1. Asset return models

Let $y = (y_1, \ldots, y_p)^\top \in \mathbb{R}^p$ denote a (random) vector of asset returns i.e., each y_i denotes the return to the ith asset in some period.

A general, statistical model of asset return posits that for some predictor vector $\psi \in \mathbb{R}^q$ and function $f : \mathbb{R}^q \to \mathbb{R}^p$ there is an approximate relationship $y \approx f(\psi)$. The error of this approximation is denoted by $\epsilon = y - f(\psi)$, and so

$$y = f(\psi) + \epsilon. \tag{1}$$

In principle, the number of dimensions, q, is arbitrary. The vector $\epsilon \in \mathbb{R}^p$ is often refered to as the the specific return, rather than an error residual. This distinction refers to the decomposition of return into a systematic (i.e., $f(\psi)$) and specific components.

A **linear asset model** posits an affine relationship between the return and the predictor variable ψ. That is, for some $p \times q$ matrix C we have $f(\psi) = \mathrm{C}\psi$. A **nonlinear asset model** allows for determinants of the asset return to be driven by interactions of several variables (e.g., the asset returns themselves), as well as higher order terms. These are often implemented in the context of deep learning. The predictor ψ models the return of some financial risk factor in the context of a factor model (Sec. 2), and is the asset return itself in the context of a graphical model (Sec. 3). The following are references for factor, graphical and deep learning models of asset returns.

- **?** : foundations of equity risk models / best practices for investment management.
- **?** develop a graphical model of asset return that accomodates the analysis, estimation and management of equity risk.
- Deep learning references to come (very recent).

The first two moments play a central role in the analysis of asset return models. We denote the expected return by $\mu = \mathrm{E}(y)$ and the covariance by $\Sigma = \mathrm{Var}(y)$. The latter is of particular importance for risk management applications.

2. Factor models

A factor model consists of an exposure matrix $\mathrm{B} \in \mathbb{R}^{q \times p}$ where q is the number of factors and we take $\mathrm{C} = \mathrm{B}$ and $\psi = x$ to write the return as

$$y = \mathrm{B}x + \epsilon \tag{2}$$

The main assumption of a factor model is that the predictor $\psi = x$ is unccorrelated with the specific return ϵ. **That is, we require** $\mathbf{Cov}(x, \epsilon) = 0$. The covariance matrix of the return $\Sigma = \mathrm{Var}(y)$ has dimensions $p \times p$ and the decomposition

$$\Sigma = \mathrm{BVB}^\top + \Omega \tag{3}$$

where $\mathrm{Var}(x) = \mathrm{V}$ and $\mathrm{Var}(\epsilon) = \Omega$ are $q \times q$ and $p \times p$ matrices respectively. This is a decomposition of the covariance into a systematic and specific risk components.

> The q columns of B contain factor exposures (or factor loadings), i.e. if $\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p$ denotes a columns of B, then β_i is the exposure of the ith asset to that factor (equivalently, the loading on that factor). We assume B is full rank (otherwise, some factor(s) is redundant).

The **specific risk** component Ω may have the following properties.

> The matrix Ω is assumed to be symmetric and positive definite. Further, its eigenvalues are assumed to be bounded in p (i.e., as more assets are added to the factor model). A **strict factor model** assumes $\Omega = \Delta$ is diagonal (i.e., uncorrelated specific returns). A more general structure (e.g., sparsity) assumption on Ω implies an **approximate factor model**.

The **systematic risk** BVB has significant freedom in its specification/interpretation.

> The covariance of the factor returns V is assumed to be symmetric and positive definite $q \times q$ matrix. It is customary to assume q is much smaller than p. Assuming that V is diagonal leads to **uncorrelated factor returns** (i.e., uncorrelated entries of $x \in \mathbb{R}^q$). This can always be assumed at the expense of orthogonally transforming the matrix B.[a] The **cannonical orientation** of the factor risk V and factor loadings B achieves uncorrelated factor returns as well as orthogonal factors:
>
> Let Ψ be any diagonal positive definite $q \times q$ matrix. Then, there is a nonsingular $q \times q$ matrix Φ such that defining $\psi = \Phi^{-1}x$ and $\Pi = \mathrm{B}\Phi$ implies the returns satisfy
>
> $$\begin{aligned} y &= \Pi\psi + \epsilon \\ \mathrm{Var}(\psi) &= \Psi \text{ (diagonal)} \\ \mathrm{Var}(y) &= \Sigma = \Pi\Psi\Pi^\top + \Omega \\ &\text{and } \Pi \text{ has orthogonal columns.} \end{aligned} \tag{4}$$
>
> [a] Let $\mathrm{BVB} = \mathrm{BO\Lambda O^\top B} = \mathrm{H\Lambda H^\top}$ for an orthogonal O (i.e., $\mathrm{O^\top O} = \mathrm{OO^\top} = \mathrm{I}$) and $\mathrm{H} = \mathrm{OB}$. This however does not guarantee orthogonal columns of H.

In the **cannonical orientation** the factor returns $\psi = (\psi_1, \ldots \psi_q)^\top$ are uncorrelated and the factor exposures $\pi^1, \ldots, \pi^q$ (the columns of Π) are orthogonal. That is, we have $\langle \pi^k, \pi^\ell \rangle = 0$ for all $k \neq \ell$ and every $\langle \pi^k, \pi^k \rangle = \frac{\alpha_k}{\Psi_{kk}}$ for some constant $\alpha_k \in (0, \infty)$. Here $\langle \cdot, \cdot \rangle$ is any inner product on $\mathbb{R}^p$. To illustrate, for a symmetric positive definite (weight) matrix W, we have $\langle u, v\rangle = u^\top \mathrm{W} v$ and the standard inner product has $\mathrm{W} = \mathrm{I}$. The (α_k) are computed via the diagonal entries of a matrix $\mathcal{A}$ as follows. Let $\mathrm{M}^{1/2}$ denote the square-root of a matrix (i.e. $\mathrm{M} = \mathrm{M}^{1/2}\mathrm{M}^{1/2}$) and let

$$\mathrm{V}^{1/2}\mathrm{B}^\top \mathrm{W}\mathrm{B}\mathrm{V}^{1/2} = \mathrm{O}\mathcal{A}\mathrm{O}^\top \tag{5}$$

where the right side is the eigen-decomposition of the matrix on the left side (i.e., the (α_k) are the eigenvalues of this matrix). Further, $\Phi = \mathrm{V}^{1/2}\mathrm{O}^\top \Psi^{-1/2}$ in (4). Then, orthonormal factor exposures (i.e., $\langle \pi^k, \pi^k\rangle = 1$) may be achieved by requiring the factor variances to satisfy $\Psi_{kk} = \alpha_k$. Alternatively, one can require unit factor variances (i.e., $\Psi = \mathrm{I}$) at the expense of only orthogonal factor exposures.

It is empirical fact that most asset returns data exhibits a factor for which the exposures are mostly all of the same sign and the variance of this factor is large (often the largest). This risk factor is refered to as the **market**. As an example in the Barra MSCI US equity model the exposures to the market factor are set as the isometric vector

$$\mathrm{e} = (1, \dots, 1)^\top \in \mathbb{R}^p \tag{6}$$

i.e., ever asset has unit exposure to the market. The volatility (square-root of the variance) of this factor is typically assumed to be around 16% annualized. In Barra models the factor return variances are typically computed by regressing asset returns onto fixed factor exposures such as e. The other Barra factors have exposures designed to have a zero average exposure. Barra factor exposures are not orthogonal.

The **canonical factor model**, defined as having uncorrelated factor returns and orthogonal factors per (4), may be written to highlight the presence of the **market factor** by decomposing the covariance $\Sigma = \Pi\Psi\Pi^\top + \Omega$ as follows.

$$\Sigma = \sigma^2\beta\beta^\top + \Gamma\Lambda\Gamma + \Omega \tag{7}$$

for market exposures $\beta \in \mathbb{R}^p$, market variance σ^2, a $p\times(q-1)$ matrix Γ of nonmarket factor exposures with the $(q-1)\times(q-1)$ covariance matrix $\mathcal{A}$ of the returns of these factors, and the specific risk matrix Ω. Here, the exposure matrix Π is Γ with an additional column β and covariance Ψ is Λ with an additional diagonal element σ^2.

The relative sizes of the variances and exposures to factors is unidentifiable. For example $\sigma^2\beta\beta^\top = (\sigma/c)^2(c\beta)(c\beta)^\top$ for any $c \neq 0$. Thus some canonical convention is needed to standardize the sizes of the variances and exposures relative to one another. Due to the interpretation of β as the market exposures we adopt the following convention:[a]

> The average market exposure is standardized to be one (which fixes the size of the market variance). Specifically, for some

weights $w \in \mathbb{R}^p$ with $\sum_{i=1}^p w_i = 1$ and every $w_i \geq 0$,

$$\mathrm{m}(\beta) = \sum_{i=1}^p w_i \beta_i = 1\,. \tag{8}$$

The standard choice is every $w_i = 1/p$ (unweighted mean).

The nonmarket factors in $\Gamma\Lambda\Gamma^\top$ may have average exposure equal to zero (the exposures Γ have mixed signs) and for this reason we standardize their variation instead (fixing the Λ). Specifically, each column γ of Γ is taken to have

$$\begin{aligned} \mathrm{s}^2(\gamma) &= \sum_{i=1}^p w_i (\gamma_i - \mathrm{m}(\gamma))^2 = 1, \\ \mathrm{m}(\gamma) &= \sum_{i=1}^p w_i \gamma_i \geq 0. \end{aligned} \tag{9}$$

for any weight vector $w \in \mathbb{R}^p$ as above.

[a]Due to the orthogonal property of the canonical factor model only one factor is expected to be market-like, i.e., large variance and exposures of mostly the same sign.

To get some intuition for these normalizations note that for any vector $v \in \mathbb{R}^p$ and equal weights $w_i = 1/p$ we have the relation the the lengh $|v| = \sqrt{\langle v, v \rangle}$,

$$|v| = \sqrt{p}(\mathrm{m}^2(v) + \mathrm{s}^2(v))\,. \tag{10}$$

In this way, all factor exposures are normalized to have a lengh proportional to $\sqrt{p}$ and either the mean $\mathrm{m}(v)$ or variation $\mathrm{s}(v)$ are standardized to unity (for β or γ).

3. Graphical models.

In this model the predictor $\psi = y$ and we write $\mathrm{C} = \mathrm{A}$ so,

$$y = \mathrm{A}y + \epsilon \tag{11}$$

Here, $p = q$ so that A is a $p \times p$ matrix with a zero diagonal (i.e. no asset is allowed to predict itself). The main assumption (can always be satified) is

$$\mathrm{Cov}(\mathrm{A}y, \epsilon) = 0\,. \tag{12}$$

The second main assumption is that:

- A is a sparse matrix.

This requirement may be in place for several reasons. It can be justified from a financial perspective, from an estimation perspective, but also to maintain a consistency with factor modeling. In a factor model the number of paramers that are estimates in much smaller than p^2. For example, for a strict factor model, only $p \times q + q + p$ parameters are estimated and q is assumed much smaller than p. This places a limit on the number of non-zero entries of A.

The covariance decomposition in a factor model takes the form

$$\Sigma = A\Sigma A^\top + (I + A)D \tag{13}$$

where $D = \mathrm{Cov}(y, \epsilon)$ which is a diagonal matrix. This decomposition mirrors that in (3) and attempts to decompose the covariance into

- $A\Sigma A^\top$ – the variance explained by the asset returns themselves. This should be compared to $BVB^\top$, the variance explained by the factors.
- The remaining variance $(I + A)D$ (not explained by the asset returns).

One implication of sparsity on A is that the precision matrix $K = \Sigma^{-1}$ is also sparse. In a graphical model we have the identity

$$K = D^{-1}\Omega D^{-1} \tag{14}$$

where $\Omega = \mathrm{Var}(\epsilon)$, which we see is also a sparse matrix.

References