CFM
INSIGHT.DATA.CLARITY.

Non-linearities and dependences in factor modeling

The case for the volatilities of factors in stock returns

Rémy Chicheporticle with Jean-Philippe Bouchaud

12 Jan. 2017

1. Modeling stock returns dependences
   - Factor models and linear correlations
   - Non-linearities
   - Description of the non-linear model
2. Empirics: parameters estimation
   - The data: range and properties
   - Parameters of the stochastic volatility of the factors
3. Conclusion

1. Modeling stock returns dependences
   - Factor models and linear correlations
   - Non-linearities
   - Description of the non-linear model
2. Empirics: parameters estimation
   - The data: range and properties
   - Parameters of the stochastic volatility of the factors
3. Conclusion

1. Modeling stock returns dependences
   - Factor models and linear correlations
   - Non-linearities
   - Description of the non-linear model

2. Empirics: parameters estimation
   - The data: range and properties
   - Parameters of the stochastic volatility of the factors

3. Conclusion

Definitions and notations

Scope = cross-sectional dependences among daily returns of stock prices

This work = mainly phenomenological/empirical contribution
(no focus on estimation techniques, statistical properties, etc.)

Minimal extension of structural factor model $\neq$ explicit copula modeling

Notations:

- X is a $T \times N$ matrix, stacking realizations of a (standard) random vector of size N
- $\rho = \frac{1}{T} X^{\dagger} X$ is the usual estimator of the $(N \times N)$ correlation matrix

Scope = cross-sectional dependences among daily returns of stock prices

This work = mainly phenomenological/empirical contribution
(no focus on estimation techniques, statistical properties, etc.)

Minimal extension of structural factor model $\neq$ explicit copula modeling

Notations:

- X is a $T \times N$ matrix, stacking realizations of a (standard) random vector of size N
- $\rho = \frac{1}{T} X^\dagger X$ is the usual estimator of the ($N \times N$) correlation matrix

Scope = cross-sectional dependences among daily returns of stock prices

This work = mainly phenomenological/empirical contribution
(no focus on estimation techniques, statistical properties, etc.)

Minimal extension of structural factor model $\neq$ explicit copula modeling

Notations:

- X is a $T \times N$ matrix, stacking realizations of a (standard) random vector of size N
- $\rho = \frac{1}{T} X^{\dagger} X$ is the usual estimator of the $(N \times N)$ correlation matrix

Definitions and notations

Scope = cross-sectional dependences among daily returns of stock prices

This work = mainly phenomenological/empirical contribution
(no focus on estimation techniques, statistical properties, etc.)

Minimal extension of structural factor model $\neq$ explicit copula modeling

Notations:

- X is a $T \times N$ matrix, stacking realizations of a (standard) random vector of size N
- $\rho = \frac{1}{T} X^{\dagger} X$ is the usual estimator of the $(N \times N)$ correlation matrix

Non-linear dependences in pairs of stock returns exhibit non-trivial patterns. F.ex. the excess joint probability

$$p_{ij} = \text{Prob}[X_{ti} < 0 \text{ and } X_{tj} < 0] - 1/4$$

is predicted to be $\sin \rho_{ij}/2\pi$ by the whole class of so-called elliptical copulas (and even beyond !).

"predicted – measured" discrepancy:

$$\Delta(\rho_{ij}) = \log[\arg\sin(2\pi p_{ij})] - \log \rho_{ij}$$

$\Delta(\rho_{ij})$ vs ρ_{ij}

Motivation: excess probabilities

Non-linear dependences in pairs of stock returns exhibit non-trivial patterns. F.ex. the excess joint probability

$$p_{ij} = \text{Prob}[X_{ti} < 0 \text{ and } X_{tj} < 0] - 1/4$$

is predicted to be $\sin \rho_{ij}/2\pi$ by the whole class of so-called elliptical copulas (and even beyond !).

"predicted – measured" discrepancy:

$$\Delta(\rho_{ij}) = \log[\arg\sin(2\pi p_{ij})] - \log \rho_{ij}$$

$\Delta(\rho_{ij})$ vs ρ_{ij}

Non-linear dependences in pairs of stock returns exhibit non-trivial patterns. F.ex. the excess joint probability

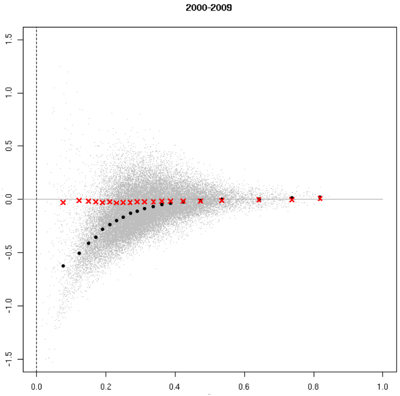
$$p_{ij} = \text{Prob}[X_{ti} < 0 \text{ and } X_{tj} < 0] - 1/4$$

is predicted to be $\sin \rho_{ij}/2\pi$ by the whole class of so-called elliptical copulas (and even beyond !).

“predicted – measured” discrepancy:

$$\Delta(\rho_{ij}) = \log[\arg\sin(2\pi p_{ij})] - \log \rho_{ij}$$

$\Delta(\rho_{ij})$ vs ρ_{ij}



Linear multi-factors model

$$X_{tj} = \sum_{k=1}^{M} \beta_{kj} F_{tk} + E_{tj}$$

F_{t1} is always some definition of "the market"

Interpretations of factor(s):

- known/exogeneous/economic vs unknown/endogenous/algebraic
- regression vs decomposition

The meaning of the "residuals" e_j?

$$X_{tj} = \sum_{k=1}^{M} \beta_{kj} F_{tk} + E_{tj}$$

F_{t1} is always some definition of "the market"

Interpretations of factor(s):

- known/exogeneous/economic vs unknown/endogenous/algebraic
- regression vs decomposition

The meaning of the "residuals" e_j?

Linear multi-factors model

$$X_{tj} = \sum_{k=1}^{M} \beta_{kj} F_{tk} + E_{tj}$$

F_{t1} is always some definition of “the market”

Interpretations of factor(s):

- known/exogeneous/economic vs unknown/endogenous/algebraic
- regression vs decomposition

The meaning of the “residuals” e_j?

Linear multi-factors model

$$X_{tj} = \sum_{k=1}^{M} \beta_{kj} F_{tk} + E_{tj}$$

F_{t1} is always some definition of “the market”

Interpretations of factor(s):

- known/exogeneous/economic vs unknown/endogenous/algebraic
- regression vs decomposition

The meaning of the “residuals” e_j?

Linear multi-factors model: estimation

- Input: standardized return series X_{ti}, number of factors M (=10 below). ***not*** F_{tk}
- Output: coefficients β_{ki}, factor series F_{tk}, residual series E_{ti}

We want to find the M most relevant *uncorrelated* and *common* unit-variance factors F ($T \times M$), and the exposures β ($M \times N$) of every stock to every factor.

$$\langle X_{ti} X_{tj} \rangle_t = \begin{cases} \sum_{k=1}^{M} \beta_{ki}\beta_{kj} & , i \neq j \\ 1 & , i = j \end{cases}$$

We look for the matrix $\beta^\dagger \beta$ of rank M that best fits the empirical correlation matrix.
We get the orthogonal series of F and E by daily cross-sectional regressions

Linear multi-factors model: estimation

- Input: standardized return series X_{ti}, number of factors M (=10 below). ***not F$_{tk}$***
- Output: coefficients β_{ki}, factor series F_{tk}, residual series E_{ti}

We want to find the M most relevant *uncorrelated* and *common* unit-variance factors F ($T \times M$), and the exposures β ($M \times N$) of every stock to every factor.

$$\langle X_{ti} X_{tj} \rangle_t = \begin{cases} \sum_{k=1}^{M} \beta_{ki}\beta_{kj} & , i \neq j \\ 1 & , i = j \end{cases}$$

We look for the matrix $\beta^\dagger \beta$ of rank M that best fits the empirical correlation matrix.
We get the orthogonal series of F and E by daily cross-sectional regressions

- Input: standardized return series X_{ti}, number of factors M (=10 below). ***not*** $\boldsymbol{F_{tk}}$
- Output: coefficients β_{ki}, factor series F_{tk}, residual series E_{ti}

We want to find the M most relevant *uncorrelated* and *common* unit-variance factors F ($T \times M$), and the exposures β ($M \times N$) of every stock to every factor.

$$\langle X_{ti} X_{tj} \rangle_t = \begin{cases} \sum_{k=1}^{M} \beta_{ki}\beta_{kj} & , i \neq j \\ 1 & , i = j \end{cases}$$

We look for the matrix $\beta^\dagger \beta$ of rank M that best fits the empirical correlation matrix.
We get the orthogonal series of F and E by daily cross-sectional regressions

Linear multi-factors model: estimation

- Input: standardized return series X_{ti}, number of factors M (=10 below). ***not*** $\boldsymbol{F_{tk}}$
- Output: coefficients β_{ki}, factor series F_{tk}, residual series E_{ti}

We want to find the M most relevant *uncorrelated* and *common* unit-variance factors F ($T \times M$), and the exposures β ($M \times N$) of every stock to every factor.

$$\langle X_{ti} X_{tj} \rangle_t = \begin{cases} \sum_{k=1}^{M} \beta_{ki}\beta_{kj} & , i \neq j \\ 1 & , i = j \end{cases}$$

We look for the matrix $\beta^\dagger \beta$ of rank M that best fits the empirical correlation matrix.
We get the orthogonal series of F and E by daily cross-sectional regressions

Linear multi-factors model: estimation

- Input: standardized return series X_{ti}, number of factors M (=10 below). ***not F_{tk}***
- Output: coefficients β_{ki}, factor series F_{tk}, residual series E_{ti}

We want to find the M most relevant *uncorrelated* and *common* unit-variance factors F ($T \times M$), and the exposures β ($M \times N$) of every stock to every factor.

$$\langle X_{ti} X_{tj} \rangle_t = \begin{cases} \sum_{k=1}^{M} \beta_{ki}\beta_{kj} & , i \neq j \\ 1 & , i = j \end{cases}$$

We look for the matrix $\beta^\dagger \beta$ of rank M that best fits the empirical correlation matrix.
We get the orthogonal series of F and E by daily cross-sectional regressions

Recall: $X = F\beta + E$, with β $(M \times N)$

Non-linear correlations of the obtained factors and residuals ?

$$\langle |F_{tk}|^p |F_{tl}|^p \rangle^{1/p^2} \quad , \quad \langle |E_{ti}|^p |E_{tj}|^p \rangle^{1/p^2} \quad , \quad p \in (0, 2]$$

Wait: “aren’t they supposed to be uncorrelated by construction ?”

UNCORRELATED BUT NOT INDEPENDENT

Recall: $X = F\beta + E$, with β $(M \times N)$

Non-linear correlations of the obtained factors and residuals ?

$$\langle |F_{tk}|^p |F_{tl}|^p \rangle^{1/p^2} \quad , \quad \langle |E_{ti}|^p |E_{tj}|^p \rangle^{1/p^2} \quad , \quad p \in (0, 2]$$

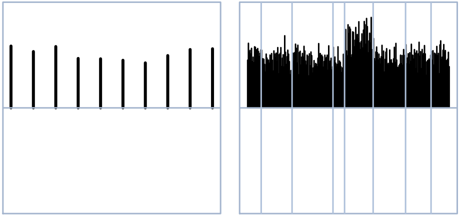
Wait: "aren't they supposed to be uncorrelated by construction ?"

UNCORRELATED BUT NOT INDEPENDENT

Recall: $X = F\beta + E$, with β $(M \times N)$

Non-linear correlations of the obtained factors and residuals ?

$$\langle |F_{tk}|^p |F_{tl}|^p \rangle^{1/p^2} \quad , \quad \langle |E_{ti}|^p |E_{tj}|^p \rangle^{1/p^2} \quad , \quad p \in (0, 2]$$



1st eigenmode

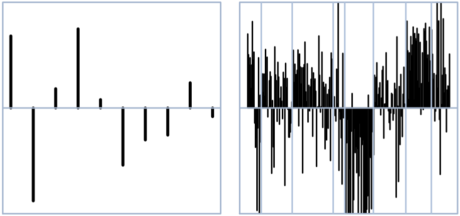
Wait: “aren’t they supposed to be uncorrelated by construction ?”

UNCORRELATED BUT NOT INDEPENDENT

Recall: $X = F\beta + E$, with β $(M \times N)$

Non-linear correlations of the obtained factors and residuals ?

$$\langle |F_{tk}|^p |F_{tl}|^p \rangle^{1/p^2} \quad , \quad \langle |E_{ti}|^p |E_{tj}|^p \rangle^{1/p^2} \quad , \quad p \in (0, 2]$$



2nd eigenmode

Wait: “aren’t they supposed to be uncorrelated by construction ?”

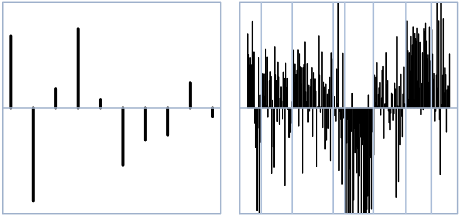
UNCORRELATED BUT NOT INDEPENDENT

Dependence structure in factors and residuals

Recall: $X = F\beta + E$, with β $(M \times N)$

Non-linear correlations of the obtained factors and residuals ?

$$\langle |F_{tk}|^p |F_{tl}|^p \rangle^{1/p^2} \quad , \quad \langle |E_{ti}|^p |E_{tj}|^p \rangle^{1/p^2} \quad , \quad p \in (0, 2]$$



2nd eigenmode

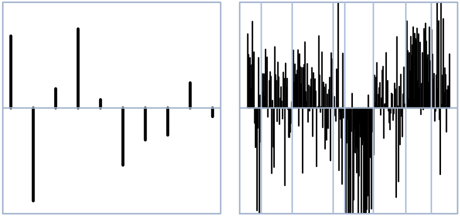
Wait: “aren’t they supposed to be uncorrelated by construction ?”

UNCORRELATED BUT NOT INDEPENDENT

Recall: $X = F\beta + E$, with β $(M \times N)$

Non-linear correlations of the obtained factors and residuals ?

$$\langle |F_{tk}|^p |F_{tl}|^p \rangle^{1/p^2} \quad , \quad \langle |E_{ti}|^p |E_{tj}|^p \rangle^{1/p^2} \quad , \quad p \in (0, 2]$$



2nd eigenmode

Wait: “aren’t they supposed to be uncorrelated by construction ?”

UNCORRELATED BUT NOT INDEPENDENT

Non-Gaussian multi-factors model

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j$$

with non-Gaussian and dependent (though uncorrelated) factors and residuals:

- One-factor model for the log-vol of linear factors f_k

$$f_k = \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k), \qquad \langle f_k^2 \rangle = 1$$

- Two-factors model for the log-vol of residuals e_j

$$e_j = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \tilde{s}_{jj}\tilde{\omega}_j), \qquad \langle e_j^2 \rangle = 1 - \sum_l \beta_{lj}^2$$

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j$$

with non-Gaussian and dependent (though uncorrelated) factors and residuals:

- One-factor model for the log-vol of linear factors f_k

$$f_k = \epsilon_k \exp(A_{k0}\Omega_0 + s_k\omega_k), \qquad \langle f_k^2 \rangle = 1$$

- Two-factors model for the log-vol of residuals e_j

$$e_j = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \tilde{s}_{jj}\tilde{\omega}_j), \qquad \langle e_j^2 \rangle = 1 - \sum_l \beta_{lj}^2$$

Non-Gaussian multi-factors model

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j$$

with non-Gaussian and dependent (though uncorrelated) factors and residuals:

- One-factor model for the log-vol of linear factors f_k

$$f_k = \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k), \qquad \langle f_k^2 \rangle = 1$$

- Two-factors model for the log-vol of residuals e_j

$$e_j = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j), \qquad \langle e_j^2 \rangle = 1 - \sum_l \beta_{lj}^2$$

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j$$

with non-Gaussian and dependent (though uncorrelated) factors and residuals:

- One-factor model for the log-vol of linear factors f_k

$$f_k = \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k), \qquad \langle f_k^2 \rangle = 1$$

- Two-factors model for the log-vol of residuals e_j

$$e_j = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j), \qquad \langle e_j^2 \rangle = 1 - \sum_l \beta_{lj}^2$$

Non-Gaussian multi-factors model

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k & = \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k) \\ e_j & = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

Stochastic (e.g. Gaussian):

- random signed ϵ_k, η_j ,
- stochastic log-volatilities $\Omega_0, \omega_k, \tilde{\omega}_j$,

Scalar parameters:

- linear weights, exposure of stock x_j to factor f_k: β_{kj}
- exposure of factor f_k to logvol Ω_0: A_{k0}
  (+ residual factor vol: s_k)
- exposure of residual e_j to logvols Ω_0, ω_1: B_{j0}, B_{j1}
  (+ residual residual vol: $\widetilde{s}_{jj}$)

Non-Gaussian multi-factors model

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k & = \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k) \\ e_j & = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

Stochastic (e.g. Gaussian):

- random signed ϵ_k, η_j ,
- stochastic log-volatilities $\Omega_0, \omega_k, \tilde{\omega}_j$,

Scalar parameters:

- linear weights, exposure of stock x_j to factor f_k: β_{kj}
- exposure of factor f_k to logvol Ω_0: A_{k0}
  (+ residual factor vol: s_k)
- exposure of residual e_j to logvols Ω_0, ω_1: B_{j0}, B_{j1}
  (+ residual residual vol: $\widetilde{s}_{jj}$)

Non-Gaussian multi-factors model

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k & = \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k) \\ e_j & = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \tilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

Stochastic (e.g. Gaussian):

- random signed ϵ_k, η_j ,
- stochastic log-volatilities $\Omega_0, \omega_k, \tilde{\omega}_j$,

Scalar parameters:

- linear weights, exposure of stock x_j to factor f_k: β_{kj}
- exposure of factor f_k to logvol Ω_0: A_{k0}
(+ residual factor vol: s_k)
- exposure of residual e_j to logvols Ω_0, ω_1: B_{j0}, B_{j1}
(+ residual residual vol: $\tilde{s}_{jj}$)

Non-Gaussian multi-factors model

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k & = \epsilon_k \exp(A_{k0}\Omega_0 + s_k\omega_k) \\ e_j & = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

Stochastic (e.g. Gaussian):

- random signed ϵ_k, η_j ,
- stochastic log-volatilities $\Omega_0, \omega_k, \tilde{\omega}_j$,

Scalar parameters:

- linear weights, exposure of stock x_j to factor f_k: β_{kj}
- exposure of factor f_k to logvol Ω_0: A_{k0} (+ residual factor vol: s_k)
- exposure of residual e_j to logvols Ω_0, ω_1: B_{j0}, B_{j1} (+ residual residual vol: $\widetilde{s}_{jj}$)

Non-Gaussian multi-factors model

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k & = \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k) \\ e_j & = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

- stochastic log-volatilities $\Omega_0, \omega_k, \tilde{\omega}_j$,

Ω_0 = dominant and common mode of log-volatility

ω_1 = log-volatility of market f_1

Non-Gaussian multi-factors model

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k & = \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k) \\ e_j & = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

- stochastic log-volatilities $\Omega_0, \omega_k, \tilde{\omega}_j$,

Ω_0 = dominant and common mode of log-volatility

ω_1 = log-volatility of market f_1

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k & = \epsilon_k \exp(A_{k0}\Omega_0 + s_k\omega_k) \\ e_j & = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

- stochastic log-volatilities $\Omega_0, \omega_k, \tilde{\omega}_j$,

Ω_0 = dominant and common mode of log-volatility

ω_1 = log-volatility of market f_1

Stock returns X_{ti}, for the companies in the SP500 continuously traded in the period.

	2000–2004	2005–2009	2000–2009
N	352	345	262
T	1255	1258	2514

Disregard ‘Basic Materials’, as mine companies are typically anti-correlated with other sectors.
Normalize each series.

- **Input**: factor series F_{tk}, residual series E_{ti}, number of factors M
- **Output**: coefficients $A_{k0}, s_k, B_{j0}, B_{j1}, \widetilde{s}_{jj}$, log-volatilities series Ω_{t0}, ω_{t1}

Taking advantage of the exponential structures in the definition of the random volatilities, predictions of arbitrary p-order absolute correlations can be expressed simply:

$$\frac{1}{p^2} \log \frac{\langle |F_{tk}|^p |F_{tl}|^p \rangle}{\langle |F_{tk}|^p \rangle \langle |F_{tl}|^p \rangle} = A_{k0} A_{l0} + \delta_{kl} \Big(\gamma(p) + s_k s_k \Big) \quad (1)$$

$$\frac{1}{p^2} \log \frac{\langle |F_{tk}|^p |E_{ti}|^p \rangle}{\langle |F_{tk}|^p \rangle \langle |E_{ti}|^p \rangle} = A_{k0} B_{i0} + \delta_{k1} A_{11} B_{i1} \quad (2)$$

$$\frac{1}{p^2} \log \frac{\langle |E_{ti}|^p |E_{tj}|^p \rangle}{\langle |E_{ti}|^p \rangle \langle |E_{tj}|^p \rangle} = B_{i0} B_{j0} + B_{i1} B_{j1} + \delta_{ij} \Big(\gamma(p) + \widetilde{s}_i \widetilde{s}_i \Big) \quad (3)$$

- Input: factor series F_{tk}, residual series E_{ti}, number of factors M
- Output: coefficients $A_{k0}, s_k, B_{j0}, B_{j1}, \widetilde{s}_{jj}$, log-volatilities series Ω_{t0}, ω_{t1}

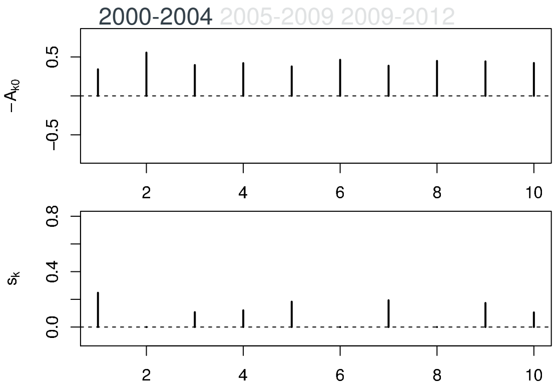
Taking advantage of the exponential structures in the definition of the random volatilities, predictions of arbitrary p-order absolute correlations can be expressed simply:

$$\frac{1}{p^2} \log \frac{\langle |F_{tk}|^p |F_{tl}|^p \rangle}{\langle |F_{tk}|^p \rangle \langle |F_{tl}|^p \rangle} = A_{k0} A_{l0} + \delta_{kl} \Big(\gamma(p) + s_k s_k \Big) \quad (1)$$

$$\frac{1}{p^2} \log \frac{\langle |F_{tk}|^p |E_{ti}|^p \rangle}{\langle |F_{tk}|^p \rangle \langle |E_{ti}|^p \rangle} = A_{k0} B_{i0} + \delta_{k1} A_{11} B_{i1} \quad (2)$$

$$\frac{1}{p^2} \log \frac{\langle |E_{ti}|^p |E_{tj}|^p \rangle}{\langle |E_{ti}|^p \rangle \langle |E_{tj}|^p \rangle} = B_{i0} B_{j0} + B_{i1} B_{j1} + \delta_{ij} \Big(\gamma(p) + \widetilde{s}_i \widetilde{s}_i \Big) \quad (3)$$

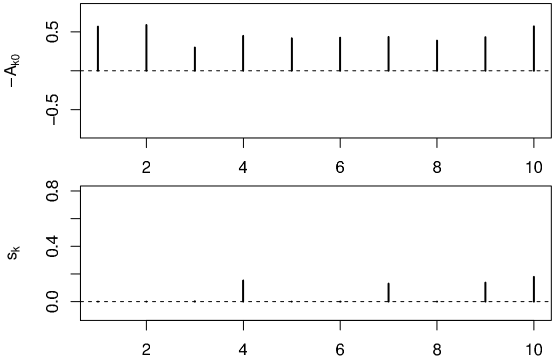
$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k &= \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k) \\ e_j &= \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \tilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$



Volatility exposures: results (A)

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k & = \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k) \\ e_j & = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

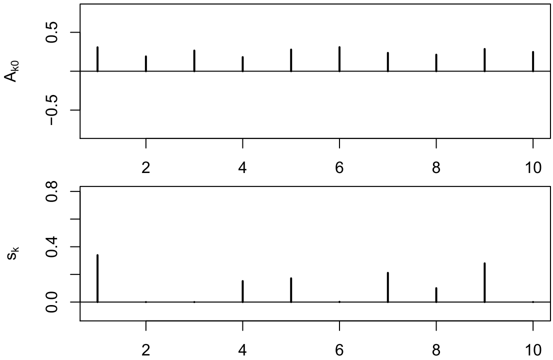
2000-2004 **2005-2009** 2009-2012



Volatility exposures: results (A)

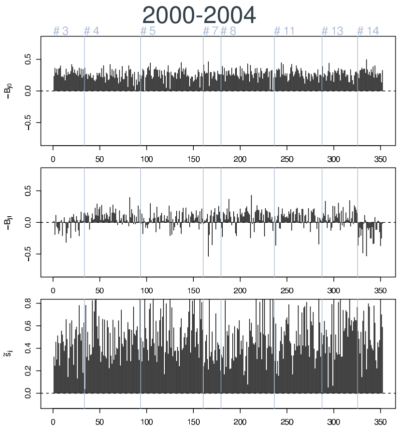
$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k & = \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k) \\ e_j & = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

2000-2004 2005-2009 2009-2012



Volatility exposures: results (B)

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k & = \epsilon_k \exp(A_{k0}\Omega_0 + s_k\omega_k) \\ e_j & = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \tilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$



Then the series of Ω_{t0}, ω_{t1} are retrieved: from

$$\log|e_j| = \Omega_0 B_{j0} + \omega_1 B_{j1} + (\tilde{\omega}_j \widetilde{s}_{jj} + \log|\eta_j|)$$

we design the linear cross-sectional regression

$$\log|E_{t\cdot}| - \langle \log|E_{t\cdot}| \rangle = \begin{pmatrix} \Omega_{t0} & \omega_{t1} \end{pmatrix} \begin{pmatrix} B_{\cdot 0} & B_{\cdot 1} \end{pmatrix}^{\dagger} + \varepsilon_{t\cdot}$$

and solve it date-by-date with a Feasible GLS.

Reconstructing log-vol series

Then the series of Ω_{t0}, ω_{t1} are retrieved: from

$$\log|e_j| = \Omega_0 B_{j0} + \omega_1 B_{j1} + (\tilde{\omega}_j \widetilde{s}_{jj} + \log|\eta_j|)$$

we design the linear cross-sectional regression

$$\log|E_{t\cdot}| - \langle \log|E_{t\cdot}| \rangle = \begin{pmatrix} \Omega_{t0} & \omega_{t1} \end{pmatrix} \begin{pmatrix} B_{\cdot 0} & B_{\cdot 1} \end{pmatrix}^{\dagger} + \varepsilon_{t\cdot}$$

and solve it date-by-date with a Feasible GLS.

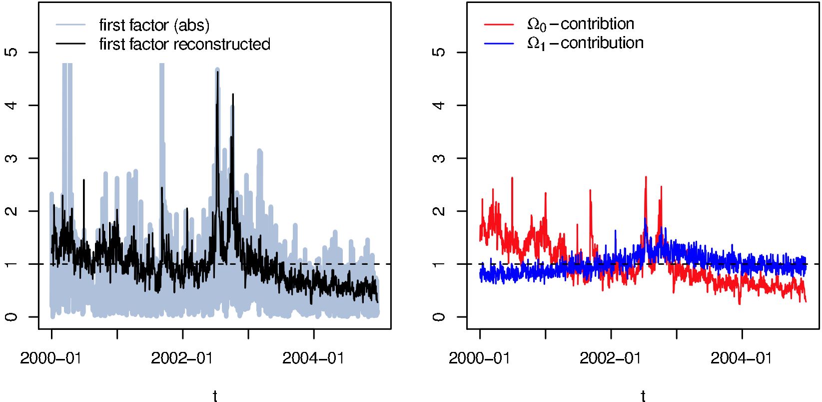
INSIGHT.DATA.CLARITY.

Reconstructing F_{t1}

1st factor of the model: $|f_1| = |\epsilon_1| \, e^{A_{11}\omega_1} \, e^{A_{10}\Omega_0}$

Stock index volatility: $\langle I(t)^2 \rangle \approx \sigma(t)^2 \rho(t)$

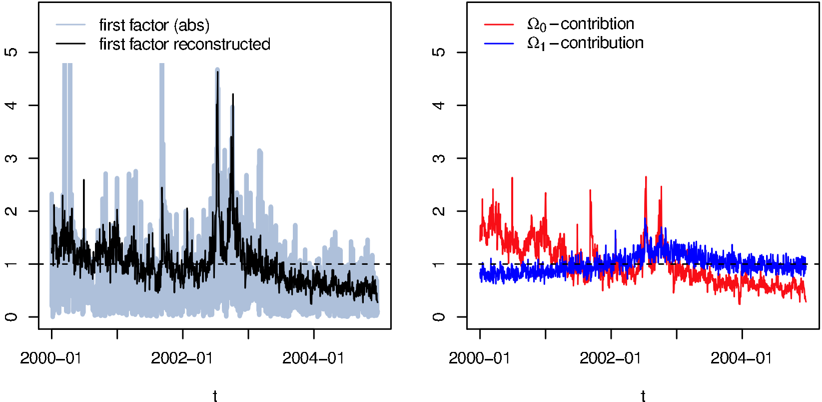
2000-2004



1st factor of the model: $|f_1| = |\epsilon_1| \, e^{A_{11}\omega_1} \, e^{A_{10}\Omega_0}$

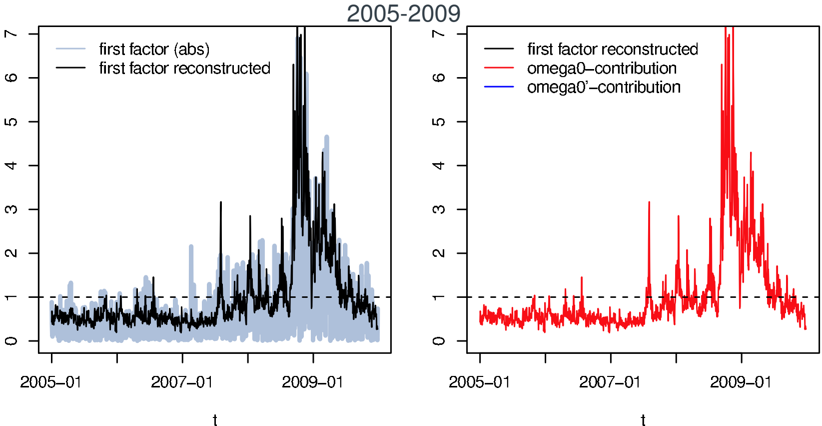
Stock index volatility: $\langle I(t)^2 \rangle \approx \sigma(t)^2 \rho(t)$

2000-2004



1st factor of the model: $|f_1| = |\epsilon_1|\, e^{A_{11}\omega_1}\, e^{A_{10}\Omega_0}$

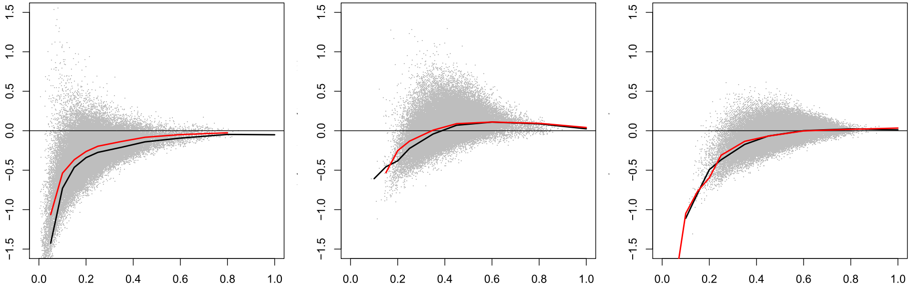
Stock index volatility: $\langle I(t)^2 \rangle \approx \sigma(t)^2 \rho(t)$



Initial motivation: excess probabilities

$$p_{ij} = \text{Prob}[x_i < 0 \text{ and } x_j < 0] - 1/4$$

$$\log[\arg\sin(2\pi p_{ij})/\rho_{ij}] \text{ vs } \rho_{ij}$$



Horizontal: elliptical copulas
Black: non-parametric fit
Red: model prediction

Take-home messages

- Stock-returns exhibit non-trivial cross-sectional non-linear dependences
- Factor models allow to account for these fine-structure effects…
- …provided factors and residuals are orthogonal but not independent
- A common mode of log-vol Ω_0 affecting all factors and residuals
- The residual log-vol of the market factor ω_1 affecting all stocks' residuals
- minimal extension of factor models = intuitive ($\neq$ abstract copulas)

Take-home messages

- Stock-returns exhibit non-trivial cross-sectional non-linear dependences
- Factor models allow to account for these fine-structure effects. . .
- . . . provided factors and residuals are orthogonal but not independent
- A common mode of log-vol Ω_0 affecting all factors and residuals
- The residual log-vol of the market factor ω_1 affecting all stocks' residuals
- minimal extension of factor models = intuitive ($\neq$ abstract copulas)

- Stock-returns exhibit non-trivial cross-sectional non-linear dependences
- Factor models allow to account for these fine-structure effects. . .
- . . . provided factors and residuals are orthogonal but not independent
- A common mode of log-vol Ω_0 affecting all factors and residuals
- The residual log-vol of the market factor ω_1 affecting all stocks' residuals
- minimal extension of factor models = intuitive ($\neq$ abstract copulas)

- Stock-returns exhibit non-trivial cross-sectional non-linear dependences
- Factor models allow to account for these fine-structure effects...
- ...provided factors and residuals are orthogonal but not independent
- A common mode of log-vol Ω_0 affecting all factors and residuals
- The residual log-vol of the market factor ω_1 affecting all stocks' residuals
- minimal extension of factor models = intuitive ($\neq$ abstract copulas)

Take-home messages

- Stock-returns exhibit non-trivial cross-sectional non-linear dependences
- Factor models allow to account for these fine-structure effects. . .
- . . . provided factors and residuals are orthogonal but not independent
- A common mode of log-vol Ω_0 affecting all factors and residuals
- The residual log-vol of the market factor ω_1 affecting all stocks' residuals
- minimal extension of factor models = intuitive ($\neq$ abstract copulas)

Take-home messages

- Stock-returns exhibit non-trivial cross-sectional non-linear dependences
- Factor models allow to account for these fine-structure effects...
- ...provided factors and residuals are orthogonal but not independent
- A common mode of log-vol Ω_0 affecting all factors and residuals
- The residual log-vol of the market factor ω_1 affecting all stocks' residuals
- minimal extension of factor models $=$ intuitive ($\neq$ abstract copulas)

INSIGHT.DATA.CLARITY.

- Jean-Philippe Bouchaud and Marc Potters.
  Theory of Financial Risk and Derivative Pricing: from Statistical Physics to Risk Management.
  Cambridge University Press, Cambridge, 2003.
- Rémy Chicheportiche and Jean-Philippe Bouchaud.
  The joint distribution of stock returns is not elliptical.
  International Journal of Theoretical and Applied Finance, 15(3):1250019–1250041, May 2012.
- Rémy Chicheportiche and Jean-Philippe Bouchaud.
  A nested factor model for non-linear dependencies in stock returns.
  Quantitative Finance, 15(11):1789–1804, 2015.

Table: Economic sectors according to Bloomberg classification, with corresponding number of individuals for each period.

Bloomberg sector	Code	2000–04	2005–09	2000–09
Communications	# 3	33	25	18
Consumer, Cyclical	# 4	60	49	40
Consumer, Non-Cyclical	# 5	67	75	53
Energy	# 7	19	21	15
Financial	# 8	57	55	37
Industrial	#11	51	50	42
Technology	#13	38	43	33
Utilities	#14	27	27	24
Total number of firms (N)		352	345	262
Total number of days (T)		1255	1258	2514

It is convenient to introduce the function

$$\Phi_l(a,b) = \frac{M_{\omega_l}(a+b)}{M_{\omega_l}(a)M_{\omega_l}(b)}$$

where $M_{\omega_l}(p) \equiv E[\exp(p\omega_l)]$ is the Moment Generating Function of ω_l.

ω_l Gaussian for the presentation: $M_{\omega_l}(p) = \exp(p^2/2)$

But in the general case, developping in cumulants, M_{ω_l} is the exponential of a polynomial. Typically, with

$$\langle \omega_l \rangle = 0 \qquad \langle \omega_l^2 \rangle = 1 \qquad \langle \omega_l^3 \rangle = \zeta_l \qquad \langle \omega_l^4 \rangle = 3 + \kappa_l$$

one gets

$$\Phi_l(a,b) = \exp\left(ab + \frac{\zeta_l}{2}(a^2b + ab^2) + \frac{\kappa_l}{12}(2a^3b + 3a^2b^2 + 2ab^3)\right)$$

It is convenient to introduce the function

$$\Phi_l(a,b) = \frac{M_{\omega_l}(a+b)}{M_{\omega_l}(a)M_{\omega_l}(b)}$$

where $M_{\omega_l}(p) \equiv E[\exp(p\omega_l)]$ is the Moment Generating Function of ω_l.

ω_l Gaussian for the presentation: $M_{\omega_l}(p) = \exp(p^2/2)$

But in the general case, developping in cumulants, M_{ω_l} is the exponential of a polynomial. Typically, with

$$\langle \omega_l \rangle = 0 \qquad \langle \omega_l^2 \rangle = 1 \qquad \langle \omega_l^3 \rangle = \zeta_l \qquad \langle \omega_l^4 \rangle = 3 + \kappa_l$$

one gets

$$\Phi_l(a,b) = \exp\left(ab + \frac{\zeta_l}{2}(a^2b + ab^2) + \frac{\kappa_l}{12}(2a^3b + 3a^2b^2 + 2ab^3)\right)$$

It is convenient to introduce the function

$$\Phi_l(a,b) = \frac{M_{\omega_l}(a+b)}{M_{\omega_l}(a)M_{\omega_l}(b)}$$

where $M_{\omega_l}(p) \equiv E[\exp(p\omega_l)]$ is the Moment Generating Function of ω_l.

ω_l Gaussian for the presentation: $M_{\omega_l}(p) = \exp(p^2/2)$

But in the general case, developping in cumulants, M_{ω_l} is the exponential of a polynomial. Typically, with

$$\langle \omega_l \rangle = 0 \qquad \langle \omega_l^2 \rangle = 1 \qquad \langle \omega_l^3 \rangle = \zeta_l \qquad \langle \omega_l^4 \rangle = 3 + \kappa_l$$

one gets

$$\Phi_l(a,b) = \exp\left(ab + \frac{\zeta_l}{2}(a^2b + ab^2) + \frac{\kappa_l}{12}(2a^3b + 3a^2b^2 + 2ab^3)\right)$$

Appendix: Technicalities

It is convenient to introduce the function

$$\Phi_l(a,b) = \frac{M_{\omega_l}(a+b)}{M_{\omega_l}(a)M_{\omega_l}(b)}$$

where $M_{\omega_l}(p) \equiv E[\exp(p\omega_l)]$ is the Moment Generating Function of ω_l.

ω_l Gaussian for the presentation: $M_{\omega_l}(p) = \exp(p^2/2)$

But in the general case, developping in cumulants, M_{ω_l} is the exponential of a polynomial. Typically, with

$$\langle\omega_l\rangle = 0 \qquad \langle\omega_l^2\rangle = 1 \qquad \langle\omega_l^3\rangle = \zeta_l \qquad \langle\omega_l^4\rangle = 3 + \kappa_l$$

one gets

$$\Phi_l(a,b) = \exp\left(ab + \frac{\zeta_l}{2}(a^2b + ab^2) + \frac{\kappa_l}{12}(2a^3b + 3a^2b^2 + 2ab^3)\right)$$

Similarly, the quantity

$$ca(p) = \frac{E\left[|\epsilon|^{2p}\right]}{E\left[|\epsilon|^{p}\right]^2} = \sqrt{\pi}\frac{\Gamma(\frac{1}{2}+p)}{\Gamma(\frac{1+p}{2})^2}$$

stands for the normalized d-moment of the abs of Gaussian variables.
The log version will be used in the following

$$\gamma(p) = \frac{1}{p^2}\log ca(p),$$

f.ex. $\gamma(2) = \log(3)/4$.

Factors and residuals: non-linear

Keep in mind:

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k &= \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k) \\ e_j &= \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

Factor-Factor:

$$\frac{E\left[|f_k|^p |f_l|^p\right]}{E\left[|f_k|^p\right] E\left[|f_l|^p\right]} = \Phi_0(pA_{k0}, pA_{l0}) \Big(ca(p)\Phi_k(ps_k, ps_k) \Big)^{\delta_{kl}} \tag{4}$$

Factor-Residual:

$$\frac{E\left[|f_k|^p |e_i|^p\right]}{E\left[|f_k|^p\right] E\left[|e_i|^p\right]} = \Phi_0(pA_{k0}, pB_{i0}) \Phi_1(pA_{11}, pB_{i1})^{\delta_{k1}} \tag{5}$$

Residual-Residual:

$$\frac{E\left[|e_i|^p |e_j|^p\right]}{E\left[|e_i|^p\right] E\left[|e_j|^p\right]} = \Phi_0(pB_{i0}, pB_{j0}) \Phi_1(pB_{i1}, pB_{j1}) \Big(ca(p)\Phi_\infty(p\widetilde{s}_i, p\widetilde{s}_i) \Big)^{\delta_{ij}} \tag{6}$$

Keep in mind:

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k &= \epsilon_k \exp(A_{k0}\Omega_0 + s_k\omega_k) \\ e_j &= \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

Factor-Factor:

$$\frac{E\left[|f_k|^p |f_l|^p\right]}{E\left[|f_k|^p\right] E\left[|f_l|^p\right]} = \Phi_0(pA_{k0}, pA_{l0}) \Big(ca(p)\Phi_k(ps_k, ps_k)\Big)^{\delta_{kl}} \tag{4}$$

Factor-Residual:

$$\frac{E\left[|f_k|^p |e_i|^p\right]}{E\left[|f_k|^p\right] E\left[|e_i|^p\right]} = \Phi_0(pA_{k0}, pB_{i0})\Phi_1(pA_{11}, pB_{i1})^{\delta_{k1}} \tag{5}$$

Residual-Residual:

$$\frac{E\left[|e_i|^p |e_j|^p\right]}{E\left[|e_i|^p\right] E\left[|e_j|^p\right]} = \Phi_0(pB_{i0}, pB_{j0})\Phi_1(pB_{i1}, pB_{j1}) \Big(ca(p)\Phi_\infty(p\widetilde{s}_i, p\widetilde{s}_i)\Big)^{\delta_{ij}} \tag{6}$$

Factors and residuals: non-linear

Keep in mind:

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k & = \epsilon_k \exp(A_{k0}\Omega_0 + s_k \omega_k) \\ e_j & = \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

Factor-Factor:

$$\frac{E\left[|f_k|^p |f_l|^p\right]}{E\left[|f_k|^p\right] E\left[|f_l|^p\right]} = \Phi_0(pA_{k0}, pA_{l0}) \Big(ca(p) \Phi_k(ps_k, ps_k) \Big)^{\delta_{kl}} \tag{4}$$

Factor-Residual:

$$\frac{E\left[|f_k|^p |e_i|^p\right]}{E\left[|f_k|^p\right] E\left[|e_i|^p\right]} = \Phi_0(pA_{k0}, pB_{i0}) \Phi_1(pA_{11}, pB_{i1})^{\delta_{k1}} \tag{5}$$

Residual-Residual:

$$\frac{E\left[|e_i|^p |e_j|^p\right]}{E\left[|e_i|^p\right] E\left[|e_j|^p\right]} = \Phi_0(pB_{i0}, pB_{j0}) \Phi_1(pB_{i1}, pB_{j1}) \Big(ca(p) \Phi_\infty(p\widetilde{s}_i, p\widetilde{s}_i) \Big)^{\delta_{ij}} \tag{6}$$

Factors and residuals: non-linear

Keep in mind:

$$x_j = \sum_{k=1}^{M} \beta_{kj} f_k + e_j \quad \text{with} \quad \begin{cases} f_k &= \epsilon_k \exp(A_{k0}\Omega_0 + s_k\omega_k) \\ e_j &= \eta_j \exp(B_{j0}\Omega_0 + B_{j1}\omega_1 + \widetilde{s}_{jj}\tilde{\omega}_j) \end{cases}$$

Factor-Factor:

$$\frac{E\left[|f_k|^p |f_l|^p\right]}{E\left[|f_k|^p\right] E\left[|f_l|^p\right]} = \Phi_0(pA_{k0}, pA_{l0})\Big(ca(p)\Phi_k(ps_k, ps_k)\Big)^{\delta_{kl}} \tag{4}$$

Factor-Residual:

$$\frac{E\left[|f_k|^p |e_i|^p\right]}{E\left[|f_k|^p\right] E\left[|e_i|^p\right]} = \Phi_0(pA_{k0}, pB_{i0})\Phi_1(pA_{11}, pB_{i1})^{\delta_{k1}} \tag{5}$$

Residual-Residual:

$$\frac{E\left[|e_i|^p |e_j|^p\right]}{E\left[|e_i|^p\right] E\left[|e_j|^p\right]} = \Phi_0(pB_{i0}, pB_{j0})\Phi_1(pB_{i1}, pB_{j1})\Big(ca(p)\Phi_\infty(p\widetilde{s}_i, p\widetilde{s}_i)\Big)^{\delta_{ij}} \tag{6}$$

Quadratic correlations

$$E[x_i^2 x_j^2] = \sum_{kl} \left(\beta_{ki}^2 \beta_{lj}^2 + 2\beta_{ki}\beta_{kj}\beta_{li}\beta_{lj}\right) \Phi_0(2A_{k0}, 2A_{l0}) \left(\tfrac{1}{3} \cdot 3 \cdot \Phi_k(2s_k, 2s_k)\right)^{\delta_{kl}}$$
$$+ (1 + 2\delta_{ij}) \left(1 - \sum_l \beta_{li}^2\right) \sum_k \beta_{kj}^2 \Phi_0(2A_{k0}, 2B_{i0}) \Phi_1(2A_{11}, 2B_{i1})^{\delta_{k1}}$$
$$+ (1 + 2\delta_{ij}) \left(1 - \sum_l \beta_{lj}^2\right) \sum_k \beta_{ki}^2 \Phi_0(2A_{k0}, 2B_{j0}) \Phi_1(2A_{11}, 2B_{j1})^{\delta_{k1}}$$
$$+ \left(1 - \sum_l \beta_{li}^2\right) \left(1 - \sum_l \beta_{lj}^2\right) \Phi_0(2B_{i0}, 2B_{j0}) \Phi_1(2B_{i1}, 2B_{j1}) \left(3\Phi_\infty(2\widetilde{s}_i, 2\widetilde{s}_i)\right)^{\delta_{ij}}$$

When all the A's and B's are zero, we get back the usual Gaussian prediction

$$E[x_i^2 x_j^2] - 1 = \begin{cases} 2\left(\beta^\dagger \beta\right)_{ij}^2 & , i \neq j \\ 2 & , i = j \end{cases} = 2\, E[x_i x_j]^2$$

$$E[x_i^2 x_j^2] = \sum_{kl} \left(\beta_{ki}^2 \beta_{lj}^2 + 2\beta_{ki}\beta_{kj}\beta_{li}\beta_{lj}\right) \Phi_0(2A_{k0}, 2A_{l0}) \left(\tfrac{1}{3} \cdot 3 \cdot \Phi_k(2s_k, 2s_k)\right)^{\delta_{kl}}$$
$$+ (1 + 2\delta_{ij}) \left(1 - \sum_l \beta_{li}^2\right) \sum_k \beta_{kj}^2 \Phi_0(2A_{k0}, 2B_{i0}) \Phi_1(2A_{11}, 2B_{i1})^{\delta_{k1}}$$
$$+ (1 + 2\delta_{ij}) \left(1 - \sum_l \beta_{lj}^2\right) \sum_k \beta_{ki}^2 \Phi_0(2A_{k0}, 2B_{j0}) \Phi_1(2A_{11}, 2B_{j1})^{\delta_{k1}}$$
$$+ \left(1 - \sum_l \beta_{li}^2\right) \left(1 - \sum_l \beta_{lj}^2\right) \Phi_0(2B_{i0}, 2B_{j0}) \Phi_1(2B_{i1}, 2B_{j1}) \left(3\Phi_\infty(2\widetilde{s}_i, 2\widetilde{s}_i)\right)^{\delta_{ij}}$$

When all the A's and B's are zero, we get back the usual Gaussian prediction

$$E[x_i^2 x_j^2] - 1 = \begin{cases} 2\left(\beta^\dagger \beta\right)_{ij}^2 & , i \neq j \\ 2 & , i = j \end{cases} = 2\, E[x_i x_j]^2$$