Factor Alignment Problems and Quantitative Portfolio Management

Sebastián Ceria, Anureet Saxena, and Robert A. Stubbs
Factor Alignment Problems and Quantitative Portfolio Management

SEBASTIÁN CERIA, ANUREET SAXENA, AND ROBERT A. STUBBS

The relationship between finance practitioners and optimization techniques can be described as troublesome at best. This is evident from the investment community’s reaction to Markowitz’s mean-variance optimization (MVO) approach, proposed almost 50 years ago. Markowitz’s [1952] seminal work brought together the concepts of risk, return, risk aversion, and diversification—the holy grail of modern portfolio theory—under a unifying framework of an optimization model. His discovery, however, divided the investment community into three distinct camps.

The first camp rejected the very hypothesis of using quantitative techniques to solve portfolio construction problems. The second camp approached his discovery with cautious optimism, analyzed the portfolios constructed on the basis of his recommendations, and soon realized that the resulting portfolios were unintuitive, unwieldy, and thus impractical. Many members of this group joined the members of the first camp and became ardent critics of the MVO approach. The third camp, composed of both academics and practitioners, decided to perform a postmortem analysis of the MVO portfolios, attempted to understand the reasons for the deviation of ex post performance from ex ante targets, and used their analysis to suggest enhancements to Markowitz’s original approach that made it possible to overcome many of these deficiencies.

One of the fundamental contributions of the third camp was the development of linear factor models to capture the sources of systematic risk and characterize the key drivers of excess returns. Indeed, over the last couple of decades, factor models have not only become an industry standard, but also the reason why many practitioners in the first two camps are re-evaluating the notion of using quantitative techniques as part of their portfolio management process.

It is worth emphasizing that factor models that are used for forecasting expected returns have different mandates from the models used to forecast risk. These differences affect the choice of specific factors that are used in the construction of factor models. While predicting expected returns is exclusively a forward-looking activity, risk prediction also focuses on explaining cross-sectional variability of the returns process, mostly by using historical data. Phrased in the jargon of statistics, the first moment of the equity returns process drives expected return modelers, while the second moment is the focus of risk modelers. These differences in the ultimate goals inevitably introduce a certain “misalignment” between the factors used to forecast expected return and risk.

While expected return and risk models are indispensable components of any active strategy, there is a third component, namely, the set of constraints used to build a portfolio.
Constraints play an important role in determining the composition of the optimal portfolio. Most real-life quantitative strategies have constraints that model desirable characteristics of the optimal portfolio. Although some of these constraints may be mandatory, for example, a client’s reluctance to invest on ethical grounds in stocks that benefit from alcohol, tobacco, or gambling activities, other constraints are the result of best practices in practical portfolio management. Examples of the latter type of constraints include industry, sector, and asset exposure constraints, turnover constraints, tax considerations, and so forth. These constraints play an important role in affecting the portfolio construction process, and their interaction with the expected return and risk model constituents naturally accrues significance.

To summarize, quantitative equity portfolio construction entails complex interaction between factors used for forecasting expected returns, risk, and the constraints. Problems that arise due to mutual discrepancies among these three entities are collectively referred to as factor alignment problems (FAP) and constitute the emphasis of this article. Our key contributions are summarized as follows:

1. The differences in the approaches that are used to build expected returns forecasts and risk models manifest themselves as misalignment between the alpha and risk factors. The presence of constraints has an exacerbating effect on the misalignment problem and introduces another layer of complexity.

2. Using an optimization tool to construct the optimal holdings has the unintended effect of magnifying the aforementioned sources of misalignment. The optimizer underestimates the systematic risk of the portion of the expected returns that is not aligned with the risk model. Consequently, it overloads the portion of the expected returns that is uncorrelated with all the user risk factors.

3. Our empirical results on a test-bed of real-life active portfolios based on client data clearly refute the validity of the assumption that the portion of alpha that is uncorrelated with all the risk factors has no systematic risk, and suggest the existence of systematic risk factors that are missing from the risk model.

4. We propose augmenting the risk model with an additional auxiliary factor to account for the effect of the missing risk factors in the risk model. The augmenting factor is constructed dynamically and takes a holistic view of the portfolio construction process involving the alpha model, the risk model, and the constraints. We provide analytical evidence to attest the effectiveness of the proposed approach.

5. Alternatively, the risk model can be augmented by adding the factors that are used to compute expected returns, and which are not represented in the risk model. The addition of these factors will provide full alignment between the risk model and the expected returns, but not necessarily handle any misalignment issues due to the use of constraints. We briefly discuss the idea of recalibrating the user risk model with emphasis on eliminating the root cause of FAP.

MISALIGNMENT PROBLEMS: SOURCES

Quantitative strategies are typically based on three key components, namely, expected returns (or alphas), a risk model, and the constraints. Each one of these constituents has a distinct role in the portfolio management process, and hence by implication, a different emphasis. The construction of expected returns, for instance, has a forecasting emphasis, and the merit of any such model is judged by its ability to predict excess returns. The risk model, however, is geared toward explaining cross-sectional variability in the historical and predicted returns. The efficacy of a risk model is judged by its ability to capture systematic risk factors and the correlation structure between their respective factor returns. To cast it in the jargon of statistics, both the expected returns and the risk model concern the stochastic nature of the returns process with the former focused on the first moment and the latter concerned with the second moment. This disparity in their respective objectives naturally affects their choice of factors thereby introducing misalignment. In this section, we give several examples of this inherent misalignment. The following section discusses how these sources of misalignment, no matter how insignificant they may appear, are magnified within an optimization framework and impact ex post performance. We refer the reader to Robinson et al. [2009] for the definition of accounting terms used in this article.

We start our discussion with one of the most fundamental attributes of an equity security, namely, its earning
potential. The earnings yield (E/P) is one of the most widely cited valuation ratios and receives wide coverage in both the academic and industry press. There are two key differences in the way earnings yields are used in the construction of expected returns and risk models.

First and foremost, being an accounting-based metric, E/P is subject to various kinds of adjustments as deemed necessary by its end-users. These adjustments can arise due to use of different accounting principles (U.S.
GAAP versus IFRS), different treatment of income sheet entries (non-recurring, unusual and infrequent items, and non-operating gains and losses) or different views on accrual accounting artifacts such as depreciation, amortization, and so on. These kinds of adjustments are necessary in the construction of valuation models used for estimating expected returns and add little, if any, value in the generation of risk models. The reason for this disparity stems from the different objectives of the risk and expected returns models. With its primary focus on explaining the cross-sectional variability of the returns process, a risk model benefits only marginally by incorporating the accounting nuances mentioned earlier. In other words, the marginal utility of improving ballpark estimates to get razor-sharp estimation of accounting entries is very small as far as the quality of the resulting risk model is concerned. Expected returns, however, are focused on exploiting return differentials between various assets and adjustments such as the ones mentioned earlier. These can be effective in differentiating assets.

For instance, consider the announcement by Dell to restate its earnings in the first quarter of 2011 and take a $100 million charge to settle the federal investigation of accounting fraud. While alpha hunters are likely to perform detailed accrual analysis of Dell’s I/B statements to assess the long-term consequences of this event and possibly revise their forward-looking earnings estimate, risk model providers are unlikely to process this event except for incorporating the effects that are already reflected in the current market price of Dell. To put it differently, this is an uninteresting problem from a risk model perspective because it is likely to affect only the specific return of Dell. Ironically, it is the same asset-specific nature of the event that makes it a remarkably attractive development from an expected return perspective. It offers active managers the unique opportunity to harvest truly asset-specific returns without taking excessive exposure to any systematic risk factor. In other words, expected return and risk modelers have different beliefs about the possible impact, or lack thereof, of various economic events on their respective mandates, and the misalignment between the alpha and risk factors is simply an inevitable manifestation of their diverse beliefs.

Second, expected return and risk model developers can at times take a completely different view on the issue of earnings potential. For instance, some alpha construction techniques use alternative valuation metrics, such as EBITDA/EV, FCFF/P, or FCFE/P, in lieu of E/P, and for good reasons. For instance, small-cap growth companies in the technology sector often continue to have negative income and positive free cash flows in their early years. The negative income is usually the result of amortization of capital investments made at the inception of the company. For such businesses, free-cash-flow-based valuation ratios are clearly better suited than earnings yield. Risk model providers, however, usually do not use the aforementioned alternatives to E/P. These different measurement choices of the same underlying fundamental metric, namely, earnings potential, leads to misalignment between the alpha and risk factors.

Another interesting source of misalignment arises from the use of the book-to-price (B/P) ratio. Roughly speaking, book value is the accounting profession’s estimate of the company’s value. Book value reflects what the company paid for the assets, except for intangible assets such as goodwill developed internally, but it includes goodwill of subsidiary companies acquired by purchase. This “cost basis” is then adjusted downward by depreciation and amortization in a highly stylized and rigid attempt to reflect the economic depreciation that actually befalls (most) assets. Off-balance-sheet items are ignored. Finally, the result is augmented by retained earnings. With book value reflecting such a melange, it is surprising that it has any explanatory power at all, but it does (Siegel [2003]). Its popularity can be gauged by the fact that it is the only factor used by S&P in differentiating value stocks from growth stocks in each market-capitalization segment.

Naturally, B/P is used extensively by all model developers. The spectrum of accounting issues discussed in the context of E/P also applies to B/P. The situation for the B/P factor, however, is far more complicated than E/P due to the presence of intangible assets, deferred taxes, and goodwill. Recall that goodwill arises when the purchase price of net assets acquired exceeds the fair market value of other net assets acquired. In recent
years, the growth of the service and knowledge sectors in the economy has changed the nature and extent of goodwill accounts created by M&A transactions. Firms increasingly derive their value from the ability to create and exploit intellectual rights and human resources rather than from the ownership of physical or financial assets. Compared to physical or financial assets, the estimate of intrinsic value of intellectual property and human resources is much more subjective, and essentially is a license to accounting imagination (Dharan [1997]). Adjustments arising due to goodwill itself can create significant misalignment between the B/P factors employed in the alpha and risk models.

As mentioned earlier, the choice of factors for constructing expected returns and risk models is governed by different objectives. For instance, risk model developers often use adjusted R², t-statistics, or other measures of statistical significance to determine the suitability of an additional risk factor. There is usually a trade-off between the accuracy of a risk model and the number of systematic risk factors that are employed. As a result of this trade-off, some factors that are used for estimating expected returns might not be considered during the risk-factor selection process, thereby further aggravating the misalignment problem. For instance, consider a scenario where a risk model provider has selected B/P and E/P as factors in the risk model and is evaluating the suitability of the S/P factor. It is possible that S/P has little marginal explanatory power beyond what is already captured by B/P and E/P, and hence, it will most likely not be included as part of the risk model. If such a risk model is used in conjunction with expected returns derived from the S/P factor, then the misalignment is inevitable.

Besides the misalignment that can occur because of the differences in the definitions of the fundamental factors, misalignment can also arise due to different definitions of technical factors such as short-term and medium-term momentum. These momentum-based factors have been extremely popular in the past 15 years, although they have come under attack since the 2008 financial crisis. It is worth noticing that no consensus exists on what constitutes an ideal momentum factor nor is there any firm theoretical guidance on how to choose one. Every model that forecasts expected returns relies on these factors, and misalignment with a similar factor used in risk model generation is commonplace (Lee and Stefek [2008]).

Until now, we have discussed misalignment issues between the expected return and risk factors. Next, we move our focus to the role of constraints in contributing to the misalignment problems. Consider the following constrained MVO model:

$$\max_{\alpha, \lambda} \alpha^T h - \frac{1}{2} \lambda h^T Qh$$

subject to

$$Ah \leq b$$

where $$\alpha$$ denotes the vector of expected returns, $$Q$$ denotes the asset-asset covariance matrix, $$\lambda$$ represents the risk aversion parameter, and $$Ah \leq b$$ represents the constraints that the portfolio is required to satisfy. While some of these constraints, such as sector/industry exposure constraints, are mandated by the client’s investment policy statement (IPS), others such as a turnover constraint are used to model the operational aspects of portfolio management. From the theory of convex optimization and first-order optimality conditions, we know that there exist $$\pi \geq 0$$ such that the optimal solution to the preceding problem is identical to the optimal solution to the following unconstrained MVO model:

$$\max (\alpha - A^T \pi)^T h - \frac{1}{2} \lambda h^T Qh$$

The vector $$\gamma = \alpha - A^T \pi$$ is referred to as implied alpha in the rest of this article. Note that implied alpha is derived by tilting the $$\alpha$$ in the direction of binding constraints. In other words, implied alpha captures the effect of constraints in determining the optimal holdings and acts as the de facto alpha for the constrained MVO model. As an immediate consequence, it follows that the extent to which implied alpha is spanned by the user risk factors will have a direct bearing on the composition of the optimal portfolio. We give two examples to illustrate this effect.

First, consider the effect of asset bound constraints, the most frequently occurring constraint in quantitative strategies. Most portfolio managers have explicit constraints that prohibit them from taking concentrated bets on a single asset. These constraints may result due to legal requirements or may represent the inherent nature of the strategy. An example of the legal requirement is the ERISA (Employment Retirement Income Security...
Act) provision that disallows more than 10% investment of the defined-benefit pension funds in a single asset. Similarly, strategies that closely track a given benchmark often impose tight bounds on individual active asset exposures. It is worth noticing that these constraints are usually binding for a large subset of the asset universe and are responsible for the deviation of the implied alpha from alpha. More importantly, the degree of deviation is determined dynamically by the first-order optimality conditions and can introduce misalignment between γ and risk factors even when the alphas are spanned by the risk factors. In other words, misalignment arising due to asset bound constraints is an unavoidable feature of constrained MVO models and cannot be eliminated even if the the alpha and risk factors are in complete consonance. We refer the reader to Jagannathan and Ma [2003] for further discussion on the impact of asset bound constraints on the composition of optimal portfolios.

Second, consider the effect of a constraint that limits the exposure of the portfolio to illiquid assets. There are several ways to model the liquidity of assets, and different choices made during the construction of a “liquidity” risk factor and a liquidity constraint can introduce misalignment in the model. For instance, the liquidity risk factor could be determined using the daily trading volume of the respective assets, while the coefficients used in the liquidity constraint could be obtained as a function of the average bid–ask spreads.

To summarize, misalignment is an integral feature of most quantitative strategies and arises due to the complex interaction between the factors used for computing expected returns, risk, and constraints. While a small amount of misalignment is innocuous in itself, when present in an optimization framework, it can lead to some unexpected consequences. The section that follows dwells on some of the undesirable effects of misalignment.

**MISALIGNMENT PROBLEMS: EFFECTS**

For ease of discussion, we focus on the following unconstrained MVO model:

$$\max \alpha^T h - \frac{\lambda}{2} h^T Q h$$

where $Q = X \Omega X^T + \sigma^2 I$ is the asset-asset covariance matrix; $X$ is the factor exposure matrix; $\Omega$ is the factor-factor covariance matrix; $\sigma^2$ is the asset-specific risk, which is assumed to be constant across the asset universe; and $\lambda$ is the risk aversion parameter. To simplify the notation, we assume that $\lambda = 2$ and $X^T X = I$. All the results presented in this section can be generalized to constrained MVO models by replacing alpha ($\alpha$) with implied alpha ($\gamma$).

Consider the following linear regression model that regresses $\alpha$ against factors in the user risk model represented by the matrix $X$,

$$\alpha = Xu + \alpha_\perp$$

$\alpha_\perp$ denotes the residuals in the regression model. Let $\alpha_\perp = Xu$ denote the portion of $\alpha$ that is explained by the user risk factors $X$. There is a subtle connection between the residuals of this regression model and the notion of misalignment discussed at length in the previous section. Specifically, if there is no misalignment, then $\alpha$ is completely explained by the risk factors resulting in vacuous residuals (i.e., $\alpha_\perp = 0$). But, if the alpha factors are not completely explained by the risk factors, then the coefficient of determination ($R^2$) of the preceding regression model captures the degree of misalignment between the two sets of factors; the higher the $R^2$, the smaller is the degree of misalignment. In view of this discussion, we defined the misalignment coefficient, $MC(\alpha)$, of alpha to be $MC(\alpha) = 1 - R^2$. The misalignment coefficient $MC(h)$ of a portfolio $h$ is defined similarly wherein the portfolio holdings $h$ are regressed against the risk factors, and the $R^2$ of the resulting regression model is used instead.

Borrowing terminology from linear algebra, we refer to $\alpha_\perp$ and $\alpha_\parallel$ as the spanned and orthogonal components of $\alpha$, respectively; $h_\perp$ and $h_\parallel$ are defined similarly. Next, we discuss the impact of using an optimizer on the relative MC of $\alpha$ and $h$.

First, consider the case when $\Omega = 0$ (i.e., there is no systematic component of risk). In this case, the optimal portfolio $h$ is given by $h = \frac{1}{\sigma^2} \alpha$. Thus, $h_\perp = \frac{1}{\sigma^2} \alpha_\parallel$, $h_\parallel = \frac{1}{\sigma^2} \alpha_\perp$, and $MC(h) = MC(\alpha)$. The optimal portfolio in this case is merely a reflection of the alpha, has the same relative composition of the spanned and orthogonal component, and hence has the same misalignment coefficient as $\alpha$. Phrased differently, it follows that in the absence of systematic risk factors, the optimizer is indifferent to the alignment between the alpha and risk factors, or lack thereof.
This situation undergoes a drastic change when $\Omega$ is different from zero, the more commonly occurring scenario. In this case, it can be shown that 

$$h_X = \frac{1}{\sigma_s^2} \alpha_X - \frac{1}{\sigma_u^2} XM^{-1} X' \alpha_X$$

$$h_\perp = \frac{1}{\sigma_s^2} \alpha_\perp$$

where $M = \sigma^2 \Omega^{-1} + I$. Note that while the orthogonal component, $h_\perp$, of the portfolio is identical to the case when $\Omega = 0$, the spanned component undergoes some kind of diminutive transformation. To understand this disparity between the spanned and orthogonal components, recall that the optimizer perceives no systematic risk in $h_\perp$, because $X' h_\perp = 0$. Thus, it exploits the systematic risk arbitrage between the spanned and orthogonal components thereby overweighting the orthogonal components relative to the spanned component. In fact, it can be easily shown that the misalignment coefficient of the optimal portfolio $h$ is always strictly greater than the misalignment coefficient of $\alpha$. In other words, the optimizer has a magnifying effect on the misalignment between the alpha and risk factors. To give the reader a better appreciation of the extent of this magnification effect, we conducted the following experiment.

We constructed a test-bed of 25 backtests derived from real alphas and strategies used by portfolio managers (PM). This test-bed of backtests comprises a fairly heterogeneous collection of active investment strategies involving long-only, long–short, short-extension, and dollar-neutral portfolios rebalanced either daily, monthly, or quarterly; the number of periods in the backtests ranges from 50 to 170. Many of the portfolios have some form of factor neutrality constraints to hedge their exposure to style or industry factors. Each backtest was run twice, once with a cross-sectional fundamental risk model and once with a statistical risk model built using asymptotic principal components.

Subsequently, for each backtest we computed the average misalignment coefficient of the alpha and optimal holdings, as reported in Exhibit 1. The consistently higher value of MC for the optimal holdings as compared to that of the alphas is not surprising, but what is surprising is the magnitude of the difference between the two values. For some of the backtests, the MC of the optimal holdings was 100% greater than that of the corresponding alphas. On average, the MC of the optimal holdings was roughly 43% higher than that of the alphas for backtests that used the fundamental risk model. The same metric for backtests that used the statistical risk model was 26%.

It is worth emphasizing that the misalignment coefficient is not a normative concept, instead, it is a descriptive statistic that captures the skewness in the composition of the optimal holdings or alphas. In other words, it cannot be used as the sole basis to determine the desirability, or lack thereof, of a given portfolio. MC becomes meaningful only when it is examined in light of the source of misalignment, and specifically the amount of residual systematic risk in the misalignment source that is not accounted for during the portfolio construction phase. We give two examples to illustrate this point.

---

**E X H I B I T 1**

**Misalignment Coefficient (MC)**

*Panel A: Fundamental Model*

*Panel B: Statistical Model*
First, consider a situation where the alpha model is based on the S/P factor, whereas the risk model does not have a S/P factor even though it includes the E/P and B/P risk factors. In this case, the orthogonal component of alpha is defined to be the residual portion of S/P that is not captured by the existing risk factors such as E/P and B/P. It is possible that the mentioned residual component has systematic risk albeit not as significant as that of E/P or B/P factors. Because the optimizer fails to perceive any systematic risk in $\alpha_\perp$, it is likely to take excessive exposure to $\alpha_\perp$, thereby increasing $MC(h)$ relative to $MC(\alpha)$. In this case, the increase in the misalignment coefficient is indeed undesirable since it signifies the presence of latent systematic risk in the portfolio that goes undetected during the portfolio construction phase.

Second, consider the example of Dell’s earnings restatement described earlier. Suppose an active manager of a large-cap growth fund uses TTM earnings yield, E/P, to define the alphas and also as a risk factor. Furthermore, in response to Dell’s earnings restatement, the manager decides to modify the exposure of her alphas to Dell but leaves the corresponding risk-factor exposures unchanged. Such a scenario is especially likely to happen if the risk model is not developed in-house but purchased from a third-party vendor, making it prohibitively expensive to re-calibrate the risk model after modifying the factor exposures. In this case, the orthogonal component of $\alpha$ is truly asset specific with exposure to exactly one asset, namely, Dell, and thus has insignificant systematic risk. Consequently, the effect of misalignment between the alpha and risk factors is innocuous and unlikely to damage the ex post performance of the optimal portfolio, the magnification of the MC notwithstanding.

Note that all of these arguments continue to hold true in the presence of constraints, provided that alpha is replaced by implied alpha. While from a conceptual standpoint this may appear to be a minor modification, it introduces a tremendous amount of complexity into the analysis. To see this, recall that implied alpha ($\gamma$) is not a “static” entity (i.e., implied alpha cannot be determined a priori independently of the constraints in the model). Furthermore, $\gamma$ depends on the optimal portfolio and parameters of the first-order optimality conditions as discussed in the previous section. It is important to note that most quantitative research into the properties of good forecasts of expected returns concentrates on alpha, and not necessarily implied alpha. In fact, no quantitative researcher we know of does research into the factor structure of implied alpha.

Next, we discuss some of the unintended consequences of the factor alignment problem. One of the most conspicuous effects of FAP is the reduced accuracy of risk forecasts (Saxena and Stubbs [2010a]). The deterioration in the quality of risk prediction is particularly acute when the orthogonal component of alpha has significant systematic risk. The mentioned case describes the situation where the fund manager is trying to reap the risk premium of a systematic risk factor that is not included in the risk model. For instance, consider using alphas derived from Pástor and Stambaugh’s [2003] liquidity factor along with a risk model defined using the Fama–French [1992, 1995, 2008] risk factors. The returns to the optimal portfolio from such a combination of alpha and risk model essentially represent the risk premium of the liquidity risk factor. Since the Fama–French model does not have a liquidity risk factor, the ex ante risk prediction of the optimal portfolio will inevitably suffer from a downward bias.

Factor misalignment also introduces a disconnect between the techniques employed in the alpha construction phase and how those alphas eventually get transformed into an optimal portfolio. For the sake of discussion, consider an active PM who is using a weighted combination of two alpha signals to construct her alphas. One of these signals, say, $\alpha_{G}$, is identical to the growth factor present in her risk model, while the second signal, say, $\alpha_{S}$, is uncorrelated with all the risk factors in her risk model. Based on inputs from her research team members, the PM concludes that $\alpha_{G}$ has roughly three times more strength than $\alpha_{S}$ and decides to use the following formula to define her alphas:

$$\alpha = \frac{3}{4}\alpha_{G} + \frac{1}{4}\alpha_{S}$$

Note that the PM’s intention is to tilt her holdings toward the growth factor, thereby introducing a style bias into her active portfolio. Based on the results presented earlier, we know that the optimizer views the two components of alpha, $\alpha_{G}$ and $\alpha_{S}$, differently; it favors $\alpha_{S}$ over $\alpha_{G}$ for lack of exposure of $\alpha_{G}$ to risk factors in the PM’s risk model. Thus, the optimal portfolio is very likely to have much higher exposure to $\alpha_{S}$ than $\alpha_{G}$, thereby nullifying the original intention of the PM.
To summarize, the optimizer cherry picks the aspects of the model of expected returns that it deems desirable when gauged on the yardstick of marginal contribution to systematic risk. In its endeavor to do so, the optimizer pushes most of the optimal holdings into the vector space, which is orthogonal to one defined by the user risk factors. In other words, the optimizer is taking an aggressive bet on the assumption that being uncorrelated with all the user risk factors is tantamount to lacking systematic risk altogether. As we discuss in the following section, this assumption turns out to be the Achilles’ heel of optimized active portfolios.

MISALIGNMENT PROBLEMS: ANALYSIS

In this section, we present an empirical analysis of the factor alignment problem. Specifically, we show that, despite being uncorrelated with all the risk factors in the user risk model, \( h \perp \) has significant systematic risk, and its cross-sectional explanatory power is comparable to that of an average fundamental or statistical risk factor. Next, we describe the experimental setup used to establish the stated hypothesis.

For each backtest in our test-bed, we performed a weighted cross-sectional regression analysis on asset returns using all the regular risk factors in the risk model and \( h \perp \). We used the optimal portfolio holdings to weight the regression model in each period, see Section 4 of Saxena and Stubbs [2010a] for more details. Subsequently, we performed a time series analysis on the factor returns of \( h \perp \) to determine the systematic risk in \( h \perp \) and also the statistical significance of \( h \perp \) viewed as an additional risk factor.

Exhibit 2 reports the root mean square (RMS) \( t \)-statistic corresponding to \( h \perp \) in the experiment. We also report the median RMS \( t \)-statistic corresponding to regular factors in the respective risk model. Note that the RMS \( t \)-statistic value for \( h \perp \) is close to the 95% threshold value of 1.96 for several backtests and often exceeds the median value corresponding to regular risk factors. This implies that not only is \( h \perp \) a statistically significant risk factor, its explanatory power is comparable to that of an average risk factor in the fundamental and statistical risk models. It is interesting to note that \( h \perp \) compares favorably with such a carefully chosen set of risk factors, thereby corroborating its significance as a systematic risk factor.

Exhibit 3 reports the annualized standard deviation of factor returns of \( \frac{1}{n} \langle h \rangle \), which is a measure of the systematic risk in \( h \perp \). Note that \( h \perp \) has roughly 20%–30% annualized volatility, which is commensurate with the volatility of an average risk factor in the fundamental risk model. These statistics clearly show that \( h \perp \) has systematic risk that is not correctly accounted for during the portfolio construction process.

Having identified the crux of the problem, the next logical step is to educate the optimizer about the systematic risk in \( h \perp \) and assist it in correctly accounting for the systematic risk during the portfolio construction phase. The section that follows develops this line of thinking, culminating with a practical and effective solution to the factor alignment problem. At this point we take a brief detour, however, and ask ourselves another pertinent question, namely, why do risk models fail to capture the systematic risk in \( h \perp \)?
The answer to this question lies in the statistical intricacies of risk model development and deserves some attention. Recall that in any multivariate linear regression model there is a trade-off between explanatory power and model accuracy. In other words, one can increase the $R^2$ of a regression model by adding another risk factor, but that increase comes at a cost, namely, higher standard errors in the factor returns computation. Risk model providers, being fully aware of this trade-off, restrict the number of factors in their risk model to a manageable number. For instance, Axioma’s U.S. fundamental risk model has 10 style factors and 68 industry factors, which are derived from the GICS classification. This “parsimony” property extends to every factor risk model and is indeed the primary argument for building factor risk models in the first place. None of these models, however, including those provided by commercial risk model vendors such as Axioma, claim to capture “all” systematic risk factors. The malady of factor alignment problems lies in the optimizer’s propensity to neutralize the exposure of the optimal portfolio to all systematic risk factors that are included in the risk model and inadvertently load up on others that are omitted. This explains the significant amount of systematic risk in $h_\perp$.

To conclude, the optimizer plays the role of devil’s advocate, exacerbating the misalignment between the alpha factors, risk factors, and constraints, and thereby pushing the optimal holdings into a territory where the factor risk model is most likely to understate systematic risk. The section that follows discusses the use of an augmented risk model in solving these problems.

**MISALIGNMENT PROBLEMS: SOLUTION**

One of the key takeaways from the previous two sections is that factor risk models include only a subset of all systematic risk factors that come to bear on the equity returns process. In other words, additional risk factors are missing from the risk model, and FAP arise when the optimizer aligns the portfolio with these missing factors. In light of these observations, we assume that the true asset-asset covariance matrix $Q_T$ is given by

$$Q_T = Q + Z\Lambda Z^T$$

where $Z$ denotes the exposure matrix of the missing factors, and $\Lambda$ denotes the covariance between the factor returns of the missing factors. For the sake of brevity of notation, we assume that $Z^TX = 0$ and $Z^TZ = I$. Our goal in this section is to evaluate the undesirable effects of the missing factors on the ex post performance of the optimized portfolios and to make recommendations to overcome them. Similar to the previous sections, we mostly restrict our focus to the unconstrained MVO model and briefly address the constrained case at the end of this section. We use the following utility function to evaluate the ex post performance of an arbitrary portfolio $h$:

$$U_p(h) = \alpha^\prime h - \frac{\lambda}{2} h^\prime Q_p h$$

We define $U(h) = \alpha^\prime h - \frac{\lambda}{2} h^\prime Q h$ to be the ex ante estimate of the utility function. Note that if $\Lambda = 0$
then the ex ante and ex post performances of \( h \) are identical. However, if \( Z \) and \( \Lambda \) are different from 0, then \( U_T(h) \leq U(h) \). Of course, the degree of ex post inefficiency of \( h \) depends on how far \( U_T(h) \) can deviate from \( U(h) \). It can be shown that if \( h \) is an optimal portfolio to MVO, then

\[
U_T(h) = U(h) - \frac{1}{2\lambda\sigma_n^2} \left( \text{SysRisk}(\alpha_{\perp}) \right)^2
\]

where \( \text{SysRisk}(\alpha_{\perp}) \) denotes the systematic risk in \( \alpha_{\perp} \). In other words, the extent of deviation of the ex post utility function of \( h \) from its ex ante prediction increases with the amount of systematic risk in \( \alpha_{\perp} \).

Our solution to FAP is based on using an augmented risk model obtained by augmenting the user risk model with an additional factor, \( y = \frac{1}{\|\alpha_{\perp}\|} \alpha_{\perp} \), which is orthogonal and uncorrelated with all the existing risk factors. Specifically, we define

\[
Q_y = Q + \nu y y^T
\]

where \( \nu \) denotes the systematic variance of \( y \), and we solve the following optimization problem to determine the optimal portfolio:

\[
\max \alpha^T h - \frac{1}{2} h^T Q_y h \quad \text{(AugMVO)}
\]

If \( h_f \) denotes the optimal solution to (AugMVO), then it can be shown that

\[
U_T(h_f) = U_T(h) + \frac{\|\alpha_{\perp}\|^2}{2\lambda\sigma_n^2} \left( \frac{\text{SysRisk}(\alpha_{\perp})}{\text{TotalRisk}(\alpha_{\perp})} \right)^2
\]

where \( \text{TotalRisk}(\alpha_{\perp}) \) denotes the total risk (i.e., including both the systematic and specific risks of \( \alpha_{\perp} \)). In other words, the ex post utility function of \( h_f \) is strictly greater than that of the portfolio constructed using a misaligned risk model provided that \( \alpha_{\perp} \) has systematic risk. Furthermore, as a byproduct of this analysis, it follows that the ex ante risk estimate of \( h_f \) exactly matches the ex post realized risk. In other words, using the augmented risk model completely eliminates the downward bias in risk prediction, one of the most serious consequences of FAP. We refer the reader to Saxena and Stubbs [2010b] for a detailed analytical treatment of these results. Next, we discuss two interesting consequences of these findings.

First and foremost, these results establish the superior ex post performance of \( h_f \) relative to \( h \). In other words, optimal portfolios derived from the augmented risk model are guaranteed to have better ex post performance than those derived from the original user risk model. This improvement in performance can be attributed to the augmented risk model’s property of correctly accounting for the systematic risk in \( \alpha_{\perp} \).

Second, the technique of using the augmented risk model can be implemented without explicit access to the missing factors. Thus, even though we used the missing factors \( Z \) to provide a theoretical structure to our analysis, the final solution to the FAP circumvents them altogether. To see this, note that there are only two additional inputs required for the implementation of the augmented risk model, namely, the factor \( y = \frac{1}{\|\alpha_{\perp}\|} \alpha_{\perp} \) and \( \nu \), the systematic variance of \( y \). While \( y \) can be estimated by using the historical returns to factor \( y \). From a statistical standpoint, the augmented risk model thus represents a compromise between the user risk model that misses some of the systematic risk factors, namely, \( Z \), and the true risk model that accounts for all the systematic risk factors but may be unwieldy to use due to the large number of factors. By augmenting the user risk model with the single factor \( y \), the augmented risk model optimally aggregates the effect of all the missing factors into one factor so as to give tangible improvement in the ex post performance of the optimal portfolio.

All of these results can be generalized to the constrained case by replacing \( \alpha \) by implied alpha (\( \gamma \)). The resulting augmenting factor, \( \frac{\|\alpha_{\perp}\|^2}{2\lambda\sigma_n^2} \), however, depends on the optimal portfolio and parameters of the first-order optimality conditions. The problem of simultaneously choosing an optimal portfolio and an augmenting factor can be phrased as an equilibrium problem that, in turn, can be reformulated as a convex second-order cone program (SOCP). The details of the mentioned reformulation are technical and outside the scope of this article (see Saxena and Stubbs [2010b]).

**CASE STUDY**

In this section, we illustrate the results presented earlier on a backtest based on real data. We use the following long-only strategy in our case study:
\[
\begin{align*}
\text{maximize} & \quad \text{Expected Return} \\
\text{st} & \quad \text{Fully invested long-only portfolio} \\
& \quad \text{Active GICS sector exposure constraint} \\
& \quad \text{Active GICS industry exposure constraint} \\
& \quad \text{Active asset bounds constraint} \\
& \quad \text{Turnover constraint (16.67\%)} \\
& \quad \text{Active Risk constraint (\(\sigma\))} \\
& \quad \text{Benchmark} = S & P 500 \\
\end{align*}
\]

We ran a monthly backtest based on this strategy in the 2001–2009 time period for various values of \(\sigma\) chosen from \{0.5\%, 0.6\%, \ldots, 3.0\%\}. For each value of \(\sigma\), we ran the backtest in two setups that were identical in all respects except one, namely, only the second setup used the Alpha Alignment Factor (AAF) methodology (\(\sqrt{\nu} = 20\%\)). We used Axioma’s U.S. fundamental medium-horizon risk model (US2AxiomaMH) to model the active risk constraint.

The expected returns—referred to as simply the alpha—for this strategy are not completely aligned with the risk factors in US2AxiomaMH. Exhibit 4 shows the time series of the misalignment coefficient of alpha, implied alpha, and the optimal portfolio. Note that almost 40\%–60\% of the alpha is not aligned with the risk factors. Interestingly, the alignment characteristics of implied alpha are significantly better than that of alpha. Among other things, this implies that the constraints of the strategy, especially the long-only constraint, play a proactive role in containing the misalignment issue. As is evident from the MC of the optimal portfolio, however, the optimizer loads up on the orthogonal component, \(\gamma_{\perp}\), of implied alpha assuming that \(\gamma_{\perp}\) has no systematic risk. Exhibit 5 refutes the validity of this assumption by showing the time series of systematic risk in \(\gamma_{\perp}\) and \(\alpha_{\perp}\). We compute the mentioned systematic risk by regressing the period asset returns against risk factors in US2AxiomaMH and \(\gamma_{\perp}(\alpha_{\perp})\), computing time series of factor returns that can be attributed to \(\gamma_{\perp}(\alpha_{\perp})\), and computing the annualized volatility of the resulting factor returns using a 12-month rolling horizon. Note that the orthogonal component of both alpha and implied alpha not only has systematic risk, but the magnitude of the systematic risk is comparable to the systematic risk associated with a median risk factor in US2AxiomaMH.

**Exhibit 4**
Case Study: Misalignment Coefficient

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implied Alpha</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal Portfolio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Exhibit 5**
Case Study: Systematic Risk in the Orthogonal Component

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implied Alpha</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Factor in US2AxiomaMH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To summarize, the primary purpose of portfolio optimization is to create a portfolio having an optimal risk-adjusted expected return. If a portion of the risk in a portfolio derived from the orthogonal component of implied alpha is not accounted for, then the resulting risk-adjusted expected return cannot be optimal.

Exhibit 6 shows the predicted and realized active risk for various risk target levels. Note the significant downward bias in risk prediction when the AAF methodology is not employed. Exhibit 7 reports the same information using the concept of the bias statistic. The bias statistic is a statistical metric that is used to measure the accuracy of risk prediction; if the ex ante risk prediction is unbiased, then the bias statistic should be close to 1.0, (see Saxena and Stubbs [2010a] for more details). Clearly, the bias statistics obtained without the aid of the AAF methodology are significantly above the 95% confidence interval, thereby showing that the downward bias in the risk prediction of optimized portfolios is statistically significant. The AAF methodology recognizes the possibility of systematic risk in $\gamma_{\perp}$ and guides the optimizer to avoid taking excessive unintended bets. Exhibits 6 and 7 illustrate the efficacy of AAF in reducing the bias in risk prediction. For every risk target level, the realized risk of the backtest that used the AAF was within the 95% confidence interval of the ex ante risk prediction.

Exhibit 8 shows the realized risk–return frontier. Interestingly, using the AAF methodology not only improves the accuracy of risk prediction but also moves the ex post frontier upward, thereby giving ex post performance improvements. The precise reasons for this performance differential are technical and are discussed elsewhere (see Saxena and Stubbs [2010b]). Nevertheless, we give some intuition for these improvements in the following discussion.

The distinguishing feature of quantitative investing as a profession is its belief in generating optimal risk-adjusted returns and its ability to access portfolios that lie on the efficient ex post frontier. As illustrated by the previous case study, and also the empirical results reported by Saxena and Stubbs [2010a], a portfolio construction approach that is agnostic to the alignment between the alpha and risk factors consistently fails to produce unbiased ex ante risk estimates. An approach that cannot even predict the risk of the portfolio correctly cannot be expected to produce optimal portfolios. The AAF approach recognizes the possibility of missing systematic risk factors and makes amends, to the extent possible without complete recalibration of the risk model, that explicitly account for the latent systematic risk in alpha factors. In the process of doing so, the AAF approach not only improves the accuracy of risk prediction, but also partly repairs the lack
of efficiency in the optimal portfolio, thereby giving the performance differential shown in Exhibit 8.

**CUSTOM RISK MODELS**

In the previous sections, we discussed various sources of FAP, their effects on optimal portfolios, and one way of circumventing the resulting problems, namely, the alpha alignment factor approach. In this section, we discuss some limitations of the AAF approach which naturally point in the direction of improving the methodology for tackling alignment problems. We conclude this article with some motivation for custom risk models (CRM) and how the combination of CRM and the AAF approach provides an ideal solution to FAP.

The AAF approach has three key limitations. First, the AAF construct is based on the assumption that the factor returns associated with the missing factors are uncorrelated with the factor returns associated with the regular factors in the user risk model. The fact that the AAF is orthogonal to the regular factors, by itself, does not imply lack of correlation of factor returns. To see this, note that even though the industry factors derived from the GICS classification scheme are mutually orthogonal, the corresponding factor returns are often correlated. By being correlation agnostic, the AAF approach fails to capture the interaction between factor returns that can be attributed to missing factors and the user risk factors. Second, the AAF approach requires calibration of the volatility parameter which presents additional practical problems. Furthermore, the temporal stationarity of the mentioned volatility parameter is not guaranteed, which introduces additional complications related to dynamic estimation of the volatility parameter. Third, the AAF approach does not use historical data to improve its representation of the missing factors. In other words, it is agnostic to the nature of residual returns that might have useful information apropos missing factors.

A natural way to circumvent these problems is to recalibrate the user risk model taking into account the possible sources of latent systematic risk. Custom risk models accomplish exactly that goal. CRM are derived from the user risk model, referred to as the base model, by introducing additional factors with the intent of eliminating various sources of misalignment. The additional factors are referred to as custom risk factors, and the resulting risk models are said to be customized. Construction of CRM involves complete recalibration of the covariance matrix by re-running the cross-sectional regressions, recomputing factor returns attributed to the user and custom risk factors, and using the resulting time series of factor and residual returns to compute the factor–factor covariance matrix and specific risk.

For instance, if the alpha model uses the S/P valuation ratio, which is missing from the suite of user risk factors, then S/P can function as a custom risk factor in the construction of CRM. Similarly, if the proprietary value signal used to define the alpha process is different from the value signal that is present in the user risk model, then the proprietary value signal can be used to augment the user risk model. In this case, it might be beneficial to drop the original value signal in the user risk model to avoid issues arising from multicollinearity. Note that using a CRM completely eliminates misalignment between the alpha and risk factors. Furthermore, by virtue of complete recalibration, CRM not only estimate the volatility of the missing factors accurately, but they also capture the correlation information.

Various other sources of misalignment discussed earlier can be addressed by appropriately choosing the custom risk factors to add to the user risk model, with only one caveat. If the misalignment results from a combination of constraints, then it might be difficult to determine a priori the right factors to use in order to augment the risk model. For instance, in a long-only strategy, the long-
only constraints get dynamically aggregated in the construction of implied alpha. Determining a custom risk factor that captures the latent systematic risk in such an aggregation of constraints can be extremely difficult and, hence, cannot be handled directly by CRM. Of course, the AAF approach can still be used to address FAP arising from such constraints. To summarize, we believe that a combination of CRM and the AAF approach offers the most practical and holistic approach to FAP.

CONCLUSION

In this article, we examine the sources of misalignment, analyze and document their effects, and present a detailed analysis, culminating with a practical and effective remedy to the factor alignment problems. Unsurprisingly, our research on the FAP brings together elements from the fields of finance, statistics, accounting, and optimization and is based on a careful understanding of how these fields interact with each other within the world of equity investments. For instance, the fundamental source of FAP is the different view alpha and risk model developers have on financial statement data, an accounting artifact. The trade-off between model accuracy and the number of factors restricts risk models from incorporating all sources of systematic risk, a statistics artifact. The combined effect of the misalignment between the alpha and risk factors, and limited representation of the systematic risk factors in the risk model results in hidden systematic risk in the orthogonal component of (implied) alpha, which is magnified and excessively represented during the construction of the optimal portfolio, an optimization artifact. All of these influences have an ominous bearing on the ex post performance, create a downward bias in the ex ante risk prediction, and thus compromise the ex post efficiency of the optimal portfolio, a finance artifact.

The augmented risk model approach strikes a careful balance among all of these influences and, in the process, creates a practical and effective solution to the factor alignment problem. It not only corrects for the risk underestimation bias of optimal portfolios, but also pushes the ex post efficient frontier upward, thereby empowering a PM to access portfolios that lie above the traditional risk–return frontier defined by the user risk model. We briefly discuss the role of custom risk models obtained by complete recalibration of the user risk model in directly addressing the root source of FAP. An empirical analysis of CRM in the context of portfolio construction, performance attribution, and risk decomposition is underway and will be the topic of a forthcoming article.

ENDNOTES

1Recent developments in risk model technology have significantly reduced the infrastructure costs involved in recalibration of commercial risk models that use custom factors. We discuss one such approach briefly in this article under the heading of custom risk models. An elaborate discussion of this topic is beyond the scope of the current article and is addressed elsewhere.

2It must be noted that the assumptions $Z^T X = 0$ and $Z^T Z = I$ can be made without loss of generality; in other words, any arbitrary factor model can be transformed so as to satisfy these properties. Of course, the assumption that the missing factor is uncorrelated with the factors in the user model is not always satisfied in reality.

REFERENCES


To order reprints of this article, please contact Dewey Palmieri at dpalmieri@iijournals.com or 212-224-3675.

Copyright © 2012 Axioma, Inc. All rights reserved. Not to be reproduced or redistributed without permission.