Axioma Robust™ Risk Model Handbook

June 2011

1 What is Risk?

There is no single answer to the question, "*what is the volatility of a given asset or portfolio?*" Economists, for example, typically associate risk with abstract notions of individual preference, whereas financial regulators may prefer a measure such as Value-at-Risk (VaR). Axioma defines risk as the standard deviation of an asset's return over time. This statistical definition is straightforward, broadly-applicable, and intuitive. An asset whose return varies wildly over time is volatile, therefore risky; another whose return remains fairly constant is relatively predictable, and thus less risky.

Throughout the following discussion, $r_{i,t}$ will represent the return to an asset i, at time t:

$$r_{i,t} = \frac{p_{i,t} + d_{i,t} - p_{i,t-1}}{p_{i,t-1}} - r_{rf,t}$$

where $p_{i,t}$ is the asset's price at time t, $d_{i,t}$ is any dividend payout at time t, and $p_{i,t-1}$ is the price at the previous time period, adjusted for any corporate actions (e.g. stock splits). The time increment t used could be days, weeks, months, or any other period.

$r_{rf,t}$ is the *risk-free rate* — the return to some minimal risk entity (e.g. LIBOR rate). Returns net of the risk-free rate are termed *excess* returns and are usually used for the purposes of risk modeling. Henceforth, unless stated otherwise, *return* will be used to mean excess return.

The risk of an asset over T time periods is thus given by

$$\sigma_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_{i,t} - \bar{r}_i)^2}$$

where $\bar{r}_i$ is the asset's mean return over time.

Investors tend to think in terms of portfolios of assets rather than individual assets in isolation. A portfolio allows for diversification and risk reduction, as illustrated by a simple example: consider a portfolio of two risky assets A and B, with weights w_A and w_B. The risk of this portfolio is

$$\sigma(w_A r_A + w_B r_B) = \sqrt{w_A^2\sigma_A^2 + w_B^2\sigma_B^2 + 2\rho_{AB}w_A w_B \sigma_A \sigma_B}$$

where $-1 \leq \rho_{AB} \leq 1$ is the correlation coefficient between the two assets. It can be seen that

$$\sigma(w_A r_A + w_B r_B) \leq w_A\sigma_A + w_B\sigma_B$$

with equality if $\rho_{AB} = 1$. When analyzing portfolio risk, it is therefore necessary to know not only the risk of each asset involved, but also the interplay or co-movement of each asset with every other.

This information is contained in a covariance matrix of asset returns:

$$Q = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & & \vdots \\ \rho_{31}\sigma_3\sigma_1 & \rho_{32}\sigma_3\sigma_2 & \sigma_3^2 & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \cdots & \cdots & \cdots & \sigma_n^2 \end{bmatrix}$$

Note that the matrix is symmetric and positive-semidefinite. The risk of a portfolio h, where

$$h = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

and w_i, $i = 1, \ldots, n$ are the weights of each asset, is simply

$$\sigma_h = \sqrt{h^T Q h}$$

2 Why a Multi-Factor Model?

Possessing an accurate estimate of the asset returns covariance matrix is the *sine qua non* of portfolio risk management. How does one calculate such a matrix in practice? The obvious solution is to build a history of asset returns and then calculate the variances and covariances directly.

Computing sample statistics directly from historical data, however, is fraught with danger. Historical returns are typically noisy; even in the absence of actual data errors, false signals and spurious relationships abound. Two assets may appear closely related when their seemingly-correlated behavior is in fact an artifact of data-mining.

Weak signals and noise aside, when a new asset enters the existing universe, there is no reliable way of calculating its relationships with the other assets, because it does not yet possess a returns history. One could construct various proxies, but such an approach is dubious at best.

Finally, data points totalling no less than the number of assets are required to accurately estimate all the variances and covariances directly. For any realistic number of assets, it is extremely unlikely that sufficient observations exist. Even with a universe of 100 assets, over $\frac{1}{2} \cdot 100 \cdot (100 + 1) > 5,000$ relationships need to be estimated. For stock markets like the U.S. (over 12,000 assets), this becomes completely infeasible.

Any one of the above problems is sufficient reason against constructing an asset returns covariance matrix directly. A better approach is to first impose some structure on the asset returns by identifying *common factors* within the market — that is, factors which drive asset returns. Returns can then be modeled as a function of a relatively small number of parameters, and estimating thousands, or tens, even hundreds of thousands, of asset variances and covariances can thus be simplified to calculating a much smaller handful of numbers.

Factors used in multi-factor models can fall into several broad categories:

- *Fundamental factors*
  - *Industry and country factors* reflect a company's line of business and country of domicile.

  - *Style factors* encapsulate the financial characteristics of an asset — a company's size, debt levels, liquidity, etc. They are usually calculated from a mixture of market and fundamental (i.e. balance sheet) data.
  - *Currency factors* represent the interplay between local currencies of the various assets within the model.
  - *Macroeconomic factors* capture an asset's sensitivity to variables such as GNP growth, bond yields, inflation, etc.

- *Statistical factors* are mathematical constructs responsible for the observed correlations in asset returns. They are not directly connected to any observable real-world phenomena, and may change from one period to the next.

An asset's return is decomposed into a portion driven by these factors (*common factor return*) and a residual component (*specific return*), producing the following model at time t:

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ b_{21} & & b_{2m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix} \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

or more succinctly, in matrix form:

$$r = Bf + u$$

where r is the vector of asset returns at time t, f the vector of factor returns, and u, the set of asset specific returns. B is the $n \times m$ exposure matrix. Its elements denote each asset's exposure to a particular factor.

Theoretical foundations for this model specification make up the Arbitrage Pricing Theory (APT), proposed by Ross (1976) as a generalization of the traditional Capital Asset Pricing Model to allow for multiple risk factors. The APT is therefore a logical starting point for building a factor risk model.

As an example, consider a simple model with four assets, a, b, c and d, and three factors — two industry factors, *IT* and *Banking*, and one style factor — *size*, which takes values of $\{-1, 0, 1\}$ to represent small, mid-cap and large companies respectively[1]. Assets a and b are IT companies, while c and d are banks. a and b are small-cap, c is mid-cap and d is large-cap. The exposure matrix is thus:

$$B = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{ccc} IT & bank & size \end{array} \\ \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \end{array}$$

r and B are known, so the system of equations can be solved for f and u.

This process is repeated over time to build two time series — one of factor returns and another of asset specific returns. Given that there is a comparatively small number of factors, it is feasible to estimate a factor covariance matrix directly from the factor returns time series. Moreover, assuming that asset specific returns are uncorrelated both amongst themselves and with factor returns, the $n \times n$ asset returns covariance matrix becomes:

$$var(r) = var(Bf + u)$$
$$\hat{Q} = B\Sigma B^T + \Delta^2$$

[1] In reality, style exposures usually take a continuum of values rather than simple scores as above.

where Σ is the $m \times m$ factor covariance matrix and Δ^2 is the diagonal matrix of specific variances.

In essence, the multi-factor model is a *dimension reduction* tool, simplifying the problem of calculating an $n \times n$ asset returns covariance matrix into calculating the variances and covariances of a much smaller number of factors, and n specific variances.

The following sections will discuss in greater detail the various parts of a risk model and the stages in its construction. Interested readers may wish to consider Grinold and Kahn (1995) or Zangari (2003) for a full exposition on factor risk models and their applications.

3 The Returns Model

Recall the linear factor model of asset returns:

$$r = Bf + u$$

There are many possible solutions to this system of equations. If factor exposures B are known, f can be estimated using cross-sectional regression analysis. With macroeconomic factors, however, f is observed, and it is B rather, that needs to be estimated, typically via time-series regression for each asset. In the case of statistical factors, neither B or f is specified, so a rotational indeterminacy exists and both parameters are determined simultaneously, albeit only up to a nonsingular transformation.

For a more thorough discussion of multi-factor models and the APT, the curious reader is encouraged to consult Campbell et al. (1997).

3.1 The Least-Squares Regression Solution

The *ordinary least-squares* (OLS) regression solution to the factor model of returns seeks to minimize the sum of squared residuals:

$$f_{ols} = arg \min_{f} \sum_{i=1}^{n} u_i^2$$

whose solution is straightforward:

$$\hat{f}_{ols} = \left(B^T B\right)^{-1} B^T r$$

3.1.1 Assumptions of the Least-Squares Solution

The field of regression theory is vast, so only the issues most relevant to risk modeling will be dealt with here. Further details can be found in any elementary econometrics textbook, such as Greene (2003, chap. 4,5).

1. *B is a $n \times m$ matrix with full column rank $\rho(B) = m \leq n$.* The OLS solution requires that $B^T B$ be invertible, which is satisfied only if the columns of B are linearly independent. Intuitively, this means the factors should all be distinct from one another.

2. *Residuals are zero-mean and independent of the factor exposures* . In order for the regression estimates to be *unbiased* — correct "on average", $E\left[u\right] = 0$ and $E\left[B^T u\right] = 0$ are required.

3. *Residuals are homoskedastic and have no autocorrelation.* These constitute the *Gauss-Markov conditions*: $var(u_i) = \sigma^2$ and $cov(u_i, u_j) = 0$ for all u_i, $i = 1, \ldots, n$, $i \neq j$ (or more compactly, $E\left[uu^T\right] = \sigma^2 I_n$) and establishes the superiority of the least-squares solution over all other linear estimators. Unfortunately, large assets tend to exhibit lower volatility than smaller ones, and homoskedastic residual returns are rarely observed. Figure 1 shows the typical relationship between asset size and returns behavior.

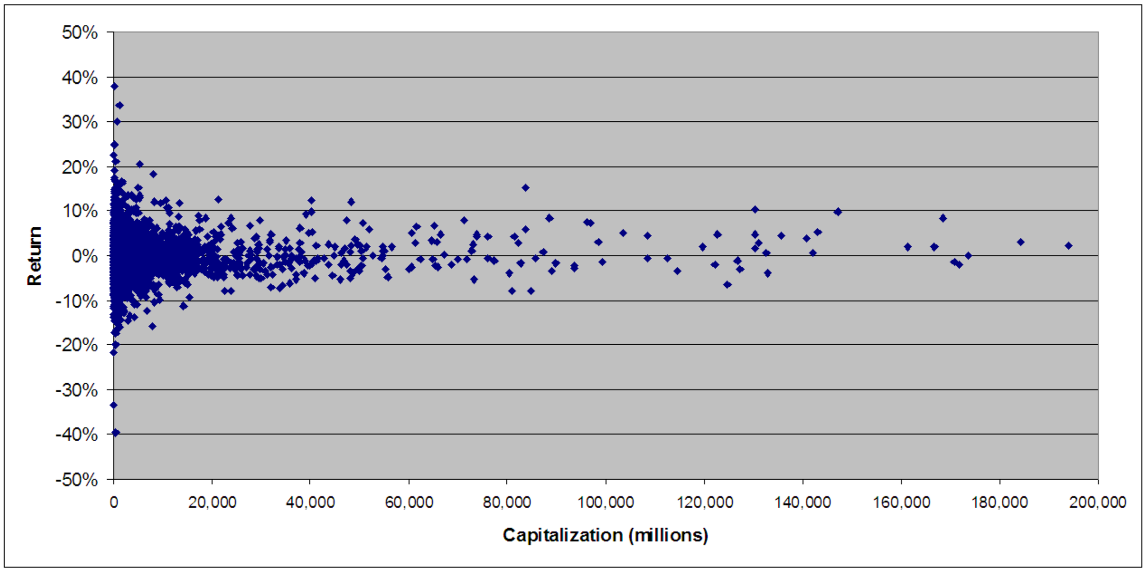


Figure 1: Daily returns vs. market capitalization of FTSE Global Index stocks, January 31, 2000

4. *Residuals are normally-distributed*: strengthening the previous assumption to $u \sim N(0, \Omega)$ where $\Omega = \sigma^2 I_n$ is not strictly required. Nevertheless, it is a convenient assumption for testing the estimators, to simplify constructing confidence intervals, evaluating hypothesis tests, and so forth.

3.1.2 Solving the Problem of Heteroskedasticity

Traditionally, one corrects for this phenomenon by scaling each asset's residual by the inverse of its residual variance, transforming the above into a *weighted least-squares* (WLS) problem:

$$W^{1/2}r = W^{1/2}Bf + W^{1/2}u$$

$$f_{wls} = arg \min_f \sum_{i=1}^{n} \frac{u_i^2}{\sigma_i^2}$$

where

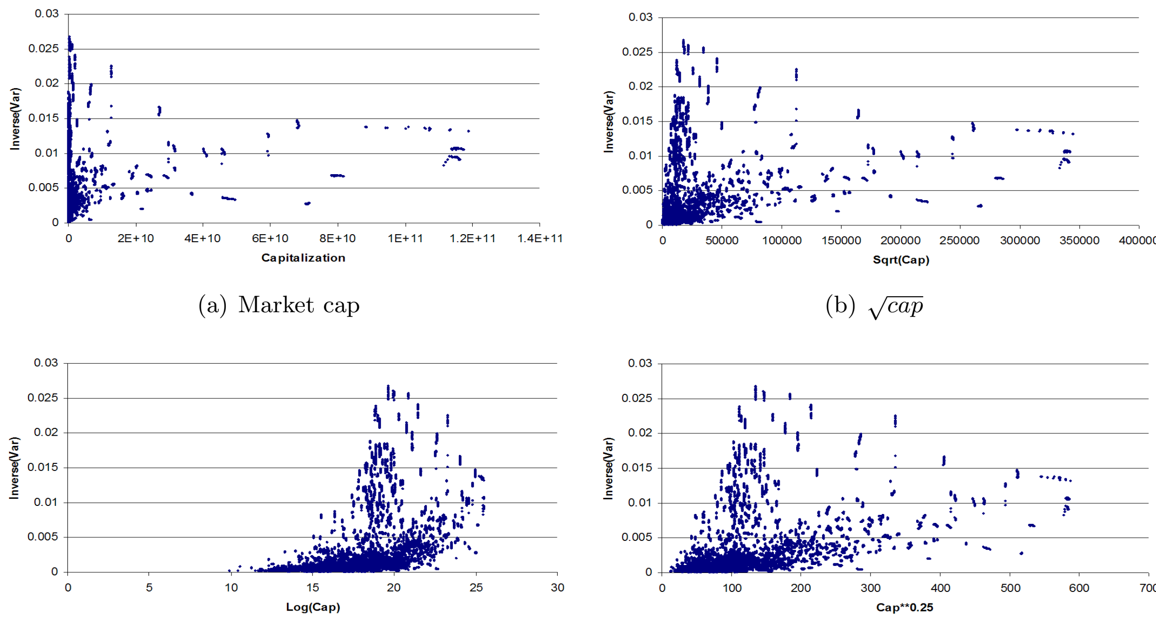
$$W^{1/2} = \begin{bmatrix} \sigma_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \sigma_n^{-1} \end{bmatrix}$$

The solution is easily shown to be

$$\hat{f}_{wls} = \left(B^T W B\right)^{-1} B^T W r$$

The challenge lies in estimating the residual variances, σ_i^2. One could calculate these directly from historical data, but such estimations are noisy and require sufficient history for each asset. As a proxy for the inverse residual variance, most Axioma models use the square-root of each asset's market capitalization.

Using 2006 data from the Axioma U.K. Risk Model, Figure 2(b) shows the typical relationship between the inverse residual variance and the square-root of market capitalization.



(a) Market cap

(b) $\sqrt{cap}$

(c) $\ln(cap)$

(d) $\sqrt[4]{cap}$

Figure 2: Relationship between various cap-based weighting schemes and inverse residual variance

Although the plot does not yield a straight line, there is certainly a stronger case than weighting by capitalization itself. Figure 2(a) demonstrates this: with the bulk of market capitalization concentrated among a small number of mega-cap assets, clearly asset capitalization would be a very poor substitute for inverse residual variance.

Figure 2(c) shows natural logarithm of asset capitalization versus residual variance. This proxy, however, is too much of a "leveler", treating everything almost equally. Finally, 2(d) looks at the fourth-root of capitalization, which lies somewhere between the square root and the log.

Through trial and error, one may find the optimal weight to be some fractional power, or more exotic function, of capitalization. The square root proxy, however, is tried and tested, simple, and largely accepted by the industry as a standard.

There is also a practical reason for adopting a weighted regression. Using a weighting scheme such as square-root capitalization "tunes" the regression estimates in favor of larger assets. Large, liquid assets constitute the bulk of most institutional investors' universes, so there is a compelling case for modeling these assets accurately, sometimes even at the expense of smaller, less important assets.

3.1.3 Outliers

Data frequently contains extreme values, or *outliers*, arising from outright errors in the data collection process, poor or unrepresentative sampling, or genuinely aberrant behavior. In particular, distributions of asset returns are known to exhibit "fat-tails" — large numbers of observations in the outer edges of the distribution, most likely attributable to economic shocks, poor liquidity, etc..

Because least-squares attempts to minimize the sum of squared residuals, outlier returns produce large residuals that have a disproportionate effect on the solution, pulling it away from the "true" solution.

As an example, consider a vector of returns

$$r = \begin{bmatrix}1.0 & 3.0 & 5.0 & 7.0 & 9.0 & 2.0 & 4.0 & 6.0 & 2.0 & 0.0\end{bmatrix}^T$$

an exposure matrix

$$B = \begin{bmatrix}1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1\end{bmatrix}^T$$

and weights

$$w = \begin{bmatrix}2.0 & 1.0 & 3.0 & 4.0 & 6.0 & 1.0 & 8.0 & 1.0 & 3.0 & 5.0\end{bmatrix}^T$$

The weighted least-squares solution is

$$\hat{f}_{wls} = \begin{bmatrix}6.375\\ 2.556\end{bmatrix}$$

Suppose the returns were altered ever so slightly, leaving exposures and weights unchanged:

$$\tilde{r} = \begin{bmatrix}10000.0 & 3.0 & 5.0 & 7.0 & 9.0 & 2.0 & 4.0 & 6.0 & 2.0 & 0.0\end{bmatrix}^T$$

An intelligent observer will easily look at these new returns and identify the first observation (being significantly larger than the rest) as an aberration and discard it from the computation. Failing to exclude the outlier yields the solution

$$\tilde{f}_{wls} = \begin{bmatrix}1256.250\\ 2.556\end{bmatrix}$$

which has obviously been substantially pulled away from the "true" solution. Note that the second element of the solution remains unchanged; this is due to the structure of B in the example — the first five assets have zero exposure to the second factor.

While this example is trivial in size and extreme in behavior, it illustrates a very real problem. Financial data is replete with outliers, legitimate and illegitimate, which obscure the genuine signals one is trying to trace. A simplistic solution may be to exclude or truncate all observations outside certain pre-determined bounds. These bounds can either be static values (*e.g. exclude returns above 200%*) or statistical-based (*e.g. exclude returns beyond 3 standard deviations from the mean*).

Hard-coded figures are inadvisable as they take no account of changing market conditions over time and the resulting changes in the distribution of the returns: an extreme return during an uneventful period may be unremarkable during a market boom or bust. Actual use of the standard deviation as a measure is not robust either, as this itself is very sensitive to outliers.

Axioma models uses *robust statistical methods* to reduce the effect of outliers, described in the section below.

3.1.4 Robust Regression

A comprehensive treatment of robust regression is far beyond the scope of this discussion[2]; rather, focus will be limited to the techniques Axioma uses in its model construction.

The weapon of choice is a *Huber M-Estimator* (where M stands for *Maximum-Likelihood*). Whereas the weighted least-squares regression seeks to minimize the sum of squared residuals

$$f_{wls} = arg\min_f \sum_{i=1}^{n} w_i u_i^2$$

[2] An thorough introduction to robust regression can be found in Fox (2002).

a robust regression estimator attempts to minimize the objective function

$$f_{robust} = arg \min_f \sum_{i=1}^{n} \rho(u_i) = arg \min_f \sum_{i=1}^{n} \rho(r_i - b_i^T f)$$

The function ρ fulfills the following criteria for all a, b:

- $\rho(a) \geq 0$
- $\rho(0) = 0$
- $\rho(a) = \rho(-a)$
- $\rho(a) \geq \rho(b) for |a| > |b|$

Denoting the derivative of ρ with respect to f as $\rho' = \varphi(a)$, the minimization program has the first-order conditions

$$\sum_{i=1}^{n} \varphi(r_i - b_i^T f) b_i^T = 0$$

Defining the *weight function* $w(a)$ as $w(a) = \varphi(a)/a$, and writing $w(u_i) = w_i$, the above becomes

$$\sum_{i=1}^{n} w_i (r_i - b_i^T f) b_i^T = 0$$

which can be solved via *iteratively re-weighted least-squares*:

1. Using ordinary least-squares, solve the transformed system

   $$\hat{r} = \hat{B} f + \hat{u}$$

   where $\hat{r} = W^{1/2} r$ and $\hat{B} = W^{1/2} B$, obtaining an initial estimate f^0.

2. For each iteration j, calculate residuals u_i^{j-1} and weights $w_i^{j-1} = w(u_i^{j-1})$.

3. Update the weighted least squares estimator

   $$f^j = \left(B^T W^{j-1} B\right)^{-1} B^T W^{j-1} r$$

4. Repeat steps 2 and 3 until successive estimates converge.

There are many different possibilities for the function ρ. Axioma models use the Huber function:

$$\rho(u_i) = \begin{cases} \frac{1}{2} u_i^2 & \text{if } |u_i| \leq k \\ k|u_i| - \frac{1}{2} k^2 & \text{if } |u_i| > k \end{cases}$$

with $k = 1.345\sigma$ where σ is the standard deviation of the residual.

Returning to the simple example in the previous section, robust regression gives the solution

$$\hat{f}_{robust} = \begin{bmatrix} 7.901 \\ 2.556 \end{bmatrix}$$

and the final regression weights are

$$w = \begin{bmatrix} 0.00106 & 1.0 & 3.0 & 4.0 & 6.0 & 1.0 & 8.0 & 1.0 & 3.0 & 5.0 \end{bmatrix}$$

The outlying return has been down-weighted by a factor of approximately 1000, thus bringing the solution back towards the original values.

3.2 Statistical Approaches

Statistical factor models deduce the appropriate factor structure by analyzing the sample asset returns covariance matrix. There is no need to pre-define factors and compute exposures, as required by fundamental factor models. The only inputs are a time-series of asset returns and the number of desired factors. There are no concrete rules specifying the appropriate number of factors. One can, for example, use a Scree plot to assess the variance explained by each additional factor.

Factor analysis is used to estimate the factor exposures B and factor returns f. *Principal components* and *maximum likelihood* estimation are two popular methods of parameter estimation. These, as well as others, are explained in detail in Johnson and Wichern (1998). Axioma risk models involving statistical factors typically employ a variation of principal components, described in detail below.

3.2.1 Principal Components

Principal components analysis (PCA) determines factors by eigendecomposition of the observed asset returns covariance matrix. Given an $n \times t$ matrix R of historical asset returns,

$$\hat{Q} = \frac{RR^T}{t} = UDU^T$$

Only the largest m eigenvalues and corresponding eigenvectors are kept, hence

$$\hat{Q} = U_m D_m U_m^T + \Delta^2$$

where Δ^2 is the $n \times n$ diagonal matrix of specific variances.

The exposure matrix B is taken to be $U_m D_m^{\frac{1}{2}}$. Factor and specific returns can then be computed by regressing, for each asset, historical returns against its exposures using ordinary or weighted least-squares:

$$\begin{aligned} F &= \left(B^T W B\right)^{-1} B^T W R \\ \Gamma &= R - BF \end{aligned}$$

Unlike cross-sectional regression, the matrices (not vectors) F and Γ ($m \times t$ and $n \times t$, respectively), contain the *entire* time-series of factor and specific returns, and will subsequently be used to compute the factor covariance matrix and specific variances.

Additionally, one could try to perform the above estimation using only a subset of assets, both for computational efficiency and to avoid noisy historical data. Exposures and specific returns for assets outside this universe can be backed-out via ordinary least-squares. For each asset i,

$$\begin{aligned} r_{i,t} &= \sum_{j=1}^{m} b_{i,j} f_{i,t} + u_{i,t} \\ b_i &= \left(FF^T\right)^{-1} F r_i \\ B &= \begin{bmatrix} b_i^T \\ \vdots \\ b_n^T \end{bmatrix} \end{aligned}$$

3.2.2 Asymptotic Principal Components

PCA requires that the number of assets n be smaller than the number of time periods t, in order for $\hat{Q}$ to be reliably estimated. In most capital markets, the number of assets far exceeds the number of time periods for which data is available, especially if returns are measured in weekly or longer intervals. If daily data were used to model the U.S. stock market, for example, one would require well over 10,000 data points — over 40 years of data! Even if such data were available, the resulting covariance matrix would be poorly estimated as noise and shocks from long in the past continue to influence current estimates.

To address this inherent shortcoming of PCA, Axioma adopted the following [3]: instead of working with the $n \times n$ covariance matrix $\hat{Q} = \frac{1}{t} R R^T$, one can shift the analysis from n- to t-space, and use the $t \times t$ covariance matrix $\tilde{Q} = \frac{1}{n} R^T R$. The raw input data, R, remains unchanged. The process begins with eigendecomposition of the covariance matrix $\tilde{Q}$:

$$\tilde{Q} = \tilde{U} \tilde{D} \tilde{U}^T = \tilde{U}_m \tilde{D}_m \tilde{U}_m^T + \Phi$$

Because there is an infinity of solutions to $R = BF + \Gamma$, one can choose $F = \tilde{U}_m^T$. Choosing the eigenvectors (right singular vectors, if using singular value decomposition) is convenient because they provide orthogonal factors. Regressing asset returns against factor returns produces B:

$$B^T = \left(F F^T\right)^{-1} F R^T = F R^T$$
$$B = R \tilde{U}_m$$

At this stage, the factor returns and exposure estimates F and B are discarded, keeping only the specific returns Γ, which are used to estimate a diagonal matrix of asset specific variances:

$$\Delta^2 = \frac{1}{t} diag \left(\Gamma \Gamma^T\right)$$

The resulting specific *risks* are then used to scale each asset's returns. This is conceptually analogous to weighting assets by inverse residual variance in the earlier regression example, whereby the influence of more volatile assets is reduced:

$$R^* = \Delta^{-1} R$$

The scaled returns are then used to compute a new $t \times t$ covariance matrix, which will undergo the eigendecomposition routine again:

$$\tilde{Q}^* = \frac{1}{n} R^{*T} R^*$$
$$\tilde{Q}^* = U_m^* D_m^* U_m^{*T} + \tilde{\Phi}$$

The top m eigenvectors U_m^{*T} are chosen as the final estimates for the factor returns history. The final exposures B are calculated by regressing these against the *unscaled* asset returns R.

The above estimation can be carried out on a subset of assets to prevent noisy returns from contaminating the eigenstructure analysis. The procedure for doing so and recovering the exposures and specific returns for the remaining assets is similar to the procedure outlined for PCA.

Finally, it is easily seen that this methodology (and PCA) is equivalent to performing least-squares regression of asset returns against factor returns (or exposures) estimates. Therefore, regression statistics such as $\bar{R}^2$, standard error, or t-statistics are applicable, though such measures are likely to be rather high, because computing them is analogous to having look-ahead bias in a least-squares estimator.

[3] This novel approach was introduced by Connor and Korajczyk (1988), specifically to address the inherent problem of applying PCA to financial markets, where assets typically outnumber time periods.

3.2.3 Maximum-Likelihood Estimation

Assuming factor and specific returns are jointly normally-distributed and *iid* over time, factor exposures B and specific variances Δ^2 can be estimated as maximizers of the likelihood function

$$L\left(B,\Delta^2\right) = (2\pi)^{\frac{-nm}{2}}\, det\left(BB^T+\Delta^2\right)^{\frac{n}{2}}\, e^{\frac{-1}{2}tr\left[\left(BB^T+\Delta^2\right)^{-1}\hat{Q}\right]}$$

where $\hat{Q} = \frac{1}{t}RR^T$, the sample asset returns covariance matrix. Imposing the constraint that $B^T\Delta^2B$ be diagonal ensures a unique solution. Considering the log-likelihood leads to a maximization of

$$\frac{-n}{2} ln\left[det\left(BB^T+\Delta^2\right) + tr\left(\left(BB^T+\Delta^2\right)^{-1}\hat{Q}\right)\right]$$

which can be solved by numerical methods, usually iteratively. Like PCA, factor returns can be recovered by time-series regression of asset returns against the factor exposures.

The optimization problem is a formidable task: even with a universe of 500 assets and 20 factors, 10,000 exposures + 500 specific variances = 10,500 parameters have to be determined simultaneously. Solving a single instance of such a problem may be trivial by today's computing standards, but because Axioma models are estimated on a daily basis, building a model history becomes prohibitively time-consuming.

The question of which factor analysis methodology works best is difficult to answer, and evidence favoring any one particular technique is mixed at best. Axioma models prefer the asymptotic principal components approach because it requires far fewer assumptions on the data and is computationally less burdensome.

3.3 Hybrid Solutions

Fundamental and statistical factors represent complementary approaches to modeling return and forecasting risk. Fundamental factors capture distinct commonalities shared only by a narrow cross-section of assets. Statistical factors may not possess a fine enough "resolution" to differentiate between semiconductor manufacturers and consumer electronics makers, but fund managers will require and appreciate that level of differentiation. On the other hand, statistical factors can quickly reveal any broad factors missed by the fundamental model, particularly useful when market conditions differ widely from those on which the model specification was originally based (such as the turbulent period in July-August 2007).

Despite creating additional computational complexity, a "hybrid" factor model may capture the best of both worlds. A set of pre-specified and intuitive factors account for most of the common variation in a portfolio, while a small number of statistical factors that pick up any remaining (possibly transient) effects in the recent returns history. Should a statistical factor turn out to be significant risk factor, a hybrid model will nevertheless fail to explain exactly what that factor represents, but the model user will at least have the comfort of knowing that all bets in his portfolio have been captured.

In practice, this is most easily implemented by keeping the fundamental factor model intact, and looking for statistical factors within its specific returns. In theory, the reverse also works, but is likely to result in a disproportionately large fraction of return attributed to a single statistical factor, thereby diminishing the power of the fundamental factors when regressed against the statistical model's residuals. A comparison of fundamental versus statistical models, and combinations thereof, is presented in Connor (1995).

4 Estimation Universe

When estimating factor returns one seldom uses the entire universe of asset returns from the market being modeled. The broad market may contain many illiquid assets as well as other potentially "problematic" assets such as investment trusts, depository receipts, foreign listings, to name a few.

Illiquid assets lack a stable, regular price, making their volatility difficult to estimate reliably. *Composite assets* such as investment trusts and ETFs are difficult to quantify in terms of their exposure to the market — if such an instrument invests in several different sectors or markets, how should one define its industry/country exposures? Similarly, is an ADR's return primarily driven by U.S. market, or by the behavior of the underlying stock?

These are difficult questions to answer and, depending on the market and the model, different assets need to be excluded from the estimation process.

Ideally, the estimation universe should contain everything that is "important", relevant within the model universe. Most crucially, the estimation universe should include sufficient assets to keep the number of "thin" factors to a minimum. These are factors to which only a small handful of assets have nonzero exposure, and are among the worst evils one can encounter while modeling return, for reasons to be explained later.

The choice of estimation universe is therefore a subjective one, depending on the market being modeled and the model itself. One tried and tested means of devising an estimation universe is to use membership in an appropriate market index as the basis for inclusion.[4] Benchmark providers typically employ sophisticated selection criteria involving price activity, liquidity, capitalization free-float, and other business logic that would be cumbersome to replicate in-house. Legal implications notwithstanding, one could leverage off their research expertise.

5 Model Factors

5.1 Market Factor or Intercept

Sometimes we may wish to include a single factor that defines the particular broad market behaviour. This could be one of the following:

1. Simple intercept: Every asset has exposure of one to this factor and the factor return is simply the regression intercept term.

2. Market beta: Each asset's exposure is calculated as its historical beta against a suitable market return (from a benchmark for instance), viz.

$$r_{i,t} = \beta_i r_{M,t} + \alpha_{i,t} \quad \text{for each asset } i = 1, \ldots, n$$

   where r_M, t are the returns of the market benchmark index.

3. Macroeconomic factor: Given an appropriate market return (again, perhaps from published benchmark figures), estimate each asset's exposure to this return via historical time-series regression.

[4] For example, the estimation universe in the Axioma U.K. Risk Model is based on rules similar to those used in construction of the FTSE All-Share Index.

5.2 Industry Factors

Returns of assets belonging to companies with similar lines of business often move together. In fact, industry factors are often the most important set of factors in a fundamental factor model, in terms of their explanatory power. Industry factors are formed from a set of mappings from each asset to one or more industries within the market. The mapping used can be a proprietary third-party scheme such as *GICS* or *ICB*, or a modification thereof, consolidating, splitting, or discarding industries where appropriate. Alternatively, one could create a new scheme altogether by investigating company fundamentals, or by statistical analysis of asset returns. Custom-tailored industry schemes may provide marginally more explanatory power but require intensive work to design. Wherever possible, Axioma prefers industry factor structures that correspond directly to classifications widely accepted among the investment community, particularly if an industry scheme is already a de-facto market standard.

Axioma generally defines industry exposures as 0/1 dummy variables. An asset has unit exposure to the industry corresponding to its company's main line of business, and zero to other industries. One could assign an asset multiple industry exposures — each exposure being a fraction representing the proportion of activity within that industry. For most markets, however, this level of detail is not easily available, and researching this information may require a significant amount of effort. Moreover, there is little or no evidence that multiple-industry schemes yield significantly more explanatory power than a simple dummy variable scheme.

An important consideration in designing industry factors is avoiding *thin industries* within the model. A thin industry contains very few or no assets, or has most of its market capitalization concentrated in a small handful of assets.

5.2.1 Thin Industries

Thinness presents a very real problem for factor returns estimation. An industry factor return is predominantly a weighted average of asset returns within the industry. With sufficient assets, the specific return is averaged out, leaving something (hopefully) close to the true factor return. If there are very few stocks, however, a large element of specific return will likely remain, over-estimating the industry factor return and inducing correlations in the specific returns. To see why, consider a simple model with two assets in industry k, equally-weighted for the sake of simplicity:

$$\begin{aligned} r_1 &= f_k + u_1 \\ r_2 &= f_k + u_2 \end{aligned}$$

This has solution $f_k = \frac{1}{2}(r_1 + r_2)$ and residuals $u_1 = -u_2$, with perfect negative correlation between the two assets' specific returns. In general, if there are n *mega-cap* assets within an industry, all with similar weights, the average correlation across their specific returns will be of the order of $\rho = -1/(n-1)$.

As stated previously, uncorrelated specific returns constitute a major assumption of the factor risk model. If this condition is violated, portfolio specific risk may be over-estimated, because the specific return correlations are not taken into account.

Because high concentrations of industry capitalization among a few assets is a common phenomenon in developed markets, the number of stocks in an industry is an inadequate measure of thinness. Many Axioma models rely on the "Herfindahl Index", an economic measure of the size of firms relative to their industry, frequently invoked in antitrust applications:

$$HI = \sum_j w_j^2$$

where w_j is stock j's weight within the industry. $\frac{1}{HI}$ then represents the "effective" number of stocks within the industry. An industry with 100 assets but 99% of its capitalization in one asset has a Herfindahl Index of 0.9801, and effective number of stocks equal to 1.0203. It is trivial to see that for an industry where every asset has equal weight, the effective and actual numbers of stocks are exactly equal.

The prevalence of concentrated industry capitalization is another reason why Axioma prefers square-root of capitalization weighting in the factor returns regression. Using capitalization weighting as-is greatly reduces the effective number of assets in model industries.

5.2.2 Treatment of Thin Industries

Thin industries can sometimes be avoided by merging similar industries within the same sector with closely correlated returns. This approach, however, is not always possible; as market structure evolves, a thin industry may become more populated as time goes by, or vice versa.

Some Axioma models introduce a dummy asset within each thin industry, with market return and unit exposure to the industry, zero elsewhere. Its inclusion will therefore only directly affect the thin industry's factor return estimate. In order to reduce the impact of the other assets within the thin industry, this asset's regression weight is defined as

$$w_{dummy,j} = \begin{cases} (\phi - 1)\,\frac{s_j^4 - \phi^4}{1-\phi^4} \cdot \frac{w_j}{s_j} & \text{if } HI_j^{-1} < \phi \\ 0 & \text{if } HI_j^{-1} > \phi \end{cases}$$

where $s_j = HI_j^{-1}$ is the reciprocal of the Herfindahl Index, w_j is the total regression weight of all assets in the industry, and ϕ is the effective number of assets below which an industry is deemed "thin". This functional form ensures that the dummy asset's weight decreases slowly as the effective number of assets increases from 1, but decays rapidly as the effective number approaches the limit ϕ. A great deal of experimentation has indicated that $\phi = 6$ is a reasonable threshold. Beyond this number returns diversify sufficiently, but this will also depend on the particular market.

Clearly, the thinner the industry, the greater the weight of the dummy asset. This pulls the factor return estimate closer to the market return as the industry becomes thinner, reflecting a decreasing certainty over the true industry factor return in the face of sparse data.

Applying this correction to the earlier example, the new factor returns are

$$\tilde{f}_k = \frac{w_{dummy}}{w_k} r_M + \left(1 - \frac{w_{dummy}}{w_k}\right)\frac{1}{2}(r_1 + r_2)$$

with specific returns

$$\tilde{u}_1 = u_1 + \frac{w_{dummy}}{w_k}\left(\frac{1}{2}(r_1 + r_2) - \tilde{f}_k\right)$$
$$\tilde{u}_2 = -u_1 + \frac{w_{dummy}}{w_k}\left(\frac{1}{2}(r_1 + r_2) - \tilde{f}_k\right)$$

5.3 Style Factors

Whereas industry factors capture trends on the broad market level, style factors capture behavior on an asset level, net of the market.

Style factor exposures are derived from market data such as return, trading volume, capitalization, and/or balance sheet (fundamental) data. For instance, to construct a style factor encapsulating the size of an asset's parent company relative to others in the market, one might consider a function of the company's market capitalization, total assets, etc. In addition to size, common style measures include historical volatility, market sensitivity, liquidity, leverage and value/growth.

Any measure that can be calculated and for which sufficient data exists can be considered as a style factor. Typically, when a model is first estimated, many candidate factors are created. These are then filtered down to an "optimal" set, based on some considerations:

- *Data availability* — there must be sufficient depth and breadth of data to be able to calculate a variable reliably.
- *Reasonableness* — a style factor must be intuitive to investors, describing a sensible characteristic of an asset or company. The average shoe size of a company's employees may carry explanatory power, but is unlikely to appeal to users.
- *Empirical performance* — a factor must be statistically significant, capable of explaining a non-negligible portion of returns variability.

5.3.1 Standardization of Style Factors

Because style factor definitions are expressed in a mixture of units, it is best to standardize them to ensure a level of consistency across the regression estimates. Failure to do so may result in scaling problems in the regression or possibly an ill-conditioned covariance matrix[5].

To make a set of factor exposures b "unit-less":

1. Using only the assets in the estimation universe portfolio h_U, calculate the capitalization-weighted mean exposure:
$$\bar{b} = b^T h_U$$
2. Calculate the equal-weighted standard deviation of the exposure values in h_U about the capitalization-weighted mean (zero, per above step):
$$\sigma_b = \sqrt{\frac{1}{N-1}\sum_{i\in U}(b_i - \bar{b})^2}$$
3. Subtract the weighted mean exposure from each asset's raw exposure and divide by the equal-weighted standard deviation:
$$\hat{b}_i = \frac{b_i - \bar{b}}{\sigma_b}$$

This ensures that the estimation universe has a capitalization-weighted mean of zero and standard deviation of one; in other words, the "market" portfolio is factor neutral.

Outliers in the raw exposure data, however, can skew the standardizing statistics, so raw exposures are typically winsorized prior to standardization. Given an upper bound b_U and lower bound b_L, for a given raw style factor exposure, b_{raw}, set:

$$\tilde{b_i} = \begin{cases} b_L & b_{raw} < b_L \\ b_{raw} & b_L < b_{raw} < b_U \\ b_U & b_{raw} > b_U \end{cases}$$

Axioma models employ different methods to pick b_U and b_L:

[5] Large spread of eigenvalues, causing numerical instability particularly when inverting.

- *Absolute values.* One may, for instance, have a prior belief that the data should lie within the interval [-5, 5], and set the bounds accordingly.
- *Standard deviation.* Compute the standard deviation of the entire set of exposures, then pick a multiple k of the standard deviations and truncate all values above or below k standard deviations from the (possibly weighted) mean.
- *Robust statistics.* Rather than using the mean and standard deviation, employ the median and *median absolute deviation* (*MAD*):

$$MAD(b) = median\,(|b - median(b)|)$$

  Then, pick a number of MADs above and below the median, and use the corresponding values to winsorize.

The first method relies upon having some idea of what "reasonable" values are, and may be suitable for data one is familiar with, such as market betas. For more complicated factor definitions, a dynamic approach may be more appropriate.

5.4 Country Factors

A risk model covering more than one market needs to distinguish between assets' countries of incorporation[6]. Types of country exposure include:

- *Dummy values* — assets have unit exposure to its country of incorporation, zero otherwise.
- *Market beta* — asset i has β_i exposure to its home country and zero elsewhere. β_i is obtained from the time-series regression:

$$r_{i,t} = \beta_i r_{M,t} + \alpha_{i,t} \quad \text{for each asset } i = 1, \ldots, n$$

  where r_M, t are the returns of a local benchmark index.
- *Multiple country exposures* — conceptually identical to multiple industry exposures; each exposure being a fraction (summing to 1) representing the proportion of the company's activity in, or revenue from, that country.

Dummy variables have the advantage of simplicity: a stock is either exposed to the market or it is not. Country betas are more computationally intensive and therefore may yield greater sensitivity and better fit, but care must also be taken when estimating betas from historical returns, which contain a great deal of noise.

As for industries, it is perfectly possible for country factors to exhibit thinness. We apply exactly the same form of correction for such as we detailed in the section on industry returns above.

5.5 Currency Factors

Currency factors are rather different to other factors in that they do not feature in the returns regression. Rather, each asset has its return calculated using its local currency, ensuring that currency effects are (as far as is possible) eliminated from the factor regression. All non-currency

[6] For some assets, e.g. ADRs, one may choose to explicitly expose the asset to the market within which it trades rather than the underlying asset's domicile.

factor returns are then computed via the regression, whilst currency returns are calculated directly from exchange and risk-free rates. Currency factor exposures and returns are then appended to the non-currency factor exposures and returns.

Suppose we have a set of m currencies, $1, 2, ..., m$. Then we define the following:

- $p_{i,t}$ is the asset price at time t in currency i,
- r_i: the asset return in local currency i from period $t-1$ to t,
- r_i^f: the risk free rate for currency i,
- $x_{ij,t}$ is the amount of currency j which one unit of currency i fetches at time t,
- r_{ij}^x is the return for an investor whose numeraire is currency j in buying currency i.

We therefore have:

$$\begin{aligned} r_i &= \frac{p_{i,t}}{p_{i,t-1}} - 1 \\ r_{ij}^x &= \frac{x_{ij,t}}{x_{ij,t-1}} - 1 \\ p_{j,t} &= p_{i,t} x_{ij,t} \end{aligned}$$

We now define what we mean by a currency return. The asset's total return from a currency j perspective is expressed exactly as

$$r_j = \frac{p_{i,t} x_{ij,t}}{p_{i,t-1} x_{ij,t-1}} - 1$$

This may be expressed more compactly as

$$r_j = (1 + r_i)(1 + r_{ij}^x) - 1$$

If we consider log returns, the above may be written as

$$r_j \approx r_i + r_{ij}^x.$$

This is exact for log returns, and should provide a good approximation for returns over a short duration (e.g. daily returns). The above is written in terms of total returns. If we consider excess returns we may write

$$r_j - r_j^f \approx r_i - r_i^f + r_{ij}^x + r_i^f - r_j^f.$$

We have therefore expressed the excess return from the perspective of currency j as the sum of the local excess return (currency i), and that which we define to be the currency factor return, viz.

$$c_{ij} = r_{ij}^x + r_i^f - r_j^f,$$

where c_{ij} represents the excess return to currency i expressed in terms of currency j. Note that when $i = j$, $c_{ij} \equiv 0$.

5.6 The Problem of Multicollinearity

If a regression contains two or more of the following:

- A market intercept
- Country dummy exposures
- Industry dummy exposures

then it will fail to find a unique solution because the exposure matrix B is singular. To simply see why this is so, consider each asset's return. Ignoring style factors and specific return, it may be written as

$$r_i = 1 \cdot f_M + 1 \cdot f_{I_j} + 1 \cdot f_{C_k}$$

where f_M is the market intercept return, f_{I_j} the asset's industry return, and f_{C_k} its country return. We may add an arbitrary amount to any one of these returns provided a similar negative amount is added to one of the other returns, and the overall fit will be unchanged. Thus there are an infinite number of solutions, all of which look equally good. This is the issue of *multicollinearity* or *linear dependence* between different sets of dummy factors.

A unique solution can be obtained if, for each set of dummy factors beyond the first, a constraint is imposed on the regression. In theory, almost anything will do. In practice, a common choice for researchers is to force one or more sets of factor returns to sum to zero. Assume henceforth that we have a market intercept, dummy industries and dummy countries, one could set the constraints

$$\lambda \sum_{j=1}^{P} w_{I_j} f_{I_j} = 0$$

$$\rho \sum_{j=1}^{Q} w_{C_j} f_{C_j} = 0$$

where w_{I_j} is the market capitalization of industry j, and w_{C_j} the market capitalization of country j. The effect of this is to "force" the broad market return into the market intercept factor, while the country and industry factors represent the residual behaviour of each net of the overall market. These constraints may be added to the regression as dummy assets and the regression performed in the normal fashion. In the Appendix we demonstrate that under easily-met conditions that the solution asymptotically approaches the "true" solution.

An alternative approach is to use a multi-stage regression. First regress the asset returns against the market factor,

$$r = X_M f_M + u$$

then regress the residual u against (for instance) the industry exposures X_I:

$$u = X_I f_I + v$$

then regress the residuals from this against the country exposures X_C:

$$v = X_C f_C + w$$

The factor returns can then be collected together, with the specific return being the final residual, w. This approach is useful when one wants a particular set of factors to extract as much power as possible from the returns, independent of the remaining factors, which will tend to be much weaker than if they were all included in one regression. In addition to a more cumbersome algorithm, however, analysis of the results becomes more complicated as different weights are used for the various regressions.

6 The Risk Model

Thus far the discussion has focused entirely on modeling returns, having said nothing whatsoever about the generation of risk forecasts. This is justified — if returns are modeled correctly and robustly then deriving risk estimates is relatively straightforward. For the mathematically-inclined, a more rigorous treatment of the subtleties can be found in Zangari (2003); De Santis et al. (2003).

If the model has been sensibly constructed with no important factors missed, then the specific returns are uncorrelated with themselves and with the factors, and the factor risk model can then be derived thus:

$$
\begin{aligned}
var(r) &= var(Bf + u) \\
\hat{Q} &= B\Sigma B^T + \Delta^2
\end{aligned}
$$

The asset returns covariance matrix $\hat{Q}$ is a combination of a common factor returns covariance matrix Σ and a diagonal specific variance matrix Δ^2.

6.1 Factor Covariance Matrix

The factor covariance matrix is calculated directly from the time-series of factor returns. Regression models estimate a set of factor and specific returns at each time period, eventually building up a returns history. Statistical models, on the other hand, generate the entire time-series anew with each iteration.

$$
F = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,T} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m,1} & f_{m,2} & \cdots & f_{m,T} \end{bmatrix}
$$

Recent events should exert more influence on the model than those long in the past, but one cannot simply curtail the history of returns and use only the most recent observations. A sufficiently long history is required to estimate all the covariances reliably. Axioma models address this dilemma by weighting the returns matrix using an exponential weighting scheme:

$$
w_t = 2^{\frac{-(T-t)}{\lambda}} \qquad t = 0, \ldots, T
$$

where T is the most recent time period. λ is the half-life parameter, the value of t at which the weight is half that of the most recent observation.

Figure 3 shows how this looks for values of t from $t = 0$ (earliest) to $t = 500$ (most recent) for a half-life of 100 days. The weights have been scaled so that the maximum value (at $t = 500$) is one. As can be seen, at $t = 400$ (one half-life), the weight is 0.5, and at $t = 300$ (two half-lives) the weight is 0.25.

Thus, given a half-life, weights $W = diag(w_1, w_2, \ldots w_T)$ are computed, and the factor returns history is weighted to yield a new matrix of values:

$$
\tilde{F} = FW^{\frac{1}{2}} = \begin{bmatrix} w_1^{\frac{1}{2}} f_{1,1} & w_2^{\frac{1}{2}} f_{1,2} & \cdots & w_T^{\frac{1}{2}} f_{1,T} \\ w_1^{\frac{1}{2}} f_{2,1} & w_2^{\frac{1}{2}} f_{2,2} & \cdots & w_T^{\frac{1}{2}} f_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ w_1^{\frac{1}{2}} f_{m,1} & w_2^{\frac{1}{2}} f_{m,2} & \cdots & w_T^{\frac{1}{2}} f_{m,T} \end{bmatrix}
$$

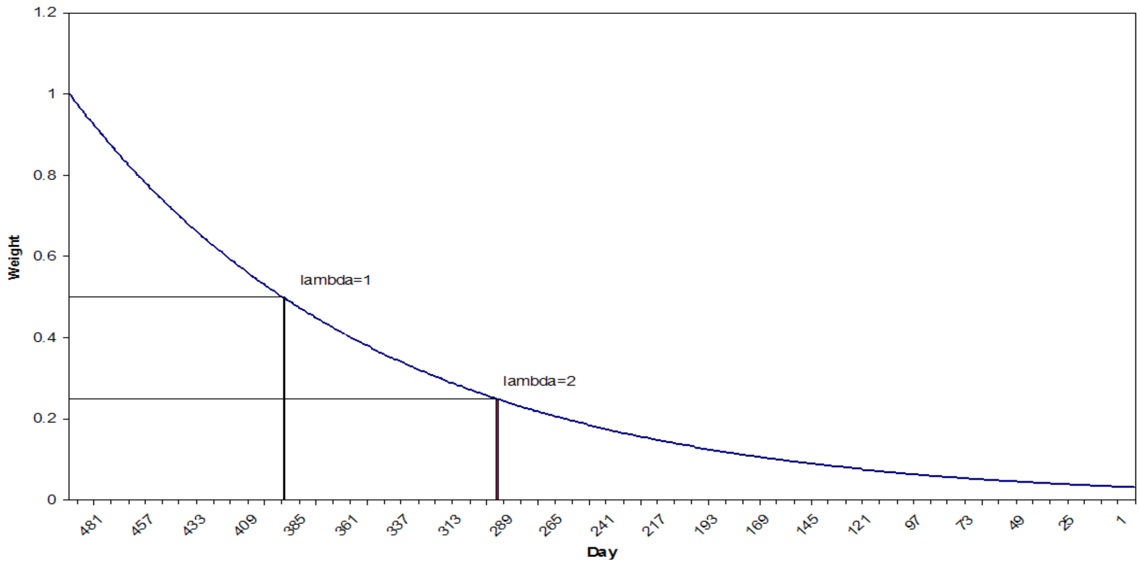


Figure 3: Typical exponential weighting scheme

and from this, the factor covariance matrix is simply calculated as

$$\Sigma = var\left(\tilde{F}\right) = \frac{FWF^T}{T-1}$$

Selecting an appropriate half-life is a major design question to which there is no definitive answer. This half-life indirectly affects the forecast horizon of the risk model. Too short a short half-life may allow for a very responsive model but creates excessive turnover for asset managers[7]; too long a half-life and the model will fail to respond sufficiently to changing market conditions. The half-life parameter therefore represents a balance between responsiveness and stability.

Figure 4 shows how the choice of half-life influences the computed volatility. The graph shows risk predictions for the S&P 500 Index using half-lives of 60 and 125 days. Interestingly, the longer half-life manages less well in "responding" to a change from high to low volatility (e.g. 2003-2004) than to a change from low to high. In the former case, smaller weights are being applied to relatively large numbers, so shocks from the past persist, whereas in the latter case, small weights are applied to small numbers.

Most Axioma risk models use multiple half-lives for greater flexibility — that is, a longer half-life for estimating the factor correlations and a shorter one for the variances. Apart from the observation that factor correlations are more stable over time than their volatilities, a relatively long half-life is necessary, given the large number of relationships to be estimated. The resulting correlations are then scaled by the corresponding variances to obtain the full covariance matrix. For even greater control over the model's forecast horizon, one could apply different half-lives for different factors (e.g. style vs. industry factors). Quite often, however, the additional computation complexity adds little or no advantage.

[7] Shortening the half-life also reduces the "effective" sample size available for estimating the factor correlations, which far outnumber the number of variances. (approximately $\frac{n^2}{2}$ vs n)

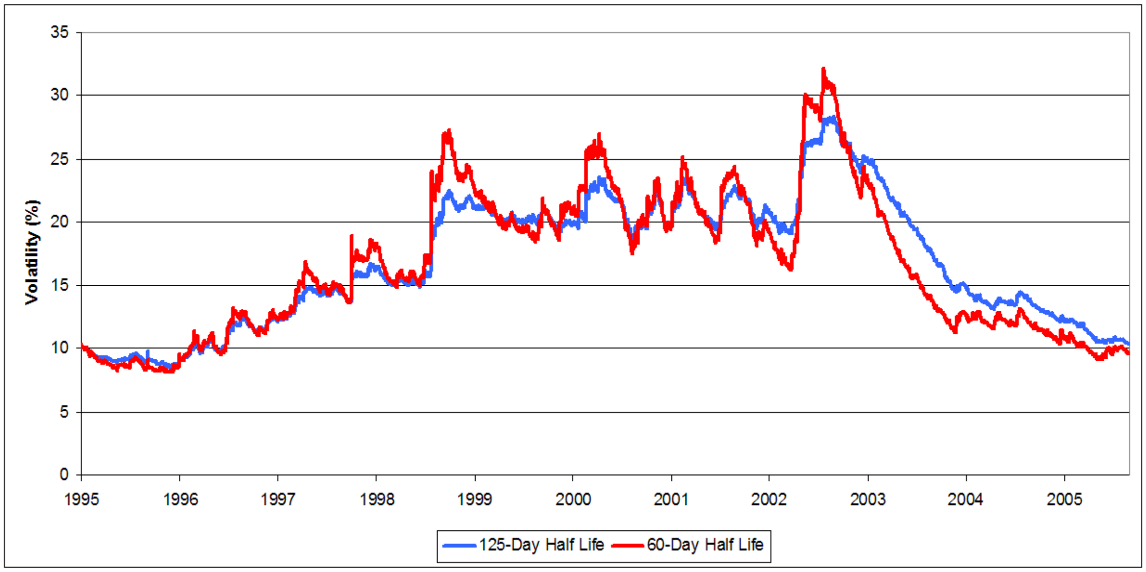


Figure 4: S&P 500 risk forecasts using 60- and 125-day half lives

6.2 Autocorrelation in the Factor Returns

If factor returns were uncorrelated across time, the above procedure alone is sufficient for calculating a factor covariance matrix. Unfortunately, the assumption of serial independence is much less true for daily asset returns than for returns over longer horizons. Over short time frames, market microstructure tends to induce lead-lag relationships that induce autocorrelation in the factor returns over time. Simply put, if an asset's price goes up one day, it is quite likely that it will go up the next day too. These "momentum" effects must be taken into consideration when aggregating risk numbers to longer horizons. For instance, one cannot simply derive a monthly risk forecast from daily numbers by the simple relationship[8]

$$\sigma_M = \sqrt{21} \cdot \sigma_d$$

Without taking autocorrelation into account, such forecasts are likely to be under- or overestimated. Axioma adjusts its factor covariance matrix estimates using a technique developed by Newey and West (1987):

A factor's return over N days may be approximated as the sum of its daily returns:

$$f(t, t+N) = \sum_{i=t}^{t+N} f(i)$$

This approximation is exact using logarithmic returns. The forecasted covariance across two factors, f and g, from t to $t+N$, is then

$$\sigma_{f,g}(t:t+N) = N\sigma_f(t)\sigma_g(t)\left[\rho_{f,g}(t) + \sum_{k=t}^{N-1}\left(1 - \frac{k}{N}\right)(\varphi_{f,g}(t, t+k) + \varphi_{g,f}(t, t+k))\right]$$

where $\varphi_{f,g}(t, t+k)$ and $\varphi_{g,f}(t, t+k)$ represent the *lagged correlations*:

$$\varphi_{f,g}(t, t+k) = corr\,(f(t), g(t+k))$$
$$\varphi_{g,f}(t, t+k) = corr\,(g(t), f(t+k))$$

[8]Assuming that there are on average 21 trading days in a month

In practice, one need not calculate lagged correlations for each lag up to N. By first analyzing serial correlation trends in the historical returns, it is reasonable to cut off the series at a point $h < N$, beyond which any correlations are insignificant.

A final note of caution: it is occasionally possible for the above to yield a covariance matrix that is not positive semi-definite. In the extremely unlikely event that it does occur, one final adjustment is made, beginning with an eigendecomposition of the factor covariance matrix,

$$\Sigma = UDU^T$$

where $D = diag(d_1, \dots, d_k)$ are the eigenvalues of Σ. If Σ is non-positive semi-definite, one or more eigenvalues will be negative. To force positive semi-definiteness, all negative eigenvalues are replaced by a positive, but very small, value. Denoting this adjusted eigenvalue matrix as $\tilde{D}$, the "corrected" factor covariance matrix is

$$\tilde{\Sigma} = U\tilde{D}U^T$$

6.3 Market Asynchronicity

Because equity markets do not open and close simultaneously, the use of daily returns to estimate relationships across assets trading in different time-zones offers a number of points of interest. Consider Tokyo: this market closes before the US market has even opened; thus, stocks trading in Japan will not reflect events on the US market that day until the Tokyo market reopens the next day. There is, therefore, the distinct possibility of a lag across asset returns' behaviour. In the worst case scenario, if one market always goes up when the other goes down, this could give the appearance of negative correlation across markets, and hence one market is a hedge for the other. In reality, this correlation is an artifact of the time-delay, and the markets are positively correlated.

To illustrate this, we compute daily returns for a selection of test markets: each is simply the weighted sum of the asset returns within the market. We then compute the correlations between these markets. We next lag the US market returns, so we compute correlations between each test market's returns and the previous day's US returns. This should indicate which markets are responding to the previous day's US market behaviour. Finally, we consolidate our daily returns and compute the correlations between weekly market returns. This will lessen the daily timing effect and give us a truer picture of market relationships. Figure 5 shows all three flavors of correlation between markets. The results are unambiguous: based on weekly returns, all markets show a significant positive correlation with the US market, with the developed markets showing the strongest relationship. This is not evident when using raw daily returns. When the US returns are lagged, we see that the Far Eastern markets are highly correlated with the previous day's US returns. Daily data do give a misleading picture of correlations across markets.

6.3.1 Synchronizing the Markets

To alleviate the problem of market asynchronicity we adjust our returns as follows. Recall that our basic returns model at a point in time, t, is as follows:

$$r_t = Bf_t + u_t. \tag{1}$$

We then make the assumption that each asset return, r_t^i may be written as

$$r_t^i = r_t^{i_m} + \gamma_t^i, \tag{2}$$

where r^{i_m} is the return to i's *local* market (the particular market on which the asset trades), and γ^i is the remainder, non-market, term.

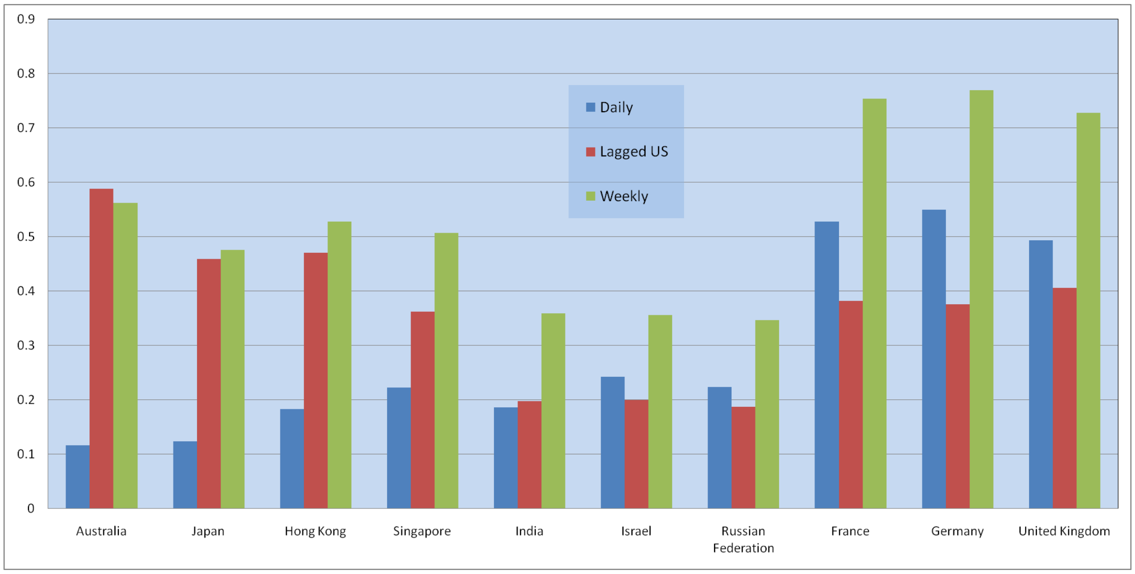


Figure 5: Correlations with the US market 2000-2009

Using the set of J local market returns, we construct a VAR model of local market returns:

$$r_t^m = N r_{t-1}^m + \epsilon_t. \tag{3}$$

We then use the coefficient matrix N above to estimate a set of forecast market returns

$$\hat{r}_t^m = N r_t^m + \epsilon_t, \tag{4}$$

and these projected market returns are then used to adjust each asset return, viz.

$$\hat{r}_t^i = \hat{r}_t^{im} + \gamma_t^i. \tag{5}$$

Using (2) we rewrite (5) as

$$\hat{r}_t^i = r_t^i + \Delta_t^{im}, \tag{6}$$

where the *returns-timing adjustment factor*, Δ_t^{im} is defined as

$$\Delta_t^{im} = \hat{r}_t^{im} - r_t^{im}. \tag{7}$$

We may then use the set of adjusted asset returns from (6) in the model factor regression

$$\hat{r}_t = B\hat{f}_t + u_t. \tag{8}$$

6.4 The Complexities of Currency Risk

Recall that our currency factor returns (if applicable) are computed separately from the other factor returns. We append the currency factor returns to the model factor returns:

$$r_j = Xf + u = \begin{bmatrix} X_q & X_m \end{bmatrix} \begin{bmatrix} f_q \\ f_m^j \end{bmatrix} + u$$

Here, $X_q \in \Re^{N \times q}$ and $f_q \in \Re^q$ are the non-currency exposures and factor returns respectively. The vector $f_m^j \in \Re^m$ is the vector of m currency returns (relative to currency j):

$$f_m^j = \begin{bmatrix} c_{1j} \\ \vdots \\ c_{mj} \end{bmatrix}$$

and $X_m \in \Re^{N \times m}$ is the matrix of currency exposures. We then compute a combined factor covariance matrix. There are a number of complications relating to currency risk, however. We detail those here.

6.4.1 Currency Covariances

Given our returns model above, to include currency risk in the risk model we could simply calculate a combined factor covariance matrix as we would any other model factor covariance matrix, viz.

$$\begin{aligned} var(r^N) &= \begin{bmatrix} X_q & X_m \end{bmatrix} var \begin{bmatrix} f_q \\ f_m^j \end{bmatrix} \begin{bmatrix} X_q^T \\ X_m^T \end{bmatrix} + var\,(u) \\ &= \begin{bmatrix} X_q & X_m \end{bmatrix} \begin{bmatrix} V_q & V_{qm} \\ V_{qm}^T & V_m \end{bmatrix} \begin{bmatrix} X_q^T \\ X_m^T \end{bmatrix} + \Delta. \end{aligned} \tag{9}$$

Here:

- V_q is the covariance matrix for the non-currency factors,
- V_{qm} is the covariance matrix block of non-currency/currency factor returns,
- V_m is the currency return covariance matrix.

The main problem with this simple approach is that currencies are noisy. Figure 6 shows

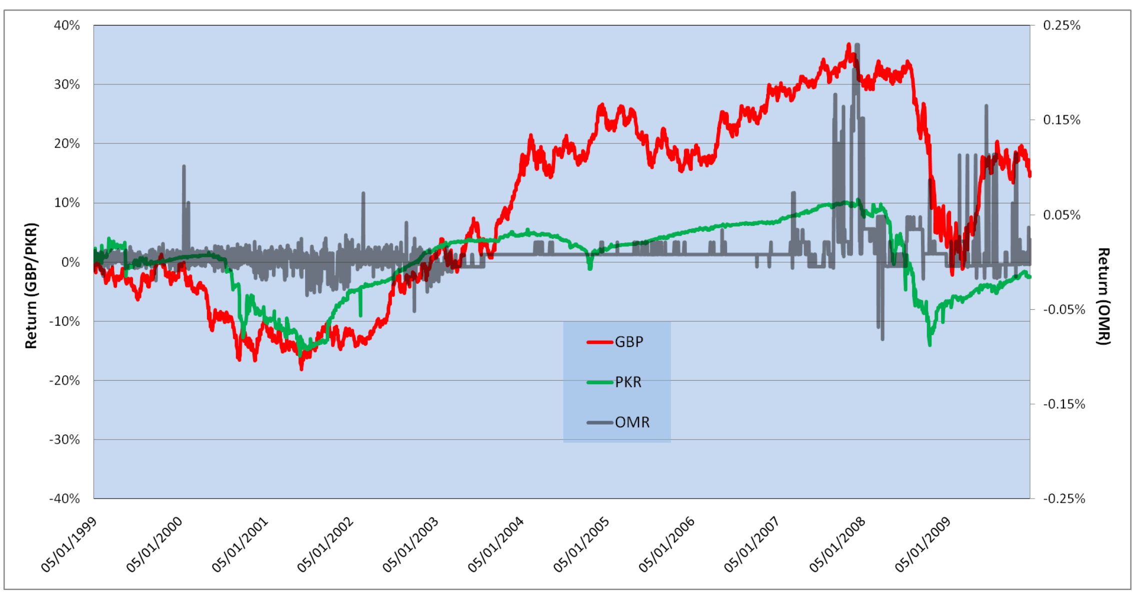


Figure 6: A Selection of Currency Cumulative Returns

cumulative currency returns for three currencies. First there is the liquid; the British Pound (GBP), then there is the fairly-illiquid; the Pakistani Rupee (PKR), and finally, the pegged; the Omani Rial (OMR). As we look at these three in order we see a progressive decrease in the amount of signal to noise in the returns; the GBP contains a lot of information, while the OMR is really little more than noise. Attempting to derive meaningful correlations between returns such as this and other factor returns, currency or non-currency, is likely to lead to some spurious, unstable results. Figure 7 demonstrates this. We compute realized 30-day correlations for the Euro and GBP, and

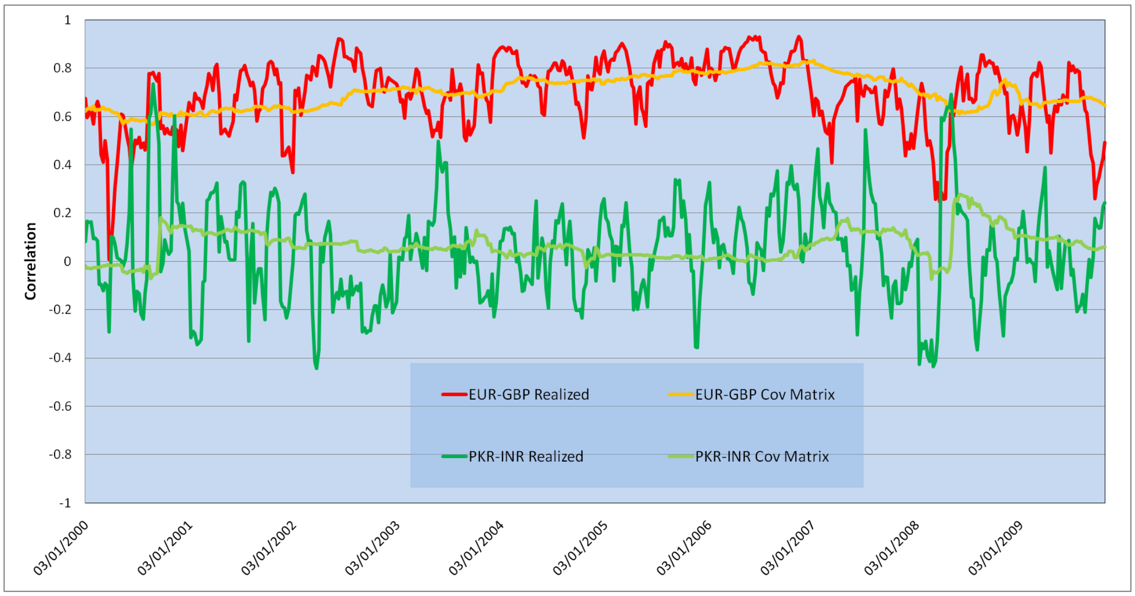


Figure 7: A Selection of Realized Currency Correlations

contrast these with those for PKR and INR (Indian Rupee). There is clearly a gulf between the two pairs in terms of meaningful information. We also show the correlations as given by our currency covariance matrix. For the EUR-GBP pair there is little to complain about; for the INR-PKR pair, we see a small but consistent correlation where there is probably no correlation in truth, but more importantly, we see how it jumps abruptly at two points in time. Noisy currencies equals noisy correlations. We see similar patterns amongst such noisy currencies and also between currencies and non-currency factors - noise getting confused with signal.

The question we ask ourselves is how can we eliminate the noise from the correlations? For numeraire-pegged currencies, we could simply force correlations with other factors to be zero, but things are never so black and white: some currencies are pegged with varying degrees of strictness, some are simply illiquid. So we look for a one-size-fits-all approach.

The solution at which we have arrived is this: given that we have a relatively small number of currencies in our model (fifty to sixty, typically), it is perfectly viable to attempt to model the currency returns via a simple Principal Components Analysis (PCA) statistical model. We thus decompose

$$C = ZG + \theta, \tag{10}$$

where:

- $C \in \Re^{m \times T}$ is a time-series matrix of currency returns, $c_m(t),\ t = 1, \ldots, T$,
- $Z \in \Re^{m \times p}$ is a matrix of exposures to p statistical factors,
- $G \in \Re^{p \times T}$ is a time-series matrix of p statistical factor returns, and

- $\theta \in \Re^{m \times T}$ is the residual from the PCA regression.

Given a time-series of currency returns we compute (10) by standard PCA decomposition.

Our particular model has one quirk; we restrict the PCA regression to an *estimation universe* of twelve currencies. These are:

- The big seven: USD, JPY, GBP, EUR, CHF, AUD, CAD. These account for the vast majority of currency trades.
- A further six from Asia and the developing world: BRL, MXN, PLN, TWD, SGD, ZAR. These are some of the next most heavily traded.[9]

The above currencies are chosen to minimise residual returns structure: some other currencies are actually more heavily traded than the latter six, (e.g. SEK, NOK, NZD), but are so highly correlated with currencies already in the estimation universe that for modeling purposes they scarcely count as separate entities. Then, we allow our PCA model to have twelve statistical factors; thus we have the same number of factors as currencies in our estimation universe, ensuring that the estimation universe currencies are modeled exactly by the PCA, and the non-estimation universe exposures can be thought of as sensitivities to the estimation universe currency returns.

To summarise the above: we model

$$C_{estu} = Z_{estu} G, \tag{11}$$

where $C_{estu} \in \Re^{12 \times T}$ is the subset of estimation universe currency returns. Here (11) is an *exact* relationship (no residual). We then "back out" non-estimation universe exposures, Z_{nonest} via

$$Z_{nonest} = C_{nonest} G^T \left(G^T G \right)^{-1}, \tag{12}$$

where $C_{nonest} \in \Re^{(m-12) \times T}$ is the subset of non-estimation universe currency returns. We combine our "estu" and "non-est" currency components to create the combined currency model (10). We assume that the residual currency returns θ are uncorrelated, both across currencies and across currency/non-currency pairs, hence $cov(\theta, f_q) = 0$, $cov(\theta, ZG) = 0$ and $var(\theta)$ is diagonal. We may thus rewrite covariance matrix (9) as

$$var(r^N) = \begin{bmatrix} X_q & X_m \end{bmatrix} \begin{bmatrix} V_q & cov(f_q, ZG) \\ cov(ZG, f_q) & var(ZG) + var(\theta) \end{bmatrix} \begin{bmatrix} X_q^T \\ X_m^T \end{bmatrix} + \Delta. \tag{13}$$

By decomposing currency returns and imposing structure onto them, we force at least some of the noise into the residual, uncorrelated component, θ, and the resulting covariance matrix will exhibit smoother and more meaningful correlations.

6.4.2 Numeraire Transformation

As it stands, our combined currency matrix gives the risk from the point of view of a particular base numeraire such as USD or Euro. However, risk model users based in other currencies may wish to look at things from the perspective of their own currency. The simple solution is to compute a number of different covariance matrices, based on each desired currency. This, however, is inflexible and wasteful. By a simple transform we may transform a covariance matrix based on one currency to the perspective of any other that we choose.

[9] The more observant will have spotted that there are actually thirteen currencies in total; however, one of these is always the numeraire, hence its returns are zero, and it is automatically excluded from the PCA decomposition

Our aim is to transform the currency returns from currency-j perspective to that of, say, currency k. Using the log-return approximation for currency returns and considering our asset to be the currency return for currency j, we can form the expression

$$r^x_{ik} \approx r^x_{ij} + r^x_{jk}$$

We can rebase the return to currency i, c_{ij}, from j to currency k as

$$c_{ik} = c_{ij} - c_{kj}$$

Thus everything can be written in terms of already-known quantities (j-numeraire currency returns). The asset returns in terms of our new base-currency, k, are

$$r_k = \begin{bmatrix} X_q & X_m \end{bmatrix} \begin{bmatrix} f_q \\ f^k_m \end{bmatrix} + u$$

The re-based currency factor returns, f^k_m, are given by

$$f^k_m = \begin{bmatrix} c_{1j} - c_{kj} \\ \vdots \\ c_{mj} - c_{kj} \end{bmatrix}$$

The returns equation can be rearranged as:

$$r_k = \begin{bmatrix} X_q & \hat{X}_m \end{bmatrix} \begin{bmatrix} f_q \\ f_m \end{bmatrix} + u$$

where the transformed currency exposure matrix is given thus

$$\hat{X}_m = X_m \left(I_m - \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \right)$$

The second term is simply a matrix of N by m zeros, save with a column of ones corresponding to the position of currency k. If we write

$$T_m = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}$$

Then our transform may be written as

$$r_k = \begin{bmatrix} X_q & X_m \end{bmatrix} \begin{bmatrix} I_q & 0 \\ 0 & I_m - T_m \end{bmatrix} \begin{bmatrix} f_q \\ f_m \end{bmatrix} + u$$

And so, from the above equation for returns, it is simple to see that we may construct our covariance matrix from the existing factor covariance matrix. We may apply the above transformation to the asset exposure matrix, considering the factor covariance matrix as fixed, or we may view the exposures as fixed and compute a new factor covariance matrix as:

$$\hat{F} = \begin{bmatrix} I_q & 0 \\ 0 & I_m - T_m \end{bmatrix} F \begin{bmatrix} I_q & 0 \\ 0 & I_m - T^T_m \end{bmatrix}$$

6.4.3 Calculating a Composite Covariance Matrix

Different risk models will utilize different sets of factor returns, perhaps with different techniques and parameters for computing the covariances. However, it would be desirable, for models with identical horizons, for the relevant currency covariances to match across models. We guarantee that this is the case by computing a single global currency covariance matrix as a distinct entity and then merging the relevant portions of this into each model covariance matrix.

We partition our q non-currency and m currency returns and exposures as

$$r = Xf + u = \begin{bmatrix} X_q & X_m \end{bmatrix} \begin{bmatrix} f_q \\ f_m \end{bmatrix} + u$$

First we compute a simple, all-in-one covariance matrix from this complete set of factor returns and exposures. We factorize this matrix into volatilities and correlations thus

$$V_S = \begin{bmatrix} D_q & \\ & D_m \end{bmatrix} \begin{bmatrix} C_q & C_{qm} \\ C_{qm}^T & C_m \end{bmatrix} \begin{bmatrix} D_q & \\ & D_m \end{bmatrix}$$

where the D blocks are diagonal matrices, each element on the diagonal being the square-root of the relevant factor volatility, and the C blocks correspond to correlations.

Next, assume that we also have calculated separate covariance matrices for the non-currency and currency factors. Each of these sub-matrices may have been calculated by differing techniques (e.g. different numbers of lags in autocorrelation correction, treatment of extreme values etc.). We stack these in a block diagonal form and factorize as follows:

$$V_C = \begin{bmatrix} \hat{D}_q & \\ & \hat{D}_m \end{bmatrix} \begin{bmatrix} \hat{C}_q & \\ & \hat{C}_m \end{bmatrix} \begin{bmatrix} \hat{D}_q & \\ & \hat{D}_m \end{bmatrix}$$

Because each covariance matrix was calculated for a particular factor group, we have no off-diagonal blocks corresponding to correlations across factor groups.

Our aim, therefore, is to combine the two covariance matrices V_C and V_S such that the composite matrix is symmetric positive-semidefinite, with its diagonal blocks being identical to the diagonal blocks of V_C.

There are two ways to effect this. The "valid" theoretical method is to define the transformation matrix

$$R = \begin{bmatrix} \hat{D}_q \hat{C}_q^{1/2} C_q^{-1/2} & \\ & \hat{D}_m \hat{C}_m^{1/2} C_m^{-1/2} \end{bmatrix}$$

Then the transformed matrix

$$V = R \begin{bmatrix} C_q & C_{qm} \\ C_{qm}^T & C_m \end{bmatrix} R^T$$

yields the desired result. Expansion will demonstrate that its diagonal blocks are equal to those of V_C, that it is symmetric is obvious, and its positive semi-definiteness follows from the properties of the central term.

However, this technique can be rather unstable. The necessity of inverting the diagonal correlation blocks of the simple correlation matrix has a tendency to boost noise, and can, in the event of very small eigenvalues of the covariance matrix, be outright dangerous. An alternative, less theoretically-pleasing, but more stable alternative is to "cut-and-paste" a correlation matrix. Define our final covariance matrix as

$$V = \begin{bmatrix} \hat{D}_q & \\ & \hat{D}_m \end{bmatrix} \begin{bmatrix} \hat{C}_q & C_{qm} \\ C_{qm}^T & \hat{C}_m \end{bmatrix} \begin{bmatrix} \hat{D}_q & \\ & \hat{D}_m \end{bmatrix}$$

One can see that we have kept the volatilities from the separately-computed block covariance matrices, and "cut-and-pasted" the correlation blocks from the two covariance matrices. Such a composite covariance matrix still retains the properties that its diagonal blocks match those of V_C, but positive-semidefiniteness is not guaranteed. However, as long as the two covariance matrices V_C and V_S were computed using techniques not too wildly different from one another, the positive-semidefiniteness condition should not be violently breached. Practical experience shows that this is the case.

We may ensure that positive-semidefiniteness is always maintained by the following. Our correlation matrix may be written as

$$C = \begin{bmatrix} \hat{C}_{11} & 0 \\ 0 & \hat{C}_{22} \end{bmatrix} + \begin{bmatrix} 0 & C_{12} \\ C_{12}^T & 0 \end{bmatrix}$$

We may perform a simultaneous eigendecomposition on these two symmetric matrices and reduce the matrix to

$$C = Z \left(\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} + \begin{bmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_n \end{bmatrix} \right) Z^T$$

We know that the eigenvalues σ_i are non-negative, what we cannot guarantee is that the eigenvalues γ_i are also non-negative. However, so long as

$$\sigma_i + \gamma_i \geq 0 \; \forall i$$

then positive-semidefiniteness of the correlation matrix is assured. It is simple to implement a numerical check to ensure the above.

6.5 Specific Risk Calculation

The final, remaining piece of the risk model is the asset specific risk. Having assumed a diagonal specific returns covariance matrix, only n specific variances require calculation. These are estimated directly from the specific returns, again applying exponential weighting, with a half-life typically equal to that used for the factor variances for consistency:

$$\sigma_i^2 = \frac{1}{T-1} \sum_{t=1}^{T} w_t \left(u_{i,t} - \bar{u}_i \right)^2$$

Unlike factor returns, specific returns are typically full of outliers, so care must be taken when estimating variances. To prevent extreme specific returns from skewing the variances, Axioma models first winsorize each asset's specific return history to lie within a certain number of median absolute deviations.

Of course, this is not feasible for assets that only have a small handful of specific returns observations, such as newly-listed assets. For such cases we generate a history of pseudo-returns based on comparison with stocks of similar size. Such an approach has the advantage that as the true history increases in length over time, the specific risk converges smoothly to the "true" risk. Further, any underprediction is swiftly corrected due to the inflationary nature of the standard deviation and exponential weighting. And overprediction merely indicates a conservative stance in the face of insufficient information.

As a final safeguard, any remaining extreme specific risks are truncated.

7 Modeling of DRs and Cross-Listings

We here consider all cases of ADRs, GDRs, foreign and cross-listed assets, assets whose *domicile* can be said to be outside the country of quotation (e.g. Israeli assets quoted in the US, South African stocks quoted in London) and any similar cases not included in the above. Henceforth we refer to these collectively as DRs. Typically, alongside the market on which a DR is quoted, there can be said to be an *underlying* market, and in many, though not all, cases, a corresponding underlying asset.

7.1 Modeling DR Return

We define the numeraire-return of *any* asset as

$$r^N = r^X + c_X^N, \tag{14}$$

where

- N is the model numeraire currency,
- X is the asset's *local* currency, i.e. that in which it is quoted,
- r^X is the asset's *local* return,
- c_X^N is the return in N for an investor buying currency X.

This relationship is exact for log-returns, but is still a very close approximation for high-frequency (such as daily) returns. We make use of log-return relationships extensively hereon. If we define X to be the currency of an underlying asset, and Y to be the currency of quotation of its DR, then we model the DR's return in terms of the underlying currency, viz.

$$r^Y = r^X + c_X^Y. \tag{15}$$

Thus, we consider the DR return to be composed of a return expressed in the underlying asset's currency, plus the return in currency Y of purchasing currency X. If we ignore effects due to differing market-hours and liquidity, this return, r^X, should be equal to that of the underlying asset. If it were not, there would be arbitrage opportunities available to the purchaser of either the DR or the underlying.

7.2 Currency Exposure

From the above we infer that a DR has exposure to two currencies; its *local* currency Y, and that of the underlying asset, X. If we write the DR return in terms of numeraire N, combining (14) and (15) we get

$$\begin{aligned} r^N &= r^Y + c_Y^N, \\ &= r^X + c_X^Y + c_Y^N. \end{aligned} \tag{16}$$

The DR would appear, therefore, to have exposure to two currency components. However, in order to fit this into the framework of a regional model, all currency returns must be expressed in terms of the numeraire. Using properties of logarithmic returns, we know

$$\begin{aligned} c_X^Y &= c_X^N + c_N^Y \\ &= c_X^N - c_Y^N. \end{aligned} \tag{17}$$

This says that for a holder of Y to buy currency X is equivalent to a holder of numeraire N doing the following:

1. Buying currency X,
2. Going short on Y.

Therefore, combining (16) and (17) the DR numeraire-return becomes

$$\begin{aligned} r^N &= r^X + c_X^N - c_Y^N + c_Y^N \\ &= r^X + c_X^N. \end{aligned} \tag{18}$$

We see that from a numeraire-perspective, the DR return is the same as the local return of the underlying asset. We therefore give DRs exposure to the underlying currency rather than the currency in which the DR is quoted.

7.3 Market Exposure

Since we have translated our DR return's currency to that of the underlying asset, it would seem logical that we give the DR exposure to the underlying asset's market rather than the market on which the DR is quoted. Thus, an ADR on a British stock will now have exposure to GBP and the British Market, rather than USD and the US market.

7.4 Effect on Style Exposures

We may divide style exposures into two kinds: issuer-based and issue-based. Issuer-based styles are those constructed from company characteristics and include size, value, growth and leverage. Issue-based styles use market data, such as returns, and typically include momentum, volatility and liquidity.

Since they represent company-wide characteristics, it makes sense that issuer-based exposures are either cloned, or defined so as to be identical between DR and underlying. Treatment of issue-based exposures is a little more fraught. If their returns series are truly different, then letting the data decide would seem the most reasonable option. However, everything we have seen so far leads us to believe that the apparent differences in returns are illusory; a result of timing, liquidity and short-term noise.

Given the assumption that, in at least the majority of cases, the behaviour of returns is actually the same for sets of linked assets, then it would be reasonable to ensure that styles dependent on returns such as momentum, market/exchange-rate sensitivity and volatility match. The odd man-out here is liquidity. Evidently, DRs and underlying assets do not trade in the same volumes. Unless we define liquidity in terms of the total trading volume across all of a company's assets, surely this should be allowed to differ?

However, in discussions with those experienced in trading with DRs, there is a strand of opinion that holds that the concept of liquidity as we usually define it does not hold for many DR-type instruments; that regardless of the market, one can always obtain access to such an asset if the will is there. While this point needs a great deal more clarification and research, it does indicate that it may not matter from a user perspective whether liquidity matches or does not between linked assets.

Given all of this, we make the decision to clone exposures, which has the effect of forcing any and all differences in linked asset returns behaviour into the assets' specific returns. We can thus model the linkage between assets purely in terms of asset specific risk and covariance. While there are good arguments for and against any particular approach, we believe this is the most robust and parsimonious. Further, not cloning asset exposures may create very strange optimization results, particularly in the presence of constraints on the factors, when mixing assets from the same issuer.

7.5 DRs and Returns-Timing

We have dealt so far with common factor and currency behaviour. One unwanted side effect of the fact that we give DRs exposure to the underlying market is to induce structure in the DR's specific return. To take an example: BP's ADR trades on the US market, which closes hours after London has closed. Therefore, the ADR contains a component of return due entirely to the US market. But, since we give the ADR exposure to the British market, this timing-difference component will tend not to be picked up anywhere in the common factor returns.

To see this, assume that we have converted the DR's return to the underlying currency. We define the underlying market of DR i as i_m, and its underlying market as i_q. We model the return of the DR as

$$r^i = g + f_{i_m} + \Psi_i + \Delta_{i_m,i_q} + u_i, \tag{19}$$

where

- g is the global market return,
- f_{i_m} is factor return to market m,
- Ψ_i is the remaining factor structure (styles, industries), of no relevance here,
- u_i is asset i's specific return,
- Δ_{i_m,i_q} is the difference in market return between markets m and q.

Unfortunately, this *DR returns-timing* component Δ_{i_m,i_q} is not taken account of in the model factor regression, since the DR has exposure only to the underlying home market. It pollutes the residual return, rendering it not truly asset-specific:

$$\hat{u}_i = u_i + \Delta_{i_m,i_q}. \tag{20}$$

u_i is the *true* specific return, but $\hat{u}_i$ is what we're *actually* left with from the regression.

To deal with this structured residual component, we write the returns-timing adjustment factor for market m relative to a reference market, R, (such as the USA) as Δ_{i_m,i_R}. We then note that, in the case of log returns, the following holds:

$$\Delta_{i_m,i_q} = \Delta_{i_m,R} - \Delta_{i_q,R}. \tag{21}$$

We recompute the DR's return as

$$\hat{r}_i = r_i - \Delta_{i_m,R} + \Delta_{i_q,R},$$

and we have, in effect, made the DR return truly local; the timing-component is removed from the total return; the factor return is genuinely that pertaining to the home market, m, whilst the specific return should by construction contain no structure due to timing differences.

7.6 ADRs and Cointegration

Experience has shown that there are a number of drivers behind the differences between the daily returns of DRs and their underlying stocks. Despite this, when we step back from noisy short-term measures, and look at longer-term behaviour, the differences begin to disappear, or at very least, do not blow up. There is obviously a profound difference between the behaviour of two linked stocks, and two that merely share similar characteristics.

Figure 8: Cumulative Returns of Petrobras, its ADR and Exxon

Figure 8 shows cumulative dollar returns of a number of assets over the course of several years. We see that of Brazilian energy company Petrobras, its ADR and US energy company Exxon. The correlation between the two Petrobras stocks' daily returns over this period is 80%, while the correlation between Petrobras and Exxon is not much lower, at 75%. Yet, despite the very close correlations, and the similar nature of the companies' business, why do the cumulative returns of the linked stocks track each other so closely, while those of the very similar Petrobras and Exxon do not?

The reason for the behaviour we see above is that the returns of the DR are mean-reverting. That is, while the price difference may be large on a given day, due to noise, market conditions etc., the price of the DR does revert to be close to the price of the underlying asset through time. Even when a price premium is present, there is still a tendency for the return to revert, as premia tend to remain relatively constant over time, and so their effect on returns appears to be negligible. When two assets behave in this way, they are said to be cointegrated. Technically, two time-series, x_t and y_t are said to be cointegrated if there exists a parameter β such that

$$u_t = y_t - \beta x_t$$

is a stationary process. The time-series u_t is stationary if it has a constant mean, variance, and autocorrelation through time.

In plain terms, it is as if there is a gravitational attraction between the two assets so that, while there may be a great deal of small-scale noise in their respective returns, the attraction prevents their prices from drifting apart over time. This is what differentiates the behaviour of linked assets from merely similar assets: small differences in returns do not blow up over time into large differences in prices - the differences remain roughly constant.

To test for cointegration, we consider the Engle-Granger (Engle and Granger, 1987) test on the log of prices. Consider the following time-series regression:

$$\log p_t = \beta \log p_t^d + e_t, \tag{22}$$

where p_t and p_t^d are the prices of the underlying and DR, respectively, and e_t is an error term. The DR and underlying are cointegrated if e_t is stationary. The augmented Dickey-Fuller (ADF) test is frequently used to test the stationarity of e_t. For now, let's assume that e_t is stationary and consider the implications on the relative returns of the DR and underlying asset.

Using (22), we get the following:

$$\log p_t - \log p_{t-1} = \beta \log p_t^d + e_t - \log p_{t-1} \tag{23}$$

$$\Rightarrow \log \frac{p_t}{p_{t-1}} = \beta \log p_t^d + e_t - \left(\beta \log p_{t-1}^d + e_{t-1}\right) \tag{24}$$

$$= \beta \log \frac{p_t^d}{p_{t-1}^d} + e_t - e_{t-1}. \tag{25}$$

If we apply (25) recursively, we get

$$\log \frac{p_T}{p_0} = \beta \log \frac{p_T^d}{p_0^d} + e_T - e_0. \tag{26}$$

Since e_t is stationary, $e_T - e_0$ has a mean of zero and a variance that is roughly twice the variance of e_t and independent of T. Equation (26) says that the cumulative return of the underlying over T periods equals β times the cumulative return of the DR plus the sum of residuals at time T and 0. If we make the additional simplification that $\beta = 1$ (in practice this is a reasonable assumption), then we see that the difference in cumulative return between the underlying and DR may be written as

$$c_T - c_T^d = e_T - e_0. \tag{27}$$

If we assume that the standard deviation of e_t is σ and that the autocovariance of e_t is zero, then this difference is distributed as $N(0, 2\sigma^2)$. The critical point is that the cumulative return over a period has a constant variance regardless of the length of the period.

Now let's consider the realized tracking error, σ_a, between the two assets. Let

$$r_t = \log p_t / p_{t-1}$$

and

$$r_t^d = \log p_t^d / p_{t-1}^d$$

be the logarithmic return of the underlying asset and DR, respectively. Then (25) says that

$$r_t = r_t^d + \varepsilon_t$$

where $\varepsilon_t = e_t - e_{t-1}$. Then the daily active variance is given as

$$\sigma_a^2 = \mathrm{E}\left[\varepsilon_t^2\right] \tag{28}$$

$$= \sum_{t=1}^{T} (e_t - e_{t-1})^2 \tag{29}$$

$$= \mathrm{E}\left[e_0^2 + e_T^2\right] + \frac{2}{T}\sum_{t=1}^{T-1} e_t^2 \tag{30}$$

$$= 2\sigma^2. \tag{31}$$

This is a daily variance figure; if we make the standard scaling assumption to convert this figure to a period-variance, we get the result

$$\sigma_a^2 = 2T\sigma^2. \tag{32}$$

According to this expected tracking error, the difference between the cumulative log return at time T is $N(0, 2T\sigma^2)$. However, according to equation (27), the difference between the cumulative log return at time T is $N(0, 2\sigma^2)$. And here is the rub: the assumption by which we scale daily variances to weekly, annual or any other period T (by multiplying by T) does not hold in the case of cointegrated assets.

For an arbitrary pair of assets, the magnitude of the differences between their weekly returns will indeed be greater than that of the daily returns, and the difference in monthly returns greater still. For cointegrated assets, (into which category our DRs and underlyings appear to fall) the theory states that this is not the case.

7.6.1 The Cointegration Algorithm for Specific Returns

We now show how we use the theory of cointegration to model the differences between linked assets. Because we have matched all common factor components, we guarantee, so long as two assets' total returns are cointegrated, that their specific returns will also be so.

Applying equation 22 recursively leads us to

$$r_t = \beta r_t^d + \varepsilon_t, \tag{33}$$

where

$$r_t = \log p_t / p_{t-1},$$
$$r_t^d = \log p_t^d / p_{t-1}^d, \text{ and}$$
$$\varepsilon_t = e_t - e_{t-1},$$

and p_t and p_t^d are the prices of the underlying and DR, respectively, and e_t is stationary.

If we assume that ε_t is uncorrelated with r_t^d, then we may write

$$var(r) = \beta^2 \cdot var(r^d) + var(\varepsilon), \tag{34}$$

and

$$cov(r, r^d) = \beta \cdot var(r^d), \tag{35}$$

and this leads us to our algorithm, viz. Given two sets of returns, r_t and r_t^d, where we assume for the sake of argument that r_t^d is the larger/more liquid of the two and $var(r^d)$ has been computed directly:

1. ... Using the daily specific returns, compute corresponding series of pseudo log prices, $\log p_t$ and $\log p_t^d$,

2. Compute β and e_t from the time-series regression (22),

3. Compute $var(r)$ from (34),

4. Compute the specific covariance between the two assets from (35).

8 Testing a Risk Model

At all stages, from initial estimation to day-to-day processing, a risk model must be tested and validated. At the estimation stage, model parameters are tuned in-sample and then validated out of sample. When running live, a model's explanatory power and risk estimates must be evaluated and compared with reality.

8.1 Testing the Returns Model

8.1.1 Overall Fit and R^2

The *coefficient of determination*, R^2, measures the proportion of returns variance that is accounted for by variation in the regressors:

$$R^2 = 1 - \frac{u^T W u}{(r - \bar{r})^T W (r - \bar{r})}$$

where $W = diag(w_1, \ldots, w_n)$ is the matrix of regression weights. $R^2 = 1$ means the model fits the data perfectly, and should the model explain precisely nothing, $R^2 = 0$. The amount of returns variance explained can vary greatly over time: during major market movements, R^2's tend to be high, as assets are moving in tandem, while quieter times typically bring low R^2's as more variability is white noise.

It is always possible to fit a model to any degree of in-sample precision simply by adding more parameters. In fact, one could fit any model to the data exactly by having as many parameters as data points. To counter over-fitting, the *adjusted* R^2 statistic penalizes the addition of factors and allows one to meaningfully compare candidate models with differing numbers of factors: The adjusted statistic is defined as:

$$\bar{R}^2 = 1 - \frac{n-1}{n-m}\left(1 - R^2\right)$$

where n is the number of assets, and m the number of factors. Note that $\bar{R}^2$ can take negative values, and approaches R^2 as n becomes very large relative to m. Note also, by altering W accordingly, both statistics can be calculated on a subset of assets — say, the estimation universe, or a specific portfolio.

8.1.2 Explanatory Power of Individual Factors

By itself, the $\bar{R}^2$ statistic is useful but limited. The explanatory power of a model may come from all the common factors, or it may be due to just one or two, with others being largely noise. To determine exactly how well each factor contributes to the model, one looks at its *t-statistic*, which measures the extent to which a factor is statistically significant.

Having assumed normally-distributed residuals earlier, the t-statistic of factor k can be computed:

$$t_k = \frac{f_k}{\sqrt{s^2 \left(B^T W B\right)^{-1}_{kk}}} \sim t(n-m)$$

$$\text{where } s^2 = \frac{u^T W u}{n-m}$$

One may then choose a degree of significance and look up the appropriate values from a t-distribution table, or simply apply an approximate rule of thumb: given a large number of observations, a t-statistic ≥ 2.0 implies that the null hypothesis $H_0 : f_k = 0$ can be rejected with 95% confidence.

Similarly, one could use a multivariate extension of the t-statistic, the *F-statistic*, to test the null hypothesis that *all* the factors are insignificant ($H_0 : f_k = 0$ for all k):

$$F = \frac{1}{m} \left[f^T s^2 \left(B^T W B\right)^{-1} f \right] \sim F(m, n-m)$$

Empirical tests suggest that a t-statistic that is significant at least 15% of the time is likely to contain true explanatory power. As market conditions evolve, however, a factor may prove significant in some periods of the model history but not others. It is therefore important to look at subsections of the model history — periods corresponding to similar market climate — to gauge whether a factor's usefulness has altered over time.

8.1.3 Factor Alphas and Betas

The above statistics are by far the most important tests of a returns model. There are, however, a number of lesser measures used to justify the model and to identify suspicious behavior. Regressing factor k's returns against market returns r_M,

$$f_{k,t} = \alpha_k + \beta_k r_{M,t} + \phi_{k,t}$$

produces a factor alpha α_k and beta β_k, and corresponding t-statistics. There are no concrete rules to apply here, but in general, factors reflecting the market (such as industries, market beta) should contain mostly beta and insignificant alpha. On the other hand, style factors, for example, should exhibit little beta and significant alpha.

8.2 Testing of Risk Predictions

8.2.1 Visual Aids

Strong explanatory power from a returns model does not necessarily translate into an accurate risk model. Low R^2's, for example, may simply indicate little common behavior within the market, not that the model is failing in any way. The ultimate test of a risk model lies in the testing of its risk forecasts against realized values.

This can be done visually by comparing *ex-ante* model predictions of risk for portfolios, individual assets, or factors with their realized (forward-looking) values. For example, figure 9 shows the 30-day realized risk for the U.S. Russell 1000 benchmark versus risk forecasts from the Axioma U.S. Short-Term Model. The model risk appears to do a reasonable job of tracking the overall trend while avoiding the short-lived shocks.

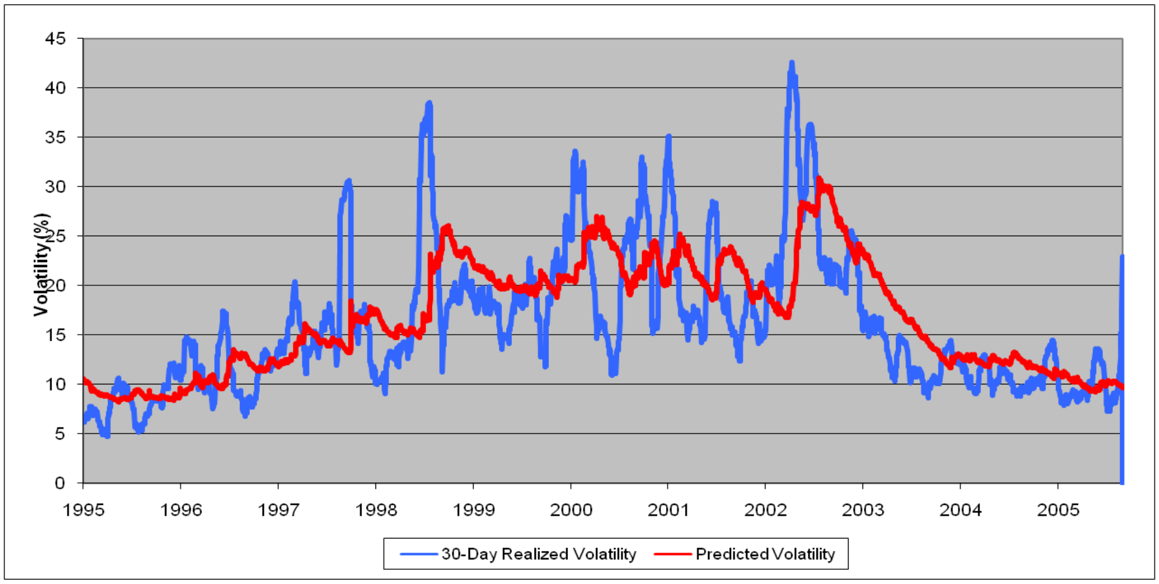


Figure 9: Predicted vs. realized volatilities, Russell 1000 Index

8.2.2 Bias Statistics

To complement visual aids, bias statistics provide hard numbers to assess how well a model is estimating risk. Testing the accuracy of each factor variance and correlation is cumbersome, if not outright infeasible. In contrast, the bias statistic measures the overall quality of the model's full covariance matrix in one go. Given a history of returns, r_t, $t = 1, \ldots, T$, and a history of predicted risks, σ_t, $t = 1, \ldots, T$, we compute the standardized return, Z_t, as follows:

$$Z_t = \frac{r_t}{\sigma_t}$$

Depending on the type of risk one is trying to test, this can be the standardized return of a single asset or a whole portfolio. It can be calculated in terms of total, common factor (or a subset thereof), or specific standardized return, in either active or absolute mode.

The bias statistic is defined as the standard deviation of the standardized returns:

$$\text{Bias Statistic} = \sqrt{var(Z_t)}$$

If the risk forecasts are accurate, then the bias statistic should be close to one. If the bias statistic is significantly below one, then the model has over-predicted risk. Conversely, if the bias statistic is significantly greater than one, then the model has under-predicted risk.

Assuming a large sample size and normally-distributed returns, the bias statistic have a 95% confidence interval of

$$\left[1 - \sqrt{2/T}, 1 + \sqrt{2/T}\right]$$

When the computed bias statistic lies within this interval, the null hypothesis that the model is unbiased cannot be rejected. Conversely, when the computed bias statistic lies outside this interval, the null hypothesis that the model is unbiased is rejected.

8.2.3 Forecast Changes and Turnover

Satisfactory bias statistics may indicate that risk forecasts have achieved some degree of accuracy, but do not ensure their stability over time. Risk predictions cannot be "too volatile"; excessive day-on-day change in forecasted risk is likely to lead to continual portfolio rebalancing with all of the expense that this implies. A useful model, therefore, should strike a balance between accuracy of risk forecasts and smoothness of volatility over time.

Forecast variability can be defined as the average relative forecast change, FC: the absolute change in forecast risk from one period to the next, divided by the forecast for the earlier period, averaged over multiple periods. Obviously, excessive changes in risk forecast have an adverse effect on turnover. A very rough rule of thumb is that the amount of turnover in an active portfolio due to forecast changes is approximately $6 \times FC$.

Ultimately, the question of what constitutes excessive forecast change and turnover is a subjective one. The ultimate authority is, of course, the model-user.

A Appendix: Constrained Regression

Our problem is as follows: calculate

$$\min \|Ax - b\|_2$$

subject to the linear constraint

$$Bx = d,$$

where $A \in \Re^{n\times k}$, $B \in \Re^{p\times k}$. We assume the following:

- $n \geq k \geq p$
- B has rank p
- A has rank q where $p + q \geq k$.

We simultaneously decompose A and B via the generalized singular value decomposition, thus:

$$A = U\Sigma X^{-1}$$

$$B = V\Delta X^{-1}$$

and partition the above as follows:

$$A = \begin{bmatrix} U_p & U_{k-p} & U_{n-k} \end{bmatrix} \begin{bmatrix} \Sigma_p & 0 \\ 0 & \Sigma_{k-p} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_p^{-1} \\ X_{k-p}^{-1} \end{bmatrix},$$

$$B = V \begin{bmatrix} \Delta_p & 0 \end{bmatrix} \begin{bmatrix} X_p^{-1} \\ X_{k-p}^{-1} \end{bmatrix}.$$

We assume without loss of generality that the singular values of A have been sorted so that all zero singular values lie within the block Σ_p. If we make the simple transformation of variables

$$X^{-1}x = y = \begin{bmatrix} y_p \\ y_{k-p} \end{bmatrix}$$

then we may rephrase the problem as

$$\min \|U_p\Sigma_p y_p + U_{k-p}\Sigma_{k-p}y_{k-p} - b\|_2$$

such that

$$V\Delta_p y_p = d.$$

There are two ways in which we may attempt to solve this.

A.1 Exact Solution

Note that the constraint is now uniquely solvable, viz.

$$y_p = \Delta_p^{-1}V^T d,$$

This may be substituted into the minimization equation, resulting in the unconstrained problem

$$\min \left\| U_{k-p}\Sigma_{k-p}y_{k-p} - \hat{b} \right\|_2$$

where

$$\hat{b} = b - U_p \Sigma_p \Delta_p^{-1} V^T d$$

This is easily shown to have the solution

$$y_{k-p} = \Sigma_{k-p}^{-1} U_{k-p}^T \hat{b}$$

And so, we arrive at the (almost) final solution

$$y = \begin{bmatrix} \Delta_p^{-1} V^T d \\ \Sigma_{k-p}^{-1} U_{k-p}^T (b - U_p \Sigma_p \Delta_p^{-1} V^T d) \end{bmatrix}$$

We convert back to the original variable, x by hitting the above on the left-hand side with the orthogonal transformation X, but there's no need to do that right now.

Note that if $p+q=k$, i.e. we have exactly the number of constraints needed to compensate for A's rank-deficiency, then the solution becomes

$$y = \begin{bmatrix} \Delta_p^{-1} V^T d \\ \Sigma_{k-p}^{-1} U_{k-p}^T b \end{bmatrix}$$

A.2 Dummy Asset Approach

Adding the constraints as dummy assets is tantamount to solving the following unconstrained problem

$$\min \left\| \begin{bmatrix} A \\ \lambda B \end{bmatrix} x - \begin{bmatrix} b \\ \lambda d \end{bmatrix} \right\|_2 .$$

And the question we ask ourselves is, how closely does the solution to this problem approach the exact solution? Making the same transformation of variables as before, we are effectively solving

$$\min \left\| \begin{bmatrix} U_p \Sigma_p & U_{k-p} \Sigma_{k-p} \\ \lambda V \Delta_p & 0 \end{bmatrix} \begin{bmatrix} y_p \\ y_{k-p} \end{bmatrix} - \begin{bmatrix} b \\ \lambda d \end{bmatrix} \right\|_2 .$$

We solve this in the usual manner, viz.

$$y = \left(\begin{bmatrix} \Sigma_p U_p^T & \lambda \Delta_p V^T \\ \Sigma_k - p U_{k-p}^T & 0 \end{bmatrix} \begin{bmatrix} U_p \Sigma_p & U_{k-p} \Sigma_{k-p} \\ \lambda V \Delta_p & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} \Sigma_p U_p^T & \lambda \Delta_p V^T \\ \Sigma_{k-p} U_{k-p}^T & 0 \end{bmatrix} \begin{bmatrix} b \\ \lambda d \end{bmatrix} .$$

After working through the linear algebra, this yields

$$y = \begin{bmatrix} \left(\Sigma_p^2 + \lambda^2 \Delta_p^2\right)^{-1} \Sigma_p U_p^T b + \lambda^2 \left(\Sigma_p^2 + \lambda^2 \Delta_p^2\right)^{-1} \Delta_p V^T d \\ \Sigma_{k-p}^{-1} U_{k-p}^T b \end{bmatrix} .$$

Note that

$$\left(\Sigma_p^2 + \lambda^2 \Delta_p^2\right)^{-1} = \frac{1}{\lambda^2} \left(\frac{\Sigma_p^2}{\lambda^2} + \Delta_p^2 \right)^{-1} ,$$

and so, asymptotically, we have that

$$y \to \begin{bmatrix} \Delta_p^{-1} V^T d \\ \Sigma_{k-p}^{-1} U_{k-p}^T b \end{bmatrix} \quad as \ \lambda \to \infty .$$

This is equal to the exact solution in the event that either $p+q=k$ or $d=0$.

A.3 Conclusion

If we have exactly the same number of constraints as the rank-deficiency of our exposure matrix, or all constraints are equal to zero, then for a suitably large dummy asset weight (the Lagrange multiplier), our solution will be consistent (in the case of ordinary least squares).

References

John Y. Campbell, Andrew W. Lo, and A. Craig MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.

Gregory Connor. The three types of factor models: A comparison of their explanatory power. *Financial Analysts Journal*, 51:42–46, 1995.

Gregory Connor and Robert A. Korajczyk. Risk and return and an equilibrium APT. *Journal of Financial Economics*, 21:255–289, 1988.

Giorgio De Santis, Bob Litterman, Adrien Vesval, and Kurt Winkelmann. Covariance matrix estimation. In *Modern Investment Management: An Equilibrium Approach*, chapter 16, pages 224–248. John Wiley & Sons, Inc., 2003.

Robert F. Engle and Clive W.J. Granger. Co-integration and error correction: Representation, estimation, and testing. econometrica. *Econometrica*, 55(2):251–276, 1987.

John Fox. *Web Appendix, An R and S-PLUS Companion to Applied Regression*. Sage Publications, 2002. URL `http://socserv.mcmaster.ca/jfox/Books/Companion/appendix.html`.

William H. Greene. *Econometric Analysis*. Prentice Hall, fifth edition, 2003.

Richard C. Grinold and Ronald N. Kahn. *Active Portfolio Management*, pages 37–66. McGraw-Hill, 1995.

Richard A. Johnson and Dean W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, fourth edition, 1998.

Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

Stephen A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13: 341–360, 1976.

Peter Zangari. Equity factor risk models. In *Modern Investment Management: An Equilibrium Approach*, chapter 20, pages 334–395. John Wiley & Sons, Inc., 2003.

Axioma

Flexible is better.®

United States and Canada: 212-991-4500

Europe: +44 (0)20 7856 2424

Asia: +852-8203-2790

New York Office
Axioma, Inc.
17 State Street
Suite 2700
New York, NY 10004
Phone: **212-991-4500**
Fax: **212-991-4539**

London Office
Axioma, (UK) Ltd.
30 Crown Place
London, EC2A 4EB
Phone: **+44 (0) 20 7856 2451**
Fax: **+44 (0) 20 3006 8747**

Atlanta Office
Axioma, Inc.
400 Northridge Road
Suite 850
Atlanta, GA 30350
Phone: **678-672-5400**
Fax: **678-672-5401**

Hong Kong Office
Axioma, (HK) Ltd.
Unit B, 17/F, Entertainment Building
30 Queen's Road Central
Hong Kong
Phone: **+852-8203-2790**
Fax: **+852-8203-2774**

San Francisco Office
Axioma, Inc.
201 Mission Street
Suite 2150
San Francisco, CA 94105
Phone: **415-614-4170**
Fax: **415-614-4169**

Singapore Office
Axioma, (Asia) Pte Ltd.
30 Raffles Place
#23-00 Chevron House
Singapore 048622
Phone: **+65 6233 6835**
Fax: **+65 6233 6891**

Geneva Office
Axioma CH
Rue du Rhone 69, 2nd Floor
1207 Geneva, Switzerland
Phone: **+33 611 96 81 53**

Sydney Office
Axioma AU
Level 4,5 & 12
95 Pitt Street, 506b
NSW 2000
Sydney, Australia
Phone: **+61 (2) 8079 2915**

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