Betas, Benchmarks and Beating the Market

Zura Kakushadze§†1 and Willie Yu♯2

§ Quantigic® Solutions LLC
1127 High Ridge Road #135, Stamford, CT 06905 3
† Free University of Tbilisi, Business School & School of Physics
240, David Agmashenebeli Alley, Tbilisi, 0159, Georgia
♯ Centre for Computational Biology, Duke-NUS Medical School
8 College Road, Singapore 169857

(May 30, 2018)

Abstract

We give an explicit formulaic algorithm and source code for building long-only benchmark portfolios and then using these benchmarks in long-only market outperformance strategies. The benchmarks (or the corresponding betas) do not involve any principal components, nor do they require iterations. Instead, we use a multifactor risk model (which utilizes multilevel industry classification or clustering) specifically tailored to long-only benchmark portfolios to compute their weights, which are explicitly positive in our construction.

---

1 Zura Kakushadze, Ph.D., is the President and CEO of Quantigic® Solutions LLC, and a Full Professor at Free University of Tbilisi. Email: zura@quantigic.com
2 Willie Yu, Ph.D., is a Research Fellow at Duke-NUS Medical School. Email: willie.yu@duke-nus.edu.sg
3 DISCLAIMER: This address is used by the corresponding author for no purpose other than to indicate his professional affiliation as is customary in publications. In particular, the contents of this paper are not intended as an investment, legal, tax or any other such advice, and in no way represent views of Quantigic® Solutions LLC, the website www.quantigic.com or any of their other affiliates.
1 Introduction and Summary

Diversified long-only portfolios consisting of many stocks are invariably exposed to broad market movements. So, is there an “optimal” way of constructing such long-only portfolios? Here we can principally distinguish two rather different cases.

In the first case, we have no detailed expectations about individual stock returns. I.e., we are trying to construct a long-only portfolio oblivious to any trading signals (or alphas). We can think of such a portfolio as a benchmark. One possibility is to use off-the-shelf (market cap weighted) broad market portfolios such as S&P 500 or Russell 3000. Another simple approach is to use a minimum-variance portfolio, whose weights \( w_i \) (i = 1, . . . , N labels N stocks in our universe), up to an overall normalization, are given by

\[
    w_i \propto \sum_{j=1}^{N} C_{ij}^{-1}
\]

There are several issues with this. First, if \( C_{ij} \) is a sample covariance matrix based on a time-series of historical stock returns, in many cases it is singular as there are not enough observations in the time-series. Second, even if it is nonsingular, the off-diagonal elements of \( C_{ij} \) (more, precisely, pair-wise correlations) are highly unstable out-of-sample, and thus so are the benchmark weights \( w_i \). Third, the betas \( \beta_i \) [Sharpe, 1963] of the individual stocks w.r.t. to this portfolio are all equal 1 (up to an overall normalization factor), so the correlations of the stock s with the benchmark end up being inversely proportional to the historical standard deviations \( \sigma_i \) of their returns, which is not what we expect for a broad market benchmark.\(^4\) Fourth, generally, some weights (1), even if computable, are negative unless we construct a minimum-variance portfolio subject to lower bounds on the weights, thereby either excluding stocks that would otherwise have negative weights\(^5\) (thereby, among other things, diminishing diversification and distorting the remainder of the weights), or assigning some ad hoc minimum weights to such stocks (and still distorting the other weights). Some of these issues can be attempted to be circumvented by using (up to an overall normalization factor) the first principal component of \( C_{ij} \) as the weights (and, consequently, the betas).\(^6\) However, the issues with negative weights and out-of-sample instability (the latter perhaps to a lesser degree) still persist.

In the second case, assume that we can forecast non-random expected returns \( E_i \) for individual stocks. We can try to construct a long-only portfolio that outperforms a benchmark portfolio. Mean-variance optimization [Markowitz, 1952] would give

\[
    w_i \propto \sum_{j=1}^{N} C_{ij}^{-1} E_j
\]

\(^4\) It has been appreciated for decades that betas are highly unstable [Fabozzi and Francis, 1978].

\(^5\) Actually, because of the off-diagonal elements in \( C_{ij} \), the stocks that are excluded are not exactly the same as those with \( w_i \leq 0 \) in (1).

\(^6\) See, e.g., [Avellaneda and Lee, 2010], [Connor and Korajczyk, 1993], [Geweke and Zhou, 1996], [Trzcinka, 1986].
This suffers from most of the issues with (1). One way around this is to: i) first build a benchmark long-only portfolio (which is oblivious to \( E_i \)); and ii) then build a long-short dollar-neutral portfolio on top of it based on \( E_i \) such that the combined portfolio is still long-only and well-diversified. The dollar-neutral portfolio can be built using standard optimization techniques by employing a well-built, stable multifactor risk model instead of the sample covariance matrix \( C_{ij} \). We emphasize that this construction is not simply a “trend-following” or “market-timing” strategy. Instead, this is a systematic way of constructing a long-only portfolio as a combination of a long-only “passive” benchmark\(^7\) and an “actively managed” dollar-neutral portfolio (e.g., based on statistical arbitrage), and the latter is expected to produce positive returns on its own. However, in the context of long-only portfolios, this dollar-neutral strategy is more constrained, as the net positions must be all long. So, this is a way of generating excess returns above that of the “passive” benchmark, the price being the risk associated with the dollar-neutral strategy (no free lunch).

In this paper we discuss a systematic approach to the above “program”. First, we give an explicit formulaic algorithm for constructing a benchmark portfolio given a set of betas, which can be chosen to have various desirable properties. The resultant benchmark weights are positive by construction. This is achieved by building a multifactor risk model \( \Gamma_{ij} \) and using it instead of the sample covariance matrix \( C_{ij} \) specifically for constructing a long-only benchmark. This \( \Gamma_{ij} \) is carefully built based on a binary multilevel industry classification (or some other clustering scheme). It does not involve any principal components, and there are no iterations (due to, e.g., lower bounds) required to obtain \( w_i \). In fact, they are given by a simple but nontrivial (and arguably elegant) formula, which is one of the main results of this paper. We give a detailed construction of \( \Gamma_{ij} \) and the benchmark weights \( w_i \) in Section 3 after reviewing some generalities pertaining to betas in Section 2. In a nutshell, the weights \( w_i \) are expressed via a product of simple algebraic quantities built from specific variances at each level in the industry classification (clustering scheme). These specific variances carry nontrivial information about the underlying stock returns. The source code for computing \( \Gamma_{ij} \) and \( w_i \) is given in Appendix A.\(^8\)

In Section 4 we then discuss the outperformance strategy based on overlaying an “actively managed” dollar-neutral strategy on top of a “passive” benchmark. Given the expected returns \( E_i \), the dollar-neutral portfolio can be constructed using standard optimization techniques with bounds. However, the nontrivial part is that the multifactor risk model \( \Gamma'_{ij} \) used in this optimization is different from \( \Gamma_{ij} \) used in constructing the benchmark. Further, this strategy is similar to the S&P 500 outperformance strategy, where a long position in the S&P 500 stocks (with dynamic weights) is offset by a short S&P futures position.\(^9\) Section 5 briefly concludes.

---

\(^7\) This “passive” benchmark can be thought of as an index. See, e.g., [Lo, 2016] for an overview.

\(^8\) The source code given in Appendix A hereof is not written to be “fancy” or optimized for speed or in any other way. Its sole purpose is to illustrate the algorithms described in the main text in a simple-to-understand fashion. Some important legalese is relegated to Appendix B.

\(^9\) Here: i) we have no futures, and ii) the benchmark is not S&P 500 based (albeit, it can be).
2 Betas

Consider a universe of $N$ stocks labeled by $i = 1, \ldots, N$. Let $R_{is}$ be a time-series of stock returns (e.g., daily, weekly, monthly, etc., close-to-close returns). Here the index $s = 1, \ldots, T$ labels trading days on which these returns are computed ($s = 1$ labels the most recent date).\(^{10}\) Let $F_s$ be another time-series of returns, which can but need not correspond to a portfolio of stocks, such as an index. A priori it can be any time-series of returns (e.g., on a commodity, bond, currency, etc.). For our purposes, we can think of $F_s$ as the returns of a benchmark portfolio (see below).

Stock betas $\beta_i$ can be thought of as measuring how closely each stock “follows” the movement of the benchmark. Betas can be formally defined via a serial regression with the intercept (and unit weights) of the returns $R_{is}$ over $F_s$:

$$R_{is} = \alpha_i \ I_s + \beta_i \ F_s + \epsilon_{is} \quad (3)$$

Here: $\alpha_i$ is the regression coefficient of the intercept $I_s \equiv 1$; $\beta_i$ is the coefficient of $F_s$; and $\epsilon_{is}$ are the regression residuals. The regression coefficients are given by

$$\alpha_i = \overline{R}_i - \beta_i \ \overline{F} \quad (4)$$

$$\beta_i = \frac{\sum_{s=1}^{T} \tilde{R}_{is} \ \tilde{F}_s}{\sum_{s=1}^{T} \tilde{F}_s^2} \quad (5)$$

where for a time-series $A_s$

$$\overline{A}_s = \frac{1}{T} \sum_{s=1}^{T} A_s \quad (6)$$

$$\tilde{A}_s = A_s - \overline{A}_s \quad (7)$$

So, the betas are given by

$$\beta_i = \frac{\text{Cov}(R_{is}, F_s)}{\text{Cov}(F_s, F_s)} \quad (8)$$

where $\text{Cov}(\cdot, \cdot)$ denotes a serial covariance. If $F_s$ are the returns of a portfolio of the stocks in our universe with weights $w_i$, i.e., if

$$F_s = \sum_{i=1}^{N} w_i \ R_{is} \quad (9)$$

then we have (here $C_{ij}$ is the sample covariance matrix of the stock returns $R_{is}$)

$$\beta_i = \frac{1}{\sigma_F^2} \sum_{j=1}^{N} C_{ij} \ w_j \quad (10)$$

$$\sigma_F^2 = \sum_{i,j=1}^{N} C_{ij} \ w_i \ w_j \quad (11)$$

$$C_{ij} = \text{Cov}(R_{is}, R_{js}) \quad (12)$$

---

\(^{10}\) Here $R_{is}$ can be raw returns or excess returns w.r.t. a risk-free rate, depending on the context.
3 Benchmarks

Suppose we now turn the tables and attempt to construct the benchmark portfolio – i.e., determine the weights \( w_i \) – given the betas \( \beta_i \). This can be done by formally solving (10) via (in defining \( \tilde{C}_{ij} \) we assume that all \( \beta_i \neq 0 \)):

\[
    w_i = \sigma_F^2 \sum_{j=1}^{N} C_{ij}^{-1} \beta_j = \frac{\sigma_F^2}{\beta_i} \sum_{j=1}^{N} \tilde{C}_{ij}^{-1}
\]

(13)

\[
    \sigma_F^2 = \left[ \sum_{i,j=1}^{N} C_{ij}^{-1} \beta_i \beta_j \right]^{-1} = \left[ \sum_{i,j=1}^{N} \tilde{C}_{ij}^{-1} \right]^{-1}
\]

(14)

\[
    \tilde{C}_{ij} = \frac{1}{\beta_i \beta_j} C_{ij}
\]

(15)

where \( C_{ij}^{-1} \) is the matrix inverse to \( C_{ij} \). However, in practice there are some issues with this formal solution. First, if \( T < N + 1 \), the sample covariance matrix \( C_{ij} \) is singular (i.e., not invertible). Second, unless \( T \gg N \), which is seldom (if ever) the case in practice, the off-diagonal elements (in particular, the correlations – the diagonal elements are relatively stable) are highly unstable out-of-sample rendering \( C_{ij} \) essentially useless (unpredictive out-of-sample). Third, even if all \( \beta_i \) are positive (e.g., \( \beta_i \equiv 1 \)), assuming \( C_{ij} \) is invertible, the weights given by (13) are not necessarily all positive due to nontrivial correlations between stocks. In many applications the benchmark is a long-only portfolio, which requires the weights \( w_i \) to be nonnegative.

3.1 Principal Components

For a moment, let us ignore the aforesaid issues with the sample covariance matrix \( C_{ij} \) and assume that it somehow – magically – is invertible and stable out-of-sample.\(^{11}\) Then there is a simple “solution” to constructing a benchmark portfolio.\(^{12}\) Thus, it is tempting to take \( \beta_i = \gamma V_i^{(1)} \), where \( \gamma > 0 \) is an overall normalization coefficient, and \( V_i^{(1)} \) is the first principal component of \( C_{ij} \):

\[
    \sum_{j=1}^{N} C_{ij} V_j^{(a)} = \lambda^{(a)} V_i^{(a)}, \quad a = 1, \ldots, N
\]

(16)

\[
    \sum_{i=1}^{N} V_i^{(a)} V_i^{(b)} = \delta_{ab}
\]

(17)

\[
    \sum_{a=1}^{N} V_i^{(a)} V_j^{(a)} = \delta_{ij}
\]

(18)

\(^{11}\) Strictly speaking, invertibility is not required here, but this is not critical.

\(^{12}\) However, as we discuss in a moment, this “solution” is intrinsically flawed.
Here $V_i^{(a)}$ denotes the $a$-th principal component of $C_{ij}$ corresponding to the eigenvalue $\lambda^{(a)}$, where the eigenvalues are organized in the descending order: $\lambda^{(1)} > \lambda^{(2)} > \cdots > \lambda^{(N)}$. Then the portfolio weights are given by
\[ w_i = V_i^{(1)}/\gamma \] (19)

More generally, we can take
\[ \beta_i = \gamma_i \ U_i^{(1)} \] (20)
\[ w_i = U_i^{(1)}/\gamma_i \] (21)
\[ \sum_{j=1}^{N} D_{ij} \ U_j^{(a)} = \tilde{\lambda}^{(a)} \ U_i^{(a)}, \ a = 1, \ldots, N \] (22)
\[ D_{ij} = \frac{1}{\gamma_i \gamma_j} C_{ij} \] (23)

Here $\gamma_i > 0$ is some $N$-vector, and $U_i^{(a)}$ are principal components of the matrix $D_{ij}$.

Let us start with (19). One immediate issue with this “solution” is that this portfolio is highly skewed, to wit, it is over-weighted with highly volatile stocks. This is because: i) the stock volatilities $\sigma_i = \sqrt{C_{ii}}$ have a skewed (roughly log-normal) distribution with a long tail for higher values of $\sigma_i$; and ii) the elements of the first principal component roughly scale as $V_i^{(1)} \propto \sigma_i$. To see this, consider a simple model where all stocks have uniform pair-wise correlations equal $\rho$:14
\[ C_{ij} = \sigma_i \sigma_j \left[ (1 - \rho) \delta_{ij} + \rho \nu_i \nu_j \right] \] (24)
where $\nu_i \equiv 1$ is the unit $N$-vector. Straightforward algebra leads to the following eigenvalue equation:
\[ \rho \sum_{i=1}^{N} \frac{\sigma_i^2}{\lambda - (1 - \rho) \sigma_i^2} = 1 \] (25)
whose $N$ solutions are the eigenvalues $\lambda = \lambda^{(a)}, \ a = 1, \ldots, N$. The eigenvectors are given by:
\[ V_i^{(a)} = \frac{\eta^{(a)} \sigma_i}{\lambda^{(a)} - (1 - \rho) \sigma_i^2} \] (26)
where $\eta^{(a)}$ are overall normalization factors (fixed via (17)). We can solve (25) iteratively via
\[ \lambda = \rho \sum_{i=1}^{N} \frac{\sigma_i^2}{1 - (1 - \rho) \sigma_i^2/\lambda} \] (27)

---

13 In theory some eigenvalues can be degenerate, but in practice positive eigenvalues are not (and here we are assuming there are no null eigenvalues). This is not critical here as we are interested in the first principal component, whose eigenvalue can be safely assumed to be non-degenerate.

14 I.e., the sample correlation matrix $\Psi_{ij} = C_{ij}/\sigma_i \sigma_j$ is replaced by a 1-factor model (see below), where $\rho$ can be taken as the average pair-wise correlation $\bar{\rho} = \sum_{i,j=1; i \neq j}^{N} \Psi_{ij}/N(N-1)$. This is an oversimplified model, but (as we discuss below) it correctly captures the point we make here.
assuming $\lambda \gg \rho \sigma_i^2$ for all $i$. In the zeroth approximation we have\(^{15}\)

\[
\lambda^{(1)} \approx \rho \sum_{i=1}^{N} \sigma_i^2 \tag{28}
\]

\[
V_i^{(1)} \approx \frac{\sigma_i}{\sqrt{\sum_{i=1}^{N} \sigma_i^2}} \tag{29}
\]

It is now evident that the portfolio given by the weights (19) is indeed skewed and overloaded with high volatility stocks. This conclusion persists even if we do not assume uniform correlations as in (24). Thus, straightforward algebra will convince the reader that this is the case for, e.g., the following model:

\[
C_{ij} = \sigma_i \sigma_j \tilde{\Psi}_{ij} \tag{30}
\]

\[
\tilde{\Psi}_{ij} = \tilde{c}_i^2 \delta_{ij} + \tilde{\lambda}^{(1)} U_i^{(1)} U_j^{(1)} \tag{31}
\]

\[
\tilde{c}_i^2 = 1 - \tilde{\lambda}^{(1)} [U_i^{(1)}]^2 \tag{32}
\]

where $U_i^{(1)}$ is the first principal component of the sample correlation matrix $\Psi_{ij} = C_{ij}/\sigma_i \sigma_j$, and $\tilde{\lambda}^{(1)}$ is the corresponding eigenvalue.\(^{16}\) Note that $0 < \tilde{c}_i^2 < 1$. Now we no longer have uniform correlations, and (in the zeroth approximation) we have $V_i^{(1)} \propto \sigma_i U_i^{(1)}$, so $V_i^{(1)}$ are skewed for large $\sigma_i$ (as $U_i^{(1)}$ are independent of $\sigma_i$ and are not skewed; in the zeroth approximation $U_i^{(1)} \approx 1/\sqrt{N}$). So the problem persists.

A simple “fix” is to take skewed betas to begin with, e.g., if we take $\gamma_i = \sigma_i$ in (20), then we have $w_i = U_i^{(1)}/\sigma_i$, where $U_i^{(1)}$ is the first principal component of the sample correlation matrix $\Psi_{ij}$. Now $w_i$ are skewed but in the opposite direction – the weights for volatile stocks are suppressed, which is a welcome feature. This would seem to solve the problem of constructing an acceptable benchmark. However, there are other issues we have been postponing dealing with. First, generally, some elements of $U_i^{(1)}$ are negative.\(^{17}\) Since for long-only benchmarks we wish to have

\(^{15}\) The actual $\lambda^{(1)}$ is even higher and the actual $V_i^{(1)}$ are even more skewed for large $\sigma_i$, which further aids the point we make here. At any rate, in practice the zeroth approximation is pretty good. E.g., for the dataset discussed in Subsection 2.6 of [Kakushadze and Yu, 2018], which consists of 21-day historical volatilities based on daily close-to-close returns for 3810 U.S. stocks, we have the following statistics for the ratio $\sigma_i^2/\sum_{j=1}^{N} \sigma_j^2$: Min = $1.70 \times 10^{-7}$, 1st Quartile = $3.48 \times 10^{-5}$, Median = $7.79 \times 10^{-5}$, Mean = $2.63 \times 10^{-4}$, 3rd Quartile = $2.00 \times 10^{-4}$, Max = $4.39 \times 10^{-2}$. Therefore, the zeroth approximation (28) is justified assuming $\rho$ is not too small, which is the case considering that the 21-day historical average pair-wise correlation for the aforesaid dataset $\rho = 0.10$. So, if we set $\rho = \rho$, then for $\kappa_i = (1 - \rho) \sigma_i^2/\rho \sum_{j=1}^{N} \sigma_j^2$ we have that $\max(\kappa_i) \approx 0.4$ and for most stocks $\kappa_i \ll 1$. Further, a relatively small number of stocks for which $\kappa_i$ is not small have even larger contributions to $\lambda^{(1)}$ and $V_i^{(1)}$ compared with the aforesaid zeroth approximation, which further exacerbates the skewness of $V_i^{(1)}$ (and thus $w_i$) for the large $\sigma_i$ stocks (see below).

\(^{16}\) In (30) the sample correlation matrix $\Psi_{ij}$ is replaced by a 1-factor statistical risk model $\tilde{\Psi}_{ij}$ based on the first principal component $U_i^{(1)}$ of $\Psi_{ij}$. See [Kakushadze and Yu, 2017] for details.

\(^{17}\) E.g., in the dataset mentioned in fn. 15, out of the 3810 elements, 575 (or over 15%) of $U_i^{(1)}$ and 668 elements of $V_i^{(1)}$ are negative. This is a sizable number.
$w_i > 0$, this poses an issue. One way to circumvent this is not to use (20) (with $\gamma_i = \sigma_i$), but instead to proceed as follows. First, instead of using the sample covariance matrix $C_{ij}$, we model it via the 1-factor statistical risk model (30) (see fn. 16). The inverse of $C_{ij}$ is given by

$$C_{ij}^{-1} = \frac{1}{\sigma_i \sigma_j} \tilde{\Psi}_{ij}^{-1}$$

(33)

$$\tilde{\Psi}_{ij}^{-1} = \frac{1}{\xi_i^2} \delta_{ij} - q \frac{U_i^{(1)} U_j^{(1)}}{\xi_i^2 \xi_j^2}$$

(34)

$$q = \left[ 1/\tilde{\lambda}^{(1)} + \sum_{k=1}^{N} [U_k^{(1)}]^2 / \xi_k^2 \right]^{-1}$$

(35)

Let \( H_+ = \{ i | U_i^{(1)} \geq 0 \} \) and \( H_- = \{ i | U_i^{(1)} < 0 \} \). Then we can set (here $\gamma$ is an overall normalization constant, while $\kappa$ is a parameter to be fixed)

$$\beta_i = \gamma \sigma_i U_i^{(1)}, \quad i \in H_+$$

(36)

$$\beta_i = \kappa \gamma \sigma_i U_i^{(1)}, \quad i \in H_-$$

(37)

$$w_i = \frac{\tilde{q} U_i^{(1)}}{\gamma \sigma_i \xi_i^2} \left[ 1/\tilde{\lambda}^{(1)} + (1 - \kappa) G_- \right]^-, \quad i \in H_+$$

(38)

$$w_i = \frac{\tilde{q} U_i^{(1)}}{\gamma \sigma_i \xi_i^2} \left[ \kappa/\tilde{\lambda}^{(1)} - (1 - \kappa) G_+ \right]^+, \quad i \in H_-$$

(39)

$$\tilde{q} = \left[ (G_+ + \kappa^2 G_-) / \tilde{\lambda}^{(1)} + (1 - \kappa)^2 G_+ G_- \right]^{-1}$$

(40)

$$G_+ = \sum_{k \in H_+} [U_k^{(1)}]^2 / \xi_k^2$$

(41)

$$G_- = \sum_{k \in H_-} [U_k^{(1)}]^2 / \xi_k^2$$

(42)

It then follows that for $i \in H_- \,$ we have: $w_i < 0$ for $\kappa_* < \kappa \leq 1; \,$ $w_i = 0$ for $\kappa = \kappa_*; \,$ and $w_i > 0$ for $\kappa < \kappa_*$. Here

$$\kappa_* = \frac{G_+}{1/\tilde{\lambda}^{(1)} + G_+}$$

(43)

Now: i) $N_+ = |H_+|$ is sizably larger than $N_- = |H_-|$ (so $G_+ \gg 1$); and ii) $1/\tilde{\lambda}^{(1)} \ll 1$, so the corresponding contribution to $\kappa_*$ in (43) is subleading. As a result, $\kappa_* \approx 1 - 1/G_+ \tilde{\lambda}^{(1)}$ is very close to 1. Also, $G_+ \gg G_-$. Naïvely, it might appear reasonable to assume that $G_+/G_- \sim N_+/N_-$. However, typically, $G_+/G_-$ is sizably larger than $N_+/N_-$. The additional skew is due to the fact that the absolute value of the average negative pair-wise correlation is sizably lower than the average positive

\[\text{In practice, generically, no element of } U_i^{(1)} \text{ is expected to be exactly 0.}\]
pair-wise correlation,\(^{19}\) which implies that average \([U_i^{(1)}]^2\) for \(i \in H_-\) is sizably lower than average \([U_i^{(1)}]^2\) for \(i \in H_+\), which in turn implies that on average \(\hat{\xi}_i^2\) are closer to 1 for \(i \in H_-\) than for \(i \in H_+\). So, let \(\kappa = 1 - \zeta/G_+\bar{\lambda}^{(1)}\). For \(\zeta = 1\) we have \(\kappa \approx \kappa_\ast\) (up to subleading corrections). For \(\zeta > 1\) we have \(w_i > 0\) for \(i \in H_-\). However, if we take \(\zeta \gg 1\), then on average the weights \(w_i\) for \(i \in H_-\) will be much higher than for \(i \in H_+\). This implies that \(\zeta \sim 1\). Then we have (up to subleading corrections)

\[
\begin{align*}
\quad w_i & \approx \frac{U_i^{(1)}}{\gamma \sigma_i \hat{\xi}_i^2 G_+}, & i \in H_+ \\
\quad w_i & \approx -\frac{U_i^{(1)}}{\gamma \sigma_i \hat{\xi}_i^2 G_+} [\zeta - 1], & i \in H_-
\end{align*}
\]

(44)

(45)

So, we can determine the value of \(\zeta = \zeta_\ast\) such that the mean (or median)\(^{20}\) of \(w_i\) is the same in \(H_+\) and \(H_-\). This way stocks from these two sets will on average be weighted more or less equally in the benchmark, with all \(w_i > 0\) (and \(\beta_i < 0\) in \(H_-\)).

Here one may argue that the solution with \(\zeta = 2\), where \(w_i \approx [U_i^{(1)}]/\gamma \sigma_i \hat{\xi}_i^2 G_+\), is more natural (notwithstanding that on average the weights \(w_i\), \(i \in H_-\), are somewhat smaller than the weights \(w_i\), \(i \in H_+\)). However, there is no magic bullet here for fixing \(\zeta\). In fact, the parametrization using a single parameter, \(\zeta\), is just one possibility out of myriad others. Now, as to \(\zeta\), it can be chosen to be pretty much anywhere between (somewhat above) 1 and (somewhat above) \(\zeta_\ast\) defined above.

So, problem solved? Well, not quite. The fact that tiny (of order \(1/G_+\bar{\lambda}^{(1)}\) on a relative basis) changes in \(\beta_i\) produce of order 1 changes in the benchmark weights \(w_i\) should sound an alarm. The culprit here is that, as already mentioned, in the zeroth approximation we have \(U_i^{(1)} \approx 1/\sqrt{N}\). This is the so-called “market mode” – the first principal component of \(\Psi_{ij}\) corresponds to the overall movement of the broad market (see, e.g., [Bouchaud and Potters, 2011]). The deviations from \(U_i^{(1)} \approx 1/\sqrt{N}\) correspond to different stocks having different \(\beta_i/\sigma_i\). However, in the simple 1-factor model (30), very different looking benchmarks have almost the same betas. To see this, consider the weights \(w_i = \omega_i/\sigma_i\), where \(0 < \omega_i \sim 1\). We then have

\[
\beta_i = \frac{\sigma_i}{\hat{\sigma}_i^\ast} \left[ \hat{\xi}_i^2 \omega_i \bar{\lambda}^{(1)} U_i^{(1)} \sum_{j=1}^N U_j^{(1)} \omega_j \right]
\]

(46)

Considering that \(\hat{\xi}_i^2 \sim 1\) and \(\bar{\lambda}^{(1)} \gg 1\), the first term due to the specific risk is

---

\(^{19}\) Thus, for the dataset mentioned in fn. 15, the median (mean) correlation in \(H_+\) is 0.214 (0.239), while in \(H_-\) it is −0.136 (−0.165). The median (mean) \([U_i^{(1)}]^2\) in \(H_+\) is \(2.40 \times 10^{-4}\) \((2.99 \times 10^{-4})\), while in \(H_-\) it is \(2.22 \times 10^{-5}\) \((5.82 \times 10^{-5})\). The median (mean) \(\hat{\xi}_i^2\) in \(H_+\) is 0.838 \((0.799)\), while in \(H_-\) it is 0.985 \((0.961)\). Also, \(G_+ = 1.637, G_- = 0.03919\), and \(\bar{\lambda}^{(1)} = 674.35\).

\(^{20}\) For the dataset mentioned in fn. 15, the value of \(\zeta\), computed based on the median (mean) is 6.988 \((6.181)\). This reflects the asymmetry between the stocks in \(H_+\) and \(H_-\) mentioned above.
subleading and the second term is dominant, so we have

$$\beta_i = \frac{\sigma_i}{\sigma_F} \tilde{\lambda}^{(1)} U_i^{(1)} \sum_{j=1}^{N} U_j^{(1)} \omega_j \left[ 1 + \mathcal{O}(1/\tilde{\lambda}^{(1)}) \right]$$

(47)

So, up to small $\mathcal{O}(1/\tilde{\lambda}^{(1)})$ corrections, $\beta_i$ are simply proportional to $\sigma_i U_i^{(1)}$, irrespective of the individual values of $\omega_i$. Therefore, all that gymnastics for constructing $w_i > 0$ is just a mirage and an exercise in self-deception. Is there a way around this?

To get betas to substantially deviate from $\beta_i \propto \sigma_i U_i^{(1)}$, we can try to include higher principal components via a $K$-factor statistical risk model:

$$C_{ij} = \sigma_i \sigma_j \tilde{\Psi}_{ij}$$

(48)

$$\tilde{\Psi}_{ij} = \tilde{\xi}^2_i \delta_{ij} + \sum_{a=1}^{K} \tilde{\lambda}^{(a)} U_i^{(a)} U_j^{(a)}$$

(49)

$$\tilde{\xi}^2_i = 1 - \sum_{a=1}^{K} \tilde{\lambda}^{(a)} [U_i^{(a)}]^2$$

(50)

Here $K > 1$ can be fixed using eRank (effective rank) of [Roy and Vetterli, 2007] or some other method – see [Kakushadze and Yu, 2017] for details. A “technical” issue with this approach is that ensuring $w_i > 0$ becomes messier. A more important, conceptual issue is that higher principal components are intrinsically unstable out-of-sample. This instability is then inherited by the benchmark weights $w_i$ and the betas $\beta_i$ constructed using such models. This is not a fruitful direction to pursue.

### 3.2 Back to a 1-factor Model

While we can and will consider multifactor models beyond statistical risk models, we are not done with 1-factor models quite yet. Let us consider a general 1-factor model:

$$C_{ij} = \sigma_i \sigma_j \tilde{\Psi}_{ij}$$

(51)

$$\tilde{\Psi}_{ij} = \tilde{\xi}^2_i \delta_{ij} + \Omega_i \Omega_j$$

(52)

$$\tilde{\xi}^2_i = 1 - \Omega_i^2$$

(53)

A priori we must only require that $\Omega_i^2 < 1$ and $\Omega_i$ are arbitrary otherwise. If we take $\Omega_i = \beta_i/\sigma_i$, straightforward algebra gives

$$w_i = \eta \frac{\beta_i}{\xi_i^2}$$

(54)

$$\xi_i^2 = \sigma_i^2 \tilde{\xi}^2_i$$

(55)

$$\eta^{-1} = \sum_{i=1}^{N} \frac{\beta_i^2}{\xi_i^2}$$

(56)
Here $\xi_i^2$ is the specific variance in the covariance matrix $C_{ij}$ (as opposed to $\tilde{\xi}_i^2$, which is the same quantity in the correlation matrix $\tilde{\Psi}_{ij}$). So, up to an overall normalization factor $\eta$, $w_i$ are proportional to $\beta_i/\xi_i^2$, which is the same result as what we would have obtained had we assumed – albeit this may appear strange at first – that $C_{ij} = \xi_i^2 \delta_{ij}$. Put differently, it is as though we ignore the factor risk altogether and only account for specific (idiosyncratic) risk. However, there is actually nothing strange about this result. Importantly, if all $\beta_i > 0$, then automatically all $w_i > 0$.

### 3.3 Interpretation

The weights (13) are nothing but the solution to maximizing the expected Sharpe ratio [Sharpe, 1966], [Sharpe, 1994] $S$ of the benchmark portfolio if we treat $\beta_i$ as the expected returns $E_i$ for our stocks:

$$E_i = \gamma \beta_i$$

where $\gamma$ is an immaterial (for our purposes here) overall normalization factor. The expected Sharpe ratio is given by

$$S = \frac{1}{\sigma_F} \sum_{i=1}^{N} E_i \ w_i$$

$$\sigma_F^2 = \sum_{i,j=1}^{N} C_{ij} \ w_i \ w_j$$

Maximizing $S$ w.r.t. $w_i$, we get (13) up to an overall normalization factor. The latter can be fixed by requiring that

$$\sum_{i=1}^{N} w_i \ \beta_i = 1$$

which is a consequence of (13) and (14). Now we can understand the issue we encountered above with the “market mode”. Maximizing the Sharpe ratio hedges$^{22}$ against the broad market going bust (i.e., all or most stocks selling off at the same time). While this is a natural thing to do when constructing dollar-neutral portfolios, it makes no sense to do this when constructing long-only portfolios. Indeed, long only portfolios are exposed to market risk by definition. Hedging against specific risk does make sense, which amounts to dropping the factor risk from $C_{ij}$ (in a 1-factor model) and simply taking $C_{ij} = \xi_i^2 \delta_{ij}$, i.e., treating it as though we only have specific risk. So, the lesson here is that we must eliminate the “market mode”.

---

$^{21}$ Here we have no transaction costs, bounds or constraints, so maximizing the Sharpe ratio is equivalent to mean-variance optimization [Markowitz, 1952]. See, e.g., [Kakushadze, 2015a].

$^{22}$ Not precisely, but approximately. However, when $N$ is large, for the 1-factor model (e.g., based on the first principal component $U_i^{(1)}$ as in (30)), the “mishedge” is suppressed by $1/N$ [Kakushadze and Yu, 2017].
3.4 Multifactor Models

Next, let us discuss a general multifactor model covariance matrix $\Gamma_{ij}$:

$$\Gamma_{ij} = \xi_i^2 \delta_{ij} + \sum_{A,B=1}^{K} \Omega_{iA} \phi_{AB} \Omega_{jB}$$  \hspace{1cm} (61)

Here: $\xi_i^2$ is the specific (a.k.a. idiosyncratic) risk; $\Omega_{iA}$, $A = 1, \ldots, K$, is the factor loadings matrix; and $\phi_{AB}$ is the factor covariance matrix. For our purposes here it will not be important to know how $\Gamma_{ij}$ is constructed.\footnote{For a general discussion, see, e.g., [Grinold and Kahn, 2000]. For an explicit open-source implementation of a general multifactor risk model for equities, see [Kakushadze and Yu, 2016a].} What matters here is that: the number of risk factors $K \ll N$ ($K$ can still be in hundreds); all $\xi_i > 0$; $\phi_{AB}$ is positive-definite (then so is $\Gamma_{ij}$); $\Gamma_{ii} = C_{ii}$ (so the sample variances are matched).

Using the inverse of $\Gamma_{ij}$

$$\Gamma_{ij}^{-1} = \frac{1}{\xi_i^2} \delta_{ij} - \sum_{A,B=1}^{K} \frac{\Omega_{iA} \Omega_{jB}}{\xi_i^2} \frac{Q_{AB}}{\xi_j^2}$$  \hspace{1cm} (62)

$$Q_{AB} = \phi_{AB}^{-1} + \sum_{i=1}^{N} \frac{1}{\xi_i^2} \Omega_{iA} \Omega_{iB}$$  \hspace{1cm} (63)

in lieu of $C_{ij}^{-1}$ in (13) and (14), we have

$$w_i = \sigma_F^2 \left[ \beta_i - \Upsilon_i \right]$$  \hspace{1cm} (64)

$$\Upsilon_i = \sum_{A,B=1}^{K} \Omega_{iA} Q_{AB}^{-1} \Lambda_B$$  \hspace{1cm} (65)

$$\Lambda_A = \sum_{j=1}^{N} \frac{\beta_j \Omega_{jA}}{\xi_j^2}$$  \hspace{1cm} (66)

$$\sigma_F^2 = \left[ \Theta - \sum_{A,B=1}^{K} \Lambda_A Q_{AB}^{-1} \Lambda_B \right]^{-1}$$  \hspace{1cm} (67)

$$\Theta = \sum_{j=1}^{N} \frac{\beta_j^2}{\xi_j^2}$$  \hspace{1cm} (68)

The weights $w_i$ can be rewritten as follows. Let

$$\tilde{\Omega}_{iA} = \frac{1}{\xi_i^2} \left[ \Omega_{iA} - \beta_i \frac{\Lambda_A}{\Theta} \right]$$  \hspace{1cm} (69)

$$\tilde{\Upsilon}_i = \sum_{A,B=1}^{K} \tilde{\Omega}_{iA} Q_{AB}^{-1} \Lambda_B$$  \hspace{1cm} (70)
Straightforward algebra results in the following expression for \( w_i \):

\[
  w_i = \frac{\beta_i}{\Theta \xi_i^2} - \sigma_F^2 \tilde{\Upsilon}_i \tag{71}
\]

Note that (thus, we have (60))

\[
  \sum_{i=1}^{N} \beta_i \tilde{\Upsilon}_i = 0 \tag{72}
\]

Therefore, assuming all \( \beta_i > 0 \), the first term in (71) is always positive; however, \( \tilde{\Upsilon}_i \) can be negative (in which case \( w_i \) is positive) or positive, in which case \( w_i \) can be negative. A deceptively “simple” way to ensure that all \( w_i > 0 \) is to take \( \Omega_{iA} \) such that they are “orthogonal” to \( \beta_i \)

\[
  \sum_{j=1}^{N} \frac{\beta_j \Omega_{jA}}{\xi_j^2} \equiv 0 \tag{73}
\]

In this case we have \( \Lambda_A \equiv 0 \) and \( \tilde{\Upsilon}_i \equiv 0 \). However, this is not practicable as a priori (i.e., before constructing the full risk model) \( \xi_i^2 \) are unknown and their dependence on \( \Omega_{iA} \) is highly nonlinear (see [Kakushadze and Yu, 2016a] for details).

On the other hand, it is also impracticable to derive conditions ensuring that \( w_i > 0 \) (or at least \( w_i \geq 0 \)) for a general multifactor model. Nonetheless, (73) has an important interpretation: it is nothing but the requirement that the risk factors be “orthogonal” to the “market mode”. In fact, we can turn this around and ask: what happens if we include the “market mode”? We already know the answer in the case of a 1-factor model: if we take the corresponding factor loading as \( \beta_i \), then the factor risk does not affect the weights \( w_i \). But for a multifactor model things are trickier.

So, let us assume that \( \Omega_{i1} = \beta_i \). Then we can always rotate the remaining \( \Omega_{iA}, A > 1 \), such that \( \Lambda_A \equiv 0 \) for \( A > 1 \). (Note that this rotation affects \( \phi_{AB} \).) We then have \( \tilde{\Omega}_{i1} \equiv 0 \), while \( \tilde{\Omega}_{iA} = \Omega_{iA}/\xi_i^2 \) for \( A > 1 \). Therefore, we have

\[
  \tilde{\Upsilon}_i = \Theta \sum_{A=2}^{K} \tilde{\Omega}_{iA} Q^{-1}_{A1} \tag{74}
\]

\[
  Q_{A1} = \phi_{A1}^{-1}, \quad A > 1 \tag{75}
\]

So, if \( \phi_{A1} \equiv 0 \) for \( A > 1 \), then for these values of \( A \), we also have \( \phi_{A1}^{-1} \equiv 0 \), \( Q_{A1} \equiv 0 \) and \( Q^{-1}_{A1} \equiv 0 \), which would imply that \( \tilde{\Upsilon}_i \equiv 0 \). It is therefore the nonzero correlations (i.e., mixing) between the “market mode” and the other risk factors that make \( \tilde{\Upsilon}_i \neq 0 \), which can lead to some negative \( w_i \). Barring setting \( \phi_{A1} \equiv 0 \) for \( A > 1 \) ad hoc (which is impracticable for the same reasons as above), there is no simple way to guarantee that \( w_i > 0 \) if we do not exclude the “market mode”. To be sure, as already mentioned, there is no practicable way to guarantee (73) either. However, our point here is that not excluding the “market mode” further exacerbates the issue. So, are we at sea? Not quite. We just need to dig deeper.
3.5 Binary “Cluster” Factors

Thus, in practice, the factor loadings are not arbitrary but relatively constrained. The columns of the factor loadings matrix $\Omega_{iA}$ typically are based on: i) industry classification (or some other clustering); ii) style factors (e.g., size, value, liquidity, volatility, etc.); and/or iii) principal components. We already discussed principal components and we will return to them a bit later. We will also come back to style factors. Here we focus on “cluster” based factors, which can be based on a fundamental industry classification or a statistical one [Kakushadze and Yu, 2016b].

First, let $\sigma_i^2 = C_{ii}$ be the total sample variances computed based on the historical time-series data. We have (see above) $\Gamma_{ii} = \sigma_i^2$. The volatilities $\sigma_i$ have a skewed (roughly log-normal) cross-sectional distribution with a long tail for higher values of $\sigma_i$. This is why in practice, instead of directly modeling $C_{ij}$ via a factor model, it makes a lot more sense to model the sample correlation matrix $\Psi_{ij} = C_{ij}/\sigma_i \sigma_j$, from which the skewness present in $\sigma_i$ has been nicely factored out. Indeed, the diagonal elements $\Psi_{ii} \equiv 1$, and the off-diagonal ones (i.e., pair-wise correlations) $|\Psi_{ij}| < 1$ ($i \neq j$). What this means in terms of the factor model for $\Gamma_{ij}$ is that

$$\Gamma_{ij} = \sigma_i \sigma_j \hat{\Gamma}_{ij}$$

(76)

$$\hat{\Gamma}_{ij} = \xi_i^2 \delta_{ij} + \sum_{A,B=1}^{K} \hat{\Omega}_{iA} \phi_{AB} \hat{\Omega}_{jB}$$

(77)

$$\hat{\xi}_i = \xi_i / \sigma_i$$

(78)

$$\hat{\Omega}_{iA} = \Omega_{iA} / \sigma_i$$

(79)

$$\hat{\xi}_i^2 + \sum_{A,B=1}^{K} \hat{\Omega}_{iA} \phi_{AB} \hat{\Omega}_{iB} \equiv 1$$

(80)

The matrix $\hat{\Omega}_{iA}$ is wholly devoid of any skewness in $\sigma_i$. This simplifies things a lot.

Now, let us consider a model where the factor loadings $\hat{\Omega}_{iA}$ are based on a binary industry classification:

$$\hat{\Omega}_{iA} = \Omega_{iA} \delta_{G(i),A}$$

(81)

$$G : \{1, \ldots, N\} \mapsto \{1, \ldots, K\}$$

(82)

Here: the $N$-vector $\Omega_i$ a priori is arbitrary; $G$ maps stocks labeled by $i$ ($i = 1, \ldots, N$) to “clusters” labeled by $A$ ($A = 1, \ldots, K$); each cluster contains one and only one stock; $J(A) = \{i|G(i) = A\}$ is the set of stocks that belong to the cluster labeled by $A$; and $N(A) = |J(A)|$ is the number of stocks in said cluster. The clusters can be, e.g., sectors, industries or sub-industries in a binary industry classification.25

24 For details, see [Kakushadze, 2015c], [Kakushadze and Yu, 2016a].

25 Such as GICS (Global Industry Classification Standard), BICS (Bloomberg Industry Classification System), SIC (Standard Industrial Classification), etc. In principle, we can also consider quasi-binary classifications where some stocks (conglomerates, whose number typically is relatively small) belong to more than one cluster. We will not do so, nor will it be critical for our purposes.
then have
\[ \xi_i^2 = 1 - \Omega_i^2 \phi_{G(i),G(i)} \] (83)
which implies that (by definition, \( \phi_{AB} \) is the factor covariance matrix, so \( \phi_{AA} > 0 \))
\[ \Omega_i^2 < 1/\phi_{G(i),G(i)} \] (84)

For \( \Omega_i \), we basically have three choices. We can take \( \Omega_i = 1/\sqrt{N(A)} \), \( i \in J(A) \), i.e., uniform within-cluster loadings.\(^{26}\) Another choice is to take \( \Omega_i = [U(A)]_i \), \( i \in J(A) \), where the \( N(A) \)-vector \( [U(A)]_i \) is the first principal component of the \( N(A) \times N(A) \) matrix \( [\Psi(A)]_{ij} = \Psi_{ij}, i, j \in J(A) \).\(^{27}\) Finally, we can take \( \Omega_i \) to be a style factor.\(^{28}\)

Once \( \Omega_i \) is specified, the factor covariance matrix \( \phi_{AB} \) and the specific risks \( \hat{\xi}_i \) can be computed.\(^{29}\) So, following our discussion in the 1-factor case, let us take \( \Omega_i = \beta_i/\sigma_i \). Then straightforward algebra gives
\[ w_i = \sigma_F^2 \frac{\beta_i}{\xi_i^2} \gamma_{G(i)} \] (85)
\[ \xi_i^2 = \sigma_i^2 \xi_i^2 \] (86)
\[ \gamma_A = 1 - \sum_{B=1}^{K} Q_{AB}^{-1} \Lambda_B \] (87)
\[ \Lambda_A = \sum_{j \in J(A)} \beta_i^2 \xi_i^2 \] (88)
\[ Q_{AB} = \phi_{AB}^{-1} + \Lambda_A \delta_{AB} \] (89)
\[ \sigma_F^{-2} = \sum_{A=1}^{K} \Lambda_A \gamma_A \] (90)

Here \( \xi_i^2 \) is the specific variance in the factor model covariance matrix \( \Gamma_{ij} \) (as opposed to \( \hat{\xi}_i^2 \), which is the same quantity in the correlation matrix \( \hat{\Gamma}_{ij} \)). So, the important lesson from (85) is that the weights \( w_i \) within each cluster are computed the same way as in the 1-factor model (i.e., by ignoring the factor risk). The normalization factors \( \gamma_{G(i)} \) are uniform within each cluster and only vary from cluster to cluster. They result from optimizing cluster returns across the \( K \) clusters (see below). In this regard, they are not guaranteed to be positive. Thus, if we assume that there is no

\(^{26}\) This is the binary risk model construction [Kakushadze and Yu, 2016a].
\(^{27}\) This is the heterotic construction [Kakushadze, 2015c], [Kakushadze and Yu, 2016a].
\(^{28}\) This is the heterotic CAPM construction [Kakushadze and Yu, 2016a]. The “style” factor here can be related to \( \beta_i \) itself (see below).
\(^{29}\) More precisely, depending on the historical lookback, \( \phi_{AB} \) may be computable as a sample covariance matrix of factor returns or itself may have to be modeled via a factor model covariance matrix as the sample factor covariance matrix may be singular or unstable out-of-sample. However, the diagonal elements \( \phi_{AA} \) are always the same as sample variances of the (appropriately defined and normalized) factor returns. See [Kakushadze, 2015c], [Kakushadze and Yu, 2016a].
mixing between the clusters, i.e., the pair-wise factor correlations vanish \((\phi_{AB} \equiv 0, A \neq B)\), then we have \(\gamma_A = 1/(1 + \phi_{AAA} \Lambda_A) > 0\). In the presence of mixing we can have some negative \(\gamma_A\). However, then all \(w_i\) in the corresponding cluster(s) would be negative as well (assuming all \(\beta_i > 0\)). This is an artifact of optimizing cluster returns, which (approximately) hedges against all clusters going bust. As in the case of a 1-factor model, for long-only portfolios it makes no sense to do this. E.g., if clusters are sectors and our broad benchmark contains stocks in all sectors (with each sector itself well-diversified), it makes no sense to hedge against all sectors going bust – the benchmark is long all sectors by definition. Mathematically, this can be understood as the factor covariance matrix containing its own “market mode” (corresponding to the overall movement of all sectors), which must be eliminated when computing the benchmark weights \(w_i\). This can be seen by modeling \(\phi_{AB}\) via a 1-factor model. The story is the same as above, with clusters in the place of stocks.

### 3.5.1 Example: 1-factor \(\phi_{AB}\)

To illustrate the discussion above, for our purposes here it suffices to consider a 1-factor model for the factor covariance matrix:

\[
\phi_{AB} = \zeta_A^2 \delta_{AB} + \chi_A \chi_B
\]

Straightforward algebra then yields

\[
Q^{-1}_{AB} = \frac{1}{\nu_A^2} \delta_{AB} + \frac{1}{\kappa} \frac{\chi_A}{\nu_A^2 \zeta_A^2} \frac{\chi_B}{\nu_B^2 \zeta_B^2}
\]

\[
\nu_A^2 = \frac{1}{\zeta_A^2} + \Lambda_A
\]

\[
\kappa = 1 + \sum_{A=1}^K \frac{\chi_A^2 \Lambda_A}{1 + \zeta_A^2 \Lambda_A}
\]

So, we have

\[
\gamma_A = \frac{\kappa^{-1}}{1 + \zeta_A^2 \Lambda_A} \left( 1 - \sum_{B=1; B\neq A}^K \left[ \frac{\chi_A}{\chi_B} - 1 \right] \frac{\chi_B^2 \Lambda_B}{1 + \zeta_B^2 \Lambda_B} \right)
\]

For generic \(\chi_A\) (i.e., \(|\chi_A/\chi_B - 1| \sim 1, B \neq A\)), some \(\gamma_A\) can be negative. Consider \(A = A^*, \chi_{A^*} = \max(\chi_A)\). If \(K \gg 1\), to avoid negative \(\gamma_{A^*}\), we would have to assume \(\chi_{A^*}^2 \Lambda_{A^*} / (1 + \zeta_{A^*}^2 \Lambda_{A^*}) \ll 1\) for most \(B \neq A^*\). If, for such \(B\), \(\zeta_B^2 \Lambda_B \gg 1\), then we have \(\chi_B^2 / \zeta_B^2 \ll 1\) and most pair-wise cluster correlations are small. If, instead, \(\zeta_B^2 \Lambda_B \ll 1\), then \(\chi_B^2 \Lambda_B \ll 1\), and \(\zeta_B \Omega_i \ll 1\) and \(|\chi_B/\Omega_i| < 1\) for \(i \in J(B)\), so, nonsensically, the within-cluster pair-wise stock correlations \(\hat{\Gamma}_{ij} = \Omega_i \Omega_j \phi_{BB} \ll 1\), \(i \neq j\), \(i, j \in J(B)\).\(^{30}\)

\(^{30}\) Further, if all \(\zeta_A^2 \Lambda_A \ll 1\), both within-cluster and inter-cluster pair-wise stock correlations \(\hat{\Gamma}_{ij} = \Omega_i \Omega_j \phi_G(i),G(j) \ll 1\), \(i \neq j\), unless all \(|\chi_A| \gg \zeta_A\), i.e., unless, nonsensically, all clusters are almost 100% (anti-)correlated. That is, in the (unrealistic) \(\zeta_A^2 \Lambda_A \ll 1\) limit \(\Gamma_{ij}\) is almost diagonal.
3.5.2 Cluster Weights

So, how can (should) we compute $w_i$ such that they are nonnegative? We can get insight by looking at (95). Let us, ad hoc, take all $\chi_A$ to be uniform: $\chi_A \equiv \chi$. Then we have

$$\gamma_A = \frac{\kappa^{-1}}{1 + \zeta_A^2 \Lambda_A} \quad (96)$$

and all $\gamma_A > 0$. So, up to an immaterial overall normalization factor, this is the same result as what we get if we assume that $\phi_{AB}$ is diagonal, in which case we have $\gamma_A = 1/(1 + \phi_{AA}^2 \Lambda_A)$. However, here $\phi_{AB}$ is not diagonal and in (96) we have the specific variance $\zeta_A^2$ in the denominator, not the total variances $\phi_{AA}$. This is because we have effectively dropped the “market mode” (i.e., the factor risk) in $\phi_{AB}$. What remains is the specific risk. So, the question is, what is the interpretation of (96)?

Let us look at each cluster independently from other clusters. We can construct the benchmark for the universe of stocks corresponding to each cluster using the 1-factor model approach. These weights are given by (see (54))

$$[w(A)]_i = \eta_A \frac{\beta_i}{\xi_i^2}, \quad i \in J(A) \quad (97)$$

$$\eta_A^{-1} = \sum_{j \in J(A)} \frac{\beta_j^2}{\xi_j^2} = \Lambda_A \quad (98)$$

Now we can construct cluster returns, i.e., the returns of the $K$ benchmark portfolios corresponding to the $K$ clusters, via

$$R_A = \sum_{i \in J(A)} [w(A)]_i R_i \quad (99)$$

Since (up to an overall normalization factor) $E_i = \beta_i$ are the expected returns for the stocks, the expected returns $E_A$ for the clusters are $E_A \equiv 1$. If we construct a “global” benchmark portfolio made of all the clusters, the corresponding weights with which we combine the clusters are (normalized such that $\sum_{A=1}^K w_A = 1$)

$$w_A = \mu \frac{E_A}{\zeta_A^2} = \mu \frac{1}{\zeta_A^2} \quad (100)$$

$$\mu^{-1} = \sum_{A=1}^K \frac{E_A}{\zeta_A^2} = \sum_{A=1}^K \frac{1}{\zeta_A^2} \quad (101)$$

The stock weights in this “global” benchmark portfolio are given by

$$w_i = w_A [w(A)]_i = \mu \frac{\beta_i}{\zeta_A^2 \Lambda_A} \frac{\beta_i}{\xi_i^2}, \quad i \in J(A) \quad (102)$$

This is precisely (85) with $\gamma_A$ of the form (96) and $\sigma^2_F = \mu \kappa$ (see (90)) in the limit where $\zeta_A^2 \Lambda_A \gg 1$. So, the question is, what is the meaning of the extra 1 in (96)?
To understand this, let us consider the opposite limit, where $\zeta_A^2 \Lambda_A \ll 1$. In this limit pair-wise stock correlations are small (see fn. 30).31 This means that the total risk approximately equals the specific risk and the factor model covariance matrix is approximately diagonal. So, in this limit the effect of the clusters is negligible and we should recover our result for the 1-factor model (54). And this is precisely what happens in the $\zeta_A^2 \Lambda_A \ll 1$ limit in (90) as $\gamma_A \approx 1/\kappa$ and is independent of $A$, so (85) correctly reduces to (54). For the intermediate values $\zeta_A^2 \Lambda_A \sim 1$, (90) smoothly interpolates between the two limits. This is what the full optimization gives, which balances the stock-specific risk and the factor risk. The only “loose end” is that in arriving at (90) we assumed uniform $\chi_A \equiv \chi$, which results in $\gamma_A$ that naively might appear as independent of $\chi$.32 However, $\Lambda_A$ depends on $\hat{\xi}_i^2 = 1 - \Omega_i^2 \phi_{AA} = 1 - \Omega_i^2 (\zeta_i^2 + \chi^2)$, $i \in J(A)$. Furthermore, $\zeta_A^2$ depends on $\chi$ (which is the factor loading in the 1-factor model (91) for $\phi_{AB}$) via a computation involving the time-series of factor returns (see [Kakushadze, 2015c], [Kakushadze and Yu, 2016a]). So, the aforesaid “loose end” is this: why are $\chi_A \equiv \chi$ uniform?

This is because the cluster expected returns are uniform: $E_A \equiv 1$ (up to an overall normalization factor). Therefore, the corresponding factor betas $\beta_A$ are uniform: $\beta_A \equiv b$. So, following our discussion in Subsection 3.2, the factor loading $\Omega_A$ in the 1-factor model for the correlation matrix $\psi_{AB} = \phi_{AB}/\sigma_A\sigma_B$ (where $\sigma_A^2 = \phi_{AA}$) is given by $\Omega_A = \beta_A/\sigma_A$, while the corresponding factor loading in the covariance matrix $\chi_A = \sigma_A \Omega_A$ is simply $\chi_A = \beta_A \equiv b$. I.e., $\chi_A$ are uniform, $\chi_A \equiv \chi$, and $\chi$ is identified with $b$. Note that while our discussion here naively may appear a bit “cavalier” w.r.t. the normalizations of $\beta_A$ and $\chi_A$, it is not. This is because, in a 1-factor model, the factor loading $\chi_A$ subsumes the $1 \times 1$ factor covariance matrix, call it $\varphi$. So, the factor model (91) actually reads:

$$\phi_{AB} = \zeta_A^2 \delta_{AB} + \chi'_A \varphi \chi'_B$$

Here $\chi'_A$ is the raw (unnormalized) factor loading and $\chi_A = \sqrt{\varphi} \chi'_A$. So, we can identify $\chi'_A$ with $\beta_A \equiv b$, which can be normalized arbitrarily, and this normalization is then subsumed in $\chi_A$ via $\varphi$, which is computed based on the time-series of factor returns and depends on $\chi'_A$. The end result is that our $\chi_A$ are uniform: $\chi_A \equiv \chi$.

### 3.5.3 A Generalization

Above we discussed a 1-factor model for the cluster factor covariance matrix $\phi_{AB}$. However, we can generalize our result to a multifactor model for $\phi_{AB}$ where the $K$ clusters labeled by $A$ can be grouped into further $F$ clusters (typically, $F \ll K$), which we will label by $a$, $a = 1, \ldots, F$. This naturally arises in binary fundamental industry classifications.33 E.g., in BICS (see above) at the most granular level we have sub-industries, which are grouped into industries, which themselves are grouped

---

31 Both within- and inter-cluster pair-wise stock correlations are small unless all $|\chi_A| \gg \zeta_A$.
32 Up to an immaterial overall normalization, that is: $\kappa$ explicitly depends on $\chi$.
33 This structure also arises in statistical industry classifications [Kakushadze and Yu, 2016b].
into sectors. This chain can be thought of as ending with the final grouping into a single cluster corresponding to the (broad) “market”. So, the 1-factor model above could, e.g., describe BICS sectors, and the single factor loading corresponds to the “market”. Alternatively, we could take sub-industries or industries and build 1-factor models for them going straight to the “market” level. However, instead, we can build multifactor models, e.g., $A$ labels sub-industries and $a$ labels industries, or $A$ labels subs-industries and $a$ labels sectors, etc. So, the idea here\footnote{This is the original nested “Russian-doll” embedding of [Kakushadze, 2015b] used in [Kakushadze, 2015c] and [Kakushadze and Yu, 2016a].} is that we have an $F$-factor model for $\phi_{AB}$:

$$\phi_{AB} = \zeta_A^2 \delta_{AB} + \sum_{a,b=1}^{F} \chi_{Aa} \varphi_{ab} \chi_{Bb}$$

(104)

$$\chi_{Aa} = \chi_A \delta_{S(A),a}$$

(105)

$$S : \{1, \ldots, K\} \mapsto \{1, \ldots, F\}$$

(106)

Here: $\chi_{Aa}$ is a $K \times F$ factor loadings matrix; $\varphi_{ab}$ is an $F \times F$ factor covariance matrix; the $K$-vector $\chi_A$ is a priori arbitrary; $S$ maps the $K$ “sub-clusters” labeled by $A$ to the $F$ “clusters” labeled by $a$, $a = 1, \ldots, F$; each cluster contains one and only one sub-cluster; $J'(a) = \{ A | S(A) = a \}$ is the set of sub-clusters that belong to the cluster labeled by $a$. Straightforward algebra gives:

$$Q^{-1}_{AB} = \frac{1}{\nu_A^2} \delta_{AB} + \chi_A \frac{\chi_B \varphi_{ab}^{1-1}}{\nu_B^{2} \nu_A^{2} \kappa_{S(A),S(B)}}$$

(107)

$$\nu_A^{2} = \frac{1}{\zeta_A^2} + \Lambda_A$$

(108)

$$\kappa_{ab} = \varphi_{ab}^{-1} + \delta_{ab} \sum_{A \in J'(a)} \frac{\chi_A^2 \Lambda_A}{1 + \zeta_A^2 \Lambda_A}$$

(109)

So, we have

$$\gamma_A = \frac{1}{1 + \zeta_A^2 \Lambda_A} \left[ 1 - \chi_A \sum_{B=1}^{K} \frac{\chi_B \Lambda_B}{1 + \zeta_B^2 \Lambda_B} \kappa_{S(A),S(B)} \right]$$

(110)

Following our logic above, we must take uniform $\chi_A \equiv \chi$. However, unlike the 1-factor case, where $\kappa$ was a number, here instead we have a matrix, $\kappa_{ab}$, which depends on the details of $\varphi_{ab}$. Consider a 1-factor model:

$$\varphi_{ab} = \rho_{a}^2 \delta_{ab} + \omega_{a} \omega_{b}$$

(111)

For the same reason as why $\chi_A \equiv \chi$ are uniform, we must take uniform $\omega_{a} \equiv \omega$.\footnote{Thus, we have $K$ sub-clusters grouped into $F$ clusters. Similarly to (97), we can construct the benchmark for the universe of sub-clusters corresponding to each cluster (recall that $E_A \equiv 1$): $[w(a)]_A = \eta_a / \zeta_A^2$, $A \in J'(a)$, where $\eta_a^{-1} = \sum_{A \in J'(a)} \zeta_A^{-2}$. We can compute the $F$ cluster returns $R_a$ using the $K$ sub-cluster returns: $R_a = \sum_{A \in J'(a)} [w(a)]_A R_A$. Then the cluster expected returns $E_a = 1$. Hence uniform $\omega_a$. Again, the foregoing holds up to immaterial overall normalizations.}
With uniform $\omega_a$, straightforward algebra yields the following simple result:

$$
\gamma_A = \frac{1}{1 + \zeta^2 A} \frac{1}{1 + \rho^2 S(A)} \frac{1}{1 + \tau} \tag{112}
$$

$$
\lambda_a = \chi^2 \sum_{A \in J'} \frac{A_A}{1 + \zeta^2 A} \tag{113}
$$

$$
\tau = \omega^2 \sum_{a=1}^{F} \frac{\lambda_a}{1 + \rho^2 \lambda_a} \tag{114}
$$

Note that the factorization in (112) occurs precisely because $\chi_A$ and $\omega_a$ are uniform.\(^{36}\)

### 3.6 General “Russian-doll” Embedding

Above we considered a 2-level clustering scheme. It is now evident how to generalize it to any $P$-level clustering scheme. We have the following sequence: Stocks (Level-0) → Level-1 Clusters → Level-2 Clusters → ... → Level-$P$ Clusters → “Market” (Level-(P + 1)). Here “Market” means the entire universe of $N$ stocks and can be thought of as the final single cluster in the above sequence.\(^{37}\) Thus, BICS is a 3-level industry classification ($P = 3$), where Level-1 Clusters = BICS Sub-industries, Level-2 Clusters = BICS Industries, and Level-3 Clusters = BICS Sectors.\(^{38}\) We have the following nested “Russian-doll” risk model construction (here $\ell = 1, \ldots, P$):

$$
\Gamma^{(\ell)}_{A^{(\ell)},B^{(\ell)}} = \left[ \zeta^{(\ell)}_{A^{(\ell)}} \right]^2 \delta_{A^{(\ell)},B^{(\ell)}} + \sum_{A^{(\ell+1)},B^{(\ell+1)}=1}^{K^{(\ell+1)}} \Omega^{(\ell)}_{A^{(\ell)},A^{(\ell+1)}} \Gamma^{(\ell+1)}_{A^{(\ell+1)},B^{(\ell+1)}} \Omega^{(\ell)}_{B^{(\ell)},B^{(\ell+1)}} \tag{115}
$$

Here: $A^{(0)}, B^{(0)} = 1, \ldots, N$ label stocks (i.e., they are the indices $i, j = 1, \ldots, N$ in the notations above); $\Gamma^{(0)}_{ij} = \Gamma_{ij}$ is the factor model covariance matrix for stocks (and $[\zeta_i^{(0)}]^2 = \xi^2_i$ are the corresponding specific variances); $A^{(\ell)}, B^{(\ell)} = 1, \ldots, K^{(\ell)}$, $\ell = 1, \ldots, P$, label the Level-$\ell$ Clusters; $\Gamma^{(\ell)}_{A^{(\ell)},B^{(\ell)}}$, $\ell = 1, \ldots, P$, are the factor covariance matrices corresponding to the Level-$\ell$ Clusters; at Level-(P + 1) we have $A^{(P+1)} = B^{(P+1)} = 1$ (i.e., these indices take only one value corresponding to the “Market”, so we have $K^{(P+1)} = 1$), and we can either have $[\zeta_1^{(P+1)}]^2 = \Gamma_{11}^{(P+1)} > 0$ (so $\Gamma^{(P)}_{A^{(P)},B^{(P)}}$ is a 1-factor model), or we can set $[\zeta_1^{(P+1)}]^2 = \Gamma_{11}^{(P+1)} = 0$ (so $\Gamma^{(P)}_{A^{(P)},B^{(P)}}$)

---

\(^{36}\) Let us emphasize that, unlike for stocks, uniform factor loadings are reasonable for clusters (e.g., industries, sectors, etc.). This is because clusters are diversified stock portfolios with much less skewed volatilities than stocks. Therefore, in this regard, very small clusters should be avoided.

\(^{37}\) Also, the last leg in the above sequence “Level-$P$ Clusters → “Market” (Level-(P + 1))” can be treated as optional, if so desired (see below).

\(^{38}\) Note that GICS (see above) has $P = 4$ levels.
is diagonal, i.e., a “0-factor model”; finally, for the factor loadings \( \Omega_{A(\ell),A(\ell+1)} \) we have (here, as above, \( \beta_i \) are the stock betas):

\[
\begin{align*}
\Omega_{i,A(1)}^{(0)} &= \beta_i \delta_{G(0)(i),A(1)} \\
\Omega_{A(\ell),A(\ell+1)}^{(\ell)} &= \chi^{(\ell)} \delta_{G(\ell)(A(\ell)),A(\ell+1)}, \quad \ell = 1, \ldots, P \\
G^{(\ell)} : \{1, \ldots, K^{(\ell)}\} &\mapsto \{1, \ldots, K^{(\ell+1)}\}, \quad \ell = 0, 1, \ldots, P
\end{align*}
\]

Here: \( G^{(\ell)} \) is a map from the Level-\( \ell \) Clusters to the Level-(\( \ell + 1 \)) Clusters, \( \ell = 0, 1, \ldots, P \); “Level-0 Clusters” = stocks; \( K^{(0)} = N \); at Level-(\( P + 1 \)) we have the “Market” (a single “cluster”). The benchmark portfolio weights are then given by:

\[
\begin{align*}
w_i &= \sigma_F^2 \frac{\beta_i}{\xi_i^2} \gamma_{G(0)(i)}^{(0)} \\
\gamma_A^{(1)} &= \prod_{\ell=1}^{P+1} \left( 1 + \left[ \zeta_A^{(\ell)} \right]^2 \Lambda_A^{(\ell)} \right)^{-1} \\
F^{(1)}(A^{(1)}) &= A^{(1)} \\
F^{(\ell+1)}(A^{(1)}) &= G^{(\ell)}(F^{(\ell)}(A^{(1)})) = G^{(\ell)}(G^{(\ell-1)}(\ldots G^{(1)}(A^{(1)}) \ldots)) \\
\Lambda_A^{(1)} &= \sum_{j \in J^{(0)}(A^{(1)})} \beta_j^2 \\
\Lambda_A^{(\ell+1)} &= \left[ \chi^{(\ell)} \right]^2 \sum_{A^{(\ell)} \in J^{(\ell)}(A^{(\ell+1)})} \Lambda_{A^{(\ell)}}^{(\ell)} \left( 1 + \left[ \zeta_A^{(\ell)} \right]^2 \Lambda_A^{(\ell)} \right)^{-1}
\end{align*}
\]

Here: \( \ell = 1, \ldots, P \) in (124); the sets \( J^{(\ell)}(A^{(\ell+1)}) = \{ A^{(\ell)} | G^{(\ell)}(A^{(\ell)}) = A^{(\ell+1)} \} \); and \( F^{(P+1)}(A^{(1)}) = 1 \).\footnote{Further, note that the factor in the product (120) corresponding to \( \ell = P + 1 \) is actually independent of \( A^{(1)} \), and it is equal 1 if \( \zeta_A^{(P+1)} = 0 \) (i.e., if \( \Gamma_{A(P),B(P)}^{(P)} \) is a “0-factor model”).} The benchmark weights (119) comprise one of our main results.

### 3.7 Risk Model Construction

Assuming all \( \beta_i > 0 \) (and all \( \chi^{(\ell)} > 0 \)), it is not all that surprising that we can construct all-positive \( w_i \) as the matrix \( \Gamma_{ij} \) has all positive elements. Thus, pursuant to the Perron-Frobenius theorem [Perron, 1907], [Frobenius, 1912], all \( V_i^{(1)} > 0 \) (or can be chosen to be such as the signs of all \( V_i^{(1)} \) can always be flipped simultaneously), where \( V_i^{(1)} \) is the first principal component of \( \Gamma_{ij} \). However, (119) is not based on principal components of some random positive covariance (or correlation) matrix \( \Gamma_{ij} \). Instead, here we construct a non-random, meaningful \( \Gamma_{ij} \) for “arbitrary” \( \beta_i > 0 \).
Now, this is where some qualifications are in order. It is clear that $\beta_i$ must be skewed similarly to $\sigma_i$, where $\sigma_i^2 = \Gamma_{ii} = C_{ii}$ are sample variances for stocks. I.e., $\hat{\beta}_i = \beta_i/\sigma_i$ is the quantity that is expected not to be skewed. Otherwise, within the same Level-1 Cluster labeled by $A^{(1)}$ stocks with large $\sigma_i$, $i \in J^{(0)}(A^{(1)})$, would have small correlations with other stocks. So, at Level-0 we have the following factor model $\hat{\Gamma}_{ij}$ for the correlation matrix

$$\Gamma_{ij} = \sigma_i \sigma_j \hat{\Gamma}_{ij}$$

(126)

$$\hat{\Gamma}_{ij} = \xi_i^2 \delta_{ij} + \hat{\beta}_i \Gamma^{(1)}_{G^{(0)}(i),G^{(0)}(j)} \hat{\beta}_j$$

(127)

$$\xi_i^2 = \xi_i^2 / \sigma_i^2$$

(128)

$$\xi_i^2 + \hat{\beta}_i^2 \Gamma^{(1)}_{G^{(0)}(i),G^{(0)}(j)} \equiv 1$$

(129)

As discussed in [Kakushadze, 2015c], [Kakushadze and Yu, 2016a], it is the condition (129) that is difficult to satisfy. In particular, for generic values of $\hat{\beta}_i$ we would have to use the method of Section 4 of [Kakushadze and Yu, 2016a], whereby we have some trial values $\hat{\beta}_i'$ and the actual values $\hat{\beta}_i$ are related to $\hat{\beta}_i'$ via a highly nonlinear combination of $\hat{\beta}_i'$ and the sample correlation matrix $\Psi_{ij}$. It is then impracticable to detangle $\hat{\beta}_i'$ from the desired values $\hat{\beta}_i$. To avoid complications with such nonlinearities, we can use the heterotic construction of [Kakushadze, 2015c], [Kakushadze and Yu, 2016a], where $\hat{\beta}_i$, $i \in J^{(0)}(A^{(1)})$, are given by the first principal component of the square block $\Psi_{ij}$, $i, j \in J^{(0)}(A^{(1)})$. However, these principal components are not guaranteed to be all positive. This can be overcome by deforming each block such that all correlations therein are positive. This is doable but somewhat convoluted and there is no unique way of doing this. At the end we would have just a single choice of $\beta_i$—subject to variability due to the choice of the deformation, that is.

Another possibility is to take $\hat{\beta}_i \equiv 1$ and use the binary construction in Subsection 3.2 of [Kakushadze and Yu, 2016a]. In this case we do not need to deform the sample correlation matrix. However, here too we have a single choice of $\beta_i$.

### 3.7.1 Computing Specific Risk

Happily, precisely because for each block we are dealing with a 1-factor model, we can use another, simple method to satisfy (129). Consider a symmetric $M \times M$ matrix

---

40 To be clear, the method of Section 4 of [Kakushadze and Yu, 2016a] is perfectly adequate if we wish to construct a factor model, e.g., for optimization purposes, and need not worry about the precise values of the factor loadings (i.e., the fact that the trial and actual loadings are not the same). However, the problem at hand is different, which is to construct a benchmark portfolio for given betas, and the actual factor loadings must coincide with these betas by construction.

41 Note that equivalently we can take $\hat{\beta}_i = b_{G^{(0)}(i)}$, where $b_{A^{(1)}} > 0$ and are otherwise arbitrary. This is because any rescaling $\hat{\beta}_i \rightarrow \tilde{\hat{\beta}}_i b_{G^{(0)}(i)}$ is simply absorbed by the corresponding rescaling of the factor covariance matrix $\Gamma^{(1)}_{A^{(1)},B^{(1)}} \rightarrow \Gamma^{(1)}_{A^{(1)},B^{(1)}}/b_{A^{(1)}} b_{B^{(1)}}$, so the factor model is unaffected.

42 Also, this method would not work for the Level-\(\ell\) Clusters, $\ell \geq 1$ (see below).
$X_{\alpha\beta}$, $\alpha, \beta = 1, \ldots, M$. For our purposes here it will suffice to assume that $X_{\alpha\beta}$ is semi-positive definite (as here we are interested in cases where $X_{\alpha\beta}$ is a covariance or correlation matrix). Suppose we wish to model it via a 1-factor model:

$$Y_{\alpha\beta} = a_\alpha^2 \delta_{\alpha\beta} + b_\alpha \vartheta b_\beta$$

subject to reproducing the diagonal elements:

$$Y_{\alpha\alpha} = X_{\alpha\alpha} = \varsigma_\alpha^2$$  

$$(131)$$

$$a_\alpha^2 + \vartheta \beta_\alpha^2 = \varsigma_\alpha^2$$  

$$(132)$$

So, we need to fit the unknown $\vartheta$ given the values of $b_\alpha$ and $X_{\alpha\beta}$. We can do this as follows. First, let us define $z_{\text{min}}$ and $z_{\text{max}}$ such that for all values of $\alpha$ we have

$$z_{\text{min}} \varsigma_\alpha \leq a_\alpha \leq z_{\text{max}} \varsigma_\alpha$$  

$$(133)$$

I.e., $z_{\text{min}}$ and $z_{\text{max}}$ define the minimum and maximum allowed values of the fraction of the total standard deviation $\varsigma_\alpha$ attributable to the specific risk $a_\alpha$. (E.g., we can set $z_{\text{min}} = 0.1$ and $z_{\text{max}} = 0.9$.) This means that

$$\vartheta_{\text{min}} \leq \vartheta \leq \vartheta_{\text{max}}$$  

$$(134)$$

$$\vartheta_{\text{min}} = (1 - z_{\text{max}}^2) / \min(\hat{b}_\alpha^2)$$  

$$(135)$$

$$\vartheta_{\text{max}} = (1 - z_{\text{min}}^2) / \max(\hat{b}_\alpha^2)$$  

$$(136)$$

$$\hat{b}_\alpha = b_\alpha / \varsigma_\alpha$$  

$$(137)$$

So, given $\varsigma_\alpha$, the values of $b_\alpha$ cannot be arbitrary but must be such that $\vartheta_{\text{min}} \leq \vartheta \leq \vartheta_{\text{max}}$ (see Appendix A for how $\vartheta_{\text{min}} > \vartheta_{\text{max}}$ cases are dealt with). Next, we can find the value of $\vartheta$ which provides the least-squares fit of the off-diagonal elements of

$$\hat{Y}_{\alpha\beta} = Y_{\alpha\beta} / \varsigma_\alpha \varsigma_\beta$$

$$(138)$$

into those of $\hat{X}_{\alpha\beta} = X_{\alpha\beta} / \varsigma_\alpha \varsigma_\beta$ (note that the diagonal elements $\hat{Y}_{\alpha\alpha} \equiv 1$ need not be fit as they are fixed via (132)):

$$\sum_{\alpha, \beta = 1; \alpha \neq \beta}^M \left[ \hat{X}_{\alpha\beta} - \hat{b}_\alpha \vartheta \hat{b}_\beta \right]^2 \rightarrow \min$$  

$$(139)$$

subject to (134). So, we have

$$\vartheta = \min(\max(\vartheta_*, \vartheta_{\text{min}}), \vartheta_{\text{max}})$$  

$$(140)$$

$$\vartheta_* = \frac{\sum_{\alpha, \beta = 1; \alpha \neq \beta}^M \hat{b}_\alpha \hat{X}_{\alpha\beta} \hat{b}_\beta}{\sum_{\alpha, \beta = 1; \alpha \neq \beta}^M \hat{b}_\alpha^2 \hat{b}_\beta^2}$$  

$$(141)$$

Note that in our context here $\hat{X}_{\alpha\beta}$ is a sample correlation matrix, so $|\hat{X}_{\alpha\beta}| \leq 1$, in fact, $|\hat{X}_{\alpha\beta}| < 1$ for $\alpha \neq \beta$. Assuming $\hat{b}_\alpha$ are tightly distributed, we can expect $\vartheta_*$ to be somewhere between $\vartheta_{\text{min}}$ and $\vartheta_{\text{max}}$ (as opposed to saturating these bounds).
3.7.2 Application to “Russian-doll” Embedding

Given $X_{\alpha \beta}, b_{\alpha}, z_{\text{min}}$ and $z_{\text{max}}$, let $\theta(X_{\alpha \beta}, b_{\alpha}, z_{\text{min}}, z_{\text{max}})$ denote the value of $\theta$ given by (140). Then we have the following procedure for computing the specific risks and factor covariance matrices in the nested “Russian-doll” embedding described above.\(^\text{43}\)

\[
X_{ij}^{(0)} = C_{ij} \quad \text{(142)}
\]

\[
b_{i}^{(0)} = \beta_{i} \quad \text{(143)}
\]

\[
b_{A}^{(l)} = \chi^{(l)}, \quad \ell = 1, \ldots, P \quad \text{(144)}
\]

\[
\Gamma_{A(\ell+1),A(\ell+1)}^{(\ell+1)} = \theta(X_{A(\ell),B(\ell)}, b_{A(\ell)}^{(\ell)}), \quad A^{(\ell)}, B^{(\ell)} \in \mathcal{J}^{(\ell)}(A^{(\ell+1)}) \quad \text{(145)}
\]

\[
X_{A(\ell+1),B(\ell+1)}^{(\ell+1)} = \tilde{X}_{A(\ell+1),B(\ell+1)} u_{A(\ell+1)}^{(\ell+1)} u_{B(\ell+1)}^{(\ell+1)} \quad \text{(146)}
\]

\[
u_{A(\ell+1)}^{(\ell+1)} = \sqrt{\frac{\Gamma_{A(\ell+1),A(\ell+1)}^{(\ell+1)}}{\tilde{X}_{A(\ell+1),A(\ell+1)}}} \quad \text{(147)}
\]

\[
\tilde{X}_{A(\ell+1),B(\ell+1)} = \sum_{A^{(\ell)} \in \mathcal{J}^{(\ell)}(A^{(\ell+1)})} \sum_{B^{(\ell)} \in \mathcal{J}^{(\ell)}(B^{(\ell+1)})} X_{A^{(\ell)},B^{(\ell)}}^{(\ell)} b_{A}^{(\ell)} b_{B}^{(\ell)} \quad \text{(148)}
\]

\[
\left[\chi_{A}^{(\ell)}\right]^{2} = X_{A^{(\ell)},A^{(\ell)}}^{(\ell)} - \left[b_{A}^{(\ell)}\right]^{2} \Gamma_{G(\ell)(A^{(\ell)}),G(\ell)(A^{(\ell)})}^{(\ell+1)} \quad \text{(149)}
\]

Here, as above, $C_{ij}$ is the sample covariance matrix of the stock returns. Also, note that the choice of $\chi^{(l)}$, $\ell = 1, \ldots, P$, is immaterial. This procedure together with (115) completely defines the risk model. All specific risks and factor covariance matrices are positive-definite by construction. And so are the benchmark weights (119) for a range of values of $\beta_i$ so long as $\tilde{\beta}_i = \beta_i/\sigma_i$ are not skewed (see above).\(^\text{44}\)

3.7.3 Some Comments

Here we can ask two questions. Why does the above construction make sense? And are there viable alternatives? Thus, above we construct the risk model in a rather specific way, grouping stocks into clusters and essentially building a 1-factor model within each cluster with the factor loadings given by the betas.\(^\text{45}\) Then we group these clusters into further clusters and repeat the procedure until we end

---

\(^{\text{43}}\) Below we suppress the $z_{\text{min}}, z_{\text{max}}$ arguments. In fact, they can be $\ell$-dependent, if so desired.

\(^{\text{44}}\) Thus, if we, e.g., take $z_{\text{min}} = 0.1$ and $z_{\text{max}} = 0.9$, then $\theta_{\text{min}} = 0.19/\min(\tilde{\beta}_i^2)$ and $\theta_{\text{max}} = 0.99/\max(\tilde{\beta}_i^2)$, so the allowed range of betas is $\max(\tilde{\beta}_i)/\min(\tilde{\beta}_i) \leq \sqrt{0.99/0.19} \approx 2.28$. Note that for the Level-$\ell$ Clusters, $\ell = 1, \ldots, P$, generally it is reasonable to expect the factor covariances and specific risks to be of order 1 and nontrivial solutions for the fitted values of $\theta$ (i.e., $\Gamma_{A^{(\ell+1)},A^{(\ell+1)}}^{(\ell+1)}$ computed via (145)) to exist. Appendix A deals with occasions when this does not hold.

\(^{\text{45}}\) This is essentially the heterotic CAPM construction of [Kakushadze and Yu, 2016a], which is similar to the heterotic construction of [Kakushadze, 2015c], except that in the latter the factor loadings are based on principal components, which are not necessarily all positive (see above).
up with a stable and positive-definite factor covariance matrix. The question we can ask is, can we take a more general risk model construction instead? Basically, there are two separate issues here. The first issue is independent of the fact that we are dealing with a long-only portfolio and pertains to the fact that: i) higher-than-first principal components are intrinsically unstable out-of-sample; ii) standard style factors (see above) are poor proxies for modeling pair-wise correlations, so – contrary to a common practice in commercial risk model offerings – using them as factor loadings is highly suboptimal\(^{46}\) (see [Kakushadze and Yu, 2016a] for details); and iii) well-constructed fundamental\(^{47}\) industry classifications are rather stable out-of-sample as companies rarely jump industries (let alone sectors). This is what justifies the above construction except for the choice of the factor loadings, which in this case are simply the betas. And this latter part of our construction is dictated by the fact that this choice is the only one that effectively removes the “market mode”. Other choices for factor loadings generically would lead to undesirable negative weights \(w_i\).

In this regard, if we take a generic multifactor risk model for \(\Gamma_{ij}\), which includes, e.g., style factors and/or principal components, some weights generically will be negative for any given choice of the betas. Let us emphasize that we can always work backwards, pick some positive weights \(w_i\), and compute the corresponding betas \(\beta_i\). However, it is unclear what any such portfolio represents. In contrast, in the above construction the meaning of the resultant benchmark portfolio is clear. Thus, assuming that the second term in the parentheses in (120) is dominant at each level \(\ell\) (typically, this is expected to be a good approximation for large clusters) we have

\[
\begin{align*}
    w_i &\approx \eta \frac{\beta_i}{\xi_i^2 \Lambda_{G(0)(i)}^{(1)}} \prod_{\ell=1}^P \left( \left[ \zeta_{F(\ell)(G(0)(i))}^{(\ell)} \right]^2 \tilde{\Lambda}_{F(\ell+1)(G(0)(i))}^{(\ell+1)} \right)^{-1} \\
    \Lambda_{A(1)}^{(1)} &= \sum_{j \in J^{(0)}(A(1))} \frac{\beta_j^2}{\xi_j^2} \\
    \tilde{\Lambda}_{A(\ell+1)}^{(\ell+1)} &\approx \sum_{A(\ell) \in J^{(\ell)}(A(\ell+1))} \left[ \zeta_{A(\ell)}^{(\ell)} \right]^{-2}, \quad \ell = 1, \ldots, P \\
    \eta^{-1} &\approx \sum_{A(1)} \prod_{\ell=1}^P \left( \left[ \zeta_{F(\ell)(A(1))}^{(\ell)} \right]^2 \tilde{\Lambda}_{F(\ell+1)(A(1))}^{(\ell+1)} \right)^{-1}
\end{align*}
\]

The interpretation of these weights (similarly to the example we discussed above) is clear: we suppress the weights by a product of specific variances at each level, with proper normalizations (such that at each level cluster betas are 1 up to immaterial overall normalization factors). Note that instead of specific variances we could use total variances, which would “overcount” the “market mode” risk, whereas specific

\(^{46}\) Also, their number is limited and fails to compete with ubiquitous industry (cluster) factors.  
\(^{47}\) Statistical industry classifications are not as stable as fundamental industry classifications, but still sizably outperform models based on principal components [Kakushadze and Yu, 2016b].
variance exclude this factor risk. And this is just as well: the overall market exposure risk (i.e., the “market mode” risk) is intrinsically present in long-only portfolios and should not be hedged against. Our construction neatly removes the “market mode”.

Finally, let us mention that a priori we could obtain a long-only portfolio for a given set of betas as follows. We can maximize the Sharpe ratio (58) (where $E_i$ are related to the betas via (57)) subject to (60) and lower bounds $w_i \geq w_i^{\text{min}}$, where $w_i^{\text{min}} \geq 0$. A priori here we can use a generic multifactor model covariance matrix $\Gamma_{ij}$ instead of the sample covariance matrix $C_{ij}$. However, for a generic $\Gamma_{ij}$ a large fraction of $w_i$ (in many cases, around 50%) can turn out to saturate the lower bounds $w_i^{\text{min}}$. Such portfolios generically can be skewed and far from “optimality”.

4 Beating the Market

Now that we have a method for constructing benchmark portfolios, we can ask a different question. Suppose we have an “alpha model”, which forecasts expected returns $E_i$. Can we construct a long-only portfolio based on these returns? If all expected returns are positive, then we can treat them as betas, i.e., as in (57), set $\beta_i = E_i/\gamma$ (where $\gamma$ is an immaterial normalization factor) and compute the weights $w_i$ via (119). There are two independent issues with this approach. First, even if all $E_i \geq 0$, the distribution of $E_i/\sigma_i$ may be too wide, so using the method (140) might be problematic.48 Second, in practice, most alpha models will not have all nonnegative $E_i$, in fact, many $E_i$ (in many cases, around 50%) may be negative.

So, how can we deal with this? Instead of constructing a long-only portfolio based on $E_i$ from scratch, we can follow a different, 2-step approach. Since we are building a long-only portfolio, we are exposed to market risk no matter what we do. So, we might as well identify a benchmark portfolio whose market exposure we are willing to live with.49 Then we can try to construct a long only portfolio that – based on out-of-sample backtests – can reasonably be expected (albeit, as with any forward-looking statements, not guaranteed) to outperform this benchmark portfolio. One way to construct this portfolio is to combine the benchmark portfolio with a dollar-neutral portfolio (such that the resultant portfolio is still long-only), where the dollar-neutral portfolio has a positive expected return and a low correlation with the benchmark portfolio. So, for the weights $w_i$ of our long-only portfolio, we have

\[
w_i = w'_i + w''_i
\]

\[
\sum_{i=1}^{N} w'_i = 0
\]

48 Note that at the level of stocks (i.e., Level-0 in the above nomenclature), we can always find $\beta_i$ such that (140) exists. Thus, we can take $\beta_i = \sigma_i$, so $\hat{\beta}_i = 1$. Then $\vartheta_* = \frac{1}{N(N-1)} \sum_{i,j=1; i \neq j}^{N} \Psi_{ij}$, where $\Psi_{ij}$ is the sample correlation matrix. That is, $\vartheta_*$ in this case is the average pair-wise correlation, which typically is positive and well-within reasonably set bounds $\vartheta_{\text{min}}$ and $\vartheta_{\text{max}}$ for a given cluster (e.g., sub-industry, industry, sector) for a well-constructed industry classification.

49 This benchmark portfolio can (but need not) be constructed as in Section 3 (see below).
where \( w'_i \) are the weights of the dollar-neutral portfolio such that

\[
\begin{align*}
    w_{i}^{\text{min}} & \leq w'_i \leq w_{i}^{\text{max}} \\
    w_{i}^{\text{min}} & \geq -w^*_i
\end{align*}
\]

and \( w^*_i > 0 \) are the benchmark weights. A priori we can take the upper bounds \( w_{i}^{\text{max}} > 0 \) to be infinity. We can also set tighter bounds, e.g., such that \( w_i \) do no deviate from the benchmark portfolio by some percentage: \( w_{i}^{\text{min}} = -z w^*_i \), \( w_{i}^{\text{max}} = z w^*_i \), where, say, \( 0 < z < 1 \). Other customizations/variations are possible.

The weights \( w'_i \) can be fixed in a variety of standard ways, e.g., via Sharpe ratio (or mean-variance – see above) optimization, which is what we will assume for the sake of definiteness (albeit this is not critical here). Then, ignoring for a moment the bounds (156) and the dollar-neutrality constraint (155),\(^50\) we have

\[
w'_i = \gamma' \sum_{j=1}^{N} \Gamma'_{ij}^{-1} E_j
\]

Here \( \gamma' \) is a normalization coefficient (to be determined), and \( \Gamma'_{ij}^{-1} \) is the inverse of \( \Gamma'_{ij} \), which is an \( N \times N \) (typically, multifactor) model covariance matrix. Note that \( \Gamma'_{ij} \) need not be the same as \( \Gamma^*_ij \), which denotes the multifactor model covariance matrix used in constructing the benchmark weights \( w^*_i \). This is because \( w'_i \) is a long-short (dollar-neutral) portfolio, so we do not have the same kinds of restrictions on \( \Gamma'_{ij} \) as on \( \Gamma^*_ij \). In fact, generally, we expect that \( \Gamma'_{ij} \) built using, say, the heterotic construction [Kakushadze, 2015c], [Kakushadze and Yu, 2016a] (which utilizes first principal components of the blocks of the sample correlation matrix corresponding to clusters) would work better than \( \Gamma^*_ij \). Then, given some \( \Gamma'_{ij} \), what should \( \gamma' \) be?

Considering that \( \Gamma'_{ij} \) is the multifactor model covariance matrix we use for modeling risk of a generic portfolio, we can compute the expected Sharpe ratio of the combined portfolio as follows:

\[
S = \frac{\sum_{i=1}^{N} E_i w_i}{\sqrt{\sum_{i,j=1}^{N} \Gamma'_{ij} w_i w_j}}
\]

Further, the expected correlation \( \rho \) between the portfolios \( w^*_i \) and \( w'_i \) is given by

\[
\rho = \frac{1}{\sigma^* \sigma'} \sum_{i,j=1}^{N} \Gamma'_{ij} w^*_i w'_j = \frac{\gamma' E^*}{\sigma^* \sigma'} = \frac{E^*}{\sigma^* \sigma'}
\]

\(^50\) A generic \( \Gamma'_{ij}^{-1} \) includes the “market mode”, so unless the returns \( E_i \) are contrivedly fine-tuned, the weights \( w'_i \), while not exactly dollar-neutral, generically are not expected to be highly skewed toward long or short positions. So, ignoring the dollar-neutrality condition is not detrimental. Without dollar-neutrality, assuming \( \sum_{i=1}^{N} w^*_i = 1 \), we no longer have \( \sum_{i=1}^{N} w_i = 1 \). However, this can be cured by simply rescaling \( w_i \) (albeit this may move \( w_i \) away from “optimality”). Ignoring the bounds poses a bigger issue. However, we will incorporate both the dollar-neutrality and bounds in a moment. Ignoring them for now serves the purpose of developing an intuitive understanding.
where $E^*$ is the expected return of the benchmark portfolio, and $\sigma^*$ and $\sigma'$ are the expected volatilities of the $w^*_i$ and $w'_i$ portfolios:

$$E^* = \sum_{i=1}^{N} E_i w^*_i \quad (161)$$

$$\left(\sigma^*\right)^2 = \sum_{i,j=1}^{N} \Gamma'_{ij} w^*_i w^*_j \quad (162)$$

$$\left(\sigma'\right)^2 = \sum_{i,j=1}^{N} \Gamma'_i w'_i w'_j = \left(\gamma'\right)^2 \left(e'\right)^2 \quad (163)$$

$$\left(e'\right)^2 = \sum_{i,j=1}^{N} \Gamma'_{-1 ij} E_i E_j \quad (164)$$

So, for the Sharpe ratio $S$ as a function of $\gamma'$ we have:

$$S(\gamma') = \frac{E^* + \gamma' \left(e'\right)^2}{\sigma(\gamma')} = \frac{\partial \sigma(\gamma')}{\partial \gamma'}$$

$$\sigma(\gamma') = \sqrt{\left(\sigma^*\right)^2 + 2 \gamma' \left(e'\right)^2} \left(\gamma'\right)^2 \quad (165)$$

The Sharpe ratio is maximized when $\gamma' \rightarrow \infty$.\footnote{Thus, we have $S(0) = E^*/\sigma^* = \rho e'$, and $S(\gamma' \rightarrow \infty) = e' > S(0)$. Also, $\partial S(\gamma')/\partial \gamma' = (\sigma^*)^2 \left(e'\right)^2 [1 - \rho^2] / \sigma^3(\gamma') > 0$.} However, in this limit we do not have a long-only portfolio. Instead, we have a long-short portfolio, which, in fact, would be dollar-neutral had we incorporated the dollar-neutrality constraint. In actuality, we must impose the bounds (156). Then, in the limit $\gamma' \rightarrow 0$, we have the long-only portfolio $w^*_i$. As we increase $\gamma'$, more and more bounds will be saturated. The bounds distort the $w'_i$ portfolio away from “optimality”. Above some value $\gamma'_{opt}$, the Sharpe ratio starts to fall off: $S(\gamma') < S(\gamma'_{opt})$ for $\gamma' > \gamma'_{opt}$. We can fix $\gamma'_{opt}$ via, e.g., the golden-section search [Kiefer, 1953] and use it as the “optimal value” of $\gamma'$.

### 4.1 Including Bounds and Constraints

In practice, in the presence of the bounds (156) it is easier to implement mean-variance optimization than maximizing the Sharpe ratio.\footnote{The two are not equivalent once bounds, costs, etc., are included, except only in one special case of establishing trades with linear costs [Kakushadze, 2015a].} In the mean-variance optimization, we maximize the objective function w.r.t. $w'_i$ (for a fixed value of $\gamma'$)

$$g(w'_i, \gamma') = \sum_{i=1}^{N} E_i w'_i - \frac{1}{\gamma'} \sum_{i,j=1}^{N} \Gamma'_{ij} w'_i w'_j \quad (166)$$

subject to the bounds (156) and the dollar-neutrality constraint (155). This optimization can be performed in a standard way (see, e.g., [Kakushadze, 2015a] for a
detailed algorithm and [Kakushadze, 2015c] for the source code). In fact, we may wish to include additional linear constraints. Thus, the expected correlation $\rho$ generally is nonzero. So, the dollar-neutral portfolio $w'_i$ has a nonzero expected beta with the benchmark portfolio. We may wish to make $\rho$ vanish. This can be achieved via the following linear homogeneous constrain on $w'_i$:

$$\sum_{i=1}^{N} q_i \cdot w'_i = 0 \quad (167)$$

$$q_i = \sum_{j=1}^{N} \Gamma'_{ij} \cdot w^*_j \quad (168)$$

Alternatively, we may wish to make the $w'_i$ portfolio simply orthogonal to the $w^*_i$ portfolio, which is achieved via the following constraint:

$$\sum_{i=1}^{N} w^*_i \cdot w'_i = 0 \quad (169)$$

More generally, we can have $p$ constraints

$$\sum_{i=1}^{N} Q_{ia} \cdot w'_i = 0, \quad a = 1, \ldots, p \quad (170)$$

Here $Q_{ia}$ is an $N \times p$ matrix with linearly independent columns, one of which is the unit $N$-vector corresponding to the dollar-neutrality constraint. We can also include neutrality w.r.t. sectors, industries, etc., or some style factors, if so desired. Etc.

### 4.1.1 A Comment

We can achieve an approximately null expected correlation $\rho$ in a different way. For generic “raw” expected returns $E_i$ we have nonzero $E^*$ and thus nonzero $\rho$. However, given an $N$-vector $E_i$, we can construct $E'_i$ orthogonal to $w^*_i$, e.g., by regressing $E_i$ (with unit weights and no intercept) over $w^*_i$ and taking the residuals (i.e., $E'_i = \epsilon_i$):

$$\epsilon_i = E_i - w^*_i \sum_{j=1}^{N} E_j \cdot w^*_j \sum_{j=1}^{N}(w^*_j)^2 \quad (171)$$

More generally, we can use a weighted regression with weights $v_i$ (and no intercept):\(^{53}\)

$$\epsilon_i = v_i \left( E_i - w^*_i \frac{\sum_{j=1}^{N} v_j \cdot E_j \cdot w^*_j}{\sum_{j=1}^{N} v_j (w^*_j)^2} \right) \quad (172)$$

If in (166) we substitute $\epsilon_i$ instead of $E_i$, then we approximately achieve (167).\(^{54}\)

\(^{53}\) Then the question is, what should the weights $v_i$ be? Basing them on volatilities $\sigma_i$ (or the corresponding specific risks) would make little sense as i) we are already optimizing the $w'_i$ portfolio and ii) both $|E_i|$ and $w^*_i$ scale linearly with $\sigma_i$. So, we can simply take unit weights $v_i \equiv 1$, or base them on quantities that are independent of $\sigma_i$ or have milder dependence thereon (e.g., $\ln(\sigma_i)$).

\(^{54}\) If it were not for the distortion caused by the bounds (156), (167) would be precisely satisfied.
5 Concluding Remarks

Let us briefly conclude with some remarks. First, in the market outperformance strategy we discuss in Section 4, the benchmark $w_i^*$ a priori can be any long-only portfolio (including S&P 500, Russell 3000, etc.), and not just built using the method of Section 3. However, market cap weighted portfolios and benchmark portfolios of Section 3 are expected to have sizable correlations. Intuitively this may appear to be evident as these are all long-only portfolios. However, this goes beyond such “zeroth-approximation” intuition. Thus, the weights $w_i^*$ in the benchmarks of Section 3 scale as $\propto 1/\sigma_i$ and $-\ln(\sigma_i)$ and $\ln(M_i)$ ($M_i$ is the market cap) are highly correlated. The “devil” then is in the details of the construction of Section 3, which aims to capture the substructure in the stock returns corresponding to the multilevel industry classification (clustering). I.e., there is more information in $w_i^*$ than in $M_i$.

In our construction above, $\beta_i$, albeit not completely arbitrary, are not fixed. Recall from Section 3 that, up to an overall normalization factor, $\hat{\beta}_i = \beta_i/\sigma_i$ are of order 1 with a tight distribution with a standard deviation also of order 1. So, what should these betas be? A simple answer is that there is no magic bullet here. We can simply pick them, backtest them out-of-sample and compare the results with those for other values.\(^{55}\) Or we can take the “observed” values $\hat{\beta}_{iobs}^*$ based on some broad market index, calculate their median value $\hat{\beta}_{median}^* = \text{median}(\hat{\beta}_{iobs}^*)$, and then cap and floor the “outliers” by $\hat{\beta}_{max}^* = \hat{\beta}_{median}^* + \kappa_{max} \text{MAD}(\hat{\beta}_{iobs}^*)$ and $\hat{\beta}_{min}^* = \hat{\beta}_{median}^* - \kappa_{min} \text{MAD}(\hat{\beta}_{iobs}^*)$, where $\kappa_{max} \sim 1$ and $\kappa_{min} \sim 1$ (we can set $\kappa_{max} = \kappa_{min}$). Fixing $\hat{\beta}_i$ this way is a “roundabout” as it uses another (cap-weighted) benchmark. But then again, there are no “first principles” that can fix $\hat{\beta}_i$ uniquely.

Let us note that, while we can take $\hat{\beta}_i \equiv 1$ in the construction of Section 3, we cannot take $\beta_i \equiv 1$. However, as mentioned above, $\beta_i \equiv 1$ would make little sense to begin with. Indeed, using (8), we have

$$
\beta_i = \sigma_i \rho_i / \sigma_F
$$

(173)

where, as before, $\sigma_F$ is the benchmark portfolio volatility, and $\rho_i$ is the sample correlation between the stock labeled by $i$ and the benchmark. Setting $\beta_i \equiv 1$ (again, up to an overall normalization factor) would imply that $\rho_i \propto 1/\sigma_i$. Similarly, in the factor model context of Section 3, we would have that, within the same cluster (sector, industry, etc.) stocks with high volatilities are almost uncorrelated to stocks with low volatilities (and each other). And this is not what we observe empirically (see, e.g., [Kakushadze, 2015c], [Kakushadze and Yu, 2016a]). We must have $\beta_i \propto \sigma_i$.

To tie up the final “lose end”, the weights $w_i'$ of the dollar-neutral portfolio in Section 4 can be computed using the bounded optimization code in Appendix C of [Kakushadze, 2015c], to wit, the function \texttt{bopt.calc.opt()}, whose arguments are: \texttt{ret} is the $N$-vector of expected returns $E_i$; \texttt{load} is the matrix of constraints $Q_{ia}$; \texttt{inv.cov} is the inverse matrix $\Gamma_{ij}^{-1}$; \texttt{upper}, \texttt{lower} are the bounds $w_i^{max}$, $w_i^{min}$.

---

\(^{55}\) This kind of “sampling” can get computationally taxing quickly.

\(^{56}\) MAD = mean absolute deviation.
A R Source Code for Benchmark Weights

In this appendix we give the R (R Package for Statistical Computing, http://www.r-project.org) source code for computing the benchmark weights \( w_i \) based on the algorithm of Section 3. This code is essentially self-explanatory and straightforward as it simply follows the formulas therein. The function `qrm.benchmark(ret, ind, beta, mkt.fac = T, z.min = 0.1, z.max = 0.9)` returns the weights \( w_i \) normalized as in (60). The input is as follows: `ret` is an \( N \times T \) matrix of returns \( R_{ts} \) (e.g., daily close-to-close returns), where \( N \) is the number of tickers, \( T \) is the number of observations in the time-series (e.g., the number of trading days), and the ordering of the dates is immaterial; ii) `ind` is a list of length \( P \), whose elements are populated by the binary matrices (with rows corresponding to tickers, so `dim(ind[[·]])[1]` is \( N \)) corresponding to the levels in the input binary industry classification in the order of decreasing granularity (for BICS `ind[[1]]` is the \( N \times K^{(1)} \) matrix \( \delta_{G(i),A^{(1)}} \) (sub-industries), `ind[[2]]` is the \( N \times K^{(2)} \) matrix \( \delta_{G'(i),A^{(2)}} \) (industries), and `ind[[3]]` is the \( N \times K^{(3)} \) matrix \( \delta_{G''(i),A^{(3)}} \) (sectors), where \( G = G^{(0)} \) maps tickers to sub-industries, \( G' = G^{(0)}G^{(1)} \) maps tickers to industries, and \( G'' = G^{(0)}G^{(1)}G^{(2)} \) maps tickers to sectors); iii) `beta` is the \( N \)-vector \( \beta_i \); iv) `mkt.fac`, where at the final step for `TRUE` we have a single industry factor (“Market”), while for `FALSE` the industry factors correspond to the least granular level in the industry classification (sectors for BICS); and v) `z.min` and `z.max` are \( z_{\min} \) and \( z_{\max} \) defined in Subsection 3.7.1.

There are two small tweaks in the source code beyond what is in Section 3. First, if Level-1 (in the nomenclature of Section 3) is very granular, for a given universe of stocks we can have one or more Level-1 Clusters each containing only one stock (e.g., single-stock sub-industries can and do arise in BICS). Typically, for long-horizon long-only portfolios using such granularity can be overkill. However, just in case, the source code deals with these situations in the internal function `calc.theta()`. Second, on occasion, it can happen that \( \theta_{\min} > \theta_{\max} \) (these quantities are defined in (135) and (136) in Subsection 3.7.1). This can happen when there are outliers with too low or too high variances. Now, here we can try to do all kinds of contrived and convoluted things. But there is a simple way of dealing with this situation. Note that by definition both \( \theta_{\min} \) and \( \theta_{\max} \) are positive. Violating \( \theta_{\max} \), i.e., having \( \theta > \theta_{\max} \), can – unacceptably – lead to negative specific variances (see Subsection 3.7.1). On the other hand, violating \( \theta_{\min} \), i.e., having \( \theta < \theta_{\min} \) is not detrimental so long as \( \theta > 0 \). Indeed, in this case we just have small factor risk. So, \( \theta_{\min} > \theta_{\max} \) can be simply dealt with by setting \( \vartheta \) as in (140) as opposed to

\[
\vartheta = \max(\vartheta_*, \vartheta_{\max}, \vartheta_{\min})
\]

which is equivalent to (140) when \( \theta_{\min} \leq \theta_{\max} \), but not when \( \theta_{\min} > \theta_{\max} \). So, the source code uses (140); see the line `t <- min(max(t, t.min), t.max)` in the function `calc.theta()`. Note that if we set \( \hat{\beta}_i \equiv 1 \), or more generally have \max(\hat{\beta}_i)/\min(\hat{\beta}_i) \leq \sqrt{(1 - z_{\min}^2)/(1 - z_{\max}^2)} \), then we are guaranteed to have \( \theta_{\min} \leq \theta_{\max} \) at Level-1. But \( \theta_{\min} > \theta_{\max} \) can arise at less granular levels due to outliers in factor variances.
qrm.benchmark <- function (ret, ind, beta, mkt.fac = T, z.min = 0.1, z.max = 0.9)
{
calc.load <- function(load, load1)
{
  x <- colSums(load1)
  load <- (t(load1) %*% load) / x
  return(load)
}

calc.theta <- function(x, b, z.min, z.max)
{
  if(length(x) == 1)
    return((1 - z.max^2) * x / b^2)
  s <- sqrt(diag(x))
  x <- t(x / s) / s
  b <- b / s
  t.min <- (1 - z.max^2) / min(b^2)
  t.max <- (1 - z.min^2) / max(b^2)
  x <- t(x * b) * b
  x <- sum(x) - sum(diag(x))
  b <- sum(b^2)^2 - sum(b^4)
  t <- x / b
  t <- min(max(t, t.min), t.max)
  return(t)
}

ind[[length(ind) + 1]] <- matrix(1, nrow(ind[[1]]), 1)
x <- cov(t(ret))
y <- list()
v <- list()
w <- b <- beta

for(lvl in 1:length(ind))
{
  if(lvl > 1)
  {
    flm <- calc.load(ind[[lvl]], ind[[lvl - 1]])
    b <- rep(1, nrow(flm))
  }
  else
    flm <- ind[[lvl]]
G <- rep(0, k <- ncol(flm))
y1 <- rep(0, nrow(flm))
v1 <- rep(0, k)

for(a in 1:k)
{
  take <- flm[, a] == 1
  if(lvl == length(ind) & !mkt.fac)
    G[a] <- 0
  else
    G[a] <- calc.theta(x[take, take], b[take],
      z.min = z.min, z.max = z.max)
y1[take] <- diag(x)[take] - b[take]^2 * G[a]
  if(lvl == 1)
    v1[a] <- sum(b[take]^2 / y1[take])
  else
    v1[a] <- sum(v[[lvl - 1]][take] /
      (1 + y1[take] * v[[lvl - 1]][take]))
}
y[[lvl]] <- y1
v[[lvl]] <- v1
x1 <- t(flm) %*% x %*% flm
u <- sqrt(G / diag(x1))
x <- t(x1 * u) * u

w <- w / y[[1]]

for(lvl in 1:(length(ind) - 1))
{
  for(a in 1:ncol(ind[[lvl]]))
  {
    take <- ind[[lvl]][, a] == 1
    w[take] <- w[take] / (1 + y[[lvl + 1]][a] * v[[lvl]][a])
  }
  w <- w / sum(w * beta)
  return(w)
}
B DISCLAIMERS

Wherever the context so requires, the masculine gender includes the feminine and/or neuter, and the singular form includes the plural and vice versa. The author of this paper (“Author”) and his affiliates including without limitation Quantigic® Solutions LLC (“Author’s Affiliates” or “his Affiliates”) make no implied or express warranties or any other representations whatsoever, including without limitation implied warranties of merchantability and fitness for a particular purpose, in connection with or with regard to the content of this paper including without limitation any code or algorithms contained herein (“Content”).

The reader may use the Content solely at his/her/its own risk and the reader shall have no claims whatsoever against the Author or his Affiliates and the Author and his Affiliates shall have no liability whatsoever to the reader or any third party whatsoever for any loss, expense, opportunity cost, damages or any other adverse effects whatsoever relating to or arising from the use of the Content by the reader including without any limitation whatsoever: any direct, indirect, incidental, special, consequential or any other damages incurred by the reader, however caused and under any theory of liability; any loss of profit (whether incurred directly or indirectly), any loss of goodwill or reputation, any loss of data suffered, cost of procurement of substitute goods or services, or any other tangible or intangible loss; any reliance placed by the reader on the completeness, accuracy or existence of the Content or any other effect of using the Content; and any and all other adversities or negative effects the reader might encounter in using the Content irrespective of whether the Author or his Affiliates is or are or should have been aware of such adversities or negative effects.

The R code included in Appendix A hereof is part of the copyrighted R code of Quantigic® Solutions LLC and is provided herein with the express permission of Quantigic® Solutions LLC. The copyright owner retains all rights, title and interest in and to its copyrighted source code included in Appendix A hereof and any and all copyrights therefor.

References

Avellaneda, M. and Lee, J.H. (2010) Statistical arbitrage in the U.S. equity market. Quantitative Finance 10(7): 761-782.

Bouchaud, J.-P. and Potters, M. (2011) Financial applications of random matrix theory: a short review. In: Akemann, G., Baik, J. and Di Francesco, P. (eds.) The Oxford Handbook of Random Matrix Theory. Oxford, UK: Oxford University Press.

Connor, G. and Korajczyk, R.A. (1993) A Test for the Number of Factors in an Approximate Factor Model. Journal of Finance 48(4): 1263-1291.
Fabozzi, F.J. and Francis, J.C. (1978) Beta as a random coefficient. *Journal of Financial and Quantitative Analysis* 13(1): 101-116.

Frobenius, G. (1912) Über Matrizen aus Nicht Negativen Elementen. In: *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, pp. 456-477.

Geweke, J. and Zhou, G. (1996) Measuring the Pricing Error of the Arbitrage Pricing Theory. *Review of Financial Studies* 9(2): 557-587.

Grinold, R.C. and Kahn, R.N. (2000) *Active Portfolio Management*. New York, NY: McGraw-Hill.

Kakushadze, Z. (2015a) Mean-Reversion and Optimization. *Journal of Asset Management* 16(1): 14-40. Available online: https://ssrn.com/abstract=2478345.

Kakushadze, Z. (2015b) Russian-Doll Risk Models. *Journal of Asset Management* 16(3): 170-185. Available online: https://ssrn.com/abstract=2538123.

Kakushadze, Z. (2015c) Heterotic Risk Models. *Wilmott Magazine* 2015(80): 40-55. Available online: https://ssrn.com/abstract=2600798.

Kakushadze, Z. and Yu, W. (2016a) Multifactor Risk Models and Heterotic CAPM. *Journal of Investment Strategies* 5(4): 1-49. Available online: https://ssrn.com/abstract=2722093.

Kakushadze, Z. and Yu, W. (2016b) Statistical Industry Classification. *Journal of Risk & Control* 3(1): 17-65. Available online: https://ssrn.com/abstract=2802753.

Kakushadze, Z. and Yu, W. (2017) Statistical Risk Models. *Journal of Investment Strategies* 6(2): 1-40. Available online: https://ssrn.com/abstract=2732453.

Kakushadze, Z. and Yu, W. (2018) Notes on Fano Ratio and Portfolio Optimization. *Journal of Risk & Control* 5(1): 1-33. Available online: https://ssrn.com/abstract=3050140.

Kiefer, J. (1953) Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society* 4(3): 502-506.

Lo, A.W. (2016) What Is an Index? *Journal of Portfolio Management* 42(2): 21-36.

Markowitz, H. (1952) Portfolio selection. *Journal of Finance* 7(1): 77-91.
Perron, O. (1907) Zur Theorie der Matrices. *Mathematische Annalen* 64(2): 248-263.

Roy, O. and Vetterli, M. (2007) The effective rank: A measure of effective dimensionality. In: *Proceedings – EUSIPCO 2007, 15th European Signal Processing Conference*. Poznań, Poland (September 3-7), pp. 606-610.

Sharpe, W.F. (1963) A simplified model of portfolio analysis. *Management Science* 9(2): 277-293.

Sharpe, W.F. (1966) Mutual Fund Performance. *Journal of Business* 39(1): 119-138.

Sharpe, W.F. (1994) The Sharpe Ratio. *Journal of Portfolio Management* 21(1): 49-58.

Trzcinka, C.A. (1986) On the Number of Factors in the Arbitrage Pricing Model. *Journal of Finance* 41(2): 347-368.