

Approximating Correlation Matrices

Mathematical Problems in Industry Workshop 2012

William Morokoff
Managing Director
Quantitative Analytics Research Group

June 11, 2012



Roadmap for Correlation Discussion

Themes and Summary of Key Points	
Standard & Poor's Overview	<ul style="list-style-type: none"> • Company profile • Quantitative analytics at S&P
Portfolio Modeling and Correlation	<ul style="list-style-type: none"> • Modeling dependence with correlation
Representing Correlation: Factor Models	<ul style="list-style-type: none"> • Properties of correlation matrices • Empirical correlation estimates • Block correlation representation • Factor models • Optimal factor model approximations
Localized Factor Models	<ul style="list-style-type: none"> • Localized Factor Models • Best approximation of localized one factor models • Semi-analytic methods for localized one factor models
Open Research Questions	<ul style="list-style-type: none"> • Optimal approximations • Computational methods

S&P Overview

S&P Overview

- **Standard and Poor's is a leading provider of independent credit analysis, and a key source for financial market intelligence.**
- **Best known for S&P 500 Index and Credit Ratings**
- **Wholly owned subsidiary of McGraw-Hill**
- **About 10,000 employees worldwide with around \$2.9 Billion in combined revenue for Standard & Poor's Ratings Services and S&P Capital IQ.**
- **Brands include Compustat, CUSIP, Capital IQ, ClariFI, Risk Solutions, GICS, RatingsDirect, IMAKE, R2**

Role of Quantitative Analytics Research Group at S&P

- **Mandate:** Support S&P Ratings Services and S&P Capital IQ with quantitative expertise to grow the businesses and develop products and services that enhance S&P's position as a respected voice in the capital markets.
- **Priorities:**
 - Quantitative Support for Ratings Methodology
 - Model Development
 - Quality and Efficiency
 - Quantitative Strategic Vision across Businesses
 - Thought Leadership
 - Business Advisory and Training

Portfolio Modeling and Correlation

Credit Portfolio Modeling

$$\Pi(t) = \sum_{i=1}^N \omega_i V_i(t)$$

$$R_{\Pi}(T_0, T_H) = \sum_{i=1}^N \hat{\omega}_i R_i(T_0, T_H)$$

$$EL = E(R_{ref} - R_{\Pi})$$

$$UL = \sigma(R_{ref} - R_{\Pi})$$

$$P(R_{\Pi} < R^*(\alpha)) = \alpha$$

Credit Portfolio Modeling

- Distribution of returns for individual exposures is a key component of determining the portfolio return distribution.
- However, the joint dependence of returns is crucial for determining the portfolio potential for large losses.
- The joint behavior is often modeled by normalizing the distribution of individual returns and estimating a multi-variate distribution for the normalized returns – often referred to as latent variables or asset returns.

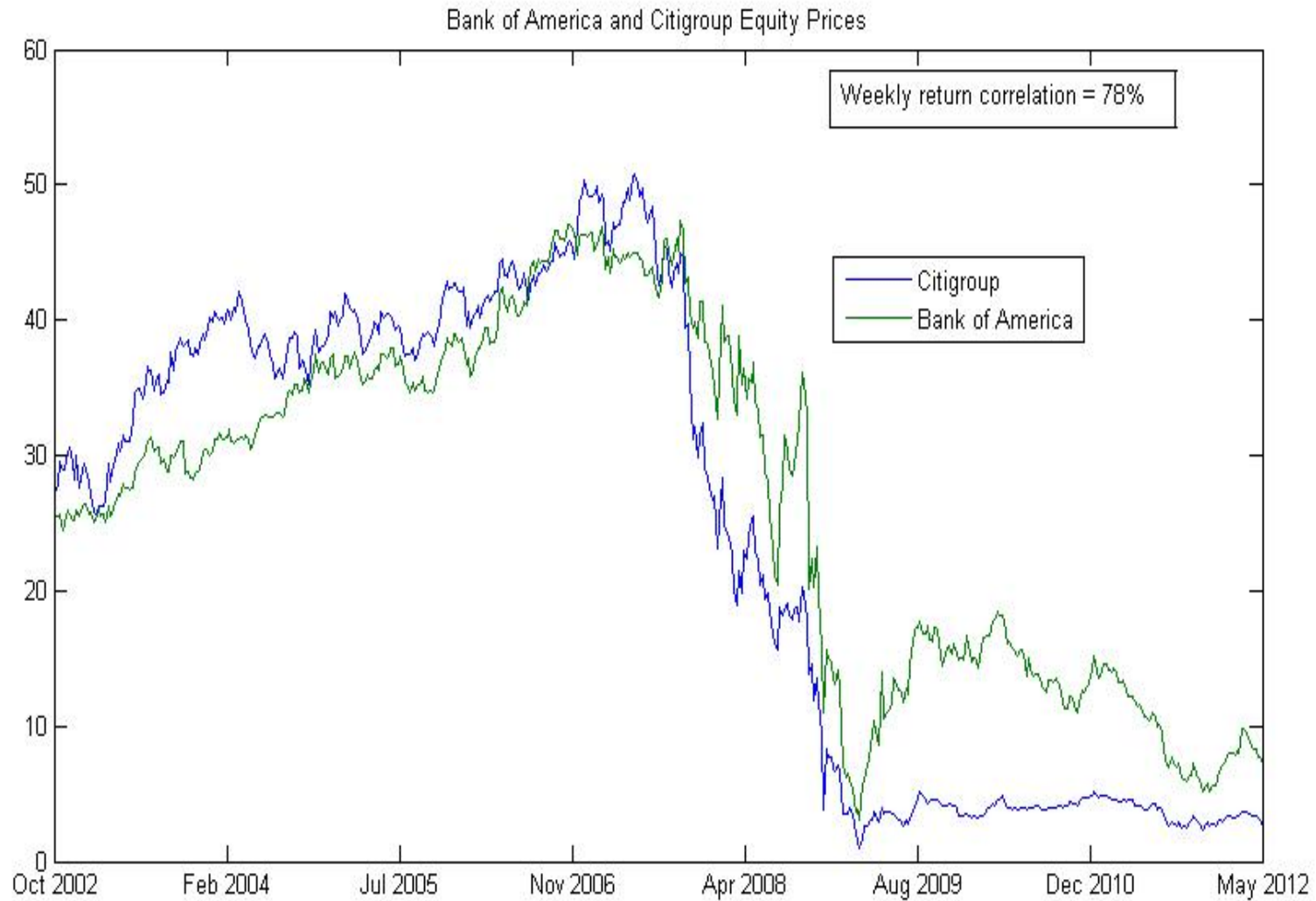
$$z_i = \Phi^{-1} \left(F_i (R_i) \right)$$

- Specifying a correlation matrix (symmetric, pos def with ones on diagonal)

$$P = E \left(z z^T \right)$$

or more generally a copula function determines the dependence of the asset returns.

Equity Return Correlation Example



Permission to reprint or distribute any content from this presentation requires the prior written approval of Standard & Poor's.

Representing Correlation: Factor Models

Correlation Matrix Properties

- **Symmetric**
- **Positive Definite**
- **Ones on the diagonal**
- **Off-diagonal elements represent pair-wise correlation of two assets in the portfolio**
- **Correlations are usually positive – particularly for firm asset returns in credit modeling**
- **Example: This doesn't work:**

$$\begin{bmatrix} 1 & .1 & .9 \\ .1 & 1 & .9 \\ .9 & .9 & 1 \end{bmatrix}$$

Empirical Correlation Estimates

- If time series of 'asset returns' is available or can be proxied by equity returns or other data, pair-wise correlations can be estimated.
- Estimate of full correlation matrix may be somewhat unstable due to missing data and large number of correlation parameters being estimated – $N*(N-1)/2$ for N firms.
- For large portfolios, describing dependence through a full correlation matrix is impractical.

Factor Models

- **Gaussian Copula Factor Model – Standard Model**

- Normalized ‘Asset Return’ to Horizon is modeled as a standard Normal random variable.
- Asset return is decomposed into systematic risk component and idiosyncratic risk component.
- Percentage of variance related to systematic risk is ‘R-squared’.
- Systematic risk is described by one or more independent standard normal variates, common to all exposures. Each exposure is assigned a set of weights on the factors. The idiosyncratic risk is modeled as standard normal, independent of all factors and other exposures.

$$z_i = \sqrt{\rho_i} \beta_i^T \varepsilon_F + \sqrt{1 - \rho_i} \varepsilon_{I,i} \quad \beta_i^T \beta_i = 1$$

$$\rho_{ij} = \sqrt{\rho_i \rho_j} \beta_i^T \beta_j$$

Factor Models

- In matrix notation, the factor model is

$$z = \Gamma^{1/2} B \epsilon_F + [I - \Gamma]^{1/2} \epsilon_I$$

- The correlation matrix can be represented as

$$P = \Gamma^{1/2} B B^T \Gamma^{1/2} + I - \Gamma$$

Block Correlation Matrix

$$P = \begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{1n}^T & \cdots & P_{nn} \end{bmatrix} \quad P_R = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1n} \\ \vdots & \ddots & \vdots \\ \rho_{1n} & \cdots & \rho_{nn} \end{bmatrix}$$

$$P_{ii} = (1 - \rho_{ii}) \mathbf{I}_{N_i} + \rho_{ii} \mathbf{E}_{N_i} \quad N_i \times N_i$$

$$P_{ij} = \rho_{ij} \mathbf{E}_{N_i \times N_j} \quad N_i \times N_j$$

The reduced correlation matrix P_R is not necessarily positive definite

$$P_R^* = P_R + \text{diag} \left(\left[\frac{1 - \rho_{11}}{N_1}, \dots, \frac{1 - \rho_{nn}}{N_n} \right] \right) > 0$$

Block Correlation Matrix – Factor Model Representation

- If exposures are classified into groups based on sector, geography, size, etc., and correlations between all members of two groups are assumed the same, the resulting correlation matrix has a block structure.
- The n diagonal blocks are correlation matrices with a constant correlation in each block. Each block has N_i exposures, $i = 1 \dots n$.
- The off-diagonal blocks have identical elements within a block.
- In order to reduce the computation burden of working with large matrices, block correlation matrices can be factored as:

$$z_{ij} = \sqrt{\rho_i + (1 - \rho_i) / N_i} \beta_i^T \varepsilon_F + \sqrt{1 - \rho_i} (\varepsilon_{ij} - \bar{\varepsilon}_i) \quad i = 1 \dots n, \quad j = 1 \dots N_i$$

$$\beta_i^T \beta_i = 1 \quad \beta_i - n \times 1 \text{ vector}$$

ε_F – $n \times 1$ vector of independent standard Normal variates

ε_{ij} – iid standard Normal variates

$$\bar{\varepsilon}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \varepsilon_{ij}$$

Permission to reprint or distribute any content from this presentation requires the prior written approval of Standard & Poor's.

Optimal Factor Model Approximation

- Andersen, Sidenius and Basu (Risk, Nov 2003) show how to find the best K factor approximation to a given correlation matrix P in the sense of minimizing the matrix difference in the Frobenius norm through principal component analysis (PCA).
- Given an estimate of the R-squared diagonal matrix, carry out the following iteration:

$$Q_i D_i Q_i^T = P - I + \Gamma_i$$

$$\Gamma_{i+1} = \text{diag} \left(Q_i D_i^{(K)} Q_i^T \right)$$

$$\Gamma_i \rightarrow \Gamma_{opt}^{(K)}$$

$$B_{opt}^{(K)} = \left(\Gamma_{opt}^{(K)} \right)^{-1/2} Q_{opt}^{(K)} \sqrt{D_{opt}^{(K)}}$$

Localized Factor Models

Localized One Factor Model

- Assume that there are one global factor and K sector factors.
- A 'localized one factor model' is a factor model where each exposure weights on the global factor and exactly one sector factor.

$$z_i = \sqrt{\rho_i} \left(\beta_i^0 \varepsilon_0 + \beta_i \varepsilon_{k(i)} \right) + \sqrt{1 - \rho_i} \phi_i$$

$$\left(\beta_i^0 \right)^2 + \left(\beta_i \right)^2 = 1$$

- If all exposures in one sector have the same R-squared value and same factor loadings, then this corresponds to a block correlation matrix. However, not every block correlation matrix can be expressed as a localized one factor model.

Localized Two Factor Model

- Assume that there are one global factor and K sector factors and J country factors for a total of (K+J+1) factors.
- A 'localized two factor model' is a factor model where each exposure weights on the global factor, exactly one sector factor and exactly one country factor.

$$z_i = \sqrt{\rho_i} \left(\beta_i^0 \varepsilon_0 + \beta_i^S \varepsilon_{k(i)}^S + \beta_i^C \varepsilon_{j(i)}^C \right) + \sqrt{1 - \rho_i} \phi_i$$

$$\left(\beta_i^0 \right)^2 + \left(\beta_i^S \right)^2 + \left(\beta_i^C \right)^2 = 1$$

- If all exposures in one sector have the same R-squared value and same factor loadings, then this corresponds to a block correlation matrix. However, not every block correlation matrix can be expressed as a localized two factor model.

Optimal Localized One Factor Approximation

- For a block correlation matrix P with K diagonal blocks and homogeneous correlations within each block, let ρ_k be the diagonal block correlations and ρ_{jk} be the off-diagonal block correlations.
- For each exposure in sector k ($k = 1 \dots K$), set the R-squared value to ρ_k
- Construct the correlation matrix corresponding to the ρ_{jk} . Find the best single factor approximation using PCA technique. Set

$$\beta_k^0 = \sqrt{\frac{\Gamma_{opt,k}^{(1)}}{\rho_k}} \quad \beta_k = \sqrt{1 - (\beta_k^0)^2}$$

- If reduced correlation matrix is not positive definite, need to modify.

Semi-Analytic Methods for Localized One Factor Model

- For a portfolio with exposures in K sectors, the total loss is

$$L = L_1 + \dots + L_K$$

- Conditional on the global factor, these sector losses are independent, and each sector is described by a single factor model, so the usual semi-analytic methods can be applied on a sector-by-sector basis:

$$P(L) = E_{\varepsilon_0} \left[\sum_{L_1 + \dots + L_K = L} P(L_1, \dots, L_K | \varepsilon_0) \right]$$

$$P(L_1, \dots, L_K | \varepsilon_0) = \prod_{k=1}^K P(L_k | \varepsilon_0)$$

$$P(L_k | \varepsilon_0) = E_{\varepsilon_k} \left[P(L_k | \varepsilon_0, \varepsilon_k) \right]$$

- **Alternative: Saddle Point Approximation**

Open Research Questions

Optimal Approximations

- For a block correlation matrix, how does the k factor PCA optimal approximation compare with the block correlation factor model representation reduced to k factors?
- For a block correlation matrix, how does the k factor PCA optimal approximation compare with the optimal localized one-factor approximation?
- Can a better localized one-factor model approximation be found that incorporates the number of assets per group?
- How can the optimal localized two factor model be determined?
- Can optimal be measured in terms of portfolio risk metrics as opposed to measured in a matrix norm?

Computational Methods

- **For a localized one factor approximation, what are the challenges and limitations of implementing a semi-analytic numerical approach? How efficient is this approach relative to a full Monte Carlo simulation?**
- **To what extent is it possible to adapt saddle point approximation methods to localized one factor models?**

STANDARD & POOR'S RATINGS SERVICES

www.standardandpoors.com

Copyright © 2012 by Standard & Poor's Financial Services LLC. All rights reserved.

No content (including ratings, credit-related analyses and data, model, software or other application or output therefrom) or any part thereof (Content) may be modified, reverse engineered, reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of Standard & Poor's Financial Services LLC or its affiliates (collectively, S&P). The Content shall not be used for any unlawful or unauthorized purposes. S&P and any third-party providers, as well as their directors, officers, shareholders, employees or agents (collectively S&P Parties) do not guarantee the accuracy, completeness, timeliness or availability of the Content. S&P Parties are not responsible for any errors or omissions (negligent or otherwise), regardless of the cause, for the results obtained from the use of the Content, or for the security or maintenance of any data input by the user. The Content is provided on an "as is" basis. S&P PARTIES DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE, FREEDOM FROM BUGS, SOFTWARE ERRORS OR DEFECTS, THAT THE CONTENT'S FUNCTIONING WILL BE UNINTERRUPTED OR THAT THE CONTENT WILL OPERATE WITH ANY SOFTWARE OR HARDWARE CONFIGURATION. In no event shall S&P Parties be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs or losses caused by negligence) in connection with any use of the Content even if advised of the possibility of such damages.

Credit-related and other analyses, including ratings, and statements in the Content are statements of opinion as of the date they are expressed and not statements of fact. S&P's opinions, analyses and rating acknowledgment decisions (described below) are not recommendations to purchase, hold, or sell any securities or to make any investment decisions, and do not address the suitability of any security. S&P assumes no obligation to update the Content following publication in any form or format. The Content should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. S&P does not act as a fiduciary or an investment advisor except where registered as such. While S&P has obtained information from sources it believes to be reliable, S&P does not perform an audit and undertakes no duty of due diligence or independent verification of any information it receives.

To the extent that regulatory authorities allow a rating agency to acknowledge in one jurisdiction a rating issued in another jurisdiction for certain regulatory purposes, S&P reserves the right to assign, withdraw or suspend such acknowledgement at any time and in its sole discretion. S&P Parties disclaim any duty whatsoever arising out of the assignment, withdrawal or suspension of an acknowledgment as well as any liability for any damage alleged to have been suffered on account thereof.

S&P keeps certain activities of its business units separate from each other in order to preserve the independence and objectivity of their respective activities. As a result, certain business units of S&P may have information that is not available to other S&P business units. S&P has established policies and procedures to maintain the confidentiality of certain non-public information received in connection with each analytical process.

S&P may receive compensation for its ratings and certain analyses, normally from issuers or underwriters of securities or from obligors. S&P reserves the right to disseminate its opinions and analyses. S&P's public ratings and analyses are made available on its Web sites, www.standardandpoors.com (free of charge), and www.ratingsdirect.com and www.globalcreditportal.com (subscription), and may be distributed through other means, including via S&P publications and third-party redistributors. Additional information about our ratings fees is available at www.standardandpoors.com/usratingsfees.

STANDARD & POOR'S, S&P, GLOBAL CREDIT PORTAL and RATINGSDIRECT are registered trademarks of Standard & Poor's Financial Services LLC.