# Forecasting Total Risk*

Greg Anderson†
*Barra, Inc.*

Lisa Goldberg†
*Barra, Inc.*

Alec N. Kercheval‡
*Florida State University*

Guy Miller†
*Barra, Inc.*

Kathy Sorge†
*Barra, Inc.*

May 1, 2003§

**Abstract**

A global model that forecasts risk for portfolios with holdings across several markets will typically disagree with the predictions of a model specifically adapted to a single market. Given a global model and a collection of single market models, we describe an optimal, consistent way to embed the single market forecasts into the global model. The method involves framing the problem as an optimization over the orthogonal group $\mathrm{O}(n)$.

*Key words: portfolio risk, total risk, orthogonal group, optimization.

†Barra, Inc., 2100 Milvia Street, Berkeley, CA 94704 USA

‡Department of Mathematics, Florida State University, Tallahassee, FL 32306-4510 USA, Phone (850) 644 8701, email `kercheval@math.fsu.edu`

§The data for the empirical studies was generously supplied by Barra, Inc. We thank Dan Stefek at Barra for originating the idea of correcting covariance blocks with a linear change of variables. Some material in this article is part of a pending patent by Barra, Inc.

# 1 Introduction

Portfolio managers need risk forecasts for many purposes, including mean-variance optimization (Markowitz (1952)) and reporting and compliance requirements. Risk forecasts depend on an estimate of the covariance matrix 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. *Risk factor models* address this difficulty by specifying a "small" set of factors that explain "most" of the asset covariances. Factor models use the sensitivities of the assets to the factors, which are specified *a priori*. These sensitivities are represented in a matrix $A$ whose $(i, j)$th entry is the sensitivity of asset $i$ to factor $j$. The vector $s$ of factor returns is then estimated from the known vector $a$ of asset returns by the linear regression

$$a = As + \epsilon\,.$$

The vector $\epsilon$ of errors is usually assumed to be normally distributed with a diagonal covariance matrix $D$ and uncorrelated with the factor returns vector $s$.

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

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

For a fuller exposition of factor risk models see, e.g., Grinold and Kahn (1999).

As quantitative risk management has evolved over the last three decades, risk factor models have become broader and more detailed. This has complicated the factor methodology, as we illustrate with two examples.

First, large models encompassing many asset classes and markets have become necessary for large firms to forecast their "total risk." The total number of factors has been forced upward to the point where, once again, more covariances are required — this time, among factors instead of assets — than can be accurately estimated from historical time series. For example, some recent commercial global equity risk models cover 50 to 60 different markets, each with 20 to 50 different risk factors, and thus collectively involve 1000 to 3000 different factors. A $1000 \times 1000$ factor covariance matrix is already too large to estimate from factor return time series.[1]

It is possible to overcome this problem by iterating the factor strategy: one can find a smaller list of "super factors" that explain most of the covariances between the original factors, and thus obtain a coarser but tractable version of the desired factor covariance matrix. However, approaches of this kind yield suboptimal results for portfolios that are not broadly diversified across markets.

[1] To 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.

Consider a portfolio of Australian equities. By adapting the data frequency and volatility model dynamics to the specific requirements of the Australian market, we can build an Australian equity model that will outperform the global model within Australian portfolios. When a firm uses both the coarser global model and the more refined Australian model, the disparity between model forecasts will make it impossible to resolve compromises between local and global risk control. We need a way to capture the resolution of the local model while maintaining the breadth of the global model.

A second difficulty is that a model may be based on factor data with different statistical properties making joint estimation difficult. Consider the problem of modelling risk for portfolios of international sovereign bonds. For portfolios unhedged against currencies, a substantial amount of the variation in portfolio return will be due to changes in currency exchange rates. A successful fixed income risk model must contain factors accounting for currency risk in addition to factors for interest rates and credit spreads.

However, currencies are highly volatile and subject to regime changes. Good risk estimates for these factors make use of high frequency data, while interest rate factors do better with lower frequency data for better stability and longer memory. We can estimate the total covariance matrix with low frequency data, and the currency matrix separately with high frequency data, but then one is faced with the problem making a consistent whole of the two.

We take a single approach to both these problems: to consistently embed the covariances of a small model within the extended covariance matrix of a larger model.

More terminology is helpful here.

**Definition 1.1** *An* aggregate factor model *is a factor model whose vector of factors $s$ decomposes into $k$ subvectors*

$$s = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{pmatrix}. \tag{1.1}$$

This decomposition imposes a block structure on the covariance matrix $\Sigma(s) = (\Sigma_{i,j})$. The $i$th diagonal block $\Sigma_{i,i}$ of $\Sigma(s)$ contains the subset of covariances between factors in the subset $x_i$. The *cross-block covariances* are the elements of $\Sigma_{i,j}$ where $i \neq j$.

Our problem is now the following. The coarse global model provides a "first draft" estimate $\Sigma$ of the total factor covariance matrix. It is our only source of information about the cross-block covariances. Single market models offer preferred estimates $\tilde{\Sigma}_{i,i}$ of the diagonal blocks. We wish to find a revised global covariance matrix $\tilde{\Sigma}$ that contains the preferred estimates as its diagonal blocks, but preserves as much of the original global cross-block information as possible.

We cannot expect to make this revision without changing the off-diagonal blocks: if we merely overwrite the draft $\Sigma_{i,i}$ with the preferred $\tilde{\Sigma}_{i,i}$, the resulting matrix typically will fail to be positive semi-definite and therefore predict negative variances for certain portfolios. The spurious presence of negative variance portfolios will, for example, spoil mean-variance optimization.

The purpose of this article is to formulate our problem precisely and to identify a mathematical technique for solving it. The basic idea is to make a linear change of variables within each market that revises its diagonal block, sending $\Sigma_{i,i} \to \tilde{\Sigma}_{i,i}$. We describe the space of all such *linear revising* transformations, and then search for the one that minimizes the disturbance to the cross-terms.

In the following sections, we develop a framework for the problem and describe its solution. We comment on the performance of numerical solutions to the resulting optimization problem. Finally, we identify a few outstanding computational problems raised by the method.

## 2 A Linear Change of Variables

Let $\Sigma$ be the covariance matrix of a random vector $s$. This means that $\Sigma$ is symmetric and positive semi-definite. Henceforth we make the simplifying mild additional assumption that our covariance matrices are all positive definite (i.e., there are no redundant variables).

Suppose the vector $s$ of factors is an aggregate of two sets of factors

$$s = \begin{pmatrix} t \\ u \end{pmatrix}. \tag{2.1}$$

The decomposition (2.1) imposes a block structure on the covariance matrix:

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

Here $\Theta$ and $\Phi$ are the covariance matrices of the vectors $t$ and $u$ respectively. The off-diagonal block $C$ contains covariances between factors in $t$ and factors in $u$. It is helpful to keep in mind the example given in the introduction of the breakdown of a global bond model into currency factors ($t$) and interest rate factors ($u$).

We now take the view that $\Sigma$ incorporates some estimation error from various sources, and assume that we have prior knowledge of the covariance matrix $\tilde{\Theta}$ of the $t$ variables from an independent and more reliable source.

Any correction of the covariance matrix $\Sigma$ can be thought of as coming from some change of variables

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

How is the covariance matrix $\tilde{\Sigma}$ for the new variables $\tilde{s}$ related to $\Sigma$? If $F$ is nonlinear, the answer will depend not just on $\Sigma$, 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 a nonsingular linear transformation. In this case the covariance matrix of $\tilde{s}$ is

$$\tilde{\Sigma} = L\Sigma L^T, \tag{2.2}$$

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

Our requirement is to convert $\Theta$ to $\tilde{\Theta}$ while respecting the decomposition of $s$ into $t$ and $u$. We can do this by making the linear change of variables

$$L = \begin{pmatrix} \tilde{\Theta}^{1/2}\Theta^{-1/2} & 0 \\ 0 & I \end{pmatrix}.$$

The exponent $1/2$ denotes the *symmetric square root*: if $A$ is a symmetric positive definite matrix, $A^{1/2}$ is the unique symmetric positive definite matrix satisfying $(A^{1/2})^2 = A$. The exponent $-1/2$ denotes the multiplicative inverse of the symmetric square root. The matrices $\tilde{\Theta}^{1/2}$ and $\Theta^{-1/2}$ exist uniquely since $\Theta$ and $\tilde{\Theta}$ are symmetric and positive definite.

Due to the block form of the matrix $L$, transformation (2.2) respects the block structure of $\Sigma$. The matrix $\Theta$ is replaced by $\tilde{\Theta}$ and $\Phi$ is unchanged. The revised covariance matrix takes the form

$$\tilde{\Sigma} \;=\; \begin{pmatrix} \tilde{\Theta} & \tilde{\Theta}^{1/2}\Theta^{-1/2}C \\ C^T\Theta^{-1/2}\tilde{\Theta}^{1/2} & \Phi \end{pmatrix}.$$

It follows that by applying a linear transformation that respects the block structure of $\Sigma$, we can overwrite a diagonal subblock without affecting the other diagonal subblocks. However, the cross-terms are altered.

This raises several issues.

1. Can we characterize the set of linear transformations that revise only one diagonal block?

2. How small can the cross-block impact of a revising transformation be made?

3. How does the situation generalize when more than one block requires revision?

We address these topics in what follows.

# 3 Simple Revising Transformations

Our goal in this section is to give a mathematical description of the linear transformations that revise a single block. First, we need to describe the space of square roots of positive definite matrices; we henceforth assume all matrices have real entries.

**Definition 3.1** *A* square root *of an $n \times n$ matrix $\Theta$ is another $n \times n$ matrix $\sqrt{\Theta}$ such that*

$$\sqrt{\Theta}\sqrt{\Theta}^T = \Theta.$$

The following standard lemma (e.g. Strang (1988)) states that any symmetric positive definite matrix $\Theta$ has infinitely many square roots, all related by multiplication with an *orthogonal matrix.*

By definition, an orthogonal matrix is a square matrix with orthonormal columns; such a matrix defines a distance-preserving linear transformation of Euclidean space. The set of $n \times n$ orthogonal matrices, denoted $\mathrm{O}(n)$, is an example of a *Lie group*, about which there is a large literature (e.g. Weyl (1966), Artin (1988)).

**Lemma 3.2** *Suppose $\Theta$ is an $n \times n$ matrix.*

1. *$\Theta$ admits a square root if and only if $\Theta$ is symmetric and positive semi-definite.*

2. *If $\Theta$ is symmetric and positive definite, the set of square roots of $\Theta$ is parametrized by the orthogonal group $\mathrm{O}(n)$ via the transformation*

$$O \mapsto \Theta^{1/2} O. \tag{3.1}$$

The relevance of matrix square roots appears in Lemma 3.6 below.

**Definition 3.3** *If $\Sigma = (\sigma_{i,j})_{i,j=1}^n$ is a square matrix, a* diagonal subblock *is a submatrix of the form $(\sigma_{i,j})_{i,j=p}^q$, where $1 \leq p \leq q \leq n$.*

**Definition 3.4** *A* simple revision problem *is a triple $(\Sigma, \Theta, \tilde{\Theta})$ where $\Sigma$ is a positive definite covariance matrix, $\Theta$ is a diagonal subblock and $\tilde{\Theta}$ is a preferred revision of $\Theta$.*

For notational convenience and without loss of generality, we assume that $\Theta$ is an upper left diagonal block of $\Sigma$. Therefore, the specification of $\Theta$ corresponds to a decomposition of the vector of factors into two subvectors,

$$s = \begin{pmatrix} t \\ u \end{pmatrix}. \tag{3.2}$$

In other words, our factor model is an aggregate where $t$ is the set of factors whose estimated covariance matrix requires revision and $u$ is the complementary set of factors.

**Definition 3.5** *A* simple revising transformation *of the simple revision problem $(\Sigma, \Theta, \tilde{\Theta})$ is a linear map $L : s \mapsto \tilde{s}$ that preserves the decomposition (3.2), that overwrites $\Theta$ with $\tilde{\Theta}$ in $\tilde{\Sigma} = L\Sigma L^T$, and that restricts to the identity on $u$.*

In this setup, we can express $\Sigma$ and the matrix representation of $L$ in two-by-two block form, where $L$ leaves alone the variables in the lower right block of $\Sigma$.

**Lemma 3.6** *Let $(\Sigma, \Theta, \tilde{\Theta})$ be a simple revision problem, and assume for convenience that $\Sigma$ takes the block form*

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

*A linear transformation $L$ is a simple revising transformation of $(\Sigma, \Theta, \tilde{\Theta})$ if and only if*

$$L = \begin{pmatrix} \sqrt{\tilde{\Theta}}(\sqrt{\Theta})^{-1} & 0 \\ 0 & I \end{pmatrix} \tag{3.3}$$

*for some choice of square roots $\sqrt{\Theta}$ and $\sqrt{\tilde{\Theta}}$. The revised matrix takes the form*

$$\tilde{\Sigma} = \begin{pmatrix} \tilde{\Theta} & \sqrt{\tilde{\Theta}}(\sqrt{\Theta})^{-1}C \\ (\sqrt{\tilde{\Theta}}(\sqrt{\Theta})^{-1}C)^T & \Phi \end{pmatrix}. \tag{3.4}$$

**Proof:** If $L$ has the form (3.3), then one easily sees that $L\Sigma L^T$ has the form (3.4).

Conversely, by definition any simple revising transformation must be of the form

$$L = \begin{pmatrix} \Lambda & 0 \\ 0 & I \end{pmatrix}, \tag{3.5}$$

and must satisfy

$$\Lambda \Theta \Lambda^T = \tilde{\Theta}. \tag{3.6}$$

Letting $\sqrt{\Theta}$ be any square root of $\Theta$, we have

$$\Lambda\sqrt{\Theta}(\Lambda\sqrt{\Theta})^T = \tilde{\Theta},$$

so $\Lambda\sqrt{\Theta} = \sqrt{\tilde{\Theta}}$ for some square root $\sqrt{\tilde{\Theta}}$ of $\tilde{\Theta}$. Hence $\Lambda = \sqrt{\tilde{\Theta}}(\sqrt{\Theta})^{-1}$. ∎

The following is an immediate consequence of Lemmas 3.6 and 3.2.

**Theorem 3.7** *Let $(\Sigma, \Theta, \tilde{\Theta})$ be a simple revision problem, and suppose $\Theta$ is $n \times n$. Then the space of simple revising transformations of $(\Sigma, \Theta, \tilde{\Theta})$ is parametrized by the orthogonal group $\mathrm{O}(n)$ via*

$$L(O) = \begin{pmatrix} \tilde{\Theta}^{1/2} O \Theta^{-1/2} & 0 \\ 0 & I \end{pmatrix}$$

*for $O \in \mathrm{O}(n)$.*

Theorem 3.7 allows us to frame the problem of finding an optimal revising transformation as an optimization over $\mathrm{O}(n)$. First we need to formulate the more general problem of revising more than one block of the covariance matrix $\Sigma$.

# 4 Revising Multiple Blocks

Given a factor covariance matrix $\Sigma$ covering many markets and asset classes, we may want to revise more than one subblock with market specific information.

Mathematically, this means that the vector $s$ of factors splits into market determined subvectors

$$s = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{pmatrix}. \tag{4.1}$$

Recall that our aggregate factor model from Definition 1.1 produces a covariance matrix with a block structure:

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

If, for $i = 1, \ldots, k$, we let $n_i$ denote the length of the $i$th subvector $s_i$, then $\Sigma_{i,j}$ is the $n_i \times n_j$ matrix of covariances between the variables in $s_i$ and $s_j$. For convenience let $\Theta_i$ denote the $i$th diagonal block $\Sigma_{i,i}$.

**Definition 4.1** *A revision problem* $(\Sigma, \{\Theta_i\}_{i=1}^k, \{\tilde{\Theta}_i\}_{i=1}^k)$ *consists of a block covariance matrix* $\Sigma = (\Sigma_{i,j})_{i,j=1}^k$*, and a collection of desired revisions* $\tilde{\Theta}_i$ *of the diagonal blocks* $\Theta_i$*.*

**Definition 4.2** *A revising transformation of the revision problem* $(\Sigma, \{\Theta_i\}_{i=1}^k, \{\tilde{\Theta}_i\}_{i=1}^k)$ *is a linear map* $L : s \mapsto \tilde{s}$ *that preserves the decomposition (4.1), that overwrites each* $\Theta_i$ *with* $\tilde{\Theta}_i$*, and that restricts to the identity on each* $s_i$ *for which* $\Theta_i = \tilde{\Theta}_i$*.*

Note that simple revising transformations are merely special kinds of revising transformations. To be pricise, a simple revising tranformation is a revising transformation for which $k = 2$ and $\Theta_i = \tilde{\Theta}_i$ either for $i = 1$ or $i = 2$. Theorem 3.7 has a straightforward generalization to revising transformations, summarized below.

**Theorem 4.3** *Let $RP = (\Sigma, \{\Theta_i\}_{i=1}^k, \{\tilde{\Theta}_i\}_{i=1}^k)$ be a revision problem, where the ith diagonal block has size $n_i \times n_i$. The set of linear revising transformations of $RP$ is parametrized by the product of orthogonal groups $O(n_1) \times O(n_2) \times \ldots \times O(n_k)$ via a block diagonal parametrization $L$, where the ith diagonal block of $L(O_1, \ldots, O_k)$ is given by*

$$L_i(O_1, \ldots, O_k) = \tilde{\Theta}_i^{1/2} O_i \Theta_i^{-1/2}.$$

*The subblocks of the revised matrix take the form*

$$\tilde{\Sigma}_{i,j} = L_i \Sigma_{i,j} L_j^T = \tilde{\Theta}_i^{1/2} O_i \Theta_i^{-1/2} \Sigma_{i,j} (\Theta_j^{-1/2})^T O_j^T (\tilde{\Theta}_j^{1/2})^T.$$

# 5 The Optimal Revising Transformation

In Section 4, we identified the space of linear revising transformations of a covariance matrix as a product of orthogonal groups

$$\mathrm{O}(n_1) \times \mathrm{O}(n_2) \times \ldots \times \mathrm{O}(n_k) \tag{5.1}$$

where $n_i$ is the dimension of the $i$th block requiring revision.

This gives us plenty of choice. For example, there is a 435 ($= 30 \times 29/2$) dimensional space of possible ways to revise a single block of dimension 30.

An obvious choice of revising transformation corresponds to choosing $O_i$ to be the identity matrix for each $i$. The resulting transformation is called the *symmetric covariance transformation* since it is built from symmetric square roots of covariance matrices. This choice generates an algorithm for computing $\tilde{\Sigma}$ that is easy to implement in practice, but is unlikely to give optimal results.

The theoretical issue of characterizing the "best" transformation is discussed in this section. The practical problem of designing and implementing an algorithm to find the best transformation is addressed in Section 6. In Section 7, we examine results from the implementation in several examples and quantify the incremental benefit of using the optimal revising transformation instead of the symmetric covariance transformation.

Recall that we have a covariance matrix $\Sigma$ with diagonal blocks $\Theta_1, \ldots, \Theta_k$, subject to estimation error. We have preferred replacements $\tilde{\Theta}_1, \tilde{\Theta}_2, \ldots, \tilde{\Theta}_k$, but $\Sigma$ is our only source of information about the cross-block covariances.

We can pose the problem of selecting a linear revising transformation as a maximum likelihood problem. If we imagine that $\tilde{\Sigma}$ is the unknown "true" covariance matrix, we can view the estimated matrix $\Sigma$ 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 $\Sigma$ 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 our 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 $\tilde{\Sigma}$. To lowest order near the peak of a smooth, generic such density, maximum likelihood is equivalent to minimizing an objective function

$$E_{\mathcal{Q}}(\tilde{\Sigma}) = \mathcal{Q}\left(\Sigma - \tilde{\Sigma}\right) , \tag{5.2}$$

where $\mathcal{Q}$ is a positive quadratic form[2] in the $n^2$ variables of $\Sigma - \tilde{\Sigma}$.

For known applications, it suffices to restrict attention to quadratic forms expressible in terms of a matrix $W$ of weights, as

$$E_W(\tilde{\Sigma}) = \left|\left|W\left(\Sigma - \tilde{\Sigma}\right)W^T\right|\right|^2 , \tag{5.3}$$

where $||\cdot||$ is the *Frobenius norm* defined as the square root of the sum of the squares of the matrix entries. If we make the simple choice $W = I$, we obtain the objective function

$$E_{\text{cov}} = \left|\left|\Sigma - \tilde{\Sigma}\right|\right|^2 \tag{5.4}$$

The weighting matrix $W$ allows us, for example, to incorporate varying levels of confidence in the estimation of the different underlying variables, or to incorporate other exogenous model-specific knowledge. The default choice $W = I$, the identity, treats all markets equally. Even though it might be unwise in practice to let covariances with a poorly characterized market (e.g., a small emerging equity market) unduly influence the revision of factors in a

[2] A quadratic form is a degree 2 homogenous polynomial

better characterized market (e.g., the Japan equity market), for convenience in this study we make the default choice $W = I$ in what follows.

The use of $||\Sigma - \tilde{\Sigma}||^2$ as an objective function also reflects an implicit bias giving greater importance to factors with larger variances. To address this, we may wish to normalize all our factors 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 $\Sigma$, and similarly for $\tilde{S}$.

Hence the objective function takes the form

$$E_{\text{corr}}(O) = \left|\left|\Sigma_{\text{corr}} - \tilde{\Sigma}_{\text{corr}}\right|\right|^2 . \tag{5.5}$$

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

By virtue of Theorem 4.3, we can think of either of these as a function of $O = (O_1, \ldots, O_k)$ ranging over the product of orthogonal groups $\mathrm{O}(n_1) \times \mathrm{O}(n_2) \times \ldots \times \mathrm{O}(n_k)$.

When we deal with correlation matrices, the revising transformation corresponding to $O_i = I$ for all $i$ will be called the *symmetric correlation transformation* since it is built from symmetric square roots of correlation matrices.

## 6 Seeking the optimal revising transformation

This section contains practical comments regarding the minimization problem set out above. We will begin with a discussion of the simple revising problem (replacing a single diagonal subblock), and note some of the complications of the general case.

For the simple revising problem we have the form of the revised matrix in equation (3.4), and our objective function is

$$\begin{aligned} E(O) &= \|\tilde{\Sigma} - \Sigma\|^2 \\ &= 2\|\tilde{\Theta}^{1/2} O \Theta^{-1/2} C - C\|^2 + K. \end{aligned} \tag{6.1}$$

In this formula $K$ is the contribution due to the diagonal subblocks, which is independent of $O$ and can be safely ignored. In this section we address the question "Can we improve the default choice of the identity for $O$?", while the next is concerned with "Does the benefit justify the computational effort?"

Since this objective function 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 $O \in \mathrm{O}(n)$ 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, the

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

We do know that this problem can be reduced to the case where $A$, $B$, and $D$ are square with $A$ and $B$ also diagonal.

Failing to produce an exact solution to the problem, we resort to numerical techniques. There is ample literature on specialized numerical methods for minimizing functions on the orthogonal group (for example Edelman, Arias, and Smith (1998) and references therein). As a practical matter, these algorithms may not be readily available or conveniently adapted to existing computing infrastructure. We describe a simple approach now for restating the problem in more standard form as a minimization of a nonlinear function of several real variables.

Let $N = n(n-1)/2$ denote the dimension of $\mathrm{O}(n)$. Our method is to carefully select a map $f$ from $\mathbb{R}^N$ to $\mathrm{O}(n)$ and then to minimize the composition of the objective function (6.1) 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 $\theta$ in the $(i, j)$ plane. This rotation, sometimes called a Jacobi map, transforms the $s_i$ and $s_j$ factors according to $s_i \to s_i \cos\theta + s_j \sin\theta$ and $s_j \to s_j \cos\theta - s_i \sin\theta$, and leaves other factors unchanged. The following fact, proved in the Appendix, is the cornerstone of this idea:

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

$$f(\theta_1, \ldots, \theta_N) \equiv R_{1,2}(\theta_1)R_{1,3}(\theta_2)\ldots R_{1,n}(\theta_n)\ldots R_{n-1,n}(\theta_N). \tag{6.2}$$

*for some angles* $\theta_1, \ldots, \theta_N$.

The form of the objective function (6.1) as a sum of squares suggests it will be suited to approach with the Levenberg-Marquardt numerical method, a hybrid of Newton's method and steepest descent (see, for example, Press et al. 2002). This method, in common with many others, requires the partial derivatives of the objective function. We selected the parameterization $f$ of equation (6.2) in part because its derivatives are tractable, allowing us to apply the Levenberg-Marquardt minimization technique to

$$E(f(\theta_1, \ldots, \theta_N))$$

in a straightforward way. If the method converges starting from the identity, the result is a value $O^* \in \mathrm{O}(n)$ 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. In conjunction with the analytic simplifications noted above, this method is fast enough to obtain the optimal revising transformation for a $30 \times 30$ diagonal subblock of a $700 \times 700$ matrix without worrying about runtime. [3]

We return to the general revising problem of Theorem 4.3. Ignoring the contribution of diagonal subblocks, which do not depend on the orthogonal matrix, our objective function in the general case is

$$E(O_1, \ldots, O_k) = \sum_{i \neq j} \left\| \tilde{\Theta}_i^{1/2} O_i \Theta_i^{-1/2} \Sigma_{i,j} (\Theta_j^{-1/2})^T O_j^T (\tilde{\Theta}_j^{1/2})^T - \Sigma_{i,j} \right\|^2 .$$

This is a substantially larger problem than the simple revision, and again an analytic solution is out of our reach. Direct application of the method above also turns out to be computationally prohibitive. Neither is it possible to reduce the general problem to a collection of simple revising problems because the $O_i$ interact in pairs. So for now we propose the more modest goal of improving on the default choice $O = I$ rather than obtaining the global minimum. To this end, we will test *sequential revision* of diagonal blocks. By this we mean to treat the general revision problem $RP = (\Sigma, \{\Theta_i\}_{i=1}^k, \{\tilde{\Theta}_i\}_{i=1}^k)$ as a sequence of simple revision problems: 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

[3] This estimate is based on our relatively crude implementation. We believe there is enormous room for performance enhancement through improvements in numerical analysis and software engineering.

the simple revision problem, and as we see below it results in an improved revising transformation.

# 7 Empirical Results

In this section, we apply the ideas developed above to the concrete example of a global equity model. Our model covers 23 markets and has 730 factors. The full details are contained 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 estimated as of April 2001 using historical monthly prices.

To make optimal use of the available information, we create a first draft of the $730 \times 730$ factor covariance matrix and then update market blocks for which there are superior estimates. We benchmark our algorithm in the simple situation where there is only one block that requires revision. Next, we move to the case where there are multiple blocks that need to be updated.[4]

For revision of a single block, we compare three simple revising transformations. The first two are the symmetric covariance and correlation transformations described in Section 5. The third is an optimal transformation that locally minimizes the objective function $E_{\text{corr}}$ given in equation (5.5), as computed by our Levenberg-Marquardt implementation. The success of these methods is evaluated with $E_{\text{corr}}$. In other words, we prefer the transfor-

[4]Due to performance considerations, we update only the smaller blocks.

| 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: Our model comprises 730 factors in 23 markets. There is one block for each market and plus the currency block.

mation that minimizes root mean square change in cross-block correlations.

Consider the 15-factor Mexico block. According to Theorem 3.7, 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.

The resulting root mean square disturbances to the correlations between Mexico factors and factors outside the Mexico block are displayed in Figure 1. Also in Figure 1 are the results of updating several other blocks. In each case,

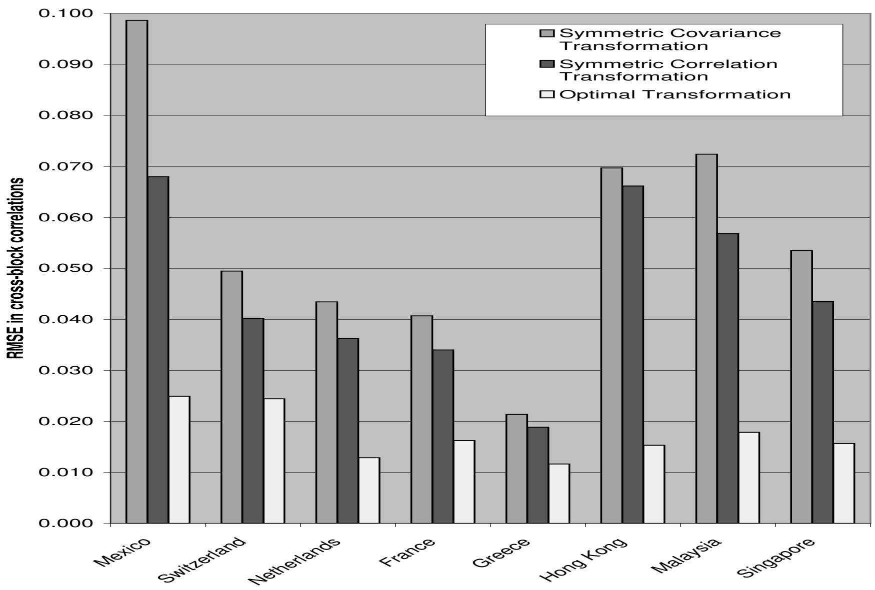


Figure 1: RMSE of cross-block correlations. Results are shown for gluing in a single country block.

only the indicated block has been changed.

The results are consistent and dramatic. In all cases, there is a benefit to using the symmetric correlation transformation instead of the symmetric covariance transformation. A greater improvement comes from our optimization. A typical case is France, where there is a relatively small improvement between the symmetric covariance and correlation transformations (.041 to .034) followed by an enormous improvement that comes from using the optimal transformation instead of the symmetric correlation transformation (.034 to .016).

In reality, we may need to update more than one block. However, as indicated at the end of Section 6, the full optimization problem is often too large even if the problem of updating individual blocks can be handled. Suppose, in the case at hand, we update all 23 blocks. According to Theorem 4.3, the optimal solution is found by minimizing over a product of 23 orthogonal groups. This problem has 13,217 parameters.

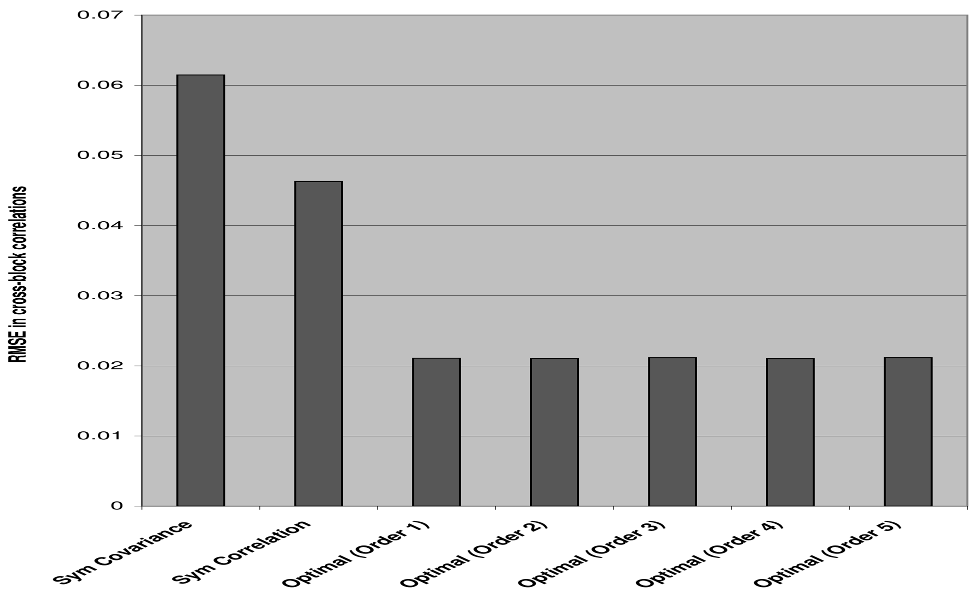


Figure 2: RMSE of cross-block correlations. Results are shown for sequentially gluing in four country blocks.

We resort to the sequential revision method described at the end of Section 6. In Figure 2 we compare the result of symmetric covariance and correlation transformations with the optimized sequential methods of updating four blocks. As in the single block case, there are dramatic cuts in the root mean square disurbance of cross-block correlations obtained by changing from the symmetric covariance transformation to the symmetric correlation transformation to the optimal transformation.

Since the sequential optimization is intrinsically order dependent, we run the algorithm several times using different orders. The results, also in Figure 2 demonstrate that the order does not have a significant impact on the average disturbance.

| | First | Second | Third | Fourth |
|---|---|---|---|---|
| Order 1 | Mexico | Switzerland | Netherlands | France |
| Order 2 | France | Mexico | Netherlands | Switzerland |
| Order 3 | Netherlands | France | Switzerland | Mexico |
| Order 4 | Switzerland | Netherlands | France | Mexico |
| Order 5 | France | Switzerland | Mexico | Netherlands |

Table 2: Order of sequentially gluing in country blocks in Figure 2.

# 8 Conclusions

In this paper we have addressed two questions. First, how can we consistently incorporate one or more refined submarket risk models into a larger but coarser total risk model? Second, given that there are many ways to do this, how can we choose the best 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 consistently embed the local models while minimizing the disturbance to the cross-market correlations.

When many refined local market model forecasts need to be embedded in the global model, it may prove computationally prohibitive to attack the problem by simultaneous optimization over the full product space, but the problem can be approached serially, optimizing one revising transformation at a time.

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.

# 9 Appendix: Surjectivity of the rotation product

Every element of the orthogonal group $\mathrm{O}(n)$ has determinant +1 or -1. The Lie subgroup of matrices with determinant +1 is called the special orthogonal group and denoted $\mathrm{SO}(n)$. The orthogonal group has two connected components, namely $\mathrm{SO}(n)$ and its coset.

Given $i, j$ such that $0 < i < j \leq n$, define $R_{i,j}(\theta)$ to be the element of $\mathrm{SO}(n)$ that rotates by angle $\theta$ in the $(i, j)$ plane (right hand rule) and leaves the other coordinates alone. There are $N = n(n-1)/2$ independent such rotations. If we choose the dictionary ordering of the pairs $(i, j)$, this determines an ordered list of $N$ functions of $\theta$,

$$R_{1,2}, R_{1,3}, \ldots, R_{n-1,n} : \mathbb{R}/2\pi\mathbb{Z} \to \mathrm{SO}(n).$$

We may then define the "rotation product" mapping

$$f : \mathbb{T}^N \to \mathrm{SO}(n)$$

by

$$f(\theta_1, \theta_2, \ldots, \theta_N) = R_{1,2}(\theta_1)R_{1,3}(\theta_2)\cdots R_{n-1,n}(\theta_N),$$

where $\mathbb{T}^N = \mathbb{R}^N/2\pi\mathbb{Z}^N$.

**Theorem 9.1** *The mapping $f$ is surjective.*

*That is, every element of $\mathrm{SO}(n)$ may be expressed as a dictionary-ordered product of the $N$ coordinate plane rotations.*

**Proof:** The proof proceeds recursively. Let $A$ be an arbitrary element of $\mathrm{SO}(n)$ and take $B = A^{-1}$.

Our task is to incrementally reduce $B$ to the identity by multiplying on the right by matrices $R_i$ in the correct order, with values of $\theta$ chosen as we go. (This technique is reminiscent of the Householder transformation and other standard matrix reduction techniques (see Press, et. al. (2002))).

Notice that the action of right multiplication by $R_{i,j}(\theta)$ is to take the $i$th and $j$th columns of $B$, $\vec{u}$ and $\vec{v}$, and replace them by $\vec{u}' = \vec{u}\cos\theta + \vec{v}\sin\theta$ and $\vec{v}' = -\vec{u}\sin\theta + \vec{v}\cos\theta$.

Letting $u_1$ denote the first component of $\vec{u}$, and so on, this means that $u_1' = u_1\cos\theta + v_1\sin\theta$ and $v_1' = -u_1\sin\theta + v_1\cos\theta$. As long as $u_1$ and $v_1$

are not both zero, we may always choose $\theta$ so that $u_1' > 0$ and $v_1' = 0$. This principle will be applied repeatedly below.

Attending to the first row of $B$, we choose $\theta_1, \theta_2, \ldots, \theta_n$ as follows. Start with the second element of the first row. If it is zero, set $\theta_1 = 0$. If not, choose $\theta_1$ as above so that right multiplication by $R_{1,2}$ zeroes out this element and leaves the first element of the first row positive.

Continue moving along the first row in this fashion, turning all the elements to zero at the expense of changing the first element to other positive values.

At this point, we have a matrix $C = BR_{1,2}(\theta_1) \cdots R_{1,n}(\theta_{n-1})$ in $\mathrm{SO}(n)$ such that the entries in the first row of $C$ are all zero except the first, which is positive. Since the rows of $C$ are orthonormal, this first entry must be equal to one. Therefore all the entries in the first column of $C$ except the first must also vanish.

We have used all the rotations in our list involving the first coordinate axis, and reduced $B$ to a matrix $C$ of the block diagonal form $I_1 \times D_{n-1}$. Now $D_{n-1}$ is a matrix in $\mathrm{SO}(n-1)$, and the remaining rotations will not touch the first row or column.

We may continue this process inductively until we have reduced $B$ to the form $I_{n-2} \times D_2$, where $D_2$ is a $2 \times 2$ matrix in $\mathrm{SO}(2)$, and used all the rotations except $R = R_{n-1,n}$. Since the first row of $D_2$ cannot be zero, we may right multiply by $R(\theta_N)$ so that the first entry of this row is one. The off-diagonal entries must then be zero, and the remaining entry must be $+1$

since $\det D_2 = 1$.

We have found angles $\theta_1, \ldots, \theta_N$ so that

$$BR_{1,2}(\theta_1)R_{1,3}(\theta_2)\cdots R_{n-1,n}(\theta_N) = I_n.$$

Hence

$$A = B^{-1} = R_{1,2}(\theta_1)R_{1,3}(\theta_2)\cdots R_{n-1,n}(\theta_N)$$

as desired, completing the proof. ∎

It's worth remarking that the mapping $f$ has critical points: if it did not, it would be a covering map from $\mathbb{T}^N$ to $\mathrm{SO}(n)$, and hence injective on the level of the fundamental group. This is impossible since $\pi_1(\mathbb{T}^N) = \mathbb{Z}^N$ and $\pi_1(\mathrm{SO}(n)) = \mathbb{Z}/2\mathbb{Z}$.

Generically, elements in $\mathrm{SO}(n)$ have $2^{(n-1)(n-2)/2}$ $f$-preimages. This is because the proof actually allows us to choose, at each step along the row, one of two possible values of $\theta$ (differing by $\pi$) to zero out the current element, until the last element is reached and $\theta$ is fixed by the requirement that the pivot entry be $+1$ rather than $-1$.

# References

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

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.

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 (1999): *Active Portfolio Management*, New York: McGraw-Hill, 2nd ed.

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.

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.

Weyl, H. (1966): *The Classical Groups*, Princeton: Princeton University Press.