

On the Aggregation of Local Risk Models for Global Risk Management

Greg Anderson
Vice President, MSCI Barra, Inc.
2100 Milvia St., Berkeley, CA 94704
greg.anderson@mscibarra.com

Lisa Goldberg
Executive Director, MSCI Barra, Inc.
2100 Milvia St., Berkeley, CA 94704
lisa.goldberg@mscibarra.com

Alec N. Kercheval[†]
Associate Professor, Department of Mathematics
Florida State University
Tallahassee, FL 32306-4510
kercheval@math.fsu.edu
850-644-8701

Guy Miller
Vice President, MSCI Barra, Inc.
2100 Milvia St., Berkeley, CA 94704
guy.miller@mscibarra.com

Kathy Sorge
MSCI Barra, Inc.
2100 Milvia St., Berkeley, CA 94704

April 25, 2005[‡]

[†]Corresponding author.

[‡]The data for the empirical studies was generously supplied by Barra, Inc. We thank Dan Stefek at Barra for early work with this problem, and the referee for many helpful suggestions. Some material in this article is part of a pending patent by Barra, Inc. This article is a revised version of an article written in 2003 under the title “Forecasting Total Risk”.

Abstract

Given a collection of single-market covariance matrix forecasts for different markets, we describe how to embed them into a global forecast of total risk.

We do this by starting with any global covariance matrix forecast that contains information about cross-market correlations, and revising it to agree with the pre-specified submarket matrices, preserving the requirement that a covariance matrix be positive semi-definite.

We characterize the ways this can be done and address the resulting numerical optimization problem.

Key words: portfolio risk, total risk, optimization, positive definite.

1 Introduction

Portfolio managers and financial firms need risk forecasts for many purposes, including mean-variance optimization (Markowitz (1952)), control of tracking error, and regulatory compliance requirements. Commonly, risk forecasts are derived from a forecast of the covariance matrix¹ of returns for the universe of available assets.

A well-known difficulty is that the asset covariance matrix is generally too large to estimate directly from a historical time series of asset returns.² Risk factor models address this difficulty by specifying a “small” set of factors that explain “most” of the asset covariances; the factor covariance matrix is then small enough to estimate from historical factor time series.

A great deal of work has been done on how to choose the best set of factors and how best to estimate the covariance matrix once the factors are chosen. Fama and French (1993) proposed three to five factors to estimate global covariance, and show that this does well for finding the minimum variance portfolio. Chan, Karceski, and Lakonishok (1999) show that more factors are needed when the objective is to minimize tracking error volatility, and in particular industry factors can be useful. Once the factors are chosen, there is still a statistical estimation problem to face. For example, Ledoit and Wolf (2003) describe a shrinking method for interpolating between a sample covariance matrix and a single-index matrix. MSCI-Barra uses a variety of proprietary methods to improve the outcome of bias tests for volatility forecasts.

In this paper we do not advocate any particular factor structure or number of factors; rather, we assume those decisions have already been made, and address the following subsequent issue.

An investment firm will often contain many individual market portfolios for a variety of markets, e.g. equities within various different countries. Each portfolio manager will have her own tailored factor model and covariance matrix forecast. However, the firm as a whole needs to understand its total

¹Sometimes this is called the “variance-covariance matrix”. It contains all the information in the correlation matrix, plus the individual variances.

²To compute an $n \times n$ sample covariance matrix without introducing spurious correlations requires a times series of length at least n for each asset; even longer series are necessary to achieve reasonable statistical depths. Nonstationarity in the returns process renders such long horizons useless. Volatility clustering on short timescales prevents solving the problem by the simple expedient of moving to high-frequency data.

risk exposure due to the sum of all its individual portfolios. This means the firm as a whole must try to understand the covariances *between* markets as well as those within markets. This can be derived from a forecast of the covariance matrix for the union of all the individual market factors.

Let m denote the total number of such factors. It is likely that the required $m \times m$ factor covariance matrix is again too large to estimate directly. For example, MSCI-Barra equity risk models use firm-specific factors, including industry factors. Because of the number of relevant industry sectors, this can lead to single-country models with anywhere from 11 to 65 factors (See Table 1). For a global model covering 50 to 60 such markets, then, one collectively has perhaps $m = 1000$ or more factors (e.g. US-oil, UK-size, JPN-financial, etc.) in the model.

There are various ways to proceed to estimate this $m \times m$ global factor covariance matrix. (For example, one could make a factor model for the factors, with individual market factors exposed to a smaller number of global factors.) However this might be done, we henceforth assume that the firm can arrive at a (necessarily crude) estimate of the global $m \times m$ covariance matrix, call it Σ .

This matrix Σ reflects greater estimation error than the individual market covariance forecasts, but we suppose it contains the best information about the cross-market covariances so important for estimating total risk.

We now face the following difficulty: the diagonal blocks of Σ corresponding to the individual markets probably are not equal to the more refined covariance matrix forecasts already developed by the individual managers. If there is a discrepancy, the global model and an individual market model will disagree about the risk of portfolios within that market. This means the firm-wide view of risk is at odds with the views of the constituent managers.

We assume in this paper that such disagreement is unacceptable to the firm: we adopt as a non-negotiable requirement that, e.g., the global and US risk models should agree on the risk of US portfolios, and likewise for the other markets.

We now arrive at the problems addressed by this paper:

1. Aggregating the given local risk models: *how can we incorporate all of the prior single market covariance forecasts into the global matrix Σ , in order to satisfy the above requirement?*
2. Minimizing loss of information caused by aggregation: *How can we*

solve this risk aggregation problem while keeping as much as possible of the cross-market covariance information already contained in Σ ?

In mathematical terms, we are asking how to find a revised global covariance matrix $\tilde{\Sigma}$ that agrees exactly with the pre-specified diagonal blocks, and agrees as closely as possible with the off-diagonal blocks of Σ .

The purpose of this paper is to highlight this issue mathematically and formulate the problem and its solution in precise terms.

In the following sections, we formalize the problem and then describe the complete set of solutions. This leads naturally to a nontrivial numerical optimization problem. We describe a simple illustrative example in Section 4, and in Section 5 we report on some results with real data.

2 Formalizing the problem

We assume a collection of asset-specific factors have been determined. This means we have specified a rectangular matrix A of asset sensitivities (or “exposures”) to the factors, where the (i, j) th element of A is the exposure of asset i to factor j .³ If a is the vector of asset returns over a given time step (say month), we denote by s the (shorter) vector of factor returns, defined by the cross-sectional linear regression

$$(2.1) \quad a = As + \epsilon,$$

where the vector ϵ of errors is assumed to have diagonal covariance matrix D and to be uncorrelated with s .

The factor covariance matrix $\Sigma(s)$ is, by design, small enough to be reliably estimated with a time series of factor returns. The asset covariance matrix $\Sigma(a)$ is then estimated as

$$\Sigma(a) \approx A\Sigma(s)A^T + D.$$

For a fuller discussion of risk factor modelling see, e.g., Grinold and Kahn (2000).

³In principle, these factors may be arbitrary, and are likely not orthogonal. For industry factors, the exposures will be 0 or 1. For other factors, (e.g. size) the numerical specification of an exposure (e.g. number of standard deviations from the mean) defines the precise meaning of the factor.

Definition 2.1 An aggregate factor model is a factor model whose m -dimensional vector of factors s decomposes into k subvectors

$$(2.2) \quad s = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{pmatrix},$$

with $m = n_1 + n_2 + \dots + n_k$, where n_i is the length of the i th subvector s_i .

We can think of the subvector s_i as representing the factors in the risk model for market i , $i = 1, \dots, k$.

This decomposition corresponds to a particular block structure on the covariance matrix

$$(2.3) \quad \Sigma = \Sigma(s) = (\Sigma_{i,j})_{i,j=1}^k,$$

where $\Sigma_{i,j}$ is the $n_i \times n_j$ matrix of covariances between the variables in s_i and s_j , called, when $i \neq j$, the *cross-block covariances*. For convenience let Θ_i denote the i th diagonal block $\Sigma_{i,i}$ – that is, Θ_i is the factor covariance matrix for the i -th sub-market.

We henceforth assume that all our covariance matrices are non-singular, hence positive definite. This means that there are no redundant variables. (Recall that positive definite means real, symmetric, and all eigenvalues positive; positive semi-definite means real, symmetric, and all eigenvalues non-negative).

We begin with the following simple version of our problem. We have a first draft factor covariance matrix Σ that incorporates some estimation error from various sources. We wish to improve it by replacing a diagonal block Θ with a more accurate one $\tilde{\Theta}$, for example coming from a more refined factor model of a submarket.

We cannot in general simply overwrite Θ with $\tilde{\Theta}$, because the resulting matrix is likely no longer to be positive semi-definite (which will lead to negative risk portfolios, spoil any portfolio optimization routines, and therefore must be strictly disallowed). (See Section 4 below for an example.)

Rebonato and Jäckel (2000) have discussed this problem of candidates for a correlation matrix that are not positive semi-definite, and presented a method for restoring positive semi-definiteness. However, their method does not respect the constraint that we fix a specified diagonal block.

The new covariance matrix $\tilde{\Sigma}$ can be viewed as the result of a change of the underlying factor variables

$$(2.4) \quad \tilde{s} = F(s).$$

In practice we expect that $\tilde{\Theta}$ will be close to Θ , so (2.4) will represent a small change of the factors in the original model, dictated by our insistence on $\tilde{\Theta}$ in place of Θ . (Since the identities of the factors are determined by the regression (2.1), they are only approximate in the first place, so the change (2.4) can be viewed as a correction via the additional information in $\tilde{\Theta}$.)

If F is nonlinear, the new covariance matrix $\tilde{\Sigma}$ depends not just on Σ , but also on higher moments of the variables s . These higher moments are usually not available in practice, so we restrict attention to the case of affine change of variables:

$$\tilde{s} = Ls + b,$$

where b is a constant vector and L is an arbitrary nonsingular linear transformation. In this case the covariance matrix of \tilde{s} is

$$(2.5) \quad \tilde{\Sigma} = L\Sigma L^T,$$

and $\tilde{\Sigma}$ is positive definite.

The matrix L now needs to be chosen to convert Θ to $\tilde{\Theta}$ while respecting the given decomposition (2.2) of s . Some terminology is convenient.

Definition 2.2 *Let Θ be a diagonal block of the $m \times m$ covariance matrix Σ for an aggregate factor model. Let $\tilde{\Theta}$ be a positive definite matrix of the same size as Θ .*

*A linear transformation $L : R^m \rightarrow R^m$ that preserves the decomposition (2.2), changes Θ to $\tilde{\Theta}$, and leaves the other variables unchanged will be called a **simple revising transformation converting Θ to $\tilde{\Theta}$** .*

That is, L is block diagonal, with the identity in each block except the one corresponding to Θ , and the diagonal block of $\tilde{\Sigma} = L\Sigma L^T$ corresponding to Θ is $\tilde{\Theta}$.

In general, there are infinitely many possible linear revising transformations for a given Σ , Θ , and $\tilde{\Theta}$. The next theorem characterizes them all.

Notation: If A is a positive semi-definite matrix, we will denote by $A^{1/2}$ the unique positive semi-definite square root of A . When A is also positive definite, $A^{-1/2}$ will denote the inverse of $A^{1/2}$.

For convenience we use the direct sum notation \oplus to represent block diagonal matrices: $A \oplus B$ is the block diagonal matrix with A and B as the diagonal blocks and zeros elsewhere. We use $O(n)$ to denote the orthogonal group of $n \times n$ real matrices O satisfying $OO^T = I$.

The proof of following theorem is in the Appendix.

Theorem 2.3 *Let Θ be an $n \times n$ diagonal block of the covariance matrix Σ with block structure (2.3). If $\tilde{\Theta}$ is an $n \times n$ positive definite matrix, then the space of simple revising transformations L converting Θ to $\tilde{\Theta}$ is parametrized by the orthogonal group $O(n)$ via*

$$(2.6) \quad L(O) = I \oplus \tilde{\Theta}^{1/2} O \Theta^{-1/2} \oplus I$$

for $O \in O(n)$, where each occurrence of I denotes the identity of the appropriate dimension.

In general the practitioner may want to revise several diagonal blocks at once – perhaps all of them. The previous theorem is easily generalized: the transformation accomplishing revision of several diagonal blocks is a composition of simple revising transformations. Since these commute for non-overlapping blocks, the order of composition doesn't matter.

Definition 2.4 *Suppose we have an aggregate factor model (2.2) with diagonal blocks $\Theta_1, \dots, \Theta_k$, and proposed replacements $\tilde{\Theta}_1, \dots, \tilde{\Theta}_k$, respectively. A **revising transformation converting the Θ_i 's to the $\tilde{\Theta}_i$'s** is a linear transformation $L : R^m \rightarrow R^m$ that preserves the decomposition (2.2), and changes the Θ_i 's to $\tilde{\Theta}_i$'s, i.e. L is block diagonal, and the diagonal blocks of $\tilde{\Sigma} = L\Sigma L^T$ are the $\tilde{\Theta}_i$'s.*

Theorem 2.5 *For $i = 1, \dots, k$, suppose Θ_i is an $n_i \times n_i$ diagonal block of the covariance matrix Σ with block structure (2.3).*

If, for each i , $\tilde{\Theta}_i$ is an $n_i \times n_i$ positive definite matrix, then the space of revising transformations L converting the Θ_i 's to the $\tilde{\Theta}_i$'s is parametrized by the product of orthogonal groups $O(n_1) \times O(n_2) \times \dots \times O(n_k)$ via the block diagonal transformation

$$(2.7) \quad L(O_1, \dots, O_k) = \bigoplus_{i=1}^k \tilde{\Theta}_i^{1/2} O_i \Theta_i^{-1/2}.$$

We remark that if some of the blocks of Σ remain unchanged, we may reduce the size of the parameter space $O(n_1) \times O(n_2) \times \dots \times O(n_k)$ by requiring L to be the identity on those factors.

3 An Optimization Problem

Any choice, such as the identity, in the parameter group

$$G = O(n_1) \times O(n_2) \times \dots \times O(n_k)$$

provides a solution of the risk aggregation problem described in the Introduction. Namely, we form a block diagonal matrix L from equation (2.7), and then $\tilde{\Sigma} = L\Sigma L^T$ is our new global covariance matrix that agrees exactly with the specified blocks $\tilde{\Theta}_i$. Moreover, all possible solutions L are of this form.

We minimize the loss of cross-market correlation information caused by this transformation by making an optimal or near-optimal choice for the orthogonal matrices in the product space G above.

This is a maximum likelihood problem. If we imagine that $\tilde{\Sigma}$ is the unknown “true” covariance matrix, we can view the estimated matrix Σ as a random sample from a distribution $P_{\tilde{\Sigma}}$ on the space of $n \times n$ matrices. Then we can seek the parameter $\tilde{\Sigma}$ maximizing the likelihood that Σ was sampled, i.e. find $\tilde{\Sigma}$ so that $p_{\tilde{\Sigma}}(\Sigma)$ is maximum, subject to the condition that $\tilde{\Sigma}$ has the specified diagonal blocks, and where $p_{\tilde{\Sigma}}$ is the density function for the distribution $P_{\tilde{\Sigma}}$.

Of course this depends on the choice of distribution $P_{\tilde{\Sigma}}$. It is reasonable to require that $p_{\tilde{\Sigma}}$ be a unimodal density peaked at $\tilde{\Sigma}$ and monotonically decreasing with distance from $\tilde{\Sigma}$. In the absence of any special information, we make the simple quadratic choice (as do Rebonato and Jäckel (2000)) corresponding to minimizing the objective

$$(3.1) \quad E_{\text{cov}} = \left\| \Sigma - \tilde{\Sigma} \right\|^2,$$

where $\| \cdot \|$ refers to the Frobenius norm, defined as the square root of the sum of the squares of the matrix entries.

A potential difficulty with the use of $\| \Sigma - \tilde{\Sigma} \|^2$ as an objective function is its implicit bias giving greater importance to factors with larger variances. To address this, we may wish to normalize all our factors in advance to have unit variance and consider only the covariance matrices of such normalized variables. This is equivalent to using the *correlation matrices* $\Sigma_{\text{corr}}, \tilde{\Sigma}_{\text{corr}}$ in place of the covariance matrices $\Sigma, \tilde{\Sigma}$, where

$$\Sigma_{\text{corr}} = S^{-1}\Sigma S^{-1}, \quad \tilde{\Sigma}_{\text{corr}} = \tilde{S}^{-1}\tilde{\Sigma}\tilde{S}^{-1}$$

and S is the diagonal matrix of standard deviations of the factors underlying Σ , and similarly for \tilde{S} .

Hence our objective function takes the form

$$(3.2) \quad E_{\text{corr}}(O) = \left\| \Sigma_{\text{corr}} - \tilde{\Sigma}_{\text{corr}} \right\|^2.$$

Unless we need to distinguish between this objective function E_{corr} and the previous objective E_{cov} , we will simply drop the subscript and use E , Σ , and $\tilde{\Sigma}$.

For the revision of a single diagonal block Θ , by permutation of the variables we may assume Θ is the lower right block of Σ , which then has the form

$$\Sigma = \begin{pmatrix} \Phi & C^T \\ C^T & \Theta \end{pmatrix}.$$

Our objective function is then

$$(3.3) \quad \begin{aligned} E(O) &= \|\tilde{\Sigma} - \Sigma\|^2 \\ &= 2\|\tilde{\Theta}^{1/2}O\Theta^{-1/2}C - C\|^2 + \|\tilde{\Theta} - \Theta\|^2, \end{aligned}$$

in which the second term is independent of O and can safely be ignored.

Since this objective is quadratic, we might hope for a simple algebraic expression for the minimum. In this direction, there is a strong resemblance between our problem and a classical problem about the orthogonal group.

Orthogonal Procrustes Problem: Given $m \times n$ matrices A and D , find $O \in O(n)$ minimizing $\|AO - D\|$.

There is a substantial literature about this problem and its relatives, e.g. Schonemann (1966); Bojanczyk and Lutoborski (1999); Golub and Van Loan (1989); Edelman, Arias, and Smith (1998). The orthogonal Procrustes problem as stated has an analytic solution in terms of singular value decompositions. Unfortunately, we know of no analytic solution for our more general problem, which we call the

Double Orthogonal Procrustes Problem: Given matrices A , B , and D of compatible sizes, find an orthogonal matrix O minimizing $\|AOB - D\|$.

This is a nontrivial numerical optimization problem. We do not attempt to fully analyze it in the present paper; however, it is not necessary for the purpose at hand. If we can simply improve the objective (3.3) inexpensively, we do.

We will show in Section 5 that we can find, for realistic data, parameters O for which the objective is significantly improved when compared with the default choice $O = I$. This will show that the optimization problem is worth addressing, and that progress toward the minimum can be made without requiring full analysis of the Double Orthogonal Procrustes Problem.

We turn to the general revising problem of Theorem 2.5. Ignoring the contribution of diagonal subblocks, which do not depend on the orthogonal matrix, our objective function for revising the first p blocks is

$$E(O_1, \dots, O_p) = \sum_{i \neq j} \|L_i \Sigma_{i,j} L_j^T - \Sigma_{i,j}\|^2,$$

where $L_i = \tilde{\Theta}_i^{1/2} O_i \Theta_i^{-1/2}$ for $i \leq p$ and $L_i = I$ otherwise.

This is a substantially larger problem than the simple revision, and again an analytic solution is out of reach. We do not attempt to minimize the objective over the full product space; rather, to improve the objective we follow a sequential revision strategy: at step i we will solve the simple revision problem consisting of the revised covariance matrix from step $i - 1$ and the i th diagonal subblock. This is easy to implement in terms of the simple revision problem, and as we see below it results in an improved revising transformation.

In the remainder of this section we discuss a simple numerical approach that improves the objective (3.3) and is easy to implement. The idea is to remove the constraint by finding a convenient parametrization of the orthogonal group and then applying an unconstrained Levenberg-Marquardt nonlinear least squares method.

Let $N = n(n - 1)/2$ denote the dimension of $O(n)$. Our method is to select a map f from \mathbb{R}^N to $O(n)$ and then to minimize the composition of the objective function (3.3) with this map f . To describe f , for $1 \leq i < j \leq n$ we denote by $R_{i,j}(\theta)$ the rotation by angle θ in the (i, j) plane. This rotation, sometimes called a Jacobi map, transforms the s_i and s_j factors according to $s_i \rightarrow s_i \cos \theta + s_j \sin \theta$ and $s_j \rightarrow s_j \cos \theta - s_i \sin \theta$, and leaves other factors unchanged. We use the following fact, whose proof is straightforward and omitted:

Every element of $O(n)$ with determinant 1 may be expressed as a product of N rotations,

$$(3.4) \quad f(\theta_1, \dots, \theta_N) \equiv R_{1,2}(\theta_1)R_{1,3}(\theta_2) \dots R_{1,n}(\theta_n) \dots R_{n-1,n}(\theta_N).$$

for some angles $\theta_1, \dots, \theta_N$.

With f in hand, we optimize $E(f(\theta_1, \dots, \theta_N))$ over \mathbb{R}^N . The form of the objective function (3.3) as a sum of squares lends itself to the Levenberg-Marquardt numerical method (see, for example, Press et al. (2002)) because the partial derivatives of $E(f(\theta_1, \dots, \theta_N))$ are easily computed. If the method converges starting from the identity, the result is a value $O^* \in O(n)$ for which the simple revising transformation $L(O^*)$ is better than $L(I)$, often, as we see in the next section, by a significant amount. Our experiments showed this method to be fast enough to treat a 30×30 diagonal subblock of a 700×700 matrix without worrying about runtime.⁴

4 Examples

We consider a small example to illustrate the ideas.

Suppose we have a 3×3 correlation matrix (say, for three world equity indices) given by

$$\Sigma = \begin{pmatrix} 1 & .9 & .7 \\ .9 & 1 & .4 \\ .7 & .4 & 1 \end{pmatrix}.$$

We suppose the risk manager has an external reason for wanting to adjust the correlation between the second and third indices from 0.4 to 0.3. In our terms, this means we wish to revise the lower right diagonal block

$$\Theta = \begin{pmatrix} 1 & .4 \\ .4 & 1 \end{pmatrix}$$

by replacing it with

$$\tilde{\Theta} = \begin{pmatrix} 1 & .3 \\ .3 & 1 \end{pmatrix}.$$

⁴This estimate is based on our relatively crude first implementation. We believe there is plenty of room for performance enhancement through refined numerical analysis.

Notice now that if we simply substitute $\tilde{\Theta}$ for Θ , it is easy to check that the resulting matrix

$$\Sigma^* = \begin{pmatrix} 1 & .9 & .7 \\ .9 & 1 & .3 \\ .7 & .3 & 1 \end{pmatrix}$$

has a negative eigenvalue, and hence does not qualify as a possible correlation matrix.

Instead, we employ Theorem 2.3, and look for a block diagonal matrix

$$L = I \oplus \tilde{\Theta}^{1/2} O \Theta^{-1/2},$$

so our revised matrix will be $\tilde{\Sigma} = L \Sigma L^T$. In this case $O \in O(2)$ is described by a single angle variable θ , and it is easy to determine numerically that the matrix L minimizing the objective $\|\tilde{\Sigma} - \Sigma\|^2$, as θ ranges between 0 and 2π , is

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1.0309 & -.0848 \\ 0 & -.0317 & 1.0122 \end{pmatrix},$$

and the corresponding correlation matrix $\tilde{\Sigma}$ is

$$\tilde{\Sigma} = L \Sigma L^T = \begin{pmatrix} 1 & .8684 & .6801 \\ .8684 & 1 & .3 \\ .6801 & .3 & 1 \end{pmatrix}.$$

Note that the lower diagonal block of $\tilde{\Sigma}$ is exactly $\tilde{\Theta}$.

For comparison, we mention that Rebonato and Jäckel (2000) address the related problem of finding the positive semi-definite correlation matrix Σ_{RJ} that is closest to Σ^* in the same sense of the Frobenius norm, but without the constraint of a specified block $\tilde{\Theta}$. In their paper they use for illustration the same matrices Σ and Σ^* .

They propose an optimization scheme over a six-dimensional space of angles, which arrives at a solution

$$\Sigma_{RJ} = \begin{pmatrix} 1 & .8946 & .6966 \\ .8946 & 1 & .3025 \\ .6966 & .3025 & 1 \end{pmatrix}$$

minimizing the Frobenius distance to Σ^* .

As expected, this is indeed closer to Σ^* than $\tilde{\Sigma}$, but at the cost of changing the correlation between the second and third indices from .3 to .3025.

5 Empirical Results

It remains to address the question of how much real benefit is gained by implementing the optimization described above. Is it possible that there is no practical difference between any two choices of revising transformation? We answer this by examining a concrete example of a commercial risk model with real data and comparing the optimized and non-optimized results.

We choose is the Barra global equity risk model, essentially as it was constituted in 2001. This model covers 23 markets and has 730 factors, as described in Table 1. The markets range from highly developed, such as UK and Japan, to emerging, such as Mexico and Taiwan. The largest market is the US, comprising 65 factors, and the smallest is the 11-factor New Zealand block. Our test data comes from factor covariances forecasts as estimated by Barra for the month of April 2001, based on historical monthly time series of prices.

The inputs to our study are a “first draft” 730×730 factor covariance matrix, estimated by a Barra method using global factors, and a collection of preferred single market blocks that have been estimated with separate models tuned to their individual markets.

We first examine the result of revising a single market block, comparing

1. the “default” simple revising transformation using $O = I$, with
2. the optimized transformation in which the choice of O is obtained by minimizing the objective (3.2).

Consider the 15-factor Mexico block. According to Theorem 2.3, the optimal revising transformation is obtained by searching over the orthogonal group $O(15)$. This generates an optimization problem with 105(= $15 \times 14/2$) parameters. We address this problem straightforwardly by using the parametrization described in Section 3, and a Levenberg-Marquardt optimization method implemented in C++ (see, e.g., Press, et. al. (2002)), using the default choice $O = I$ as a starting value.

Our results show that this optimization results in a substantial improvement of the objective function. That is, cross-market correlation changes are noticeably reduced when the revising transformation is optimized.

In Figure 1 we illustrate this finding for the Mexico block, and also for several other markets. The results show the root mean square change in cross-market correlations when a single target block of the global 730×730 correlation matrix is revised to match our prior specialized single-country correlation matrix forecast. The root mean square change provides us with a kind of average change per entry, and is proportional to our objective function.

We can see from this chart that there is a significant improvement in the use of an optimized revising transformation as compared to the default, generally by a factor of two to four. Typical average correlations change by around 0.05 with the default transformation ($O = I$), while the optimized transformation leads to average correlation changes of around 0.02. Note that we do not attempt to globally optimize our non-convex objective function, simply to inexpensively improve it by waiting until our Levenberg-Marquardt routine converges. Therefore these results should be interpreted as only a lower bound on the amount of improvement potentially available from a solving the optimization problem at hand.

The full problem requires us to update more than one diagonal block in the rough draft global covariance matrix – perhaps all of them. Suppose, in the case at hand, we update all 23 blocks. An optimized solution would require minimizing over a product of 23 orthogonal groups, which in this case has dimension 13,217 — far too large for us to handle with our current method.

Instead, we may simply try updating the blocks one at a time, where the input to the next simple revising problem is the output of the previous one. The resulting covariance matrix depends on the order, but our experiments show that dependence is very small.

For performance reasons, we choose four smaller blocks to update: Mexico, Netherlands, Switzerland, and France, sequentially updating the diagonal blocks in five different orders. For each chosen order, we performed the four simple revising transformations sequentially, first using the default choice $O = I$ for each block, and second by performing our optimization technique to chose the value of O for each block in the sequence.

Interestingly, our results did not significantly depend on the sequence order in which the four blocks were revised. The default choice led to a

resulting root mean square correlation change of about 0.062 in the global correlation matrix, while with optimization we found an RMS change of about 0.021 no matter what sequence order was chosen. This represents an improvement by about a factor of three.⁵

In summary: we have taken a 730×730 global factor correlation matrix forecast and replaced four smaller diagonal blocks (Mexico 15×15 , Netherlands 15×15 , Switzerland 20×20 , and France 21×21) with separately estimated substitutes. In general, making this replacement will force the cross-market correlations to change in order to preserve positive semi-definiteness, but, by optimizing as described in this paper, we can accomplish this “model aggregation” at the cost of only about 0.021 RMS change in cross-market correlations.

6 Conclusions

In this paper we have addressed two questions. First, how can we incorporate one or more refined submarket risk models into a larger but coarser total risk model, without violating positive semi-definiteness of the resulting covariance matrix? Second, given that there are many ways to do this, how can we choose a good way?

The first question is addressed by means of a linear change of variables of the underlying factors (in practice usually close to the identity). The space of all such changes of variables is identified as a product of orthogonal groups.

We answer the second question by posing it as a maximum likelihood problem, and deriving an objective function for numerical optimization. The objective function is defined on a product of orthogonal groups. The optimal transformation will embed the local models while simultaneously preserving positive semi-definiteness and minimizing the disturbance to the cross-market correlations.

The optimization problem, as we have framed it, is unconstrained, but is typically large and non-convex. However, the true global minimum has no special properties: we win if we can practically improve on the default choice at reasonable cost.

⁵Note that the average change is a little larger than results for the best single block update, because successive updates tend to degrade a little the results of previously updated blocks.

With real world data, our numerical optimization tests show that, by optimization, we can reduce the root mean square change forced on the cross-block correlations by a factor of three to five. We view this improvement to be sufficiently significant to justify implementing such optimization in commercial models of total risk.

We have not yet attempted to solve the underlying numerical problem; it is based on the Double Orthogonal Procrustes problem stated earlier, which is apparently open.

Another promising approach to the numerical problem begins with the observation that the positive semidefinite matrices form a convex cone. Rather than optimizing over a space of orthogonal groups, one could reformulate the question as a constrained convex optimization problem, known as “semidefinite least squares”. See Malick (2004) for a good discussion of this.

7 Appendix: Proof of Theorem 2.3

It is straightforward to verify that if $L = L(O)$ is given by (2.6) then $\tilde{\Sigma} = L\Sigma L^T$ has the same diagonal blocks as Σ except that Θ is replaced by $\tilde{\Theta}$.

Conversely, If L is a simple revising transformation it must have the diagonal block form $L = I \oplus \Lambda \oplus I$, where

$$\Lambda\Theta\Lambda^T = \tilde{\Theta},$$

or equivalently,

$$\Lambda\Theta^{1/2}(\Lambda\Theta^{1/2})^T = \tilde{\Theta}.$$

We want to show that $\Lambda = \tilde{\Theta}^{1/2}O\Theta^{-1/2}$ for some orthogonal O .

A standard lemma of linear algebra (see, e.g., Strang (1988)) says that A is symmetric positive definite if and only if there is a nonsingular matrix B such that $BB^T = A$, and if $BB^T = A$ then $B = A^{1/2}O$ for some orthogonal matrix O . Hence, from above, $\Lambda\Theta^{1/2} = \tilde{\Theta}^{1/2}O$ for some orthogonal O , whence the conclusion. ■

References

Bojanczyk, A.J., and A. Lutoborski (1999): The procrustes problem for orthogonal Stiefel matrices, *SIAM J. Scientific Computing*, 21, 1291–1304.

- Chan, L. K. C., J. Karceski, and J. Lakonishok (1999): On portfolio optimization: Forecasting covariances and Choosing the Risk Model, *Rev. of Financial Studies*, 12 (5), 937–974.
- Edelman, A., T.A. Arias, and S.T. Smith (1998): The geometry of algorithms with orthogonality constraints, *SIAM J. Matrix Analysis and Applications*, 20, 303–353.
- Fama, E. F., and K. R. French (1993): Common risk factors in the returns on stocks and bonds, *J. of Financial Economics*, 43, 153–193.
- Golub, G.H., and C.F. Van Loan (1989): *Matrix Computations*. Baltimore: Johns Hopkins Univ. Press, 2nd ed.
- Grinold, R.C. and R. N. Kahn (2000): *Active Portfolio Management*, New York: McGraw-Hill, 2nd ed.
- Malick, J. (2004): A dual approach to semidefinite least-squares problems, *SIAM J. Matrix Anal. Appl.*, 26, 272–284.
- Markowitz, H. (1952): Portfolio Selection, *Journal of Finance*, 7.1, 77–91.
- Press, W. et. al. (2002): *Numerical Recipes in C++*, Cambridge: Cambridge Univ. Press.
- Rebonato, R. and P. Jäckel (2000): The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes, *J. of Risk*, 2(2), 17–27.
- Schonemann, P.H. (1966): A generalized solution of the orthogonal Procrustes problem, *Psychometrika*, 3, 1-10.
- Strang, G. (1988): *Linear Algebra and its Applications*, Orlando: Harcourt Brace Jovanovich, Inc.

Country	Block Size
Currency	40
Australia	33
Brazil	29
Canada	30
France	21
Greece	23
Germany	27
Hong Kong	23
Indonesia	31
Japan	52
Korea	35
Malaysia	24
Mexico	15
Netherlands	15
New Zealand	11
South Africa	53
Singapore	27
Sweden	30
Switzerland	20
Thailand	41
Taiwan	35
United Kingdom	50
United States	65
TOTAL	730

Table 1: For illustration, this table reports the number of factors in each of 22 single-country equity risk models that were used by Barra circa 2001. (In each market, most of the factors are Industry factors.) Combined with a 40-factor currency model, the resulting global model contains 730 factors total. The corresponding 730 x 730 factor covariance matrix is too large for direct estimation from time series.

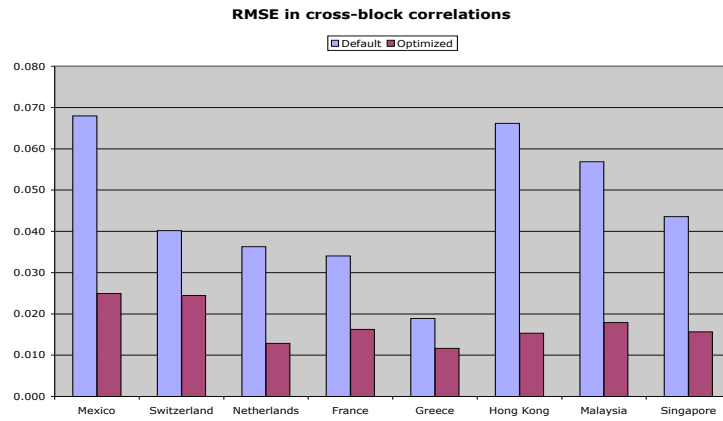


Figure 1: RMSE of cross-block correlations. Results are shown for gluing in a single country block.