STANDARD
&POOR'S
RATINGS SERVICES

Mathematical Problems in Industry Workshop
University of Delaware
June 2012

Approximating Correlation Matrices

06 2012

William Morokoff
Managing Director
Quantitative Analytics
212 438 4828
william_morokoff@sandp.com

Liming Yang
Senior Director
Quantitative Analytics
212 438 1870
liming_yang@sandp.com

Saiyid Islam
Director
Quantitative Analytics
301 634 4284
saiyid_islam@sandp.com

Abstract

We pose here several questions related to modeling the joint dependence structure of financial assets in a portfolio. Portfolio models intended to describe properties such as the concentration of risk or the probability of large losses depend directly on the joint behavior of assets in the portfolio, often described by a correlation matrix. Approximating a correlation matrix with a simple underlying structure can lead to faster computation times and clearer economic interpretation. We consider here how to find the optimal approximation, how to measure the impact of the approximation, how to effectively implementation fast algorithms based on the approximation, and other questions related to the approximation.

***Disclaimer*:** The mathematical framework and questions posed here represent a research project intended to study the advantages and limitations of certain economically intuitive approximations to more general correlation structures. No inferences should be made with regard to Standard & Poor's credit ratings or any current or future criteria or models used in the ratings process for credit portfolios or any type of financial security.

1 Introduction

Assessing the potential for gains and losses on portfolios of financial assets as well as understanding the drivers of realized returns are important parts of performance tracking and risk management operations for banks, asset managers, insurance companies, etc. Understanding systematic exposure to macro-economic, regional and sector factors across the portfolio helps explain concentration risks and predict the potential for large losses. Banks often use the modeling of portfolio loss distributions to help establish capital requirements to survive periods of extreme credit stress, while portfolio managers employ similar models to track the performance of their portfolios against benchmark portfolios.

One common approach to portfolio return modeling is to combine a process for changes in individual exposure returns with a dependence structure across exposures that captures the joint evolution. In some models, dependence structure is represented by a normalized Gaussian distribution (or other multivariate distribution) used to model joint changes in the underlying value of the exposures; such distributions are often characterized by a correlation matrix - a symmetric positive definite matrix with ones on the diagonal.

A general correlation matrix is often derived empirically from historical asset returns, and working directly with pair-wise correlations provides a framework for stress testing views in a natural way - e.g., assessing the impact of in an increase in correlation among U.S. banks. However, for large portfolios with tens or hundreds of thousands of exposures, working directly with an explicit correlation matrix of extremely large dimension is impractical, particularly as a Monte Carlo simulation is often required to compute the portfolio return distribution. Hence approximating a correlation matrix with a much lower dimensional factor model structure is of interest.

The solution to the question of how to optimally approximate a given correlation matrix with a fixed number of factors is known. However, the solution is expressed in terms of principal components of the matrix and does not necessarily have an intuitive economic interpretation. If the original correlation matrix has a block structure (with constant correlation within a block), it is also known how to exactly represent the matrix in terms of a factorization into much lower dimensional components; again, however, the interpretation of the decomposition is not intuitive.

We would like to study the estimation of the optimal factor loadings when a specified factor structure is imposed. Specifically, we consider 'localized one or two factor models'. This means that each exposure in the portfolio can weight on a shared global factor and on either one or two additional factors that would typically represent the sector and/or region of the exposure. For example, there may be ten possible sectors, leading to an eleven factor model (one global and one for each sector), but any specific exposure would only have weight on the global factor and one of the sector factors.

Assuming we start with either a general correlation matrix or a block correlation matrix, questions of interest are:

1) How can the optimal localized one or two factor model be computed?

2) Can bounds for the impact of the approximation error on portfolio risk measures be estimated?

Another related question arises when we are given a sector correlation matrix and a regional correlation matrix. Here we would like a method to estimate a localized two factor model that is consistent with the given sector and regional correlation matrices.

2 The Factor Model Framework

A basic question for risk management of portfolios of financial assets is to describe the distribution of potential returns on the portfolio over a time horizon T as viewed from today t_0. If the portfolio has N assets, each with a dollar position ω_i and a return over the horizon $R_i(t_0, T)$, then the portfolio return on this period is simply

$$R_\pi(t_0, T) = \sum_{i=1}^{N} \omega_i R_i(t_0, T). \tag{2.1}$$

The return on an individual asset clearly depends on that asset's characteristics. The portfolio return also depends strongly on the joint behavior, i.e. dependence, of the assets on underlying common factors such as macro-economic conditions, regional effects and sector conditions. This dependence across assets is often loosely referred to as correlation. A well-diversified portfolio seeks to minimize the correlation effects, while concentration risks are associated with assets with higher correlations.

For portfolios of equities, a reasonable first cut approximation is to model the set of asset returns as multi-variate Normal, so that the portfolio return distribution is also Normal and determined by the mean vector and covariance matrix of the multi-variate Normal distribution. More realistic properties observed in portfolio performance data such as fat tails and skew can be readily incorporated in various ways, but often a correlation matrix remains a key determinant of the portfolio model.

For credit portfolios (bonds, loans, credit derivatives, etc.), the portfolio return distribution tends to be heavily skewed with limited up-side (with high probability) and potentially very large down-side (with low probability). Often the time horizon for credit portfolios (particularly bank loan portfolios) is much longer than for equity portfolios, and the focus is on potential losses due to default. In the simplest default/no-default model, the assets are assumed at par (hence zero return or loss) if default does not occur, and valued at the recovery price if a default occurs. The portfolio return distribution is then determined by the probability distribution of the number of defaults in the portfolio. The greater the default risk concentration, the higher the probability of multiple defaults, and hence the higher the probability of large losses. More sophisticated credit portfolio models allow for credit migration to more than one state; account for changes in asset value due to credit migration, interest rates, coupon cash flow and options; and incorporate current pricing data into the return calculation. However, at the heart of a credit portfolio model is the need to describe the dependence of changes in asset credit quality on common underlying factors. While the change in credit quality can be modeled in various ways, in general for a

given firm, this variable can be transformed to a standard Normal variate (the transformation is often calibrated to historical data and may be parametric or empirically estimated). Often this transformed variable is interpreted as the firm's asset return, i.e. the return on the fundamental value of the firm. This concept arises in the structural model for debt valuation (see de Servigny and Renault [2004] for a general summary of credit models and Bohn and Stein [2009] for further discussion of credit portfolio models). In the transformed variable space, dependence across firms in often introduced by assuming a multi-variate Normal distribution to describe all the asset returns for the portfolio. More generally a Gaussian or other copula can be applied; for our purposes, we will work with multi-variate Normal asset returns. Since the asset returns are normalized (mean zero, variance one), then the distribution is fully characterized by a correlation matrix.

2.1 Factor Models

While the distributions associated with portfolio returns are often linked to a correlation matrix, when the number of assets in the portfolio is in the tens or hundreds of thousands, it is generally impractical to work directly with a full correlation matrix. Often a more economically intuitive approach is employed in which the underlying factors that drive the pair-wise correlations are explicitly modeled. In this case, each asset return in the portfolio weights on a relatively small number of factors. Generally the number of factors stays fixed as the size of the portfolio increases, so that the computational work and data storage requirements scale linearly with the size of the portfolio. Factors can be identified with global, regional and sector economic effects and are often based on time series of returns of composite indices. The factor distributions are generally standardized and orthogonalized, so that a normalized asset return can be expressed as

$$z_i = \sqrt{\rho_i}\left(\sum_{k=1}^{N_F} \beta_{i,k}\epsilon_{F,k}\right) + \sqrt{1-\rho_i}\epsilon_{I,i}. \tag{2.2}$$

Here N_F is the number of factors and $\epsilon_{F,k}$ are assumed to be independent, identically distributed standard Normal variates represented the standardized, orthogonalized factor returns. The column vector $\vec{\beta_i}$ provides the factor loadings and is normalized so that $\vec{\beta_i}^T\vec{\beta_i} = 1$. The quantity $\epsilon_{I,i}$ represents the idiosyncratic risk not explained by the factors and is assumed independent of the factor returns; generally the $\epsilon_{I,i}$ are also assumed independent across all the assets in the portfolio (although some models also allow for correlation here, we will not consider this). The quantity ρ_i is the percentage of asset return variance associated with systematic risk. Under this modeling framework, the correlation of asset returns between two assets i and$j$ can be expressed as

$$\rho_{ij} = \sqrt{\rho_i\rho_j}\vec{\beta_i}^T\vec{\beta_j}. \tag{2.3}$$

The factor model is specified by providing ρ_i and $\vec{\beta_i}$ for each asset in the portfolio.

In matrix notation, the vector of N asset returns $\vec{z}$ can be expressed as

$$\vec{z} = \Gamma^{1/2}B\vec{\epsilon_F} + (I-\Gamma)^{1/2}\vec{\epsilon_I} \tag{2.4}$$

where Γ is the $N \times N$ diagonal matrix with entries ρ_i on the diagonal and B is an $N \times N_F$ matrix with rows $\vec{\beta_i}^T$. The associated correlation matrix P can then be expressed as

$$P = \Gamma^{1/2}BB^T\Gamma^{1/2} + I - \Gamma. \tag{2.5}$$

2.2 Factor Model Approximations

While the factor model described above implies a specific correlation matrix, not all correlation matrices can be expressed as a factor model of this form with a specified number of factors. The question of the best possible K factor approximation was addressed by Andersen et al. [2003]. In this paper the authors propose an iterative procedure for computing the closest factor model correlation matrix to the given correlation matrix P as measured in the Frobenius norm, thereby minimizing the element-wise square error. The idea is to use the first K principal components of systematic risk portion of correlation matrix, $P-(I-\Gamma)$, as the factors. The iteration is required because Γ is not known. However, starting from an initial guess Γ_0, the iteration

$$Q_i D_i Q_i^T = P-(I-\Gamma_i) \tag{2.6}$$

$$\Gamma_{i+1} = \text{diag}\left(Q_i D_i^{(K)} Q_i^T\right) \tag{2.7}$$

converges to the required solution for Γ. Here QDQ^T is the orthonormal eigenvector decomposition of the symmetric matrix $P-(I-\Gamma)$ ordered from largest to smallest eigenvalues, and $D^{(K)}$ is the clipped diagonal matrix with all diagonal entries beyond the K^{th} set to zero. Once the optimal matrix $\Gamma_{\text{opt}}^{(K)}$ is determined, the corresponding matrix B can be constructed as

$$B_{\text{opt}}^{(K)} = \left(\Gamma_{\text{opt}}^{(K)}\right)^{-1/2} Q_{\text{opt}}^{(K)} \sqrt{D_{\text{opt}}^{(K)}} \tag{2.8}$$

where $Q_{\text{opt}}^{(K)}$ is the $N \times K$ matrix of the first K orthonormal eigenvectors of $P-(I-\Gamma_{\text{opt}}^{(K)})$.

2.3 Block Correlation Matrices

As an alternative to factor models, assets are often grouped together based on various classification combinations including region, sector, size, etc. A single correlation parameter is assigned for every pair of groups, implying the same pair-wise correlation for any pair of assets chosen from the two groups. For two assets within the same group, another correlation is assumed, the same across all pairs within the group. In this model, the number of correlation parameters scales like the square of the number of groups, as opposed to the number of assets, and the asset return correlation matrix assumes a block structure. Although the association of correlations with region and sector has a similar flavor to a factor model, there is not in general a direct representation of a block correlation matrix in the factor model structure. The challenge is that the reduced matrix of correlations, one entry for each group-pair and intra-group correlations on the diagonal, is not necessarily positive definite, even when the original correlation matrix has this property. If the reduced matrix of correlations is positive definite, then a direct extension of 2.2 can be applied, with the number of factors set equal to the number of groups.

However, Huang and Yang [2010] have shown that there is a very similar expression to equation 2.2 for computing asset returns consistent with a given block correlation matrix that works for all situations where the original block correlation matrix is positive definite. The method also allows for fast verification that the original correlation matrix has this property. If we assume that there are n distinct groups, each with N_i assets $(i=1,\ldots,n)$ for a total of $N=N_1+\ldots N_n$ assets in the portfolio, then we can express the asset return of the j^{th} asset in the i^{th} group $(1 \le j \le N_i)$ as

$$z_{i,j} = \sqrt{\rho_i + (1-\rho_i)/N_i}\left(\vec{\beta_i}^T \vec{\epsilon_F}\right) + \sqrt{1-\rho_i}\left(\epsilon_{I,i,j} - \bar{\epsilon_i}\right). \tag{2.9}$$

Here ρ_i is the correlation of assets within group i, $\vec{\beta_i}$ is an $n \times 1$ vector of 'factoring loadings' on the n independently, identically distribution standard Normal factor $\vec{\epsilon_F}$. The $\epsilon_{I,i,j}$ are also iid standard Normal variates and $\bar{\epsilon_i}$ is defined

as

$$\bar{\epsilon_i} = \frac{1}{N_i} \sum_{j=1}^{N_i} \epsilon_{I,i,j}. \tag{2.10}$$

It is easy to check that the 'reduced' $n \times n$ correlation matrix P_g with off-diagonal elements specified by the cross-group correlations (and ones on the diagonal) can be expressed as

$$P_g = \Gamma_g^{1/2} B B^T \Gamma_g^{1/2} + I - \Gamma_g \tag{2.11}$$

where the diagonal matrix Γ_g has elements $\rho_i + (1-\rho_i)/N_i$ and B is a square $n \times n$ matrix with rows $\vec{\beta_i}^T$. A key result of Huang and Yang is that if the original $N \times N$ matrix is a true correlation matrix (i.e. positive definite), then the matrix $P_g - I + \Gamma_g$ must also be positive definite. From equation 2.11 it follows that the symmetric $n \times n$ matrix $P_B = BB^T$ must also be positive definite and satisfies

$$P_B = \Gamma_g^{-1/2} \left(P_g - I + \Gamma_g \right) \Gamma_g^{-1/2}. \tag{2.12}$$

Thus P_B depends only on the intra- and inter-group correlations. The factor loading matrix B can be selected as any factorization of P_B that satisfies $P_B = BB^T$. Thus the quasi-factor model representation is not unique and the factors do not have an economic interpretation. For example, a Cholesky decomposition or an eigenvector decomposition could be used to determine B. The advantage of the quasi-factor model representation 2.9 of a block correlation matrix is that asset return samples for a Monte Carlo simulation can be generated in order $N \times n$ operations as opposed to N^2 operations if the full correlation matrix had to be factored.

An advantage of the block correlation approach of working directly with correlation values is that certain stress scenarios can be more easily expressed than in a factor model framework. For example, it is easy to study the impact of increasing the correlation between the insurance and banking sectors. However, care must be taken when changing individual values within a correlation matrix to ensure that the resulting matrix remains positive definite.

2.4 Semi-Analytic Methods and Single Factor Models

For credit portfolios, the return for each component of the portfolio (bond, loan, etc.) is generally a non-linear function of the associated 'asset return' being modeled by the factor model. Thus the portfolio return is a fairly complex high dimensional non-linear function of the underlying factors in the factor model. For this reason, Monte Carlo simulation is generally used to determine the portfolio return distribution, with correlated asset returns being sampled then mapped to the price returns of the bonds and loans, thereby generating the portfolio return for that sample.

For the case when there are a small number of factors, alternatives to Monte Carlo exist, generally known as 'Semi-Analytic' methods. In practice these are most commonly applied to the case of a single factor model, although various extensions have been considered. For our purposes, we will focus on the single factor case.

The basic idea of a semi-analytic methods for a single factor model is that conditional on the single factor, the return on each instrument in the portfolio is independent of all other returns. Thus the portfolio return, conditional on the single factor draw, is the sum of independent random variables. If the conditional density function for each instrument return is known, then through convolution, the conditional density function for the portfolio can be obtained. The final step is to integration the conditional portfolio distribution against the distribution of the single factor to produce the unconditional portfolio distribution. Depending on the assumptions made about the instruments in the portfolio, often all but the last step can be carried out analytically without numerical methods,

and only the final one-dimensional integration over the single factor needs to be handled numerically with a one-dimensional quadrature. Hence the name semi-analytic. A common application of the semi-analytic approach is in determining the potential losses due to default in a portfolio of credit default swaps. If it is assumed that all recoveries given default will be the same across the portfolio, then the loss is determined by the probability of having a specified number of defaults. The probability of each name defaulting, conditional on the single factor, can be determined analytically, and it is relatively straight forward to compute the conditional probability of each number of defaults by using the fact that the defaults are (conditionally) independent.

Mathematically, if the single factor is z, which we can assume to be Normally distributed, the portfolio return distribution can be described in terms of the expectation with respect to z as

$$P\left(R_\pi < L\right) = E_z\left(P\left(R_\pi < L|z\right)\right) \tag{2.13}$$

In the special case that the portfolio instruments are homogenous and we consider the default/no-default case for portfolio loss with constant recovery given default r, there is a well known limiting loss distribution first derived by Vasicek for the limit where the number of assets N goes to infinity while each asset has weight $1/N$. This is known as the asymptotic single factor loss model. In this case, conditional on the single factor draw value z, the loss on the infinitely large portfolio of independent assets converges with probability one to $(1-r)PD(z)$ where $PD(z)$ is the conditional probability of default for an asset, which is analytically known. The expectation of this quantity with respect to z can be computed analytically, resulting in the loss distribution

$$F(L) = \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(L/(1-r)) - \Phi^{-1}(PD)}{\sqrt{\rho}}\right) \tag{2.14}$$

where $\Phi(\cdot)$ is the cumulative Normal distribution function and PD is the unconditional probability of default. This result can be extended to the case where there are n distinct asset classes, each with its own set of homogeneous parameter values for probability of default, recovery and correlation. In the limit where the number of assets in each class goes to infinity such that the proportion of assets in each class remains fixed, the limiting loss distribution can be analytically derived. In particular, if the asset class weights are ω_k and $L_k(q)$ is the solution of equation 2.14 using the parameters for asset class k for $F(L) = q$ for specified quantile q, then the full portfolio loss satisfies

$$L(q) = \sum_{k=1}^{n} \omega_k L_k(q). \tag{2.15}$$

This results states that in the asymptotic single factor framework, the tail loss for a portfolio consisting of several asset classes is simply the weighted sum of the tail losses for each asset class. This specifically means that there is no additional diversification benefit of mixing two asset classes - the total loss is just the sum of the individual losses. This model forms the basis of the Basel accord concerning bank capital requirements; the implication is that there is no capital benefit to lending over a range of countries or sectors as it is assumed that the are all subject to only one global single factor.

3 Localized Factor Models

Localized factor models attempt to capture and extend the advantages of a block correlation structure while working in with a more economically intuitive factor model framework. The idea is that for each asset in the portfolio, the asset return can be expressed as in equation 2.2, with the constraint that each asset weight on one common factor shared by all assets (the global factor) and only one other factor drawn from a group of n factors.

With this constraint the factor model has a total of $n+1$ factors, but each asset weights on only two of the factors. This is helpful if the assets can be divided into n sectors; the model then specifies that each asset return is influenced by a global economic factor, a sector-specific factor and its idiosyncratic risk. Mathematically, for any asset j in sector i this can be expressed as

$$z_{i,j} = \sqrt{\rho_{i,j}}\left(\beta^G_{i,j}\epsilon_G + \beta^S_{i,j}\epsilon_i\right) + \sqrt{1-\rho_{i,j}}\epsilon_{I,i,j} \tag{3.1}$$

with the constraint that

$$\left(\beta^G_{i,j}\right)^2 + \left(\beta^S_{i,j}\right)^2 = 1. \tag{3.2}$$

If all assets within a given sector have homogeneous properties, i.e. $\beta^S_{i,j}$ and $\rho_{i,j}$ are constant for all j within the same sector i, then it follows that the associated correlation matrix has a block structure. However, it should be noted that not all block correlation matrices can be represented as a localized factor model.

The localized factor model described above is more precisely a localized one-factor model. It is also of interest to consider localized two factor models. For example, we can consider the case where we have both n_S sector factors and n_C country factors, with each asset in the portfolio being mapped to a unique sector and country. With the inclusion of the global factor, this creates a model with $n_S + n_C + 1$ factors for which any specific asset weights on exactly three of the factors. Mathematically, for any asset j in associated with sector-country pair i, this can be expressed as

$$z_{i,j} = \sqrt{\rho_{i,j}}\left(\beta^G_{i,j}\epsilon_G + \beta^S_{i,j}\epsilon^S_{S(i)} + \beta^C_{i,j}\epsilon^C_{C(i)}\right) + \sqrt{1-\rho_{i,j}}\epsilon_{I,i,j}. \tag{3.3}$$

Here $S(i)$ is the index of the sector and $C(i)$ is the index of the country associated withe the country-sector pair i. We have the constraint that

$$\left(\beta^G_{i,j}\right)^2 + \left(\beta^S_{i,j}\right)^2 + \left(\beta^C_{i,j}\right)^2 = 1. \tag{3.4}$$

Again, if all the assets with a country-sector pair have homogenous parameters, this leads to a block correlation matrix; but of course a block correlation matrix need not correspond to a localized two factor model.

3.1 Optimal Localized One Factor Approximation

Given a block correlation matrix, it is natural to ask if there is an optimal localized one factor model approximation. If the reduced matrix of correlations, one entry for each group-pair and intra-group correlations on the diagonal, is positive definite, by adaption the work in Andersen et al. [2003], the answer is a yes. In this case, for a block correlation matrix with n groups, the following algorithm will produce an optimal localized one factor model:
1) Set ρ_i equal to the intra-group correlation for group i, $1 \le i \le n$.
2) Consider the reduced $n \times n$ correlation matrix P_g with off-diagonal elements given by the inter-group correlations and ones on the diagonal. If the original correlation matrix is positive definite, then this reduced matrix is also positive definite. Use the Andersen, Sidenius and Basu algorithm above to determine the best single factor approximation to P_g. The result is an $n \times 1$ vector $\vec{v}$ such that

$$P_g \approx \vec{v}\vec{v}^T + I - \text{diag}\left(\vec{v}\vec{v}^T\right). \tag{3.5}$$

3) For group i, define the $(n+1) \times 1$ vector $\vec{\beta}_i$ through

$$\beta^G_i = v(i)/\sqrt{\rho_i} \tag{3.6}$$

$$\beta^{S(i)}_i = \sqrt{1-\left(\beta^G_i\right)^2} \tag{3.7}$$

$$\beta^{S(k)}_i = 0 \qquad k \neq i. \tag{3.8}$$

The resulting values of ρ_i and $\vec{\beta_i}$ can then be used in the equation 3.1 to produce a localized one factor model with a corresponding block correlation matrix which is close to the original block correlation matrix. The approximation is close in the sense that the associated reduced correlation matrix is the optimal single factor approximation of the original reduce correlation matrix (optimal in the Frobenius norm). Note that the single factor approximation of the reduced correlation matrix P_g works here because a pair of assets from two different groups share only the one global factor in common, so that P_g depends only on the loadings on the global factor.

This method breaks down when the reduced matrix of correlation with intra-group correlations on the diagonal is not positive definite. The problem is that β_i^G as defined here may be larger than one. In this case it is necessary to adapt the localized one factor model to the method described by Huang and Yang [2010]. Equation 3.1 must be modified as

$$z_{i,j} = \sqrt{\rho_i + (1-\rho_i)/N_i}\left(\beta_{i,j}^G \epsilon_G + \beta_{i,j}^S \epsilon_i\right) + \sqrt{1-\rho_i}\left(\epsilon_{I,i,j} - \bar{\epsilon_i}\right). \tag{3.9}$$

where $\bar{\epsilon_i}$ is defined as in 2.10. To determine the optimal one factor approximation, step 3 of the above procedure should be modified as

$$\beta_i^G = v(i)/\left(\sqrt{\rho_i + (1-\rho_i)/N_i}\right) \tag{3.10}$$

$$\beta_i^{S(i)} = \sqrt{1-\left(\beta_i^G\right)^2} \tag{3.11}$$

$$\beta_i^{S(k)} = 0 \qquad k \neq i. \tag{3.12}$$

3.2 Extending Asymptotic Single Factor Models to Asymptotic Localized Factor Models

A key advantage of the Asymptotic Single Factor Framework is that equation 2.14 can be used to determine the tail quantile associated with capital requirements for each asset class for which capital must be held, and the dependence of the capital on the risk assumptions of the asset class is clear. This transparency is valuable for regulatory rules. A disadvantage, as discussed above, is that the single factor framework does not allow for diversification across the asset classes, with the consequence that the total capital required is the sum of the capital required for each individual asset class. Intuitively it is desirable to allow for some diversification under the assumption that not all sectors and regions are going to experience extreme losses in a perfectly correlated fashion. The localized one and two factor models provide an interesting extension to the single factor framework that allows for non-perfect correlation across sectors even in the asymptotic limit of number of assets per sector going to infinity.

For example, consider the case where every loan in a bank's corporate loan portfolio can be associated with a single sector out of n_S possible sectors and a single country out of n_C possible countries. Thus there are a total of $i = 1, \ldots, n_S \cdot n_C$ possible sector-country pairs. We form the localized two factor model of the form 3.3 for each loan in each sector-country pair, assume that within a sector-country pair, all parameters are homogeneous and take the limit as the number of loans in each sector-country pair goes to infinity, keeping ω_i, the overall percentage allocation of the portfolio to sectors and countries, fixed. Assuming the default/no-default and fixed recovery model, the resulting portfolio loss can be expressed as

$$L\left(\epsilon_G, \epsilon_1^S, \ldots, \epsilon_{n_S}^S, \epsilon_1^C, \ldots, \epsilon_{n_C}^C\right) = \sum_{i=1}^{n_S \cdot n_C} \omega_i (1-r_i) PD_i\left(\epsilon_G, \epsilon_{S(i)}^S, \epsilon_{C(i)}^C, \rho_i, PD_i\right). \tag{3.13}$$

Monte Carlo simulation can be used to generate a loss distribution over the $n_S + n_C + 1$ factors and determine the tail quantiles. In general there will be a point such that for all quantiles further out in the tail, the portfolio loss at that quantile is less than the sum of the losses for each individual sector-country pair at that quantile.

It is also possible to create associated localized single factor models by aggregating all sector-country pair exposures by either country or sector. For example, the country model would be based on aggregating over all sectors within a given country. Correlations could be estimated by looking at country equity indices. There should be a relationship between the localized two factor model and the two associated localized one factor models.

3.3 Numerical Methods for Localized Factor Approximations

We now consider applying semi-analytic numerical methods to the localized one factor approximation. Although this model has $n+1$ factors, the fact that only the global factor is shared in common suggests that we may be able to apply the single factor semi-analytic approach iteratively, first over the local sector factor, then over the global factor. Consider the n sectors, each with an associated loss L_k so that the total portfolio loss is

$$L = \sum_{k=1}^{n} L_k. \tag{3.14}$$

If we assume that losses occur in discrete quantities associated with discrete events (e.g. the default/no-default model or extensions to discrete multi-state credit models), then there are a finite number of that the loss L_k can occur for sector k and a finite number of ways the total loss L can occur. We can the express the probability of loss L as

$$P(L) = E_{\epsilon_G}\left[\sum_{L_1+\ldots L_n=L} P(L_1,\ldots,L_n|\epsilon_G)\right] \tag{3.15}$$

$$P(L_1,\ldots,L_n|\epsilon_G) = \prod_{k=1}^{n} P(L_k|\epsilon_G) \tag{3.16}$$

$$P(L_k|\epsilon_G) = E_{\epsilon_k}\left[P\left(L_k|\epsilon_G,\epsilon_k\right)\right]. \tag{3.17}$$

Here the sum in the first equation is taken over all possible distinct ways in which the total loss can be L. The second equation arises from the fact that conditional on the global factor, all the sectors in the localized one factor model are independent.

4 Research Questions

A number of interesting questions around optimal correlation matrix approximations, numerical implementations and applications of localized factor models remain to be addressed. For the 2012 Mathematical Problems in Industry Workshop we pose the following open problems.

4.1 Optimal Approximations

1) For a given block correlation matrix with n groups, consider the exact n factor model representation given by equation 2.9. If the factor loadings in this representation are obtained from an eigenvector decomposition of the matrix P_B in equation 2.12, then a k factor approximation is possible by using only the largest k principal components. How does this approximation compare with the optimal k factor approxmiation obtained from the PCA analysis described by algorithm 2.7 from Andersen et al. [2003]?
2) For a given block correlation matrix with n groups, how does the localized one factor approximation compare

to a k factor PCA approximation 2.7 of the original block correlation matrix?
3) The algorithm presented here for determining the optimal localized one factor model approximation of a block correlation matrix does not take into account the relative size of the blocks in determining the best approximation - that is, it only depends on the reduced correlation matrix for determining the one factor approximation. Is it possible to incorporate the number of assets N_i in each group, possibly as a weight, to achieve a better approximation to the original block correlation matrix?
4) How can the optimal localized two factor model be determined for a block correlation matrix? In this case the number of groups n must be factored into the two types of factors (e.g. $n = n_C \times n_S$ if there are n_C countries and n_S sectors), or alternatively an algorithm could be developed to search for the optimal break down into the two factors. However, it is likely that the original block correlation matrix will have a natural factor identification.
5) The optimality of the correlation matrix approximation has been considered in terms of how close the approximating matrix is to the original matrix in the Frobenius norm. However, it would be good to understand the sensitivity of the associated portfolio risk measures to the approximation error. This is complicated by the many portfolio-specific factors including the characteristics of each individual asset (exposure size, probability of default, bond characteristics, etc.). However, under some simplifying assumptions, such as a homogeneous portfolio with default/no-default valuation, it may be possible to translate correlation approxmiation error to portfolio standard deviation error or quantile error.

4.2 Computational Methods

6) 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?
7) An alternate numerical method that may work for the localized one factor approximation is to adapt the saddle point approximation methods described in Huang et al. [2007]. To what extent is this approach applicable and effective?

4.3 Estimating a Localized Two Factor Model from Two Localized One Factor Models

8) Consider two asymptotic localized one factor modesl, one based on aggregating across sectors and the other based on aggregating across countries. One implies an $n_C \times n_C$ intra-country correlation matrix, while the other is assocaited with an $n_S \times n_S$ intra-sector correlation matrix. Is it possible to derive an optimal estimation procedure for creating a localized two factor model in terms of sector-country pairs that is consistent with the given country and sector correlation matrices?

References

L. Andersen, J. Sidenius, and S. Basu. All your hedges in one basket. *RISK*, November:67–72, 2003.

J. Bohn and R. Stein. *Active Credit Portfolio Management in Practice*. Wiley Finance, 1st edition edition, 2009.

A. de Servigny and O. Renault. *The Standard & Poor's Guide to Measuring and Managing Credit Risk*. McGraw-Hill Books, 2004.

J. Huang and L. Yang. Correlation matrix with block structure and efficient sampling methods. *J. Comput. Finance*, 14(1):81–94, 2010.

X. Huang, C. Oosterlee, and J. Van der Weide. Higher order saddlepoint approximations in the vasicek portfolio credit loss model. *Journal of Computational Finance*, 11(1):93–113, 2007.

For more information, visit us at www.standardandpoors.com or call:

Americas	1	212.438.7280
Australia	61	1300.792.553
Europe	44	20.7176.7176
Japan	81	3.4550.8711
Singapore	65	6239.6316

Standard & Poor's Ratings Services
55 Water Street
New York, NY 10041
www.standardandpoors.com