Flexible is better.™

Axioma Research Paper
No. 001

April, 2005

Computing Return Estimation Error Matrices for Robust Optimization

Robert A. Stubbs, PhD
Pamela Vance, PhD

A practical look at methods for computing effective error estimates to be used as inputs for robust optimization.

TABLE OF CONTENTS:

Introduction

Recently, a new technology known as robust portfolio optimization has been introduced to handle some of the ill-effects of classical mean-variance optimization caused by treating the expected return estimates as exact values. Because of the statistical noise inherent in these inputs and the resulting negative effects that classical mean-variance optimization has on these statistical errors, mean-variance optimization is often referred to as error maximization. Robust optimization overcomes these shortcomings by explicitly accounting for the error in the expected return estimates by considering a confidence region around the estimates rather than treating the estimates as exact values.

Specifically, in robust portfolio optimization, we consider all values of the expected return vector, α, that lie in the confidence region $(\alpha - \overline{\alpha})^T \Sigma^{-1} (\alpha - \overline{\alpha}) \le \kappa^2$, where $\overline{\alpha}$ is a point estimate of expected returns, κ is a constant scalar giving the size of the confidence region, and Σ is a symmetric positive semidefinite matrix whose size is equal to the number of assets for which we have estimates of expected return. Robust portfolio optimization therefore requires two additional inputs, Σ and κ, which are an asset-asset covariance matrix and a confidence-controlling scalar, respectively. We will discuss both of these parameters in the remainder of this document.

There is not much science involved in choosing the value of κ to use in robust portfolio optimization. The good news is that it is a single parameter and it is therefore very easy to experiment with potential values. The κ parameter controls the size of the confidence region considered in the robust optimization problem. Logically, the larger the value of κ, the larger the confidence region, and therefore the more likely it is that the true mean of the expected return distribution lies inside the region. Note if $\kappa = 0$, robust optimization is equivalent to classical mean-variance optimization where the only information considered about the alphas are their point estimates. In this case, the point estimate is treated as if it were the true mean of the expected return distribution.

The size of the confidence region and the parameter κ are related to more familiar confidence-interval terminology such as 95% confidence region, 50% confidence region, etc. When thinking in these more traditional terms, the value of κ depends on the level of confidence considered and the distribution of the expected return estimator. Rather than worry about the explicit distribution of the estimator (and its associated error), we prefer to speak directly in terms of κ. This also allows us to view the value of κ as an estimation-error aversion parameter.

In general, it is difficult to get an accurate estimate of the asset-asset covariance matrix, Σ. The good news is that an accurate estimate is not required. In our testing, we have found benefits from using simple diagonal estimation error matrices that can be computed relatively easily. Before we describe some techniques for computing simple and effective covariance matrices, we will describe a method for computing an exact estimation error matrix for a particular alpha construction technique in order to gain a better understanding of exactly what an estimation error matrix is and isn't.

Suppose that the return process is stationary and we are computing our estimate of $\overline{\alpha}$ by the average of a time-series of realized returns of length *T*. In this case, it is well known that the estimation error matrix is *(1/T)Q* where *Q* is the covariance matrix of the time-series of returns. In this example, the estimation error matrix is a constant scalar of the covariance matrix of returns. This alpha construction process is not used in practice and is not recommended. We only use this as an example to illustrate the difference between the covariance of returns and the matrix defining the covariance of errors around the expected return estimates.

Some alpha-generation processes, for example Bayesian and regression-based methods, provide estimation error information as a byproduct of the alpha-estimation process. Other alpha-building approaches use non-linear combinations of signals to produce alphas. For signal-blending approaches, statistical analysis of historical alpha-predictions and realized returns is the best approach for gathering information about the error in the estimation process.

In this paper, we give an overview of methods for computing practical, effective error estimates to be used as inputs for robust optimization.

Using Historical Data to Compute Error Estimates

If alpha estimates are generated using linear transformations of signals with known variance and covariance information, it is possible to derive the estimation error information from the mathematical forms used. However, this is seldom the case. Often, either the statistical properties of the inputs are not well-known or there are nonlinear transformations (for example, conversions of rankings to z-scores) used in the process. When this is the case, analysis of historical data is the most tractable approach for computing the error estimates.

Suppose we have historical data in the form of alpha estimates α^t and realized returns r^t for each security for the past *T* time periods. Note that α^t and r^t are n-dimensional vectors where *n* is the number of securities. For each time period *t*, we estimate the error in the alpha estimate for security *i* as $(\alpha_i^t - r_i^t)$. If we assume that the estimation errors are normally distributed over the time horizon with constant mean and variance, we can compute the variance of the sampling distribution for security *i* as:

$$\hat{\sigma}_i^2 = \frac{\sum_{t=1}^{T} [(\alpha_i^t - r_i^t) - \hat{\mu}_i]^2}{T(T-1)}$$

where

$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T}(\alpha_i^t - r_i^t).$$

Covariances for each pair of securities can be estimated similarly. However, computing a complete asset-asset estimation error matrix including covariances requires an enormous number of observations in order to get a meaningful full-rank matrix. Specifically, for n assets, at least n time periods are needed in order to obtain a full-rank estimation error matrix. This much information is rarely available and the approach is not recommended even if the data is available. Such an approach suffers in several ways just as it does when computing an asset-by-asset risk covariance matrix. Asset-by-asset covariance matrices are rarely used in large-scale equity portfolio risk measurement and should not be used here for the same reasons.

If you wish to work with error information at the asset level, we recommend omitting the covariances. The variance terms described above will specify a diagonal matrix Σ.

Computing Error Distributions Based on Fundamental Factors

In many cases, sufficient data is not available to compute estimates for all securities in the universe on a security-by-security basis. In particular, some securities may have far less historical data available than others making it difficult to assess the alpha-building process for those securities. Also, as noted previously, if off-diagonal elements are to be included in the covariance matrix, the data requirements are prohibitive. In these situations, computing estimation errors at the factor level (e.g., industries) is appropriate.

Note that this type of analysis can be done for any classification of the assets such that each asset is a member of exactly one subset in the classification scheme. So, for example, the same analysis can be done based on country or region membership. In the discussion that follows, we refer to industries but bear in mind that the same computations apply for sectors, countries, regions, or other classification schemes in which each security belongs to exactly one subset.

Suppose we have historical data in the form of alpha estimates α^t and realized returns r^t for the past T time periods along with a classification, $c(i)$, of the securities into industries. Let n_k^t be the number of securities in industry k with alpha and return data available in time period t. For each time period t, we compute the variance of the estimate for industry k as:

$$\hat{\sigma}_{kt}^2 = \frac{\sum_{i:c(i)=k}[(\alpha_i^t - r_i^t) - \hat{\mu}_k^t]^2}{n_k^t(n_k^t - 1)}$$

where:

$$\hat{\mu}_k^t = \frac{1}{n_k^t} \sum_{i:c(i)=k} (\alpha_i^t - r_i^t).$$

If we assume that the estimation errors are normally distributed over the time horizon with constant mean and variance, we can use the sample mean:

$$\hat{\mu}_k = \frac{1}{T} \sum_{t=1}^{T} \hat{\mu}_k^t$$

to estimate the error in alpha for industry *k*. The variance of the sample mean for industry *k* is then computed as:

$$\hat{\sigma}_k^2 = \frac{\sum_{t=1}^{T} \hat{\sigma}_{kt}^2}{T^2}.$$

Covariances for each pair of industries can be estimated as:

$$\hat{\sigma}_{kj}^2 = \frac{\sum_{t=1}^{T} [\hat{\mu}_k^t - \hat{\mu}_k][\hat{\mu}_j^t - \hat{\mu}_j]}{T(T-1)}.$$

These variance and covariance values can be substituted directly into the asset-asset covariance matrix Σ using the variance for industry *c(i)* in place of the variance for security *i* and using the covariance for the industry pair *c(i)*, *c(j)* in place of the asset pair *i, j*. Alternatively, a factor-based model can be produced by creating an industry-industry covariance matrix, and an asset-industry exposure matrix The exposure matrix would have a column for each asset and a row for each industry. The column for security *i* would contain all zeroes except that it would have a 1 in row *c(i)*. This model could be augmented by the addition of a diagonal asset-specific variance matrix, if desired... and a diagonal asset specific variance matrix... and a diagonal asset specific variance matrix. A

Computing the Estimation Error Matrix for Least Squares Regression

If cross-sectional regression models are used to predict asset returns, estimation error information is provided as an output of the regression. In this section, we present the basics of calculating estimation error parameters for multivariate multiple linear regression (i.e., *r* predictor variables, and a *m*-dimensional response vector, *Y*).

Let $z_1, z_2, \ldots, z_r$ be the predictor variables and $Y=(y_1, y_2, \ldots, y_m)$ be the response for a linear regression model. Suppose n observations of the predictors and responses were used to fit the regression model. Let Y be a matrix with n rows and m columns containing the values of the response variables for the n observation, let Z be a matrix with n rows, and $r+1$ columns containing the values of the predictor variables (the first column contains 1's to represent the intercept term), and let $\hat{\beta}$ be the matrix of regression coefficients provided by least squares. That is:

$$Y = \begin{bmatrix} Y_{11} & Y_{12} & \ldots & Y_{1m} \\ Y_{21} & Y_{22} & \ldots & Y_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ Y_{n1} & Y_{n2} & \ldots & Y_{nm} \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & z_{11} & z_{12} & \ldots & z_{1r} \\ 1 & z_{21} & z_{22} & \ldots & z_{2r} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & z_{n1} & z_{n2} & \ldots & z_{nr} \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_{01} & \hat{\beta}_{02} & \ldots & \hat{\beta}_{0m} \\ \hat{\beta}_{11} & \hat{\beta}_{12} & \ldots & \hat{\beta}_{1m} \\ \ldots & \ldots & \ldots & \ldots \\ \hat{\beta}_{r1} & \hat{\beta}_{r2} & \ldots & \hat{\beta}_{rm} \end{bmatrix}.$$

The estimation error for the mean response corresponding to the predictor $\boldsymbol{z_0} = (z_{01}, z_{02}, \ldots, z_{0r})$ can be calculated as:

$$z_0^T (Z^T Z)^{-1} z_0 \hat{\Sigma}$$

where

$$\hat{\Sigma} = \frac{1}{n}(Y - Z\hat{\beta})^T (Y - Z\hat{\beta}).$$

Computing the Estimation Error Matrix for Bayesian Methods

One methodology used to compute expected return estimates is Bayesian statistics. Bayesian methods assume that we are computing an estimate of a distribution of expected returns rather than a single point estimate. This distribution is obtained by combining a "prior" with additional information. The prior estimate of expected returns is generally a stable estimate such as the global cross-sectional mean of expected returns over a time horizon or the market-implied returns. These priors are then combined with investor information such as recent sample means or the investor's own views. We will discuss two of the more common Bayesian techniques found in the literature for computing estimates of expected returns, Bayes-Stein estimation and the Black-Litterman model, as well as a more general Bayesian framework.

The Bayes-Stein Estimator

The Bayes-Stein estimator was developed by Jorion. It uses the average return of minimum variance portfolios as the prior expected return for each asset. Assume that asset returns are normally distributed with known covariance matrix, Q. (We assume this in order to focus on the estimation error in the expected returns.) Suppose that we have a sample of T observations of realized returns having sample mean $\overline{R}$.
Jorion assumes a prior on the expected returns to be:

$$\mu \sim N\left(\mu_0 e, \frac{1}{\tau} Q\right)$$

where e is a vector of all ones, τ is a scalar based on the confidence in the prior, and:

$$\mu_0 = \frac{e^T Q^{-1} \overline{R}}{e^T Q^{-1} e}$$

is a scalar giving the average expected return of the minimum variance portfolio. Then the posterior pdf of the expected returns is $\mu \sim N(\mu_T, \Lambda_T)$ where:

$$\mu_T = \frac{\mu_0 \tau}{\tau + T} e + \frac{T}{\tau + T} \overline{R} \text{ and } \Lambda_T = \frac{1}{\tau + T} Q.$$

Using this approach for computing expected returns, we would let the estimation error covariance matrix:

$$\Sigma = \Lambda_T.$$

The Black-Litterman Model

A more recent Bayesian technique for computing expected return estimates is the so-called Black-Litterman model. The Black-Litterman model combines equilibrium expected returns, or implied expected returns, with investor views to compute an expected return estimate. Let, P be the matrix of investor views, q be the right-handside of these views, and Ω be the covariance matrix of the views. Most papers on the Black-Litterman model assume that the investor's views are independent which implies that Ω is a diagonal matrix. Let π the vector of equilibrium expected returns, Q be an estimate of the covariance of returns, and τ be a scalar measuring the confidence in the market prior such that the distribution of π has covariance (τQ).

The Black-Litterman model combines the market prior and the views to compute an expected return estimate, α, according to:

$$\alpha = \left[(\tau Q)^{-1} + P^T \Omega^{-1} P\right]^{-1} \left[(\tau Q)^{-1} \pi + P^T \Omega^{-1} q\right].$$

The expected return estimate is then assumed to have a Gaussian distribution with mean α and covariance $\left[(\tau Q)^{-1} + P^T \Omega^{-1} P\right]^{-1}$. Therefore, if you use the Black-Litterman model to compute your expected return estimate, then you can use:

$$\Sigma = \left[(\tau Q)^{-1} + P^T \Omega^{-1} P\right]^{-1}.$$

General Bayesian Methods

For Bayesian approaches in general, assume the asset returns are normally distributed with unknown mean, μ, and known covariance matrix, Q. (We assume this in order to focus on the estimation error in the expected returns.) Suppose that we have a sample of T observations of realized returns having a sample mean of $\overline{R}$. Also suppose that we have a conjugate prior defined by $\mu \sim N(\mu_0, \Lambda_0)$. Then the sample and the prior are combined to compute a posterior pdf that is multivariate normal with mean μ_T and covariance Λ_T where:

$$\mu_T = \left(\Lambda_0^{-1} + TQ^{-1}\right)^{-1} \left(\Lambda_0^{-1} \mu_0 + TQ^{-1} \overline{R}\right)$$

and:

$$\Lambda_T = \left(\Lambda_0^{-1} + TQ^{-1}\right)^{-1}.$$

If you are using this approach for computing an expected return estimate, you should let:

$$\Sigma = \Lambda_T.$$

Flexible is better.™

United States and Canada: 212-991-4500
Europe: +44 20 3170 8745
Asia: +852-8203-2790

New York Office
Axioma, Inc.
17 State Street
Suite 800
New York, NY 10004

Phone: **212-991-4500**
Fax: **212-991-4539**

Atlanta Office
Axioma, Inc.
8800 Roswell Road
Building B, Suite 295
Atlanta, GA 30350

Phone: **678-672-5400**
Fax: **678-672-5401**

San Francisco Office
Axioma, Inc.
201 Mission Street
Suite #2230
San Francisco, CA 94105

Phone: **415-614-4170**
Fax: **415-614-4169**

London Office
Axioma, (UK) Ltd.
288 Bishopsgate
London EC2M 4QP

Phone: **+44 20 3170 8745**

Hong Kong Office
Axioma, (HK) Ltd.
Unit C, 17/F
30 Queen's Road Central
Hong Kong

Phone: **+852-8203-2790**
Fax: **+852-8203-2774**

Singapore Office
Axioma, (Asia) Pte Ltd.
30 Raffles Place
#23-00 Chevron House
Singapore 048622

Phone: **+65 6233 6835**
Fax: **+65 6233 6891**

Sales: sales@axiomainc.com
Client Support: support@axiomainc.com
Careers: careers@axiomainc.com