models

To determine the coefficients that determine the statistical model, we follow our leading principle (5) of maximizing the r-square of the factor model, which in this context reads

$$(\mathbf{B}^*, \mathbf{G}^*) \equiv \underset{(\mathbf{B},\mathbf{G})}{\operatorname{argmax}} R^2_{\mathbf{W}} \{\mathbf{BGX}, \mathbf{X}\} . \tag{48}$$

This maximization can be solved explicitly. First, we perform the spectral decomposition of the covariance matrix

$$\operatorname{Cov}\{\mathbf{X}\} \equiv \mathbf{E}\mathbf{\Lambda}\mathbf{E}'. \tag{49}$$

In this expression $\mathbf{\Lambda}$ is the diagonal matrix of the decreasing, positive eigenvalues of the covariance:

$$\mathbf{\Lambda} \equiv \operatorname{diag}\left(\lambda_1^2, \ldots, \lambda_N^2\right) ; \tag{50}$$

and $\mathbf{E}$ is the juxtaposition of the respective eigenvectors, which are orthogonal and of length 1 and thus $\mathbf{EE}' = \mathbf{I}_N$:

$$\mathbf{E} \equiv \left(\mathbf{e}^{(1)}, \ldots, \mathbf{e}^{(N)}\right) , \tag{51}$$

Next, we define a $N \times K$ matrix as the juxtaposition of the first K eigenvectors

$$\mathbf{E}_K \equiv \left(\mathbf{e}^{(1)}, \ldots, \mathbf{e}^{(K)}\right) . \tag{52}$$

Then the solution of (48) reads

$$\mathbf{B}^* = \mathbf{E}_K, \quad \mathbf{G}^* \equiv \mathbf{E}'_K, \tag{53}$$

see Appendix ??? which relies on results in Brillinger (2001).

The r-square (48) provided by the principal component solution (53) reads

$$R^2_{\mathbf{W}} \{\mathbf{E}_K \mathbf{E}'_K (\mathbf{X} - \operatorname{E}\{\mathbf{X}\}) , \mathbf{X}\} = \frac{\sum_{k=1}^K \lambda_k^2}{\sum_{n=1}^N \lambda_n^2}, \tag{54}$$

see Meucci (2005). As a consequence, the steeper the spectrum of the covariance matrix, i.e. the larger the discrepancy among the first eigenvalues, the lower the number K of factors required to obtain a large r-square.

To illustrate, we discuss the swap market. In this context $\mathbf{X}$ are the yet-to-be realized changes of $N \equiv t$ key points of the swap curve, namely $1y$, $2y$, $3y$, $5y$, $7y$, $10y$, $15y$, from the current date to the investment horizon of, say, one week. We model the distribution of $\mathbf{X}$ by means of the empirical distribution

$$f_{\mathbf{X}} \equiv \frac{1}{T} \sum_{t=1}^{T} \delta^{(\mathbf{x}_t)}, \tag{55}$$

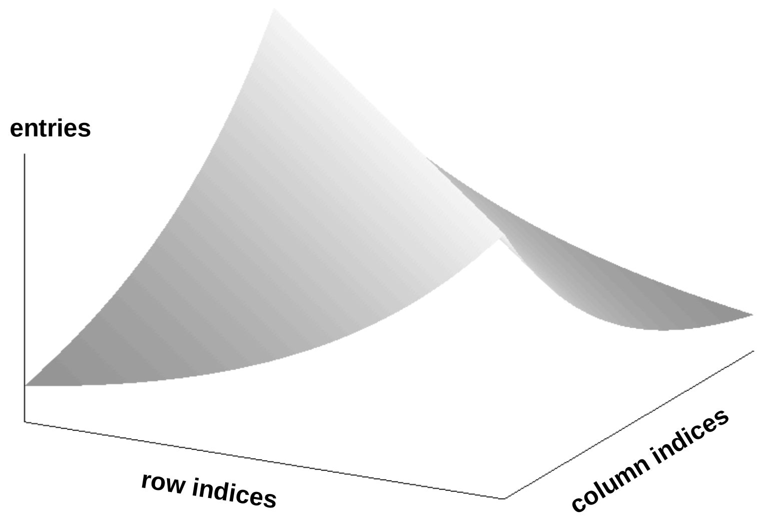

Figure 3: Toeplitz structure of the covariance matrix of interest rate changes

where $\{\mathbf{x}_t\}_{t=1,\ldots,T}$ is the time series of the historical realizations of the weekly swap rate changes, and $\delta^{(\mathbf{y})}$ denotes the Dirac-delta, which concentrates a unit probability mass on the generic point $\mathbf{y}$, see Appendix A.1 for more details and for a visual interpretation.

The covariance matrix stemming from the empirical distribution (55) is the sample covariance of the time series $\{\mathbf{x}_t\}_{t=1,\ldots,T}$ of the swap rate changes, which has the Toeplitz structure plotted in Figure 3. This implies that the eigenvectors (51) in the spectral decomposition (49) are shaped as a Fourier basis, i.e. sinusoidal waves of higher frequency, see Meucci (2005).

In particular, we perform a dimension reduction to $K \equiv 3$ factors ???

$$\mathbf{e}^{(1)} = ()' \tag{56}$$

$$\mathbf{e}^{(2)} = ()' \tag{57}$$

$$\mathbf{e}^{(3)} = () \tag{58}$$

The entries of the first eigenvector (56) are all equal and thus represent a wave with an infinite cycle; the entries of the second eigenvector (57) are monotonic and thus represent a quarter of a cycle; and the entries of the third eigenvector (58) are hump-shaped and thus represent a half cycle. These eigenvectors give rise to respectively to a parallel shift, a tilt and a twist in the swap curve, see Figure 4 and refer to Meucci (2009a) for more details and for the MATLAB code.

The factor loadings $\mathbf{B}^*$ and the factor extraction matrix $\mathbf{G}^*$ in (48) then

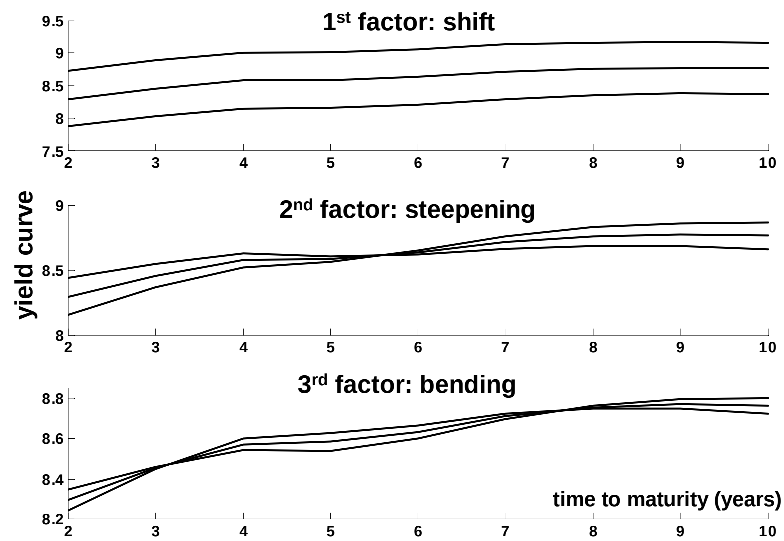

Figure 4: Sinusoidal eigenvectors of the covariance matrix of rate changes and their financial interpretation

read $\mathbf{B}^{*} = \left(\mathbf{e}^{(1)}, \mathbf{e}^{(2)}, \mathbf{e}^{(3)}\right) = \mathbf{G}^{*\prime}$. The constant vector in (61) is approximately null because the expectations of the rate changes that follows from the empirical distribution (55), which is the sample mean during the observation period, is approximately zero: $\mathrm{E}\{\mathbf{X}\} = \widehat{\boldsymbol{\mu}}_X \approx \mathbf{0}$.

Then we specify the constant in the factor extraction formula (46) in such a way that the factors have null expectation

$$\mathbf{d}^{*} \equiv -\mathbf{E}_K' \,\mathrm{E}\{\mathbf{X}\}. \tag{59}$$

Therefore the statistical factors are constructed as

$$\mathbf{F} \equiv \mathbf{E}_K' \left(\mathbf{X} - \mathrm{E}\{\mathbf{X}\}\right). \tag{60}$$

Next, we set the constant vector in the statistical factor model (45) in such a way that the expectation of the residuals be null

$$\mathbf{a}^{*} \equiv \mathrm{E}\{\mathbf{X}\}. \tag{61}$$

From the definition of the statistical factor model (45), the constant vector (59) and the maximum r-square statistical factors (60), the approximation to the market $\mathbf{X}$ provided by the statistical factors is

$$\mathbf{X} \approx \mathrm{E}\{\mathbf{X}\} + \mathbf{E}_K \mathbf{E}_K' \left(\mathbf{X} - \mathrm{E}\{\mathbf{X}\}\right). \tag{62}$$

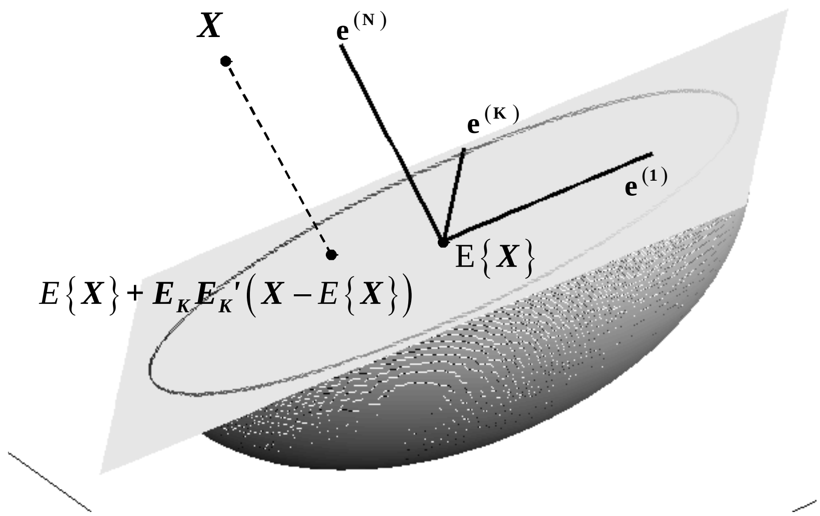

Figure 5: Geometric interpretation principal component - based "statistical" factor models

As we show in Meucci (2005) this approximation has an intuitive geometrical interpretation: it is the orthogonal projection of the market $\mathbf{X}$ onto the hyperplane spanned by the first K eigenvectors. Equivalently, it is the orthogonal projection of the market $\mathbf{X}$ onto the hyperplane spanned by the K longest principal axes of the location-dispersion ellipsoid defined by the market expectation $\mathrm{E}\{\mathbf{X}\}$ and the market covariance $\mathrm{Civ}\{\mathbf{X}\}$, see Figure 5 and refer to Meucci (2010c) for more details.

The residuals of the statistical model (47) are the difference between the market $\mathbf{X}$ and the approximation to the market provided by the statistical factors. From (62) we readily obtain

$$\mathbf{U} \equiv (\mathbf{I} - \mathbf{E}_K \mathbf{E}_K')(\mathbf{X} - \mathrm{E}\{\mathbf{X}\}). \tag{63}$$

These residuals are not uncorrelated among each other, i.e. $\mathrm{Cor}\{U_n, U_m\} \neq 0$ and thus the residuals are not idiosyncratic. However, as we show in Meucci (2005) the residuals are uncorrelated with the factors, i.e. $\mathrm{Cor}\{U_n, F_k\} = 0$. As in the unconstrained "time series" model discussed in Section 4.2, it is remarkable that all these NK equalities are satisfied ex-post without being imposed ex-ante in the r-square maximization (48).

We continue with our stock market case study. The optimal constant vectors

(59) and (61) implied from the empirical distribution of the market (55) read

$$\mathbf{d}^* \equiv -\mathbf{E}_K' \widehat{\boldsymbol{\mu}}_X \tag{64}$$
$$\mathbf{a}^* \equiv \widehat{\boldsymbol{\mu}}_X . \tag{65}$$

Using (60) we can extract the statistical factors

$$\mathbf{f}_t \equiv \mathbf{E}_K' \left(\mathbf{x}_t - \widehat{\boldsymbol{\mu}}_X \right) . \tag{66}$$

Using (63) we can compute the realized residuals

$$\mathbf{u}_t \equiv \left(\mathbf{I} - \mathbf{E}_K \mathbf{E}_K' \right) \left(\mathbf{x}_t - \widehat{\boldsymbol{\mu}}_X \right) \tag{67}$$

Then we can compute the joint distribution of the statistical factors and the residuals

$$f_{\mathbf{F},\mathbf{U}} \equiv \frac{1}{T} \sum_{t=1}^{T} \delta^{(\mathbf{f}_t,\mathbf{u}_t)} . \tag{68}$$

From this distribution we can compute all the cross-correlations. In Figure ??? we report the correlations of the residual with the factors and among each other. Please refer to Meucci (2009a) for more details and for the MATLAB code.

6.3 Hybrid factor models with general constraints

More in general, we can determine the loadings $\mathbf{B}$ and the factors $\mathbf{F}$ factors in a "statistical" approach by applying our general principle (5) under a fully general set of constraints $\mathcal{C}$. Then (48) becomes

$$(\mathbf{B}^*, \mathbf{G}^*) \equiv \underset{(\mathbf{B},\mathbf{G}) \in \mathcal{C}}{\operatorname{argmax}} R^2_{\mathbf{W}} \left\{ \mathbf{BGX}, \mathbf{X} \right\} . \tag{69}$$

Notice that hybrid models include as special cases all the "cross section" approaches discussed in Section (5), where we fully constrain the loadings. Furthermore, after extending the market $\mathbf{X}$ to include a set of exogenous factors $\mathbf{F}$, hybrid models include as special cases all the "time series" approaches discussed in Section 4, where the factors are fully constrained. However, hybrid models are much more flexible, because they allow for partial and joint constraints on both factors and loadings.

By imposing that the hybrid factors have null expectation we generalize (59) as follows

$$\mathbf{d}^* \equiv -\mathbf{G}^* \operatorname{E} \left\{ \mathbf{X} \right\} , \tag{70}$$

obtaining the hybrid factors

$$\mathbf{F} \equiv \mathbf{G}^* \left(\mathbf{X} - \operatorname{E} \left\{ \mathbf{X} \right\} \right) . \tag{71}$$

By imposing that the residuals expectations be null we obtain as in (61)

$$\mathbf{a}^* \equiv \operatorname{E} \left\{ \mathbf{X} \right\} . \tag{72}$$

Then the residuals (28) of the hybrid model read

$$\mathbf{U} \equiv (\mathbf{I} - \mathbf{B}^{*}\mathbf{G}^{*})(\mathbf{X} - \mathrm{E}\{\mathbf{X}\}). \tag{73}$$

Hybrid models are not of systematic-plus-idiosyncratic type because the residuals (73) are not uncorrelated with the factors, i.e. $\mathrm{Cor}\{U_n, F_k\} \neq 0$ and they are not uncorrelated among each other, i.e. $\mathrm{Cor}\{U_n, U_m\} \neq 0$. As discussed at the end of Section 2.4, this happens because the conditions $\mathrm{Cor}\{U_n, F_k\} = 0$ and $\mathrm{Cor}\{U_n, U_m\} = 0$ for all the factors and the residuals correspond to $NK + N(N-1)/2$ equations, whereas the unknowns in the r-square maximization (69) are at most $2NK$, without considering the constraints $\mathcal{C}$.

To illustrate ???

6.4 Factor analysis, or over-constrained models

All the approaches to factor modeling discussed so far are instances of the r-square maximization (5) under different constraints: total constraints for the "pure residual" models in Section 3; constraints on factors for the "time series" models in Section 4; constraints on exposures for the "cross section" models in Section 5; no constraints for the PCA-based "statistical" models in Section 6.2; and mixed generalized constraints for the hybrid models in Section 6.3. As discussed at the end of Section 2.4, all the models fail to be of systematic-plus-idiosyncratic type, because the conditions $\mathrm{Cor}\{U_n, F_k\} = 0$ and $\mathrm{Cor}\{U_n, U_m\} = 0$ represent too many constraints.

In this section we discuss the approach known as "factor analysis", which over-constrains the market to achieve a systematic-plus-idiosyncratic factorization. More precisely, factor analysis aims at approximating the covariance of the market $\mathbf{X}$ in terms of a $N \times K$ matrix $\mathbf{B}$ and a diagonal $N \times N$ matrix $\boldsymbol{\Delta}$ as follows

$$\mathrm{Cov}\{\mathbf{X}\} \approx \mathbf{B}\mathbf{B}' + \boldsymbol{\Delta}, \tag{74}$$

Several methods are available to perform this approximation, see e.g. Rencher (2002).

The formulation (74) appears to be compatible with a statistical factor model (45) where the factors are uncorrelated and have unit variance, i.e. $\mathrm{Cov}\{\mathbf{F}\} = \mathbf{I}$; the statistical factors are truly systematic because they are uncorrelated with the residuals, i.e. $\mathrm{Cor}\{U_n, F_k\} = 0$; and the residuals are truly idiosyncratic, i.e. $\mathrm{Cor}\{U_n, U_m\} = 0$ if $m \neq n$, with $\boldsymbol{\Delta}$ being the variances of the residuals.

However such a factor model cannot be recovered: as we prove in Appendix A.3, it is not possible to extract the hidden factors $\mathbf{F}$ and the residuals $\mathbf{U}$ from the market $\mathbf{X}$ in such a way that the above conditions are satisfied.

To illustrate this phenomenon, we consider an artificial market that is normally distributed

$$\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X), \tag{75}$$

100%	30%	25%	8%	-56%
	100%	-46%	-69%	-80%
		100%	-9%	63%
			100%	12%
				100%

Figure 6: Correlations among "idiosyncratic" residuals from factor analysis

where

$$\boldsymbol{\Sigma}_X \equiv \mathbf{B}\mathbf{B}' + \boldsymbol{\Delta}^2. \tag{76}$$

We generate a large number of scenarios $\{\mathbf{x}_t\}_{t=1,\ldots,T}$ from the distribution (75) with matching sample moments $\widehat{\boldsymbol{\mu}}_X = \boldsymbol{\mu}$ and $\widehat{\boldsymbol{\Sigma}}_X = \boldsymbol{\Sigma}$ as discussed in Meucci (2009c).

Then we run the built-in function `factoran`, which estimates the values $\widehat{\mathbf{B}}$ and $\widehat{\boldsymbol{\Delta}}$ as well as the realizations of the hidden factors $\{\mathbf{f}_t\}_{t=1,\ldots,T}$. From these we can compute the residuals $\mathbf{u}_t \equiv \mathbf{x}_t - \widehat{\mathbf{B}}\mathbf{f}$.

Given the normality of the market and the moment-matching feature, the estimated values match exactly the true parameters, i.e. $\widehat{\mathbf{B}} = \mathbf{B}$ and $\widehat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}$. As we see in Figure 6, the sample correlation of the residuals with the factors is close to zero. However, the sample correlations among the residuals are very high in absolute value. Please refer to Meucci (2009a) for more details and for the MATLAB code.

From the above discussion it follows that factor analysis does not give rise to a true systematic-plus-idiosyncratic factor model. Furthermore, since factor analysis does not attempt to maximize the r-square of the model, in principle the ensuing factor model is not of dominant-plus-residual type. In practice, the routines to compute the approximation (74) are based on iterations of principal component analysis and thus yield a dominant-plus-residual decomposition.

7 Linear factor models, asset pricing and estimation

In this section we connect linear factor modeling with the theory of asset pricing and with estimation theory. In Section 7.1 we present a refresher on the pillars of asset pricing theory, namely the CAPM and the APT. In Section 7.2 we review the concept of invariance, which lies at the very foundation of estimation theory. In Section 7.3 we connect the dots between asset pricing, estimation, and the theory of factor modeling discussed in Sections 2-6.

7.1 Review of asset pricing theory

In order to discuss the financial theory of asset pricing we recall some definitions and simple results, see e.g. Meucci (2005).

The yet to be realized linear returns $\mathbf{R} \equiv (R_1, \ldots, R_N)'$ of N securities from the current time T to a given future investment horizon $T + \tau$ are defined as

$$R_n \equiv \frac{P_{n,T+\tau} - P_{n,T}}{P_{n,T}}, \quad n = 1, \ldots, N, \tag{77}$$

where $P_{n,t}$ denotes the value of one unit of the generic n-th security at time t.

The current portfolio weights $\mathbf{w} \equiv (w_1, \ldots, w_N)'$ in a portfolio that contains u_n units of the generic n-th security are defined as

$$w_n \equiv \frac{u_n P_{n,T}}{\sum_{m=1}^{N} u_m P_{m,T}}, \quad n = 1, \ldots, N. \tag{78}$$

Then the linear return of a portfolio is the weighted average of the securities linear returns

$$R_{\mathbf{w}} \equiv \mathbf{w}'\mathbf{R} \tag{79}$$

and the full investment budget-constraint reads

$$\mathbf{w}'\mathbf{1} \equiv 1. \tag{80}$$

7.1.1 The CAPM

Consider a market of a large set N of non-risk-free securities. We introduce the maximum-Sharpe-ratio portfolio achievable under the budget constraint (80) as follows

$$\mathbf{w}_{SR} \equiv \underset{\mathbf{w}'\mathbf{1}\equiv 1}{\operatorname{argmax}} \left\{ \frac{\mathrm{E}\{R_{\mathbf{w}}\}}{\mathrm{Sd}\{R_{\mathbf{w}}\}} \right\}, \tag{81}$$

where from (79) the portfolio return expectation and standard deviation read

$$\mathrm{E}\{R_{\mathbf{w}}\} = \mathbf{w}'\,\mathrm{E}\{\mathbf{R}\}, \tag{82}$$

$$\mathrm{Sd}\{R_{\mathbf{w}}\} = \sqrt{\mathbf{w}'\,\mathrm{Cov}\{\mathbf{R}\}\,\mathbf{w}}. \tag{83}$$

Assume that there exists a security that is risk-free over the given investment horizon. We denote by r its linear return over the horizon.

It can be proved that the following identity holds for the expected linear returns of the securities

$$\mathrm{E}\{\mathbf{R}\} = r + \mathbf{b}_{\mathbf{w}_{SR}} \left(\mathrm{E}\{R_{\mathbf{w}_{SR}}\} - r\right), \tag{84}$$

where

$$\mathbf{b}_{\mathbf{w}} \equiv \frac{\mathrm{Cov}\{\mathbf{R}, R_{\mathbf{w}}\}}{\mathrm{Var}\{R_{\mathbf{w}}\}} = \frac{\mathrm{Cov}\{\mathbf{R}\}\,\mathbf{w}}{\mathbf{w}'\,\mathrm{Cov}\{\mathbf{R}\}\,\mathbf{w}}, \tag{85}$$

see e.g. Ingersoll (1987). We emphasize that (84) always holds in full generality.

The Capital Asset Pricing Model (CAPM) by Sharpe (1964) and Lintner (1965) states that if all the investors maximize a subjective trade-off between a portfolio expectation $\mathrm{E}\{R_{\mathbf{w}}\}$ and its standard deviation $\mathrm{Sd}\{R_{\mathbf{w}}\}$, then the aggregate market portfolio $\mathbf{w}_M$ is the maximum-Sharpe-ratio portfolio (81), see e.g. Ingersoll (1987).

As a result the following equilibrium relationship must hold for the expected linear returns of the securities

$$\mathrm{E}\{\mathbf{R}\} = r\mathbf{1} + \mathbf{b}_{\mathbf{w}_M}\left(\mathrm{E}\{R_{\mathbf{w}_M}\} - r\right). \tag{86}$$

7.1.2 The APT

The Arbitrage Pricing Theory (APT) by Ross (1976) assumes that the yet to be realized linear returns $\mathbf{R}$ are of systematic-plus-idiosyncratic type, i.e. as in (1)

$$\mathbf{R} = \mathbf{a} + \mathbf{BF} + \mathbf{U}, \tag{87}$$

where $\mathbf{a} \equiv (a_1, \ldots, a_N)'$ are N constants, $\mathbf{F} \equiv (F_1, \ldots, F_K)'$ are K factors, i.e. yet to be realized random variables, $\mathbf{B}$ is a $N \times K$ matrix of coefficients, $\mathbf{U} \equiv (U_1, \ldots, U_N)'$ are the N residuals that make (87) an identity, and as in (6)-(6)

$$\mathrm{Cor}\{U_n, F_k\} = 0, \quad n = 1, \ldots, N, k = 1, \ldots, K \tag{88}$$

$$\mathrm{Cor}\{U_n, U_m\} = 0, \quad n \neq m = 1, \ldots, N. \tag{89}$$

Define K factor-mimicking portfolios $\{\mathbf{w}_{(k)}\}_{k=1,\ldots,K}$ that have unit exposure to the respective factor, but null exposure to all the other factors

$$\begin{aligned} \mathbf{w}_{(1)}\mathbf{B} &\equiv (1, 0, \ldots, 0) \\ &\vdots \\ \mathbf{w}_{(K)}\mathbf{B} &\equiv (0, 0, \ldots, 1) \end{aligned} \tag{90}$$

Notice that these portfolios, which can be computed as follows

$$\begin{pmatrix} \mathbf{w}_{(1)} \\ \vdots \\ \mathbf{w}_{(K)} \end{pmatrix} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B} \tag{91}$$

do not satisfy the budget constraint (80).

Then the APT states that the expected linear returns of the securities satisfy

$$\mathrm{E}\{\mathbf{R}\} \approx r\mathbf{1} + \mathbf{b}_1\left(\mathrm{E}\left\{R_{\mathbf{w}_{(1)}}\right\} - r\right) + \cdots + \mathbf{b}_K\left(\mathrm{E}\left\{R_{\mathbf{w}_{(K)}}\right\} - r\right), \tag{92}$$

see e.g. Ingersoll (1987).

7.2 Estimation and invariance

The "quest for invariance" is at the foundation of risk estimation. Here we provide a quick refresher, we refer the reader to Meucci (2005) for more details.

We denote by $\mathbf{Y}_t \equiv (Y_{t,1}, \ldots, Y_{t,S})'$ the potentially large set of all the S risk drivers that affect a given market at the generic time t. These risk drivers are driven by the invariants, which are shocks $\boldsymbol{\epsilon}_t \equiv (\epsilon_{t,1}, \ldots, \epsilon_{t,S})'$ that are identically and independently distributed.

The most commonly assumed dynamic process that connects the risk drivers with the invariants is the random walk

$$h(\mathbf{Y}_t) = h(\mathbf{Y}_{t-1}) + \boldsymbol{\epsilon}_t, \tag{93}$$

where h is a suitable invertible deterministic function. Refer to Meucci (2009b) for an overview of more complex dynamics that include autocorrelations, stochastic volatility, long memory, etc.

To illustrate, consider one specific stock. Then the driver Y_t is the price itself, and the invariant ϵ_t is the compounded return

$$\epsilon_t \equiv \ln Y_t - \ln Y_{t-1}, \tag{94}$$

which defines a random walk as in (93).

As a second example, consider a corporate bond. Then the drivers are the interest rates of a reference curve Rc_t^y, where y denotes the rate term such as one month, six months, one year, and the spreads Sp_t^y over those rates

$$\mathbf{Y}_t \equiv \left(Rc_t^{1m}, \ldots, Rc_t^{30y}, Sp_t^{1m}, \ldots, Sp_t^{30y}\right)'; \tag{95}$$

and the invariants $\boldsymbol{\epsilon}_t$ are the changes in curve and spreads

$$\boldsymbol{\epsilon}_t \equiv \mathbf{Y}_t - \mathbf{Y}_{t-1}. \tag{96}$$

Again, this is a random walk as in (93).

Finally, consider a European call option with a given strike and expiry on a given underlying. Then the drivers are the underlying, which trades at the price U_t, and the entries $\sigma_t^{m,\tau}$ of the implied volatility surface for that underlying

$$\mathbf{Y}_t \equiv \left(U_t, \sigma_t^{\underline{m},\underline{\tau}}, \ldots, \sigma_t^{\overline{m},\overline{\tau}}\right)', \tag{97}$$

where $[\underline{m}, \ldots, \overline{m}] \times [\underline{\tau}, \ldots, \overline{\tau}]$ denotes the points of a grid of moneyness and time-to-expiry values. In this case the invariants are the compounded returns of the underlying and the volatility surface

$$\boldsymbol{\epsilon}_t \equiv \ln \mathbf{Y}_t - \ln \mathbf{Y}_{t-1}, \tag{98}$$

which is a random walk as in (93).

The invariants are random variables whose behavior is similar across time and therefore they allow us to learn about the future from the past. Since the invariants are i.i.d., their distribution $f_{\boldsymbol{\epsilon}}$ is independent of the time t. This distribution is unknown, and needs to be estimated. One of the main results of statistics, the Law of Large Numbers states that the unknown distribution $f_{\boldsymbol{\epsilon}}$ can be approximated by the empirical distribution stemming from a time series $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_T$ of past realizations of the invariants

$$f_{\boldsymbol{\epsilon}} \approx \frac{1}{T} \sum_{t=1}^{T} \delta^{(\boldsymbol{\epsilon}_t)}. \tag{99}$$

In this expression $\delta^{(\mathbf{y})}$ denotes the Dirac-delta, which concentrates a unit probability mass on the generic point $\mathbf{y}$, see Appendix A.1 for more details and for a visual interpretation.

The approximation (99) can be refined with a variety of advanced estimation techniques, which include generalized maximum likelihood, robustness, Bayesian approaches, etc. However, the main message is clear in (99): in order to learn about the future risk in a given market we have to perform a "quest for invariance".

7.3 Relationships with linear factor models

The theory of linear factor models reviewed in Sections 2-6 applies in full generality to any set of random variables $\mathbf{X}$. The most generic factor model (1), which we report here for convenience

$$\mathbf{X} = \mathbf{a} + \mathbf{BF} + \mathbf{U}, \tag{100}$$

is a linear attribution exercise of a set of random variables $\mathbf{X}$ to a set of K factors $\mathbf{F}$. When the attribution is performed by maximizing the multivariate r-square (5) as we did throughout this article, the attribution is shift invariant. In other words, for the theory of factor modeling, the value of the expectation $\mathrm{E}\{\mathbf{X}\}$ is irrelevant.

The theory of asset pricing reviewed in Section 7.1 only applies to the linear returns $\mathbf{R}$ on a set of securities and therefore it is much more restrictive in scope that factor modeling. The APT *assumes* a systematic-plus-idiosyncratic model for the returns (87) and derive the result (100) on the returns expectations. The CAPM assumes mean-variance preferences in the market and derives the result (86) on the returns expectations. To summarize, the theory of asset pricing is about constraints on the expectation $\mathrm{E}\{\mathbf{R}\}$ of a restrictive market of linear returns.

All the results discussed in this article are undisputable mathematical identities that hold in full generality, rather than questionable empirical results. To achieve this, we clearly separated modeling, which applies to yet to be realized random variables with arbitrary distributions, from estimation, which applies to observed time series of realized data.

The theory of estimation applies to the invariants $\boldsymbol{\epsilon}_t$, i.e. variables that are identically and independently distributed across time. When the number of invariants is large, we can use a factor model on the invariants

$$\boldsymbol{\epsilon}_t = \mathbf{a}_t + \mathbf{B}_t \mathbf{F}_t + \mathbf{U}_t. \tag{101}$$

Then we can use the factor model to impose structure on the estimation. To be completed ???

References

Brillinger, D. R., 2001, *Time Series: Data Analysis and Theory* (Society for Industrial and Applied Mathematics, Classics in Applied Mathematics).

Fama, E. F., and K. R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Ingersoll, E. J., 1987, *Theory of Financial Decision Making* (Rowman and Littlefield).

Lintner, J., 1965, The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.

Menchero, J., A. Morozov, and P. Shepard, 2008, The Barra global equity model (GEM2), *Barra Research Notes* pp. 1–79.

Meucci, A., 2005, *Risk and Asset Allocation* (Springer).

———, 2009a, Exercises in advanced risk and portfolio management - with step-by-step solutions and fully documented code, *Free E-Book* available at http://ssrn.com/abstract=1447443.

———, 2009b, Review of discrete and continuous processes in finance: Theory and applications, *Working paper* Available at http://ssrn.com/abstract=1373102.

———, 2009c, Simulations with exact means and covariances, *Risk* 22, 89–91 Extended version available at http://ssrn.com/abstract=1415699.

———, 2010a, Common misconceptions about 'beta' - hedging, estimation and horizon effects, *GARP Risk Professional - "The Quant Classroom"* June, 42–45 Available as "Quant Nugget 3" at http://ssrn.com/abstract=1619923.

———, 2010b, Factors on Demand - building a platform for portfolio managers risk managers and traders, *Risk* 23, 84–89 available at http://ssrn.com/abstract=1565134.

———, 2010c, Visualizing Principal Component Analysis, *GARP Risk Professional - "The Quant Classroom"* October, 84–87 Available as "Quant Nugget 5" at http://ssrn.com/abstract=1650603.

Rencher, A. C., 2002, *Methods of Multivariate Analysis* (Wiley) 2nd edn.

Ross, S., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Sharpe, W. F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Straumann, D., and T. Garidi, 2007, Developing an equity factor model for risk, *Riskmetrics Journal* 7, 89–128.

A Appendix

In this appendix we discuss technical results that can be skipped at first reading.

A.1 The Dirac delta and the empirical distribution

In this appendix we follow Meucci (2005).

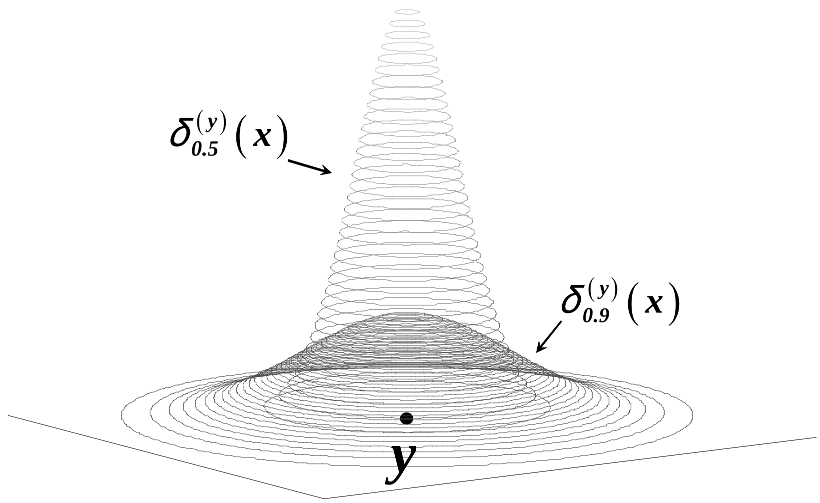

Figure 7: The Dirac delta is the zero-volatility limit of the normal pdf

To introduce the Dirac delta, we first recall the expression of the normal pdf centered in $\mathbf{y}$ with volatility of the order of ϵ

$$\delta_{\epsilon}^{(\mathbf{y})}(\mathbf{x}) \equiv \frac{1}{(2\pi)^{\frac{N}{2}}\epsilon^{N}} e^{-\frac{1}{2\epsilon^{2}}(\mathbf{x}-\mathbf{y})'(\mathbf{x}-\mathbf{y})}. \tag{102}$$

The Dirac delta centered in a generic point $\mathbf{y} \in \mathbb{R}^N$, which we denote by $\delta^{(\mathbf{y})}$, is the limit of the normal pdf centered in $\mathbf{y}$ as the volatility tends to zero, see Figure 7

$$\delta^{(\mathbf{y})} \equiv \lim_{\epsilon \to 0} \delta_{\epsilon}^{(\mathbf{y})} \tag{103}$$

As ϵ approaches zero the pdf becomes taller and thinner around the peak $\mathbf{y}$. Therefore, intuitively, the Dirac delta is a function which is zero everywhere, except at the point $\mathbf{y}$ where it becomes infinity, and its integral over $\mathbb{R}^N$ equals 1. Therefore, for a generic smooth function g the Dirac delta satisfies

$$\int_{\mathbb{R}^N} g(\mathbf{x})\,\delta^{(\mathbf{y})}(\mathbf{x})\,d\mathbf{x} \equiv g(\mathbf{y}). \tag{104}$$

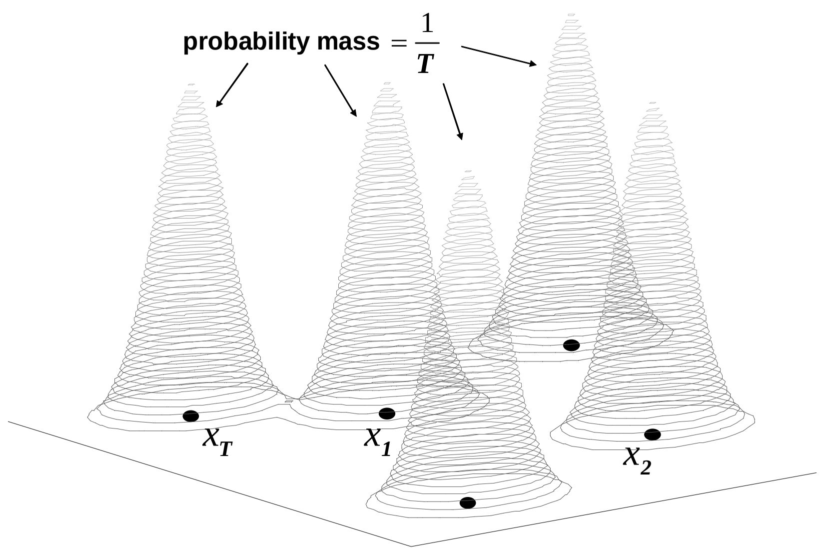

Figure 8: Visualization of the empirical distribution

The empirical distribution represents the simplest way to model the basic assumption of statistics that we can learn about the future realization of a random variable $\mathbf{X}$ from the past realizations

$$i_T \equiv \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}. \tag{105}$$

More precisely, under this distribution any of the past occurrences is an equally likely potential outcome for $\mathbf{X}$, whereas other values cannot take place. Therefore the pdf f_{i_T} of the empirical distribution spikes on the past realizations and is zero otherwise. Furthermore, the spikes have probability mass equal to $1/T$. We can express the pdf of the empirical distribution in terms of the Dirac delta

$$f_{i_T}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} \delta^{(\mathbf{x}_t)}(\mathbf{x}). \tag{106}$$

In Figure 8 we visualize the empirical pdf

A.2 Results on "cross-sectional" factor models

From the definition of the generalized r-square (4) we minimize

$$\begin{aligned} M &\equiv \mathrm{E}\left\{(\mathbf{X}-\mathbf{BGX})'(\mathbf{X}-\mathbf{BGX})\right\} \qquad (107)\\ &= \sum_{n,k,m,j,p} \mathrm{E}\left\{(X_n - B_{nk}G_{km}X_m)(X_n - B_{nj}G_{jp}X_p)\right\} \\ &= \sum_n \mathrm{E}\left\{X_n^2\right\} - 2\sum_{n,k,m} B_{nk}G_{km}\,\mathrm{E}\left\{X_m X_n\right\} \\ &\quad + \sum_{n,k,m,j,p} B_{nk}B_{nj}G_{km}G_{jp}\,\mathrm{E}\left\{X_p X_m\right\}. \end{aligned}$$

If the constraint set $\mathcal{C}$ is empty, we consider the first order conditions with respect to G_{sl}

$$0_{sl} = \frac{\partial M}{\partial G_{sl}} = -2\sum_n B_{ns}\,\mathrm{E}\left\{X_n X_l\right\} + 2\sum_{n,j,p} B_{ns}B_{nj}G_{jp}\,\mathrm{E}\left\{X_p X_l\right\}. \qquad (108)$$

In matrix notation this reads

$$\mathbf{0}_{K\times N} = -2\mathbf{B}'\,\mathrm{E}\left\{\mathbf{XX}'\right\} + 2\mathbf{B}'\mathbf{BG}\,\mathrm{E}\left\{\mathbf{XX}'\right\} \qquad (109)$$

or $\mathbf{G}^* = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$.

A.3 Results on factor analysis

More precisely, the approximation (74) is consistent with a joint covariance of factors and residual that reads

$$\mathrm{Cov}\left\{\begin{pmatrix}\mathbf{F}\\ \mathbf{U}\end{pmatrix}\right\} = \begin{pmatrix}\mathbf{I}_{K\times K} & \mathbf{0}_{K\times N}\\ \mathbf{0}_{N\times K} & \mathbf{\Delta}^2\end{pmatrix}, \qquad (110)$$

which has rank K (the number of factors $\mathbf{F}$) $+N$ (the dimension of the market $\mathbf{X}$). However, assume we could find the extraction matrix $\mathbf{G}$ such that

$$\mathbf{F} = \mathbf{GX}. \qquad (111)$$

Then from (2) we obtain

$$\begin{pmatrix}\mathbf{F}\\ \mathbf{U}\end{pmatrix} = \mathbf{QX}, \qquad (112)$$

where

$$\mathbf{Q} \equiv \begin{pmatrix}\mathbf{G}\\ \mathbf{I}-\mathbf{BG}\end{pmatrix}. \qquad (113)$$

This implies that

$$\mathrm{Cov}\left\{\begin{pmatrix}\mathbf{F}\\ \mathbf{U}\end{pmatrix}\right\} = \mathbf{Q}\,\mathrm{Cov}\left\{\mathbf{X}\right\}\mathbf{Q}', \qquad (114)$$

which has rank N (the number of observables) and thus contradicts (110).