

# AUGMENTED RISK MODELS TO MITIGATE FACTOR ALIGNMENT PROBLEMS

Anureet Saxena\* and Robert A. Stubbs<sup>†</sup>

Construction of optimized portfolios entails a complex interaction between three key entities, namely, the risk factors, the alpha factors and the constraints. The problems that arise due to mutual misalignment between these three entities are collectively referred to as Factor Alignment Problems (FAP). Examples of FAP include risk underestimation of optimized portfolios, undesirable exposures to factors with hidden and unaccounted systematic risk, consistent failure in achieving ex-ante performance targets, and inability to harvest high quality alphas into above-average IR. In this paper, we give a detailed analysis of FAP and discuss solution approaches based on augmenting the user risk model with a single additional factor y. For the case of unconstrained mean-variance optimization (MVO) problems, we develop a generic analytical framework to analyze the ex-post utility function of the corresponding optimal portfolios, derive a closed-form expression of the optimal factor volatility value and compare the solutions for various choices of y culminating with a closed-form expression for the optimal choice of y. Augmented risk models not only correct for risk underestimation bias of optimal portfolios but also push the ex-post efficient frontier upward thereby empowering a portfolio manager (PM) to access portfolios that lie above the traditional risk-return frontier. We corroborate our theoretical results by extensive computational experiments, and discuss market conditions under which augmented risk models are likely to be most beneficial.



## 1 Introduction

Factor models play an integral role in quantitative equity portfolio management. Their applications extend to almost every aspect of quantitative investment methodology including construction of alpha models, risk models, portfolio construction, risk decomposition, and performance attribution. Given their pervasive presence in the field and the natural trend toward specialization, it comes as no surprise that different groups of researchers are often involved in developing factor models for each one of the aforementioned applications.

For instance, a team of quantitative portfolio managers (PM) can develop an in-house model for

<sup>\*</sup>Vice President, Analyst, Allianz Global Investors, 600 West Broadway, San Diego, CA 92101, USA.

<sup>&</sup>lt;sup>†</sup>Vice President, Research, Axioma, Inc., 400 Northridge Road, Suite 850, Atlanta, GA 30350, USA.

alpha generation, and procure a factor model for the purposes of risk management from a thirdparty risk model vendor. Subsequently, they can combine the two models within the framework of Markowitz mean-variance optimization (MVO) framework to construct optimal portfolios. A completely different factor model can then be used for the purposes of performance attribution to identify the key drivers and detractors of performance. Notably, the choice of factors in each one of these factor models need not be identical, thereby introducing incongruity in the portfolio management process. To further complicate the matters, the constraints in the quantitative strategy can introduce additional systematic risk exposures that are not captured by the risk model. The problems that arise due to the interaction between the alpha model, the risk model, and constraints in an MVO framework are collectively referred to as Factor Alignment Problems (FAP). Detailed theoretical investigation of FAP leading to a solution methodology in the form of augmented risk models constitute the emphasis of this paper. Next we give a brief survey of existing research on FAP, and highlight our key contributions.

The primary purpose of portfolio optimization is to create a portfolio having an optimal riskreturn tradeoff. If a portion of systematic risk exposure of the portfolio is inadequately captured by the risk model then the resulting portfolio cannot be expected to be optimal ex-post, its exante optimality notwithstanding. In other words, FAP symbolize the difficulties that a PM faces in ensuring the ex-post optimality of a portfolio that is deemed to be optimal ex-ante in the MVO framework. Examples of FAP include risk underestimation of optimized portfolios, undesirable exposures to factors with hidden and unaccounted systematic risk, consistent failure in achieving ex-ante performance targets, and inability to harvest high quality alphas into aboveaverage IR.

Several authors have examined FAP recently and have proposed various solution techniques. Saxena and Stubbs (2013) conducted an empirical case study to understand the risk underestimation problem, a prominent symptom of FAP. The authors used real-world data and a battery of backtests to demonstrate the perverse and pervasive nature of FAP. They demonstrated that all optimized portfolios share a common property, namely, they have exposure to certain kinds of latent systematic risk factors that are uncorrelated with factors of the risk model that was used to generate them. Ceria et al. (2012) examine potential sources of the mentioned systematic risk factors and suggest that proprietary definitions of certain style (B/P, E/P, etc.) and technical factors can introduce them. Lee and Stefek (2008) illustrate a similar idea by using two different definitions of a momentum factor to define alpha and risk factors, and argue that the optimizer is likely to load up on the difference between the two thereby taking unintended bets. Saxena and Stubbs (2012) discuss a detailed empirical case study on FAP using the USER model (see Guerard et al., 2012a); among other things, they quantify the portion of unaccounted systematic risk that can be attributed to the constituent factors of the USER model and constraints present in the strategy. Unlike previous studies which have investigated FAP from an empirical standpoint, we pursue a theoretical exploration of this topic. We go back to the roots of mean-variance optimization, and demonstrate analytically the tendency of the optimizer to adversely exploit inconsistencies between the alpha and risk models, thereby compromising the efficiency of the resulting portfolios. Our analysis not only yields diagnostic tools to identify the presence of FAP, but also provides a natural remedy to FAP in the form of augmented risk models. The rest of this paper is organized as follows.

Section 2 discusses a prototypical quantitative strategy with the aim of identifying some of

the common symptoms of FAP, such as the risk underestimation problem, and undesirable and unintended exposure to systematic risk factors which are not captured by the risk model. The aim of this section is to provide a simple but comprehensive practical example which can be used to put the theoretical results presented in the later part of the paper in context. We return to the example discussed in Section 2 again in Section 7 wherein we show how our proposed methodology address FAP, and improves riskadjusted returns. Section 3 lays out the theoretical model which is used in the rest of the paper. We discuss the assumptions we make in our analytical derivations, and provide theoretical and empirical justifications to them. Among other things, we introduce the notion of an augmented risk model which is used throughout the paper to remedy FAP. Given an arbitrary asset-asset covariance matrix Q, and a risk factor y, an augmented risk model is defined by the covariance matrix  $Q_{y} = Q + vyy^{T}$ . In other words, if Q was derived from a factor model,  $Q_{y}$  is derived from an enhanced factor model that has all the original factors, and an additional augmenting factor y which is assumed to be uncorrelated with the original suite of risk factors.

In Section 4 we focus on the unconstrained MVO model, and develop a generic analytical framework to analyze the ex-post utility function of the corresponding optimal portfolios, derive a closedform expression of the optimal factor volatility and compare the solutions for various choices of *y* culminating with a closed-form expression for the optimal choice of *y*. Among other things, we show that using an augmented risk model with an appropriately chosen volatility parameter  $\nu$  not only solves the risk underestimation problem but also improves the ex-post utility function. The key result in this paper, referred to as the "Pushing Frontier Theorem" shows that using an augmented risk model shifts the ex-post efficient frontier upward, thereby allowing the PM to access portfolios that are not attainable using the traditional MVO approach. In Section 5, we extend these results to a constrained MVO model. We employ the concept of implied alpha to allow us to make a smooth transition from unconstrained MVO to its constrained counterpart. Recall that implied alpha is obtained by tilting the alpha in the direction of binding constraints, and acts as the de facto alpha for constrained MVO models.

Application of augmented risk models requires two key parameters, namely, the choice of an augmenting factor y and an estimate of its volatility  $\nu$ . In Section 6, we discuss an exponentially weighted moving average (EWMA) volatility model to calibrate the volatility of augmenting factors. The emphasis in this section is on simplicity and practicality of the proposed approach, and we discuss a model that meets both of these criteria. Section 7 has a threefold emphasis. First, we give extensive empirical results that corroborate the theoretical findings discussed in the preceding sections. Second, we introduce the notion of "frontier spreads" to capture improvements in risk-adjusted returns that result due to the application of augmented risk models. Subsequently, we study the impact of various strategic parameters (turnover limits, asset bounds, etc.) on the frontier spreads, and also seek to identify market regimes where using an augmented risk model is most likely to yield significant improvements. Finally, we present computational results with a wide variety of alpha models to attest the robustness of the proposed approach. Section 8 concludes the paper with some closing remarks.

# 2 A practical active strategy

The focus of this section is twofold. First, we want to use a very simple value momentum strategy to illustrate a classic symptom of FAP, namely, the risk underestimation problem. Second, we show how the risk underestimation problem can be traced to certain hidden systematic risk factors that are not captured by the factor structure of the base risk model (BRM). Among other things, this sets the stage for the theoretical model discussed in the following section which assumes the existence of systematic risk factors missing from the BRM.

We used the following strategy in our experiments.<sup>(B,1)</sup>

maximize Expected Return s.t. Fully invested long-only portfolio Active GICS<sup>®</sup> sector exposure

Active GICS<sup>®</sup> sector exposure  
constraint (
$$\pm 20\%$$
)  
Active GICS industry exposure  
constraint ( $\pm 10\%$ )  
Active asset bounds constraint  
( $\pm 2\%$ )  
Turnover (two-way) constraint  
(16%)  
Active Risk constraint ( $\sigma\%$ )  
Benchmark = Russell<sup>®</sup> 3000.

We used a fundamental risk model as our BRM in defining the active risk constraint.<sup>2</sup> The expected returns were derived using an equal weighted combination of the BP variable in the USER model (Guerard *et al.*, 2012b) and the

medium-term momentum factor in the BRM. We ran monthly backtests based on the above strategy in the 1999–2009 time period for various values of  $\sigma$  chosen from {1.0%, 1.1%, ..., 5.0%}.

We use the notion of the bias statistic to identify statistically significant biases in risk prediction. If the ex-ante risk prediction is unbiased, then the bias statistic should be close to 1.0 (see Saxena and Stubbs, 2013 for more details). A bias statistic value which is significantly above (below) 1.0 indicates downward (upward) biases in risk prediction. Figure 1 reports the bias statistics of the portfolios for various risk target levels. Clearly, the bias statistics are significantly above the 95% confidence interval, thereby confirming the statistical significance of the downward bias in predicted risk estimates. We next focus on optimal portfolios that were generated when a risk target of 3.0% was employed.

At  $\sigma = 3.0\%$ , the optimal portfolios had realized active risk of 3.81%. The bias statistic for these portfolios was 1.27 which clearly lies outside the 95% confidence interval [0.87, 1.13]. Figure 2 further corroborates this phenomenon by showing the time series of realized risk of the optimal portfolios computed using 24 period realized returns on a rolling horizon basis; we also show the predicted risk of the portfolios



Figure 1 Bias statistic (active risk).



**Figure 2** Time series of realized (24-period rolling) and predicted risk of optimal portfolios constructed using our BRM.

for the sake of comparison. While the degree of under-prediction might have varied, the realized risk was consistently above the predicted risk in most of the periods. It is tempting to believe that the risk models used in construction of the optimized portfolios were themselves biased, and the risk underestimation problem is simply an artifact of the bias in the BRM. Saxena and Stubbs (2013) examined this issue and demonstrated that the base risk model, in fact, produces unbiased risk estimates for random portfolios. Consequently, the bias depicted in Figure 1 is peculiar to optimized portfolios.

We introduce two additional concepts to assist us in tracing the sources of the risk underestimation problem. Let X denote the  $n \times m$  exposure matrix associated with the BRM; n denote the number of assets; and m denote the number of factors in the BRM. Given an arbitrary factor  $\alpha$ , consider the following linear regression model that regresses  $\alpha$  against factors in the base risk model,

$$\alpha = Xu + \alpha_{\perp};$$

the residual  $\alpha_{\perp}$  in the above regression model is referred to as the *orthogonal* component of  $\alpha$ , whereas  $\alpha_X = Xu$  is referred to as the *spanned* component of alpha. Mathematically,

$$\alpha_X = X(X^T X)^{-1} X^T \alpha; \quad \alpha_\perp = \alpha - \alpha_X.$$

By virtue of being uncorrelated with all the factors included in the BRM,  $\alpha_{\perp}$  has no systematic risk exposure with respect to the BRM; in other words, the BRM assumes that  $\alpha_{\perp}$  has only idiosyncratic risk. This assumption can be problematic if certain systematic risk factors are missing from the BRM and  $\alpha_{\perp}$  has overlap with some of the missing systematic risk factors. As we will soon discover, this indeed turns out to be the case, thus providing a key insight into the risk underestimation problem.

Next we introduce the notion of augmented regressions which can be used to determine if the orthogonal component of a given factor has overlap with systematic risk factors missing from the BRM. Consider a linear regression model that regresses asset returns against factors in the BRM, represented by the matrix X, and the normalized orthogonal component  $y = \frac{1}{\|\alpha_{\perp}\|} \alpha_{\perp}$  of  $\alpha$ . If  $\alpha_{\perp}$ was truly idiosyncratic in nature then the factor returns associated with y, denote by  $f_y$ , in the above regression model should not be statistically significant. Alternatively, if  $f_v$  is indeed statistically significant and has nontrivial volatility then we can be assured that  $\alpha_{\perp}$  has overlap with systematic risk factors missing from the base risk model.



Figure 3 Time series of *t*-statistics (absolute values) associated with the orthogonal component.

Figure 3 shows the time series of (absolute) *t*-statistics associated with the orthogonal component of alpha ( $\alpha$ ) and implied alpha ( $\tilde{\alpha}$ ) in the corresponding augmented regression model;  $\alpha_{\perp}$  ( $\tilde{\alpha}_{\perp}$ ) was found to be statistically significant (90% cf) in 40% (20%) of the periods. Given that a median factor in our BRM is statistically significant in about 20–30% of the periods, these statistics imply that  $\alpha_{\perp}$  and  $\tilde{\alpha}_{\perp}$  are as significant as half of the factors in the BRM. Figure 4 reports the annualized volatility of factor returns associated with  $\alpha_{\perp}$  and  $\tilde{\alpha}_{\perp}$  computed using a rolling 24-period window. As evident from the chart, not only are  $\alpha_{\perp}$  and  $\tilde{\alpha}_{\perp}$  statistically significant, but their factor returns also exhibit significant volatility. To put these numbers in perspective, note that a median normalized<sup>3</sup> factor in the BRM has annualized volatility of roughly 30%.

All of these results indicate that the orthogonal component of  $\alpha$  and  $\tilde{\alpha}$  does carry a significant amount of systematic risk which is not accounted for during the process of portfolio construction. The section that follows builds on these observations; specifically we propose a theoretical model that explicitly accounts for systematic risk factors missing from the BRM, and use it to assess the marginal cost of FAP.



**Figure 4** Time series of realized systematic risk of the orthogonal component computed using a rolling 24-period window.

#### **3** Theoretical model

In this section, we describe a theoretical model that is used in the rest of the paper. We discuss various underlying assumptions, and provide theoretical or empirical justifications wherever possible. We assume that the returns process is governed by the following stationary stochastic process,

$$r = Xf + Zg + u,$$

where  $r \in \mathbb{R}^{n \times 1}$  denotes the vector of asset returns, *n* denotes the number of assets;  $X \in \mathbb{R}^{n \times m}$  denotes the exposure matrix associated with common systematic risk factors that are present in the BRM; and  $Z \in \mathbb{R}^{n \times p}$  denotes the exposure matrix associated with systematic risk factors missing from the BRM. For the sake of brevity, we refer to systematic risk factors represented by *X* and *Z* matrices as *X* and *Z* factors, respectively. Let  $f \in \mathbb{R}^{m \times 1}$  and  $g \in \mathbb{R}^{p \times 1}$  denote the random factor returns associated with *X* and *Z* factors, whereas  $u \in \mathbb{R}^{n \times 1}$  denotes the vector of random asset-specific returns. Without loss of generality, we assume that  $X^TX = I$ ,  $X^TZ = 0$ and  $Z^TZ = I$ . Additionally, we assume that

$$(f,g,u) \sim N\left(\begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}\Omega & 0 & 0\\0 & \Lambda & 0\\0 & 0 & \sigma_s^2I\end{bmatrix}\right).$$

Let  $Q_T = E[rr^T]$  denote the asset-asset covariance matrix; the risk model associated with  $Q_T$ is referred to as the *true* risk model. It is easy to verify that  $Q_T = X\Omega X^T + Z\Lambda Z^T + \sigma_s^2 I$ .

Next, consider a base risk model (BRM) that is calibrated by using only the X-factors. Let Q denote the covariance matrix associated with such a risk model, and let  $\hat{\Omega} \in \mathbb{R}^{m \times m}$  and  $\hat{\Delta} \in \mathbb{R}^{n \times n}$  denote the corresponding factor-factor covariance matrix and the diagonal matrix of asset-specific variances, respectively; we use Qas a proxy for the BRM. Using standard error analysis arguments from linear regression theory (see Greene, 2002) it can be shown that,<sup>4</sup>

$$\hat{\Omega} = \Omega + \sigma_u^2 I, \quad and$$
$$\hat{\Delta} = Diag(Z\Lambda Z^{\mathrm{T}}) + \sigma_s^2 Diag(I - XX^{\mathrm{T}}).$$

Diag(M) denotes the diagonal matrix obtained by retaining only the diagonal entries of a square matrix M. In our analysis we make the following two assumptions,

(1) 
$$\hat{\Omega} = \Omega.$$
  
(2)  $\hat{\Delta} = \sigma_s^2 I$ 

Next, we give theoretical and empirical results to justify the above assumptions.

It is instructive to examine how the systematic risk that arises by virtue of the exposure to the *Z*-factors is captured by the BRM. For the sake of illustration, consider the scenario where there is exactly one missing systematic risk factor, say *z* with ||z|| = 1. Let  $\Lambda = [\sigma_z^2]$ . Let  $\sigma_T^2(z) = z^T Q_T z$  and  $\sigma_Q^2(z) = z^T Q z$  denote the true and estimated variance of factor *z*. Note that

$$\begin{aligned} \sigma_Q^2(z) &= z^{\mathrm{T}} Q z \\ &= z^{\mathrm{T}} \hat{\Delta} z \\ &= \sigma_s^2 + \sigma_z^2 z^{\mathrm{T}} Diag(z z^{\mathrm{T}}) z \\ &- \sigma_s^2 z^{\mathrm{T}} Diag(X X^{\mathrm{T}}) z \end{aligned}$$

$$\begin{aligned} &= \sigma_s^2 + \sigma_z^2 \sum_{i=1}^n z_i^4 - \sigma_s^2 \sum_{i=1}^n \sum_{j=1}^m X_{ij}^2 z_i^2 \end{aligned}$$

$$< \sigma_s^2 + \sigma_z^2 \sum_{i=1}^n z_i^2 \end{aligned}$$

$$\begin{aligned} &= \sigma_s^2 + \sigma_z^2 \\ &= z^{\mathrm{T}} Q_T z \\ &= \sigma_T^2(z). \end{aligned}$$

The above result has far-reaching consequences that we discuss next. By virtue of missing the Z

THIRD QUARTER 2015

factors, the BRM differs from the true risk model  $Q_T$  in key respects. First, the BRM assumes that any factor (portfolio) which is uncorrelated with the X factors has only idiosyncratic risk. For instance, the systematic variance of z as computed by the BRM is given by  $z^{T}X\hat{\Omega}X^{T}z = 0$ . In other words, the risk model Q fails to account for systematic risk of portfolios that arises by virtue of exposure to the Z factors. Second, there is a small increment in the specific risk estimate of Z as determined by the BRM; essentially a part of the systematic volatility that can be attributed to Z factors is captured by the specific risk component, namely  $\Delta$ , of Q. Unfortunately, the increase in the specific risk of zis very small as compared to its true systematic risk. In other words, the specific risk component of the BRM fails to capture the true systematic risk that arises by virtue of exposure to the Zfactors resulting in the risk underestimation problem. Furthermore, by only utilizing information contained in the diagonal terms of  $Z\Lambda Z^{T}$ , the BRM fails to capitalize the nondiagonal terms, a phenomenon whose impact grows quadratically with the number of assets. To see this, note that

$$\hat{\Delta}_{ii} = \sigma_s^2 + \sum_{j,k=1}^p Z_{ij} \Lambda_{jk} Z_{ki} - \sigma_s^2 \sum_{j=1}^m X_{ij}^2.$$

As  $n \to \infty$ ,  $Z_{ij} \to 0$ , and  $X_{ij} \to 0$  assuming that X and Z factors have nonzero exposure to a

sizable subset of the asset universe. Thus, unless the Z factors have unusually high volatilities we can assume that  $\hat{\Delta}_{ii} = \sigma_s^2$ . To provide empirical support to this assumption we conducted the following experiment.

We generated a custom risk model (CRM) that includes all factors in the BRM and also the orthogonal component of the BP variable (see Section 2). In other words, we enlarged the suite of risk factors in the BRM by adding  $BP_{\perp}$ , used the enlarged suite of risk factors in cross-sectional regressions to generate time series of factor and asset-specific returns which were subsequently used to generate the CRM. We also generated the time series of optimal portfolios using the setup described in Section 2, and computed the time series of predicted active systematic and specific risk estimates for the resulting portfolios using the BRM and CRM; recall that the portfolios were constructed using the BRM. Figure 5 reports the key results. Two remarks are in order.

First, the estimates of systematic risk produced by the CRM are consistently higher than those produced by the BRM. This confirms the inability of the BRM to correctly gauge the systematic risk that arises by virtue of exposure to  $\alpha_{\perp}$  (also see Section 2). Second, the difference between the specific risk estimates is roughly two orders of magnitude smaller than the difference between the systematic risk estimates. All in all, these





JOURNAL OF INVESTMENT MANAGEMENT

results indicate that the specific risk estimates produced by the two risk models are materially indistinguishable as far as predicted active risk computation is concerned, and hence provides credence to our assumption  $\hat{\Delta}_{ii} = \sigma_s^2$ .

Next we move our focus to  $\hat{\Omega}$ . Note that due to estimation errors the factor-factor covariance matrix  $\hat{\Omega}$  associated with the BRM differs from the true covariance matrix  $\Omega$  by  $\sigma_s^2 I$ . While these estimation errors can be extremely important for risk-based strategies (minimum variance, risk parity, etc.), their significance for active strategies is considerably reduced by several factors. For example, most portfolio managers (PM) often impose fairly tight bounds on exposure to common or well-known systematic risk factors due to compliance reasons and also to avoid unintended beta exposure. Furthermore, volatilities of common systematic risk factors tend to be significantly higher than specific risk estimates thus mitigating the impact of estimation errors. For instance, Figure 6 shows the ratio of average variance of systematic risk factors in a particular statistical risk model<sup>5</sup> and the average asset-specific variance obtained using the same risk model. Note that the factor variances are roughly 30–50 times larger than the asset-specific variances. Based on

these observations, and given that the focus of this paper is on model specification error and not on estimation error, we assume that  $\hat{\Omega} = \Omega$ . To summarize, we assume that the BRM is given by  $Q = X\Omega X^{T} + \sigma_{s}^{2}I$ .

Throughout this paper, we work with portfolios generated as an optimal solution to the following MVO problem.

$$\max_{\substack{s.t.\\ h \in \mathcal{C},}} \alpha^{\mathrm{T}} h - \frac{\lambda}{2} h^{\mathrm{T}} \mathcal{Q} h$$

where C represents a suitable set of constraints. For instance, for a PM managing a long-only portfolio, C would represent the universe of all fully invested long-only portfolios with active asset, industry, and sector exposure constraints. Let Q denote the risk model used during portfolio construction and h(Q) denote the resulting optimal portfolio. Our objective in this paper is to evaluate and compare portfolios generated for various choices of the risk model Q. Naturally, we need a common metric of comparison; we use the following utility function for this purpose.

$$U(h) = \alpha^{\mathrm{T}} h - \frac{\lambda}{2} h^{\mathrm{T}} Q_{T} h.$$

The above utility function can be regarded as an expected-return realized-risk utility function,



Figure 6 Ratio of average factor variance and average asset-specific variance in our statistical risk model.

and is referred to as the "realized" utility function so as to differentiate it from the "predicted" utility function given by  $\alpha^{T}h - \frac{\lambda}{2}h^{T}Qh$ . We would like to emphasize that we used the above utility function only to get analytical insights into FAP. The insights thus obtained are equally applicable and meaningful when applied in a realistic setting wherein the actual realized returns, and not expected returns, are used for performance evaluation (see Section 7).

Finally, we introduce the concept of augmented risk models, a key device that is used throughout this paper to devise solutions to FAP. Given an arbitrary factor *y* satisfying  $y^{T}X = 0$  and ||y|| = 1, let  $Q_{\nu} = Q + \nu y y^{T}$  where  $\nu$  denotes the volatility parameter which is left unspecified. We will see later the impact of different choices of  $\nu$  on the resulting optimal portfolios  $h(Q_v)$ , and how  $\nu$  can be calibrated in practice.  $Q_{\nu}$  is referred to as an *augmented* risk model. Our solution to FAP entails replacing the base risk model Q by an augmented risk model  $Q_{y}$ ; we compare and contrast portfolios obtained for various choices of augmenting factors y, and derive a closedform expression for the optimal augmenting factor.

With this background, we initiate our analysis with the case when  $C = \phi$ , i.e., an unconstrained MVO model.

# 4 Augmented risk models: Unconstrained MVO

Throughout this section, we assume that  $C = \phi$ ; additionally, unless otherwise stated we also assume that  $\alpha_{\perp} \neq 0$ . Let *y* denote an arbitrary augmenting factor satisfying  $y^{T}X = 0$  and ||y|| = 1. The proposition that follows sheds light on the effect of using an augmented risk model on the structure of optimal holdings.

### **Proposition 1**

$$h(Q_y) = h(Q) - \frac{1}{\lambda} \left( \frac{\nu y^{\mathrm{T}} \alpha_{\perp}}{\sigma_s^2 (\sigma_s^2 + \nu)} \right) y$$
  
$$y^{\mathrm{T}} h(Q_y) = \frac{y^{\mathrm{T}} \alpha_{\perp}}{\lambda (\sigma_s^2 + \nu)}.$$
 (1)

First, note that if  $y^{T}\alpha_{\perp} = 0$ , then  $h(Q_{y}) = h(Q)$ regardless of the value of v. In other words, using an augmenting factor y will materially change the structure of the optimal holdings only if  $\alpha_{\perp}$  has a nonzero exposure to y. This observation reiterates the integral role of  $\alpha_{\perp}$  in any solution technique to FAP that is based on using an augmented risk model.

Second, assuming that  $y^{T}\alpha_{\perp} \neq 0$  and  $\nu \neq 0$ , the above proposition shows that using an augmenting factor y is equivalent to tilting the optimal portfolio h(Q) in the direction away from y. Furthermore, by controlling the volatility  $\nu$  of y in the augmented risk model, the end user can carefully manage the extent of tilting. For instance, while using a value of  $\nu = 0$  annuls the effect of the augmenting factor, choosing  $\nu = \infty$  completely eliminates the exposure of the optimal portfolio  $h(Q_y)$  to factor y. Naturally, a portfolio manager will choose a value of  $\nu$  in between these extreme values so as to manage the exposure of the optimal portfolio to factor y depending on the IC and systematic risk of y.

Third, when  $y = \frac{1}{\|\alpha_{\perp}\|} \alpha_{\perp}$ , it can be shown that

$$h(Q_y) = \frac{1}{\lambda} \left( Q^{-1} \alpha_X + \frac{1}{\sigma_s^2 + \nu} \alpha_\perp \right).$$

Thus by varying the parameter  $\nu$  we can control the relative weight of  $\alpha_{\perp}$  in the optimal holdings. Since overloading of  $\alpha_{\perp}$  is the main source of risk underestimation (see Saxena and Stubbs, 2013) it follows that the bias in the risk prediction can be completely eliminated by using an appropriate value of  $\nu$ ;  $y = \frac{1}{\|\alpha_{\perp}\|} \alpha_{\perp}$  is known as the Alpha Alignment Factor (AAF; see Saxena and Stubbs, 2012, 2013). But,  $\alpha_{\perp}$  is not the only factor that can be used to remedy the risk underestimation problem.

Fourth, consider the case when  $y = \frac{1}{\|P\alpha_{\perp}\|} P\alpha_{\perp}$ , where  $P = Z(Z^{T}Z)^{-1}Z^{T} = ZZ^{T}$  is the projection matrix associated with Z; thus  $P\alpha_{\perp}$  denotes the projection of  $\alpha_{\perp}$  to the space of missing factors. There is a subtle but important distinction between using  $\alpha_{\perp}$  and  $P\alpha_{\perp}$  as the augmenting factor. Recall that the act of augmenting the base risk model amounts to tilting the optimal portfolio in the direction away from the augmenting factor. Thus while using  $\alpha_{\perp}$  reduces the exposure of the portfolio to the entire orthogonal component of  $\alpha$ , using  $P\alpha_{\perp}$  reduces the exposure to only that portion of  $\alpha_{\perp}$  that has systematic risk. Since the portion of  $\alpha_{\perp}$  that has no systematic risk, namely  $\alpha_{\perp} - P\alpha_{\perp}$ , does not contribute to the FAP,  $P\alpha_{\perp}$ should be able to outperform  $\alpha_{\perp}$  as an augmenting factor.

As mentioned in Section 3, we will use the following utility function to evaluate and compare different portfolios,

$$U(h) = \alpha^{\mathrm{T}} h - \frac{\lambda}{2} h^{\mathrm{T}} Q_T h.$$

Since our focus in this section is on optimal portfolios derived from augmented risk models, we define U(y, v) to be  $U(h(Q_y))$  for ease of notation. To present the results in this section we introduce some additional notations defined below.

- Let  $\delta^2 = \frac{\alpha_{\perp}^T Z \Lambda Z^T \alpha_{\perp}}{\|\alpha_{\perp}\|^2}$  denote the systematic variance of  $\frac{1}{\|\alpha_{\perp}\|} \alpha_{\perp}$ .
- Let  $\epsilon^2 = y^T Z \Lambda Z^T y$  denote the systematic variance of y.
- Let  $\gamma = \frac{\alpha_{\perp}^{T} Z \Lambda Z^{T} y}{\|\alpha_{\perp}\|}$  denote the joint systematic covariance of  $\frac{1}{\|\alpha_{\perp}\|} \alpha_{\perp}$  and y.
- Let  $\eta = \frac{\nu(y^{\mathrm{T}}\alpha_{\perp})}{\sigma^2(\nu+\sigma^2)}$ .

THIRD QUARTER 2015

The proposition that follows gives a closed-form expression for U(y, v) in terms of U(h(Q)).

#### **Proposition 2**

$$U(y, \nu) = U(h(Q)) + \frac{1}{2\lambda} \left( \frac{2\eta\gamma \|\alpha_{\perp}\|}{\sigma^2} - \eta^2 (\sigma^2 + \epsilon^2) \right).$$

Note that if  $Z^{T}\alpha_{\perp} = 0$  then  $\gamma = 0$  and  $U(y, v) \leq U(h(Q))$ . In other words, if the orthogonal component of  $\alpha$  has no systematic risk then there is no marginal benefit of using an augmented risk model. This observation highlights the fact that it is the component  $ZZ^{T}\alpha_{\perp}$  of  $\alpha_{\perp}$  that has systematic risk which holds the clues to the solution of FAP. Our experiments with a collection of real-life alphas of portfolio managers from many different firms show that  $\alpha_{\perp}$  not only has systematic risk, but also its magnitude is comparable to the systematic risk of standard factors in fundamental risk models (see Saxena and Stubbs, 2013). In the rest of this section, we assume that  $Z^{T}\alpha_{\perp} \neq 0$ .

Recall that we have not yet fixed the volatility  $\nu$  of the augmenting factor y. The proposition that follows gives a closed-form expression for  $\nu$  that maximizes  $U(y, \nu)$ .

**Proposition 3** For a given factor y, U(y, v) is maximized at

$$v_{opt} = \frac{\sigma^2}{\frac{(\sigma_s^2 + \epsilon^2)(y^{\mathrm{T}}\alpha_{\perp})}{\gamma \|\alpha_{\perp}\|} - 1}.$$

Furthermore, 
$$U(y, v_{opt}) = U(h(Q)) + \frac{\|\alpha_{\perp}\|^2}{2\lambda\sigma_s^4} \frac{\gamma^2}{\sigma_s^2 + \epsilon^2}$$

First, note that if  $\gamma = 0$  then  $U(y, v_{opt}) = U(h(Q))$  and there is no marginal benefit of augmenting the base risk model with factor y. Thus, using an augmented risk model results in portfolios with better ex-post performance only if the

systematic covariance of the augmenting factor y and  $\alpha_{\perp}$  is nonzero; in other words, an augmenting factor is useful only if it can capture at least some portion of the systematic risk in  $\alpha_{\perp}$ . Second, it can be easily verified that if  $y = \frac{1}{\|\alpha_{\perp}\|} \alpha_{\perp}$  or  $y = \frac{1}{\|P\alpha_{\perp}\|} P\alpha_{\perp}$  then  $\nu = y^T Z \Lambda Z^T y$ , the true systematic variance of y. This statement, however, cannot be generalized for any arbitrary factor y.

Let  $\Phi(y) = \frac{\gamma^2}{\sigma_s^2 + \epsilon^2}$ . By Proposition 3, it follows that  $U(y, v_{opt}) = U(h(Q)) + \frac{\|\alpha_{\perp}\|^2}{2\lambda\sigma_s^4}\Phi(y)$  and our quest for an optimal augmenting factor y that maximizes the utility function  $U(y, v_{opt})$  reduces to solving the following problem,

max 
$$\Phi(y)$$
  
s.t.  $||y|| = 1$  (2)  
 $X^{T}y = 0.$ 

Before we present the solution to the above problem, note the following inequality which can be easily verified.

$$\Phi\left(\frac{1}{\|\alpha_{\perp}\|}\alpha_{\perp}\right) \le \Phi\left(\frac{1}{\|P\alpha_{\perp}\|}P\alpha_{\perp}\right). \quad (3)$$

The above inequality confirms the aforementioned conjecture, namely that  $P\alpha_{\perp}$  is a better augmenting factor than  $\alpha_{\perp}$ . Naturally, we are interested in knowing if there is a better factor than  $P\alpha_{\perp}$ . The proposition that follows settles this question.

**Proposition 4** The optimal solution to (2) is given by,  $y_{opt} = \frac{Z(I+\sigma^2\Lambda^{-1})^{-1}Z^{T}\alpha_{\perp}}{\|Z(I+\sigma^2\Lambda^{-1})^{-1}Z^{T}\alpha_{\perp}\|}$ . Furthermore, when the base risk model is augmented with  $y_{opt}$  using the optimal volatility value given in Proposition 3, the resulting optimal holdings are identical to those obtained by using the true risk model. Thus  $h(Q_{y_{opt}}) = h(Q_T)$ .

Several comments are in order. First, the above proposition not only gives a closed-form

expression for the optimal choice of the augmenting factor but also shows that using the corresponding factor, in fact, replicates the effect of using the true risk model. It is indeed remarkable that the overall impact of using the true risk model can be captured by a single augmenting factor.

Second, consider the case when there is exactly one missing factor, i.e., Z has exactly one column, say z. In this case, it can be shown that  $y_{opt} =$ z. Thus when exactly one factor is missing from the base risk model, it is optimal to augment the base risk model with the missing factor. Third, the results presented so far can be summarized as

$$U(\alpha_{\perp}) \le U(P\alpha_{\perp}) \le U(y_{opt}) = U(h(Q_T)),$$
(4)

where we have dropped the condition ||y|| = 1and defined  $U(y) = U(y, v_{opt})$ , where  $v_{opt}$  is the optimal value of v given in Proposition 3. The above inequality shows that our search for an augmenting factor and use of  $Q_y$  in portfolio optimization problems is not in vain. While we may not currently have the technology to compute  $ZZ^{T}\alpha_{\perp}$  without the explicit knowledge of Z, augmenting the base risk model with  $\alpha_{\perp}$  is a step in the right direction.

We started this section by mentioning the risk underestimation problem. After having established some interesting facts about augmenting factors, it is instructive to revisit the risk underestimation problem and understand whether the augmenting factors ( $\alpha_{\perp}$ ,  $P\alpha_{\perp}$ ,  $y_{opt}$ , etc.) which lead to superior ex-post performance also give rise to unbiased risk estimates.

Given an arbitrary augmenting factor y and its associated optimal volatility  $v_{opt}$  given by Proposition 3, y is said to be a *risk-unbiased* factor if the predicted risk of  $h(Q_y)$  is equal to its true risk, i.e.,

$$h(Q_y)^{\mathrm{T}}Q_yh(Q_y) = h(Q_y)^{\mathrm{T}}Q_Th(Q_y)$$

Thus, augmenting the base risk model with a riskunbiased factor completely eliminates the bias in risk prediction. The proposition that follows gives examples of risk-unbiased factors.

**Proposition 5**  $\alpha_{\perp}$ ,  $P\alpha_{\perp}$ , and  $y_{opt}$  are riskunbiased factors.

The above proposition has an important practical ramification apropos factor alignment problems that we discuss next. The fact that there is a multitude of risk-unbiased factors which give rise to optimal portfolios with widely varying ex-post performance (see Equation (4)) shows that risk underestimation is only a symptom of a much bigger and complex ailment, namely the factor alignment problem. Just as eliminating a symptom of a disease is not necessarily the same as curing the disease, simply circumventing the risk underestimation problem does not necessarily shield a PM from the ill-effects of the FAP. Many practitioners fail to acknowledge this aspect of portfolio optimization and resort to various kinds of ad hoc techniques that resolve the risk underestimation problem the symptom, without actually addressing the factor alignment problem-the disease. We examine one such ad hoc approach in detail and compare it with using an augmented risk model.

One of the widely used solutions to risk underestimation problems is to artificially modify the parameters of the strategy in anticipation of the bias in risk prediction. For instance, a PM can reduce the maximum allowable risk limit in the presence of a risk constraint or increase the risk aversion parameter when the risk/variance term is present in the objective function. Both of these approaches are equivalent to scaling the covariance matrix Q by a factor  $\tau \ge 1$ . Next we show that this *scaling approach* is outperformed by the AAF approach. For  $\tau > 0$ , let  $h(\tau)$  denote the optimal solution to the following problem,

$$\max_{h} \alpha^{\mathrm{T}} h - \frac{\tau \lambda}{2} h^{\mathrm{T}} Q h.$$

Let  $\tau^*$  denote the value of  $\tau$  that maximizes the ex-post utility of  $h(\tau)$  given by,

$$U(h(\tau)) = \alpha^{\mathrm{T}} h(\tau) - \frac{\lambda}{2} h(\tau)^{\mathrm{T}} Q_{T} h(\tau).$$

The proposition that follows shows that the expected ex-post utility function obtained by using the AAF is at least as good as the optimal utility function that can be obtained by varying the risk aversion parameter and using the base risk model.

#### **Proposition 6**

$$U(h(\tau^*)) \leq U(h(Q)) + \frac{\|\alpha_{\perp}\|^2}{2\lambda\sigma^4} \Phi\left(\frac{1}{\|\alpha_{\perp}\|}\alpha_{\perp}\right).$$

Furthermore, the above inequality holds strictly unless  $\alpha_{\perp} = 0$  or  $\alpha_X = 0$ .

Until now we have discussed and presented various results that show how augmented risk models can be used to derive MVO portfolios with better ex-post utilities. A related, and equally important, consideration is the effect of these models on the ex-post risk-return frontier. Does using an augmented risk model simply move the chosen portfolio on the risk-return frontier or does it actually push the frontier upward thereby allowing the PM to access previously inaccessible portfolios? The discussion that follows answers this question.

Given an arbitrary covariance matrix Q, one can define a risk–return frontier that is specified by the associated risk model. For instance, consider the following MVO problem parameterized by  $\lambda$ 

and Q,

$$\max_{h} \alpha^{\mathrm{T}} h - \frac{\lambda}{2} h^{\mathrm{T}} \mathcal{Q} h$$

Let  $h(Q, \lambda)$  denote the optimal solution to the above problem, and let  $R(Q, \lambda) = \alpha^{T}h$  and  $\sigma(Q, \lambda) = \sqrt{h^{T}Q_{T}h}$  denote the expected return and realized risk of h.<sup>6</sup> By choosing various values of the risk aversion parameter  $\lambda$ , we can derive a risk-return frontier associated with Q. An interesting question is, how does the risk return frontier derived from the base risk model Q compare with the one derived using an augmented risk model  $Q_y$ . The theorem that follows answers this question.

**Theorem 7** (Pushing Frontier Theorem) If  $(\sigma_f, r_f)$  lies on the risk frontier associated with the base risk model then there exists a point  $(\sigma_f, \hat{r}_f)$  on the ex-post risk-return frontier associated with  $Q_y$  where  $y = \frac{1}{\|\alpha_{\perp}\|} \alpha_{\perp}$  such that  $r_f \leq \hat{r}_f$ . Furthermore,  $r_f < \hat{r}_f$  unless  $\alpha_X = 0$  or  $\alpha_{\perp} = 0$ .

*Proof.* The existence of a point  $(\sigma_f, \hat{r}_f)$  on the ex-post risk-return frontier associated with  $Q_y$  which has an ex-post risk of  $\sigma_f$  is straightforward. We next show that the ex-post return  $\hat{r}_f$  of the associated portfolio is greater than or equal to  $r_f$ .

Let  $\hat{h}$  denote the optimal portfolio associated with  $(\sigma_f, \hat{r}_f)$  and let  $\hat{\lambda}$  denote the risk aversion parameter that gives rise to  $\hat{h}$  in conjunction with the risk model  $Q_y$ . Similarly, let h denote the optimal portfolio associated with  $(\sigma_f, r_f)$ . By Propositions 3 and 6 it follows that the ex-post utility function of  $\hat{h}$  is greater than or equal to the ex-post utility function of h evaluated at the risk aversion parameter  $\hat{\lambda}$ . Consequently,

$$U(h) \le U(\hat{h}),$$
$$\alpha^{\mathrm{T}}h - \frac{\hat{\lambda}}{2}h^{\mathrm{T}}Q_{T}h \le \alpha^{\mathrm{T}}\hat{h} - \frac{\hat{\lambda}}{2}\hat{h}^{\mathrm{T}}Q_{T}\hat{h},$$

$$lpha^{\mathrm{T}}h - \frac{\hat{\lambda}}{2}\sigma_{f}^{2} \le lpha^{\mathrm{T}}\hat{h} - \frac{\hat{\lambda}}{2}\sigma_{f}^{2}$$
 $lpha^{\mathrm{T}}h \le lpha^{\mathrm{T}}\hat{h}.$ 

Next, we move our attention to constrained MVO models.

# 5 Augmented risk models: Constrained MVO

In this section we extend the results of the previous section to constrained MVO problems of the form,

$$\max_{h} \quad \alpha^{\mathrm{T}}h - \frac{\lambda}{2}h^{\mathrm{T}}Qh$$
  
s.t. 
$$Ah = b.$$
 (5)

Although we focus our discussion on equality constrained problems, all of the results presented in this section can be easily generalized to MVO problems with arbitrary convex constraints.

We would like to remind the reader that the gradients of constraints as represented by rows of the matrix *A* can themselves be considered as factors. Thus when we talk about spanned (orthogonal) component of a constraint, we refer to the spanned (orthogonal) component of the corresponding constraint gradient. For instance, if one of the constraints enforces beta neutrality of the portfolio, then the corresponding constraint gradient is simply the well-known market sensitivity or beta factor. Indeed, not all constraint gradients have such simple interpretation. For instance, the gradient associated with a turnover constraint cannot usually be associated with a known style or industry factor.

The key concept that allows us to make a transition from the unconstrained to constrained setting is that of *implied alpha*. Recall that implied alpha,  $\tilde{\alpha}$ , is the expected return vector which when used in an unconstrained setting gives rise to the same optimal holdings as derived from Equation (5). In other words, if  $h^*$  is the optimal solution to Equation (5) then  $h^*$  is also an optimal solution to,

$$\max_{h} \quad \tilde{\alpha}^{\mathrm{T}}h - \frac{\lambda}{2}h^{\mathrm{T}}Qh.$$

Clearly  $\tilde{\alpha} = \lambda Q h^*$ . Note that the implied alpha depends on the optimal portfolio, and hence by implication, on  $\alpha$  and Q. Thus if an augmented risk model, say  $Q_y$ , is used in Equation (5) then the implied alpha is given by  $\tilde{\alpha} = \lambda Q_y h^*$ .

To give the reader some appreciation of the influence of constraints in determining implied alpha we give an alternative derivation of  $\tilde{\alpha}$ . Recall that any optimal solution to Equation (5) satisfies the following first-order optimality conditions:

$$\alpha - \lambda Qh - A^{\mathrm{T}}u = 0$$
$$Ah = b,$$

where *u* denotes the optimal Lagrange multipliers associated with the equality constraints. From the above equations it follows that  $\tilde{\alpha} = \alpha - A^T u$ . In other words, the implied alpha is obtained by tilting the alpha in the direction of constraints as determined by the optimal Lagrange multipliers. The above expression for  $\tilde{\alpha}$  has an important practical implication that we discuss next.

Consider the case when  $\alpha_{\perp} = 0$ , i.e., the expected return vector is spanned by the base risk factors. In the unconstrained setting this would imply the absence of FAP. Factor alignment could still be an issue in the constrained setting, however, if some of the constraints are not spanned by the regular risk factors (X) and have nonzero Lagrange multipliers associated with them; in this case,  $\tilde{\alpha}_{\perp} \neq 0$ and the optimizer will tend to overload on  $\tilde{\alpha}_{\perp}$ . To conclude, it is the implied alpha,  $\tilde{\alpha}$ , and not alpha,  $\alpha$ , that determines the key characteristics of the optimal portfolio in a constrained setting, and hence serves as a watershed between constrained and unconstrained MVO problems. We also like to emphasize that many of the constraints in real-world quant strategies that are binding at the optimal portfolios do have significant orthogonal components. For instance, most quant strategies impose bounds on maximum asset holdings for each asset, and the corresponding constraint is usually never spanned by the regular risk factors; to see this note that it is extremely unlikely that the base risk model will have a factor with exposure to only one asset. The same argument applies to long-only constraints, turnover constraints, factor exposure constraints derived from secondary risk models, etc.

Having established the vital role of implied alpha in determining the optimal holdings, we now proceed to solution techniques to FAP in a constrained setting. We limit our discussion to the augmenting factor  $y = \frac{1}{\|\tilde{\alpha}_{\perp}\|} \tilde{\alpha}_{\perp}$ ; the analysis of other factors introduced in the previous section is more complicated and goes beyond the scope of this paper. We also assume that the volatility  $\nu$  of *y* is prespecified and fixed. Thus we need to solve the following MVO problem,

$$\max_{h} \quad \alpha^{\mathrm{T}}h - \frac{\lambda}{2}(h^{\mathrm{T}}Qh + \nu(y^{\mathrm{T}}h)^{2})$$
  
s.t. 
$$Ah = b \qquad (6)$$
$$y = \frac{1}{\|\tilde{\alpha}_{\perp}\|}\tilde{\alpha}_{\perp}.$$

Note that Equation (6) is an equilibrium problem since  $\tilde{\alpha}$  depends on the Lagrange multipliers, which in turn are determined by the optimality conditions. An iterative approach to solving Equation (6) involves starting with an initial approximation of  $\tilde{\alpha}$ , say  $\tilde{\alpha}_0$ , solving the approximation of Equation (6) obtained by fixing  $\tilde{\alpha}$  to  $\tilde{\alpha}_0$ , extracting the implied alpha of the resulting problem, updating the approximation of  $\tilde{\alpha}$  and repeating the entire process until it converges. As intuitive as this iterative approach may seem, it has two shortcomings. First, it is not immediately clear if the procedure converges. Second, even if it does converge it may take an excessive

THIRD QUARTER 2015

number of iterations to do so. The proposition that follows establishes an important property of feasible solutions to Equation (6) which is used subsequently to solve the equilibrium problem as a single non-iterative convex optimization problem.

**Proposition 8** If  $(h, y, \tilde{\alpha})$  is a feasible solution to Equation (6) then  $\frac{1}{\|h_{\perp}\|}h_{\perp} = \frac{1}{\|\tilde{\alpha}_{\perp}\|}\tilde{\alpha}_{\perp}$ .

By Proposition 8 we get the following equivalent formulation of Equation (6).

$$\max_{h} \quad \alpha^{\mathrm{T}}h - \frac{\lambda}{2}(h^{\mathrm{T}}Qh + \nu(\|h_{\perp}\|)^{2})$$
  
s.t. 
$$Ah = b.$$
 (7)

It can be easily shown that Equation (7) is a convex optimization problem and hence can be solved efficiently. We refer to Equation (7) as the AAF approach to FAP in the rest of this paper.

## 6 Volatility calibration model

Until now we have assumed that  $\nu$  is already prespecified or is chosen by the PM based on the IC and systematic risk inherent in  $\tilde{\alpha}_{\perp}$ . In this section, we discuss a simple model that can be used for dynamic calibration of  $\nu$ . Notably, the emphasis of this section is not on determining the best volatility model for  $\nu$ . Instead, we want to develop a model which is simple, intuitive, implementable, and which can be used to attest the theoretical results discussed earlier.

We used the following volatility model in our experiments.

$$\nu_t = \xi \nu_{t-1} + (1 - \xi) r_{\nu}^2.$$
(8)

 $v_t$  denotes the value of v to be used in the *t*th period;  $r_y$  denotes the factor return of the augmenting factor  $y = \frac{1}{\|\tilde{\alpha}_{\perp}\|} \tilde{\alpha}_{\perp}$  determined using the augmented regression apparatus discussed in Section 2;  $\xi$  is the standard parameter used associated with Exponentially Weighted Moving Average

(EWMA) models such as Equation (8) (see Hull, 2009). In our experiments we choose  $\xi = 0.94$ ; in Section 7 we give empirical results to illustrate the sensitivity of the optimal portfolios to the choice of the  $\xi$  parameter.

As for the choice of the initial volatility  $v_0$  of y, there are two competing alternatives. First, we can run daily augmented regressions prior to the first date of rebalancing and use the variance estimator which is to estimate  $\Omega$  to determine  $\nu_0$ . Alternatively, we can start with a fixed choice of the initial estimate  $v_0$ , and let the exponential weighing mechanism adjust the estimate based on the factor returns determined using the augmented regressions. We believe that the first option is the best alternative, and comes close to building a custom risk model that uses regular factors (X) and the augmenting factor y. However, it is also more cumbersome to implement and requires access to fairly sophisticated volatility estimators. For the sake of simplicity, we chose the second option and used an initial value of  $\sqrt{\nu_0} = 30\%$  in our experiments. We illustrate the impact of using different values of  $v_0$  on the optimal portfolios in the section on computational results.

# 7 Computational results

In this section, we emphasize three important points. First, we give extensive empirical results that corroborate the theoretical findings discussed in the preceding sections. Second, we introduce the notion of "frontier spreads" to capture improvements in risk-adjusted returns that result due to the application of augmented risk models. Subsequently, we study the impact of various strategic parameters (turnover limits, asset bounds, etc.) on the frontier spreads, and also seek to identify market regimes where using an augmented risk model is most likely to yield significant improvements. Finally, we present computational results with a wide variety of alpha models to attest the robustness of the proposed approach. Throughout this section we assume that the AAF  $y = \frac{1}{\|\tilde{\alpha}_{\perp}\|} \tilde{\alpha}_{\perp}$  is used as the augmenting factor, and refer to the resulting methodology for portfolio construction as the AAF approach.

We reran the backtests discussed in Section 2 using the augmented risk models. Figures 7 and 8 report the key results. As it is evident from Figure 7, using the AAF approach completely eliminates the bias in risk prediction yielding unbiased ex-ante risk estimates (see Proposition 5). Furthermore, using the AAF approach also improved the risk-adjusted returns thus illustrating the "Pushing Frontier" phenomenon (see Proposition 7). For a given level of realized active risk, the difference in annualized active returns of portfolios generated using the AAF approach and the BRM is referred to as the *frontier spread*. Figure 9 shows the frontier spreads for various levels of realized active risk.

Notably, the frontier spread curve has a humped shape. In other words, the frontier spreads tend to be compressed at very low and very high active risk levels, and attain their maximum at an intermediate level. At very low active risk levels, the optimized portfolio is tightly tied to the benchmark thereby mitigating the adverse effects of FAP, and hence the ability of the AAF approach to produce significant improvements. At very high active risk levels, auxiliary constraints such as turnover constraint, asset bounds constraints, etc., start to play a more dominant role rendering a secondary status to the active risk constraint.



Figure 7 Bias statistic (active risk): AAF versus BRM.



Figure 8 Realized risk–return frontier.



Figure 9 Frontier spreads.

Consequently, the risk model has a subdued effect on the optimal holdings at high active risk levels reducing the severity of FAP, and by extension, the marginal improvements that result by application of the AAF approach. To summarize, the tightness or looseness of auxiliary constraints in the strategy can have a significant impact on the amount of improvement that can be garnered by applying the AAF approach. We conducted some additional experiments to understand this phenomenon.

Figure 10 reports the frontier spread curve for three levels of turnover limits. Note that frontier spreads compress significantly when the turnover limit is reduced to 8%; this is to be expected since at very low levels of turnover budget, the active risk constraint plays a diminished role in determining the optimal portfolio holdings, thereby reducing the impact of FAP. Figure 11 reports the same results when the asset bounds are varied between 13%. Once again, using loose constraints gives rise to maximum improvements that can be obtained via the AAF approach.

Note that the primary foundation of the theoretical results presented earlier is the unaccounted systematic risk in the orthogonal component of implied alpha which goes undetected during the process of portfolio construction. Indeed



Figure 10 Frontier spreads: Varying turnover budget.



Figure 11 Frontier spreads: Varying asset bounds.

Proposition 2 shows that the improvement in the utility function that result by application of augmented risk models increases monotonically with the unaccounted systematic variance of the orthogonal component. Next we present empirical results to highlight this phenomenon. Figure 12 reports the return differential between portfolios generated using the AAF approach and the BRM, computed over a rolling 24 months window; the portfolios were chosen so as to have a realized active risk of 3%. For the sake of discussion, we also report the volatility of factor returns associated with  $\frac{1}{\|\tilde{\alpha}_{\perp}\|}\tilde{\alpha}_{\perp}$ , computed using a rolling 24 months window. Note that the two time series shown in Figure 12 are highly correlated with each other; in fact, over the time period depicted in the figure the correlation between these two time series was 84%. Figure 13 reports the same information using a scatter plot. As evident from these figures, the AAF approach tends to produce maximum improvements in periods when the unaccounted systematic risk associated with  $\tilde{\alpha}_{\perp}$  is at its highest level.

Next we move our attention to the volatility calibration model that is used to determine  $\nu$ . Recall that the mentioned model has two key parameters, namely,  $\xi$  and the initial variance estimate  $\nu_0$ . In all of the experiments discussed until now, we used  $\xi = 0.94$  and  $\nu_0 = 0.09$ ; next we examine



Figure 12 Performance differential between portfolios generated using the AAF approach and BRM.



Performance differential vs Latent Volatility (Implied Alpha)

**Figure 13** Performance differential between portfolios generated using the AAF approach and BRM (scatter plot).

the impact of changing these parameters on the resulting optimal portfolios. We refer to  $\sqrt{\nu}$  as the volatility of the AAF.

Figure 14 depicts the evolution of the volatility of AAF through time for various choice of the initial volatility. As it is evident from the figure, volatilities of the AAF computed using different initial choices soon converge to the same time series by virtue of exponentially decaying impact of  $v_0$  in an EWMA model. To further understand the impact of the choice of  $v_0$ , we reran the backtests for various values of  $v_0$ ; Figures 15 and 16 report the key results. For each choice of  $v_0$ , we chose a point on the realized risk-return frontier that had a realized active risk of 3%, and report the statistics associated with the corresponding portfolio. Figure 15 reports the bias statistic for portfolios generated using the AAF approach for values of  $\sqrt{\nu_0}$  chosen from {10%, 20%, ..., 100%}. As evident from the figure, regardless of the choice of the initial volatility estimate, the AAF approach produced unbiased risk estimates. Figure 16 reports the realized IR for the same set of portfolios; for the sake of comparison we also report the IR of optimal portfolios generated using the BRM. The AAF approach produces significant improvements in IR regardless of the initial choice of  $\sqrt{\nu_0}$ . Notably, the realized IR increases monotonically with the



Figure 14 Time series of AAF volatility computing using different choices of initial volatility estimate.



**Figure 15** Bias statistics of portfolios generated using the AAF approach using different choices of initial volatility estimate.



Figure 16 Realized IR of portfolios generated using the AAF approach using different choices of initial volatility estimate.

initial volatility estimate  $\nu_0$ . We believe that this is the result of low or possibly negative IC of the orthogonal component as compared to the spanned component. A detailed analysis of this phenomenon goes beyond the scope of this paper. Next we move our focus to the  $\xi$  parameter.

Similar to the previous experiment, we reran the backtests for various choices of the  $\xi$  parameter. Figure 17 reports the movement of the frontier spread curve for various choices of  $\xi$  parameter. Note that using a value of  $\xi$  in [0.8, 0.99] interval yields portfolios that have similar characteristics. However, using a value of  $\xi \leq 0.7$  significantly compresses the frontier spreads especially at

higher active risk levels. Essentially, using a small value of  $\xi$  significantly increases the weight that is assigned to the most recent observation, thereby increasing the estimation error in the volatility calibration model, thereby adversely affecting the quality of the augmented risk model.

We conclude this section by reporting computational results using a larger class of alpha models; all of these alpha models were constructed using the constituent factors in the USER model (see Guerard *et al.*, 2012b); Table 1 reports the key results. The first four alpha models were obtained by taking equal weighted combination of a valuation factor (B/P, E/P, C/P, and S/P) in the USER



**Figure 17** Impact of modifying the  $\xi$  parameter in the EWMA volatility model on frontier spreads.

model and the medium-term momentum (MTM) factor in the fundamental BRM used in Section 2. The next four alpha models used the same valuation factors but employed the momentum factors (RBP, REP, RCP, and RSP) in the USER model in lieu of the MTM factor. The last four risk models were obtained by adding an additional factor (CTEF) along with the valuation and momentum factors. The results are reported for portfolios that have a realized active risk of 3%.

**Table 1** Computational results to illustrate robust-ness of the AAF methodology.

	Bias statistic		
Alpha model	BRM	AAF	Frontier spread
BP+MTM	1.27	1.00	0.75%
EP+MTM	1.18	1.00	-0.02%
CP+MTM	1.10	0.93	0.31%
SP+MTM	1.20	0.96	0.64%
BP+RBP	1.61	1.22	0.49%
EP+REP	1.11	0.90	0.46%
CP+RCP	1.43	1.05	0.19%
SP+RSP	1.21	0.91	-0.01%
BP+RBP+CTEF	1.52	1.18	0.37%
EP+REP+CTEF	1.17	0.91	0.10%
CP+RCP+CTEF	1.33	1.02	0.18%
SP+RSP+CTEF	1.29	0.96	-0.01%

The AAF approach produced unbiased risk estimates for 11 out of 12 models: the 95% confidence interval for the bias statistics associated with this backtest is (0.87, 1.13). The only exception is the BP + RBP alpha model wherein the risk estimates produced using the AAF approach have a small downward bias; even in this case, the bias in risk prediction is reduced significantly from 1.61 to 1.22. Furthermore, the AAF approach produced positive frontier spreads, i.e., improvements in risk adjusted returns for 9 out of 12 cases; the worst frontier spread in the remaining three cases was -0.02%. Overall these results confirm the robustness of the AAF approach, and provide strong empirical evidence to support the theoretical results presented in the previous sections.

#### 8 Conclusion

The MVO approach to portfolio construction has a history spanning more than six decades. Despite coming under criticism from both practitioners and academics, it has survived the test of time and has become an industry standard for quantitative approaches to equity portfolio management. Nevertheless, the ambitious goal of constructing a truly optimal portfolio, as originally envisaged by Markowitz, remains a daunting challenge primarily due to the disparity between the ex-ante and ex-post characteristics of optimized portfolios. In this paper we set out to bridge this gap by focusing on specification error, i.e., existence of systematic risk factors which are missing from the risk model despite having significant overlap with the optimized portfolios.

We argued that exposure to such factors creates unintended and undesirable systematic bets which eventually compromise the efficiency of MVO portfolios. Based on a detailed analytical investigation, we demonstrated the usefulness of augmented risk models in addressing this issue. Our results indicate that augmenting the base risk model with an appropriate augmenting factor not only remedies the risk underestimation problem but also improves risk-adjusted returns. We presented extensive computational results to corroborate our findings.

Among other things, these results suggest strong synergistic advantages of integrating alpha and risk research processes. In other words, we need to abandon the "one-size-fits-all" approach to risk management and take a more nuanced approach that is sensitive to the specific requirements of a PM. Ultimately, the primary responsibility of a risk model is to capture all undiversifiable (i.e., systematic) sources of risk that are relevant to a given investment process. A risk model that is constructed in a manner which is agnostic to the very factors that the PM is betting on, cannot be expected to accomplish that goal. We believe that augmented risk models partly accomplish this goal, and should act as precursor to fully customized risk models that shed the artificial barrier between alpha and risk research, and take a holistic view of the investment process.

#### Notes

<sup>1</sup> GICS is a registered trademark of McGraw-Hill and MSCI Inc. Russell 3000 is a registered trademark of Russell Investments.

- <sup>2</sup> We used Axioma's medium-horizon fundamental risk model (US2AxiomaMH) as the choice of base risk model (BRM).
- <sup>3</sup> Since the cross-sectional factor exposures of different factors can have varying norms, we normalize the factors so as to have a  $l_2$ -norm of 1.0, and modify the factor variances accordingly. In other words we scale each factor f so as to satisfy ||f|| = 1.0.
- <sup>4</sup> We limit our discussion to cross-sectional risk models. The extension of these results to risk models constructed using alternative methodologies (for instance, timeseries regressions) is technical and goes beyond the scope of this paper.
- <sup>5</sup> Here, we used Axioma's medium-horizon statistical risk model (US2AxiomaMH-S) as our choice of statistical model.
- <sup>6</sup> We assume that the true risk model gives the realized risk of the portfolio; in other words, the realized risk is measured over a sufficiently long time horizon to eliminate the effects of estimation error.

#### References

- Ceria, S., Saxena, A. and Stubbs, R. A. (2012). "Factor Alignment Problems and Quantitative Portfolio Management," *The Journal of Portfolio Management* **38**(2), 29–43.
- Greene, W. H. (2002). *Econometric Analysis*. Prentice Hall; 5th edition (August 22, 2002).
- Guerard Jr, J., Krauklis, E. and Kumar, M. (2012a). "Further Analysis of Efficient Portfolios with the User Data," *Journal of Investing* **21**(1), 81–88.
- Guerard Jr, J., Xu, G. and Gültekin, M. (2012b). "Investing with Momentum: The Past, Present, and Future," *Journal of Investing* **21**(1), 68–80.
- Hull, J. (2009). *Options, Futures, & Other Derivatives*. Prentice Hall; 7th edition.
- Lee, J.-H. and Stefek, D. (2008). "Do Risk Factors Eat Alphas," *The Journal of Portfolio Management* **34**(4), 12–25.
- Saxena, A. and Stubbs, R. A. (2012). "An Empirical Case Study of Factor Alignment Problems Using the User Model," *Journal of Investing* 21(1), 25–43.
- Saxena, A. and Stubbs, R. A. (2013). "Alpha Alignment Factor: A Solution to the Underestimation of Risk for Optimized Active Portfolios," *The Journal of Risk* **15**(3), 3–37.

*Keywords*: Factor models; factor alignment problems (FAP); mean variance optimization (MVO)