

Fundamentals of Efficient Factor Investing

Roger Clarke, Harindra de Silva, CFA, and Steven Thorley, CFA

Combining long-only-constrained factor subportfolios is generally not a mean-variance-efficient way to capture expected factor returns. For example, a combination of four fully invested factor subportfolios—low beta, small size, value, and momentum—captures less than half (e.g., 40%) of the potential improvement over the market portfolio's Sharpe ratio. In contrast, a long-only portfolio of individual securities, using the same risk model and return forecasts, captures most (e.g., 80%) of the potential improvement. We adapt traditional portfolio theory to more recently popularized factor-based investing and simulate optimal combinations of factor and security portfolios, using the largest 1,000 common stocks in the US equity market from 1968 to 2015.

Equity investors increasingly view their portfolios as not only a collection of securities but also a bundle of exposures to the factors that drive security returns. The recent growth in factor-based strategies under a variety of names indicates that many investors now view managing factor exposures on a par with traditional asset allocation. For example, Kahn and Lemmon (2016) suggested that factor-based investing strategies often labeled “smart beta” represent a disruptive innovation in the asset management industry. In the equity market, investors focus on such well-known factors as size, value, and momentum, although the factor framework can also be applied to other asset classes. One unifying theme is that factor-replicating subportfolios may allow investors to effectively manage factor risk and return trade-offs without having to trade directly in individual securities.

Roger Clarke is chairman of Analytic Investors, Los Angeles. Harindra de Silva, CFA, is president of Analytic Investors, Los Angeles. Steven Thorley, CFA, is the H. Taylor Peery Professor of Finance at the Marriott School of Management, Brigham Young University, Provo, Utah.

Editor's note: This article was reviewed and accepted by Executive Editor Stephen J. Brown.

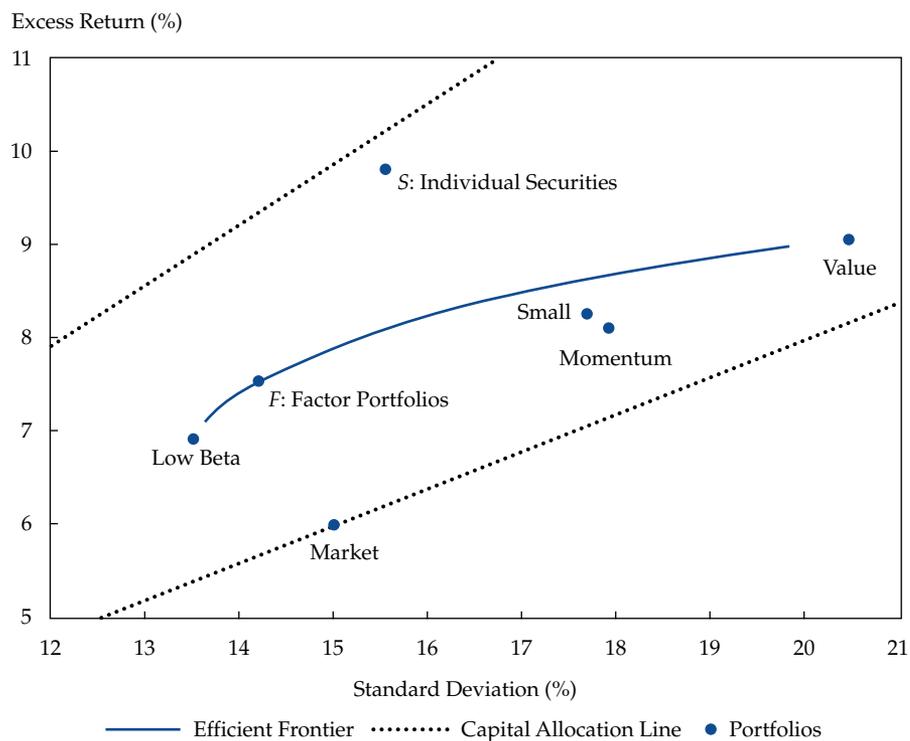
Editor's note: Steven Thorley, CFA, became co-editor of the *Financial Analysts Journal* after the article was submitted but before it was accepted for publication. He was recused from the peer review and acceptance processes. All the necessary measures were taken to prevent Dr. Thorley from accessing any information related to the submission, including the identity of the reviewers. The reviewers were also unaware of his and his co-authors' identities. For information about the current conflict-of-interest policies, see www.cfapubs.org/page/faj/policies.

The central finding of our study is that portfolios built directly from individual securities capture most of the potential gain from exploiting a small set of factors, whereas combinations of specialized factor portfolios capture only a fraction of that potential. The increase in mean-variance efficiency comes from the wider latitude in portfolio construction afforded by the cross-sectional variation of the security exposures to the factors. Other important concepts are that secondary exposures in factor subportfolios do not impose a material reduction on the expected Sharpe ratio (Sharpe 1964)—if measured and incorporated into the final portfolio weights—as well as the relatively minor impact of security-specific or idiosyncratic risk.

Consider the capitalization-weighted market portfolio and four other portfolios that tilt toward low-beta, small-size, value, and momentum stocks. **Figure 1** plots the location of the market and factor subportfolios in terms of their expected returns and risks, given the factor exposures of the largest 1,000 stocks in the US equity market in 2016. Point *S* is the Sharpe ratio-optimal portfolio of long-only positions in individual stocks and lies on a long-only-constrained efficient frontier of other optimal security portfolios. Point *F* is the Sharpe ratio-optimal long-only combination of the four factor subportfolios and lies on an efficient frontier of other optimal subportfolio combinations.

The efficient frontier of factor subportfolio combinations in **Figure 1** stretches from a large weight on the low-beta portfolio through portfolio *F* to a large weight on the value portfolio. The locations of the subportfolios and associated efficient frontier curves in **Figure 1** are specific to (1) a given set of investor expectations, (2) the factor exposure correlation structure built into the 2016 US equity market,

Figure 1. Optimal Security and Factor Portfolios



and (3) the impact of secondary (i.e., unintended) exposures in the subportfolios. As explained later in the article, the impact of these secondary exposures makes the location of the factor portfolios in Figure 1 different from what the expected return and risk of each factor alone would dictate.

One important concept behind Figure 1 is that the optimal subportfolio combinations would include a short position in the market portfolio *if not for the long-only constraint*. Each of the subportfolios already has ample market-factor exposure, and “squeezing in” enough simultaneous exposure to the nonmarket factors to make a material difference in the risk-adjusted return requires a large hedge on the market portfolio. In fact, the number of nonmarket factors and the factor information ratio magnitudes needed to motivate active versus passive investing *increase* the size of that market hedge. The higher dotted line in Figure 1 shows the potential Sharpe ratio if the market hedge could be deployed, and the lower dotted line shows the expected Sharpe ratio of the market portfolio. In contrast, the long-only optimal portfolio *S* of individual securities lies on an efficient frontier that comes close to the maximum factor potential without having to short the market portfolio or any securities.

In our study, we used the mathematics of multifactor portfolio theory that originated with Treynor and Black (1973)—extended to accommodate correlated factor returns and secondary factor exposures—to compare the mean-variance efficiency of

security versus factor portfolio combinations. Using a set of well-known factors in the US equity market over 1968–2015, we measured the magnitude of the loss in efficiency from combining factor subportfolios. We did not address the set of factors that best explains the covariance structure of individual stocks or that offers the best prediction of long-term returns. We were agnostic about what led these particular factors to be identified in the historical return data—whether rational returns to systematic risk, behavior- or friction-induced market anomalies, or simply extensive data mining. We assumed only that investors selected some small set of equity market factors as the primary driver of returns in well-diversified portfolios.

Our results have parallels to Kritzman and Page (2003), who showed that portfolios constructed from individual securities present a greater opportunity set for skilled investors than choosing among asset classes, economies, or sectors. This article, however, is not about the potential of investor skill to select among alternative securities or asset classes. The underlying drivers of returns are the same, whether the final portfolio is formed from individual securities or subportfolios. Our focus here is the mean-variance efficiency of long-only portfolios formed from individual securities versus portfolios formed from factor subportfolios.

Econometric advances such as those of Ledoit and Wolf (2003) and Fan, Fan, and Lv

(2008)—together with commercially available factor risk models by Barra and Axioma, among others—have made the application of the portfolio theory of Markowitz (1952) a practical reality for many investors. Solutions to such large-scale portfolio construction issues as the “curse of dimensionality” and “error optimization” have also been provided over time, with frameworks for quantitative active portfolio management established by Grinold (1989) and Black and Litterman (1992).

Although such implementation issues as turnover, transaction costs, and managerial fees are important considerations in designing investment products and services, this article focuses on potentially more material issues associated with multiple levels of portfolio optimization and secondary factor exposures under the long-only constraint. In this article, we illustrate several alternative methodologies for weighting securities in factor subportfolios, but we do not specify the “best” way to construct such portfolios. Capitalization weighting, single-factor optimization, and heuristic sorting on factor exposure all produce final portfolios with lower Sharpe ratios than security-based portfolios.

Readers who are primarily interested in the long-term (1968–2015) backtests can skip directly to the final section, but the theory sections that come first help explain *why* multifactor portfolios built directly from individual securities maintain such a significant advantage.

Multifactor Portfolio Theory

Our notation for the well-known linear factor model of the return on the i th (out of N) risky assets is

$$r_i = \alpha_i + \beta_{i,1} R_1 + \dots + \beta_{i,K} R_K + \varepsilon_i, \quad (1)$$

where R_j is the realized return on the j th (out of K) factors, $\beta_{i,j}$ is the “exposure” of asset i to factor j , and ε_i is asset i 's idiosyncratic return. In subsequent equations, we use the bolded notation \mathbf{B} for the N -by- K matrix of factor exposures $\beta_{i,j}$ and ignore both security-specific expected returns, α_i , and the source of security-specific idiosyncratic risk, ε_i , to focus on the returns and risks of factor exposures. An important new result derived in Appendix A is that the N -by-1 vector of security weights that maximize the unconstrained factor Sharpe ratio of security portfolio S is

$$\mathbf{w}_S = \left(\frac{\sigma_S}{\text{SR}_S} \right) \mathbf{B} (\mathbf{B}'\mathbf{B})^{-1} \mathbf{V}^{-1} \mathbf{U}, \quad (2)$$

where \mathbf{V} is the K -by- K factor return covariance matrix and \mathbf{U} is the K -by-1 vector of forecasted factor returns.

The scalar multiplier at the beginning of Equation 2 is simply the risk of the optimal portfolio,

σ_S , divided by the optimal portfolio's Sharpe ratio, SR_S . The terms at the end of Equation 2 capture investor expectations of factor risks, \mathbf{V} , and factor returns, \mathbf{U} , similar to other mean–variance-optimal solutions. The innovative part of Equation 2 is the middle term, which includes the factor exposure matrix \mathbf{B} and the inverse factor exposure correlation matrix $(\mathbf{B}'\mathbf{B})^{-1}$. Applying Equation 2 with factor subportfolios as the underlying assets rather than individual securities gives the K -by-1 vector of optimal subportfolio weights as

$$\mathbf{w}_F = \left(\frac{\sigma_F}{\text{SR}_F} \right) \mathbf{B}_F (\mathbf{B}_F' \mathbf{B}_F)^{-1} \mathbf{V}^{-1} \mathbf{U}, \quad (3)$$

where \mathbf{B}_F is the K -by- K matrix of subportfolio factor exposures.

Because the weights in Equations 2 and 3 are unconstrained (i.e., may be positive or negative but sum to 100%), portfolios S and F both achieve the maximum possible *factor* Sharpe ratio, defined as the expected factor-driven return divided by factor-driven risk (i.e., without consideration of idiosyncratic risk). As shown in Appendix A for generic portfolio P , when these unconstrained weights are used, the optimal portfolio's factor Sharpe ratio squared is

$$\text{SR}_P^2 = \mathbf{R}' \mathbf{\Pi}^{-1} \mathbf{R}, \quad (4)$$

where \mathbf{R} is a K -by-1 vector composed of the market Sharpe ratio and the nonmarket factor information ratios, and $\mathbf{\Pi}$ is the K -by- K factor return correlation matrix. If the factor returns are uncorrelated (i.e., $\mathbf{\Pi}^{-1} = \mathbf{\Pi} = \mathbf{I}$), Equation 4 collapses to a property first identified by Treynor and Black (1973).¹ The Treynor–Black rule is that the maximum possible unconstrained (i.e., long–short portfolio) Sharpe ratio squared is equal to the market portfolio Sharpe ratio squared plus the sum of the squared information ratios of the other $K - 1$ factors:

$$\text{SR}_P^2 = \text{SR}_M^2 + \sum_{j=2}^K \text{IR}_j^2. \quad (5)$$

Although we use information ratios to parameterize investor views on factor risks and returns, we do not use the portfolio's overall information ratio as the primary performance statistic. Characterizing portfolio performance by a single information ratio implicitly assumes that the market and purely active portfolios can be separated, whereas limits on shorting the market are in fact critical to the differences in mean–variance efficiency of the security versus subportfolio combinations that we studied. In addition, a portfolio's information ratio, as opposed to its Sharpe ratio, does not account for the *optimal*

amount of active risk. As shown in Appendix A, the optimal amount of active risk declines with the portfolio's unconstrained information ratio, as well as the Clarke, de Silva, and Thorley (2002) transfer coefficient, owing to the reduced potential for active management to add value. Thus, our primary numerical expression of a portfolio's mean–variance efficiency is the percentage of potential Sharpe ratio capture, $(SR - SR_M) / (SR_P - SR_M)$, where SR is the Sharpe ratio of the portfolio being measured, SR_P is the unconstrained optimal Sharpe ratio, and SR_M is the Sharpe ratio of the market portfolio.

Factor Return Parameters. Table 1 reports statistics on the set of factor returns in the US equity market over 1968–2015 that motivated our choice of parameter values in the previous illustration, as well as the numerical examples in this section. These well-known factors were identified over time by, among others, Jegadeesh and Titman (1993); Fama and French (1996); Carhart (1997); Chan, Karceski, and Lakonishok (1998); Clarke, de Silva, and Thorley (2010); and Frazzini and Pedersen (2014). As shown in Table 1, the average return in excess of the risk-free rate for the cap-weighted market portfolio (largest 1,000 US common stocks) was 5.73%, with a risk of 15.56%. The *incremental* (i.e., in excess of market) return and risks for the four other factors produce information ratios that range from 0.180 for the small-size factor to 0.634 for the momentum factor.

We estimated the factor returns in Table 1 using monthly cap-weighted multivariate regressions on the cross section of security returns, as explained in Clarke, de Silva, and Thorley (2014).² One implication of the multivariate regression framework is that the realized factor returns in Table 1 are less correlated with each other, in contrast to factor returns based on univariate or bivariate sorts (e.g., the Fama–French factors HML, or high minus low, and UMD, or up

minus down). The notable exception to the general pattern of small-magnitude correlations is the -0.644 correlation of the low-beta factor return with the market return. Because of this large negative correlation, the risk-adjusted information ratio provides a better perspective on that factor's potential than the simple quotient of active return to active risk. Adjusting for the realized market beta, the "alpha" of the low-beta factor is 2.73%, with active risk of 4.96%, giving a risk-adjusted information ratio of $2.73/4.96 = 0.551$ (shown at the bottom of Table 1).

In the next sections, we emphasize several important concepts based on the optimal portfolio weights specified in Equations 2 and 3. First, the generalized Treynor–Black result in Equation 4 is achievable only if the security weights are unconstrained, meaning that short positions in individual securities are allowed in portfolio S and shorting subportfolios is allowed in portfolio F . Using the simple case of *one* factor in addition to the market, we show that the reduction in the long-only portfolio F factor Sharpe ratio increases with the factor's information ratio and decreases with its active risk.

Second, using a more involved case of the market and *two* additional factors, we show how secondary factor exposures in matrix B_F contribute to the construction of portfolio F . Appendix A specifies the design of pure factor-replicating subportfolios, with zero exposure to all but one nonmarket factor, but such portfolios require multivariate optimization and short selling that may be costly to implement in practice. Third, the reduction in the factor Sharpe ratio of portfolio F becomes larger as more factors are used—although less so with positively correlated factor exposures and more so with negatively correlated exposures.

The Market Plus One Factor. Consider the simple case of optimally combining two portfolios: the market portfolio M and one other fully invested

Table 1. Annualized Factor Returns, 1968–2015

	Market	Low Beta	Small	Value	Momentum
Average	5.73%	1.19%	0.67%	0.92%	3.89%
Standard deviation	15.56%	6.48%	2.67%	4.16%	6.14%
Average/Standard deviation	0.368	0.184	0.250	0.222	0.634
<i>Correlation with</i>					
Market	1.000	-0.644	0.199	-0.082	-0.026
Low beta	-0.644	1.000	-0.159	0.109	0.071
Small	0.199	-0.159	1.000	-0.152	0.064
Value	-0.082	0.109	-0.152	1.000	-0.153
Momentum	-0.026	0.071	0.064	-0.153	1.000
Market beta	1.000	-0.268	0.034	-0.022	-0.010
Market alpha	0.00%	2.73%	0.47%	1.05%	3.95%
Active risk	0.00%	4.96%	2.62%	4.15%	6.14%
Information ratio	0.000	0.551	0.180	0.252	0.643

portfolio with unit exposure to some factor A with a return that is uncorrelated with the market. The Treynor–Black result (Equation 5) for the maximum possible factor Sharpe ratio in this case is

$$SR_F^2 = SR_M^2 + IR_A^2, \quad (6)$$

where the required weights for the market portfolio and the factor A portfolio are given by

Equation 3. Specifically, $\mathbf{B}_F = \begin{bmatrix} 1.0 & 0.0 \\ 1.0 & 1.0 \end{bmatrix}$ such that

$$\mathbf{B}_F (\mathbf{B}'_F \mathbf{B}_F)^{-1} = \begin{bmatrix} 1.0 & -1.0 \\ 0.0 & 1.0 \end{bmatrix}, \text{ the assumption of uncor-}$$

related factor returns gives $\mathbf{V}^{-1} \mathbf{U} = \begin{bmatrix} SR_M / \sigma_M \\ IR_A / \sigma_A \end{bmatrix}$, and

the budget constraint gives $SR_F / \sigma_F = SR_M / \sigma_M$. With these substitutions in Equation 3, the required weight for subportfolio A is

$$w_A = \frac{\sigma_M IR_A}{SR_M \sigma_A}, \quad (7)$$

and the required weight for the market portfolio is $w_M = 1 - w_A$. Note that although the investor's view about factor A is parameterized by that factor's information ratio and active risk, the weight specified in Equation 7 applies to a fully invested portfolio (i.e., the constituent security weights may be positive or negative but sum to 100%). In other words, with the additional exposure to the market factor, the total expected return on subportfolio A is $\mu_M + \mu_A$ and the total factor risk of subportfolio A is $(\sigma_M^2 + \sigma_A^2)^{1/2}$.

Suppose that the parameters for the market factor are $\mu_M = 6\%$ and $\sigma_M = 15\%$ ($SR_M = 6.0/15.0 = 0.400$) and that the parameters for factor A are $\mu_A = 0.9\%$ and $\sigma_A = 3.0\%$ ($IR_A = 0.9/3.0 = 0.300$). For these numerical values, the maximum possible Sharpe ratio in Equation 6 is $(0.400^2 + 0.300^2)^{1/2} = 0.500$. To obtain that Sharpe ratio, however, the weight for subportfolio A in Equation 7 must be $w_A = (15.0 \times 0.300) / (0.400 \times 3.0) = 375\%$, meaning that a large short position of -275% in the market portfolio is required for the combined portfolio to have weights that sum to 100%. With those weights, the combined portfolio has an expected return of $3.75(6.9) - 2.75(6.0) = 9.4\%$ and a factor risk of 18.8%, yielding the specified Sharpe ratio of $9.4/18.8 = 0.500$. Without the ability to short the market portfolio, the next-best Sharpe ratio in this simple case is achieved with a 100% investment in subportfolio A , which has an expected return of $6.0 + 0.9 = 6.9\%$ and a risk of $(15.0^2 + 3.0^2)^{1/2} = 15.3\%$. Thus, the Sharpe ratio of the long-only-constrained optimal solution is only $6.9/15.3 = 0.451$, about halfway between the passive market Sharpe ratio of 0.400 and the maximum possible Sharpe ratio of 0.500.

Figure 2 plots the unconstrained and long-only-constrained Sharpe ratios for a range of factor information ratios and three levels of active risk. For instance, the assumed numerical values in the previous example plot on the “Long-Only 3% Active Risk” curve at a factor information ratio of 0.300. Moving from left to right in Figure 2, the Treynor–Black promise of value added from using a nonmarket factor embedded in a fully invested portfolio requires a short position in the market portfolio that increases with the information ratio. The larger required short positions in the market portfolio lead to larger reductions in the long-only Sharpe ratio, as shown by the gap between the unconstrained and long-only-constrained lines.

As specified in Equation 7, lower values for factor risk, σ_A , holding the information ratio constant, also lead to a larger reduction in the long-only portfolio F Sharpe ratio. The intuition is that a factor with lower risk requires a larger position in the factor portfolio to adequately affect the Sharpe ratio—and thus a larger short position in the market portfolio to meet the fully invested budget constraint. For lower values of the single nonmarket factor information ratio, however, the constrained lines in Figure 2 do *not* gap below the unconstrained optimal solution, meaning that the factor A portfolio weight is less than 100% and shorting the market portfolio is not required.

Before moving on to the case of the market plus two factors, note that the factor return in the case of the single nonmarket factor could be correlated *ex ante* with the market return, which would make the math in Equations 6 and 7 more involved. Most factor portfolios that have been examined in practice—for example, SMB (small minus big) for the small-cap factor (Fama and French 1996) and HML (high minus low book-to-market ratio) for the value factor—are designed to be approximately uncorrelated with the market return. In contrast, the more recently introduced “beta” factors—for example, VMS (volatile minus stable) in Clarke, de Silva, and Thorley (2010) and BAB (betting against beta) in Frazzini and Pedersen (2014)—are by design highly correlated with the market. Using the generalized Treynor–Black result in Equation 4, with a nonzero correlation of ρ_{MA} between the market and factor A returns, we see that the optimal possible Sharpe ratio is

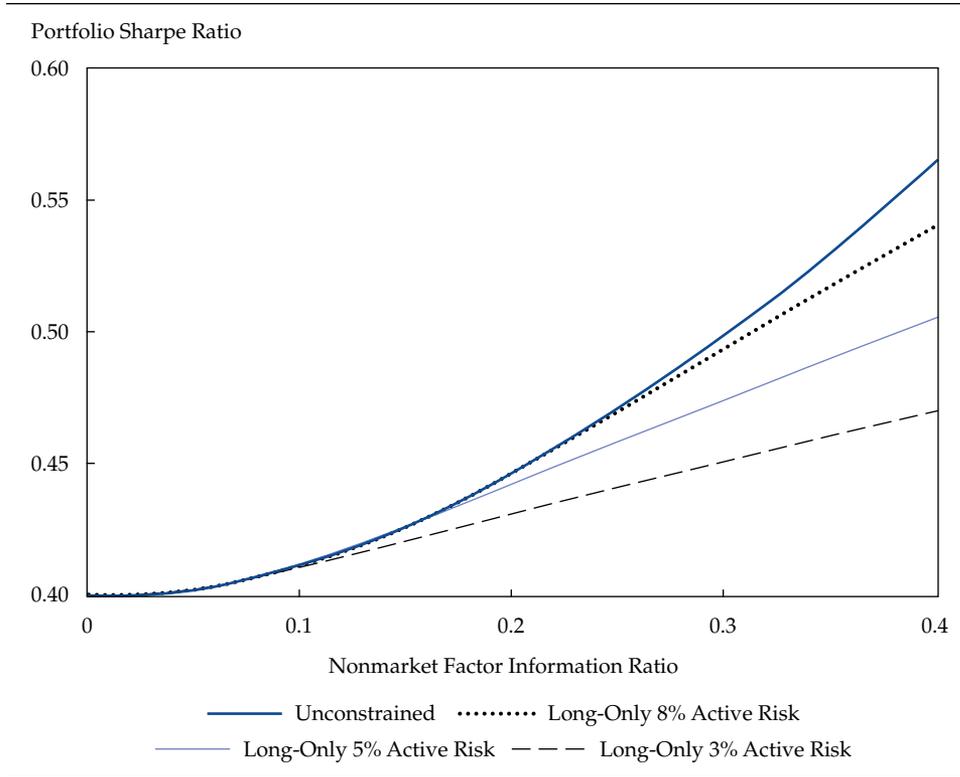
$$SR_F^2 = \frac{(\rho_{MA} SR_M + IR_A)^2}{1 - \rho_{MA}^2} + IR_A^2 \quad (8)$$

instead of Equation 6, and the required weight for subportfolio A is

$$w_A = \frac{\sigma_M (\rho_{MA} SR_M + IR_A)}{(\rho_{MA} SR_M + IR_A) \sigma_A} \quad (9)$$

instead of Equation 7.

Figure 2. Optimal Factor Sharpe Ratio with One Nonmarket Factor



Suppose that A is the low-beta factor with an expected return of $\mu_A = 0.0\%$, risk of $\sigma_A = 10.0\%$, and a market correlation of $\rho_{MA} = -0.500$. Even though the simple information ratio for this factor is $0.0/10.0 = 0$, the maximum possible Sharpe ratio in Equation 8 is $0.400/(1 - 0.500^2)^{1/2} = 0.462$, a material improvement over the market Sharpe ratio of 0.400. In other words, although factor A has an expected return of zero, optimal deployment allows for a hedge on market risk without lowering the expected return. To construct that hedge, the optimal weight for subportfolio A in Equation 9 is $w_A = 0.15(0.000 + 0.500 \times 0.400) / [(0.400 - 0.000) \times 0.10] = 75\%$. In fact, for the parameter values in this illustration, the expected incremental return of factor A would have to be -2.0% for w_A to be zero in Equation 9, consistent with the prediction of the traditional CAPM.

The Market Plus Two Factors. Now consider the case of the market factor plus *two* additional factors, A and B . First, assume that factors A and B are represented by pure factor-replicating portfolios, meaning that subportfolio A has no exposure to factor B and subportfolio B has no exposure to factor A . In addition, both subportfolios are fully invested, with market factor exposures of exactly 1. Specifically, assume that the subportfolio factor exposures used in Equation 3 are

$$\mathbf{B}_F = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 1.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 1.0 \end{bmatrix} \quad (10)$$

The nondiagonal elements of 0.0 in the first *row* of Equation 10 indicate that the market portfolio has no positive or negative incremental exposure to the nonmarket factors. Alternatively, the nondiagonal elements of 1.0 in the first *column* of Equation 10 indicate that the factor subportfolios have full market factor exposure.

If the returns to factors A and B are uncorrelated with each other and are uncorrelated with the market factor, the maximum possible Sharpe ratio (according to the Treynor–Black result) is

$$SR_F^2 = SR_M^2 + IR_A^2 + IR_B^2 \quad (11)$$

The weight for subportfolio A required to achieve the result in Equation 11 is still given by Equation 7, with a similar form for subportfolio B 's weight. Suppose that the market parameters are $\mu_M = 6\%$ and $\sigma_M = 15.0\%$ ($SR_M = 6.0/15.0 = 0.400$) but that the parameters for factor A are more modest at $\mu_A = 1.0\%$ and $\sigma_A = 5.0\%$ ($IR_A = 1.0/5.0 = 0.200$). Given the same values for factor B , the maximum possible Sharpe ratio in Equation 11 is $(0.400^2 + 0.200^2 + 0.200^2)^{1/2} = 0.490$. To obtain that Sharpe ratio, however, the weight for subportfolio A must be

$w_A = (15.0 \times 0.20) / (0.40 \times 5.0) = 150\%$, with the same weight for subportfolio B , requiring a short position in the market portfolio of -200% . Alternatively, the long-only-constrained optimal solution is 50% weights for subportfolios A and B , resulting in a Sharpe ratio of 0.454—again, about halfway between the unconstrained optimal Sharpe ratio of 0.490 and the market benchmark Sharpe ratio of 0.400.

The market-plus-two-factor case allows for an examination of the impact of correlated nonmarket factor returns as well as the impact of nonzero secondary factor exposures. First, suppose that the factor subportfolios are pure, without secondary factor exposures, but the factor returns are thought to have some nonzero correlation value ρ_{AB} . The generalized Treynor–Black result in Equation 5 gives the maximum possible Sharpe ratio as

$$SR_F^2 = SR_M^2 + \frac{IR_A^2 + IR_B^2}{1 + \rho_{AB}} \quad (12)$$

(rather than Equation 11) and the required weight for subportfolio A as

$$w_A = \frac{\sigma_M}{SR_M(1 - \rho_{AB}^2)} \left(\frac{IR_A}{\sigma_A} - \rho_{AB} \frac{IR_B}{\sigma_B} \right) \quad (13)$$

(rather than Equation 7), with a similar form for subportfolio B 's optimal weight.

For example, if $\rho_{AB} = 0.200$, the maximum possible factor Sharpe ratio in Equation 12 is $SR_F = 0.476$. The potential Sharpe ratio is lower than the simple Treynor–Black value of 0.490 because the factors are not independent. The required weights for subportfolios A and B in Equation 13 are 125% each, lower than the 150% in the uncorrelated case, so the long-only constraint is not as binding. But if the factor returns are thought to be *negatively* correlated—say, $\rho_{AB} = -0.200$ —the maximum possible Sharpe ratio in Equation 12 is 0.510, the required weights for subportfolios A and B in Equation 13 are 187.5% each, and the long-only constraint is *more* binding.

Now assume that factor portfolio B has a secondary exposure to factor A of 0.2 instead of 0.0 but the factor returns are uncorrelated. In other words, the matrix of subportfolio factor exposures is

$$\mathbf{B}_Q = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 1.0 & 1.0 & 0.0 \\ 1.0 & 0.2 & 1.0 \end{bmatrix} \quad (14)$$

instead of Equation 10. Although the maximum possible Sharpe ratio—given uncorrelated returns—still conforms to the Treynor–Black result of 0.490, the weights required to obtain that Sharpe ratio must account for B_{BA} , the ancillary exposure of subportfolio B to factor A :

$$w_A = \frac{\sigma_M}{SR_M} \left(\frac{IR_A}{\sigma_A} - B_{BA} \frac{IR_B}{\sigma_B} \right). \quad (15)$$

The optimal weight formula in Equation 15 is more involved than Equation 7 to adjust for the fact that subportfolio B provides some of the desired exposure to factor A . The required weight for subportfolio B is still $w_B = 150\%$, but the required weight for subportfolio A is now only $w_A = 120\%$, so the required short position in the market portfolio is -170% instead of -200% . As a result, the imposition of a long-only constraint is less binding than when the factor portfolios are pure.

Alternatively, suppose that subportfolio B has a negative exposure to factor A of -0.2 , so the matrix of subportfolio factor exposures is

$$\mathbf{B}_F = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 1.0 & 1.0 & 0.0 \\ 1.0 & -0.2 & 1.0 \end{bmatrix}. \quad (16)$$

With the values shown in Equation 16, the negative exposure of subportfolio A to factor B must now be offset by the weight for subportfolio B . The required weight in Equation 15 is $w_A = 180\%$, so the required weight for the market portfolio is -230% instead of -200% . In other words, the imposition of a long-only constraint is more binding than when the factor portfolios are pure.

In summary, our analysis of the market-plus-one-factor and market-plus-two-factor cases shows that the reduction in the expected Sharpe ratio of portfolio F (i.e., a long-only combination of factor subportfolios) increases with

1. the number of nonmarket factors,
2. the magnitude of factor information ratios,
3. lower levels of nonmarket factor risk,
4. negative correlations between nonmarket factor returns, and
5. negative correlations between nonmarket factor exposures.

As more nonmarket factors are considered, reductions in the potential factor Sharpe ratio of portfolio F become a complex function of the assumed correlations between factor returns, as well as any secondary exposures within the factor subportfolios. Alternatively, as we show in the next section, the optimal long-only security portfolio S still captures most of the potential factor Sharpe ratio by assigning larger weights to securities that simultaneously have high exposures to multiple factors.

Ex Ante Empirical Results for 2016

Here we illustrate the portfolio theory developed in Appendix A and reviewed in the last section, with

data from the CRSP and Compustat at the beginning of calendar year 2016. In our study, we constructed both unconstrained and long-only-constrained portfolios, using individual securities and factor subportfolios, and examined the reduction in factor Sharpe ratios caused by long-only constraints. Specifically, we used the market-cap weights of the largest 1,000 stocks as well as data on four stock characteristics: 60-month market beta, negative log market capitalization, book-to-market ratio, and 11-month price momentum.

Table 2 provides summary statistics on the values that populate the last four columns of the 1,000-by-5 factor exposure matrix **B** in Equation 2. Note that the first column of matrix **B** (not reported in Table 2) is populated by all 1s because all securities have unitary exposure to the market factor. The nonmarket factor exposures are adjusted to have cap-weighted averages of 0.000 (first row of Table 2) and are standardized to have cross-sectional variances of 1.000 (third row of Table 2). For example, the equally weighted mean of 1.656 for the small-cap factor (second row of Table 2) is due to the highly skewed nature of that factor's exposures. The lower half of Table 2 reports the correlations of the factor exposures across the 1,000 stocks. At the beginning of 2016, high-value securities tended to have low momentum exposure, as shown by the relatively large negative correlation of -0.313 . Similarly, low-beta securities tended to be larger stocks, as shown by the negative correlation of -0.173 with the small factor.

Although the factor exposures summarized in Table 2 are dictated by the set of securities available to investors in 2016, different investors will have

different views on the expected factor returns, risks, and correlations. **Table 3** reports an investor's view of the factor returns and risks at the beginning of calendar year 2016. Specifically, the return on the market in excess of the risk-free rate is expected to be 6.00%, with a risk of 15.00% and a Sharpe ratio of 0.400. The four other factors have expected active returns (in excess of the market) and active risks that yield information ratios of 0.300 for the small and value factors and 0.200 for the momentum factor.

The small and value factors have the same information ratio, but the active return of the small factor is half the magnitude of that of the value factor. The expected active return on the low-beta factor is zero, but the low-beta factor is assumed to have a negative correlation of -0.500 with the market, providing potential for that factor to hedge market exposure. Given the set of investor expectations in Table 3, the maximum possible factor Sharpe ratio for an actively managed portfolio, calculated directly from Equation 4, is

$$\left(\frac{(0.400 - 0.500 \times 0.000)^2}{1 - 0.500^2} + 0.300^2 + 0.300^2 + 0.200^2 \right)^{1/2} = 0.658,$$

compared with a Sharpe ratio of 0.400 for the market portfolio.

Before examining the more common practice of combining factor subportfolios that have secondary exposures, we consider the simpler case of using "pure" factor portfolios. As specified in Appendix A, pure factor portfolios generally require some shorting of individual securities, as shown in the last row

Table 2. Statistics on Nonmarket Factor Exposures

	Low Beta	Small	Value	Momentum
Cap-weighted average	0.000	0.000	0.000	0.000
Equal-weighted average	-0.279	1.656	0.093	-0.104
Standard deviation	1.000	1.000	1.000	1.000
<i>Exposure correlations</i>				
Low beta	1.000	-0.173	-0.131	0.102
Small	-0.173	1.000	0.083	-0.028
Value	-0.131	0.083	1.000	-0.313
Momentum	0.102	-0.028	-0.313	1.000

Table 3. Investor Views on Expected Factor Returns

	Active Return	Active Risk	Information Ratio
Low beta	0.00%	5.00%	0.000
Small	0.75	2.50	0.300
Value	1.50	5.00	0.300
Momentum	1.00	5.00	0.200

Notes: Market excess return = 6.00%, risk = 15.00%, and Sharpe ratio = 0.400. Factor returns are assumed to be uncorrelated, except that the low-beta factor has a negative correlation of -0.500 with the market.

Table 4. Pure (Long–Short) Factor Portfolios

	Market	Low Beta	Small	Value	Momentum	Long–Short Optimal “F”
Return	6.00%	6.00%	6.75%	7.50%	7.00%	12.19%
Risk	15.00%	13.23%	15.21%	15.81%	15.81%	18.52%
Sharpe ratio	0.400	0.435	0.444	0.474	0.443	0.658
<i>Factor exposures</i>						
Market	1.000	1.000	1.000	1.000	1.000	1.000
Low beta	0.000	1.000	0.000	0.000	0.000	1.500
Small	0.000	0.000	1.000	0.000	0.000	3.375
Value	0.000	0.000	0.000	1.000	0.000	1.688
Momentum	0.000	0.000	0.000	0.000	1.000	1.125
Weight in F	–668.8%	150.0%	337.5%	168.8%	112.5%	100.0%
N (securities)	1,000	1,000	1,000	1,000	1,000	1,000
Sum of longs	100.0%	114.4%	101.7%	115.8%	115.0%	243.1%
Sum of shorts	0.0%	–14.4%	–1.7%	–15.8%	–15.0%	–143.1%

Note: Nonmarket portfolio exposures to the primary factor of interest are in boldface.

of **Table 4**. But the exposure to the factor of interest in these portfolios is exactly 1.000 and the exposure to all other nonmarket factors is exactly 0.000, as shown in **Table 4**. The unconstrained optimal weights for the pure subportfolios in the combined portfolio *F* (calculated from Equation 3) match the exposure of portfolio *F* to each factor. For example, the optimal weight for the value subportfolio is 168.8%, as reported near the bottom of **Table 4**, and the exposure of portfolio *F* to the value factor is 1.688 (last column of **Table 4**). The positive weights assigned to the four nonmarket factors lead to a large negative weight for the market portfolio but provide the expected portfolio return and factor risk that yield the 0.658 *ex ante* Sharpe ratio predicted by the generalized Treynor–Black result (last column of **Table 4**).

Although pure factor subportfolios have the advantage of avoiding secondary factor exposures, they require shorting and so are more difficult to implement in practice. One natural alternative is to maximize the subportfolio’s Sharpe ratio for the factor of interest under the long-only constraint, as in the previous illustration and as shown in Panel A of **Table 5**. The maximum Sharpe ratio subportfolios in Panel A have the large intended exposures to the primary factor (bolded numbers), but with secondary factor exposures that are also material. The largest secondary exposures are to the small factor in each of the three other nonmarket factor portfolios. The low-beta subportfolio has a small-factor exposure of 1.450, the value subportfolio has a small-factor exposure of 1.810, and the momentum portfolio has a small-factor exposure of 1.799. With a Sharpe ratio of 0.531, the combined portfolio *F* in Panel A captures only $(0.531 - 0.400)/(0.658 - 0.400) = 51\%$ of the unconstrained potential Sharpe ratio improvement.

One result of the large exposures to the small factor (which is assumed to have a positive information ratio) is that the subportfolios in Panel A of **Table 5** have higher expected returns and risks than the primary factor of interest alone would warrant. To illustrate, **Figure 3** shows the same portfolios as **Figure 1** but includes the positions of the pure factor portfolios from **Table 4** as well as the factor portfolios in Panel A of **Table 5**. Another result is that the small subportfolio itself does not come into the long-only solution for portfolio *F* (see the weight of zero for the small subportfolio in Panel A) because ample small-factor exposure is already provided by the other subportfolios. In fact, in an unconstrained solution, portfolio *F* would short both the small subportfolio and the market portfolio to avoid “doubling up” on the small factor.

Panel B of **Table 5** reports on a third factor subportfolio construction methodology: sorting securities into factor exposure quintiles and then forming equally weighted portfolios from the 200 of 1,000 stocks in the largest quintile. To allow meaningful capitalization of the small subportfolio, the top four quintiles of all investable securities (i.e., 800 of 1,000) are included for that factor. As shown in **Table 5**, this subportfolio construction methodology also leads to large small-factor exposures in the other factor subportfolios—and thus no direct exposure to the small subportfolio. The large weighting on the small factor in other factor portfolios may be one of the reasons Blitz (2015) found attractive empirical results for equal-weighted security positions. With a Sharpe ratio of 0.519, the combined portfolio *F* in Panel B captures $(0.519 - 0.400)/(0.658 - 0.400) = 46\%$ of the unconstrained potential Sharpe ratio improvement.

Table 5. Long-Only Factor Subportfolios

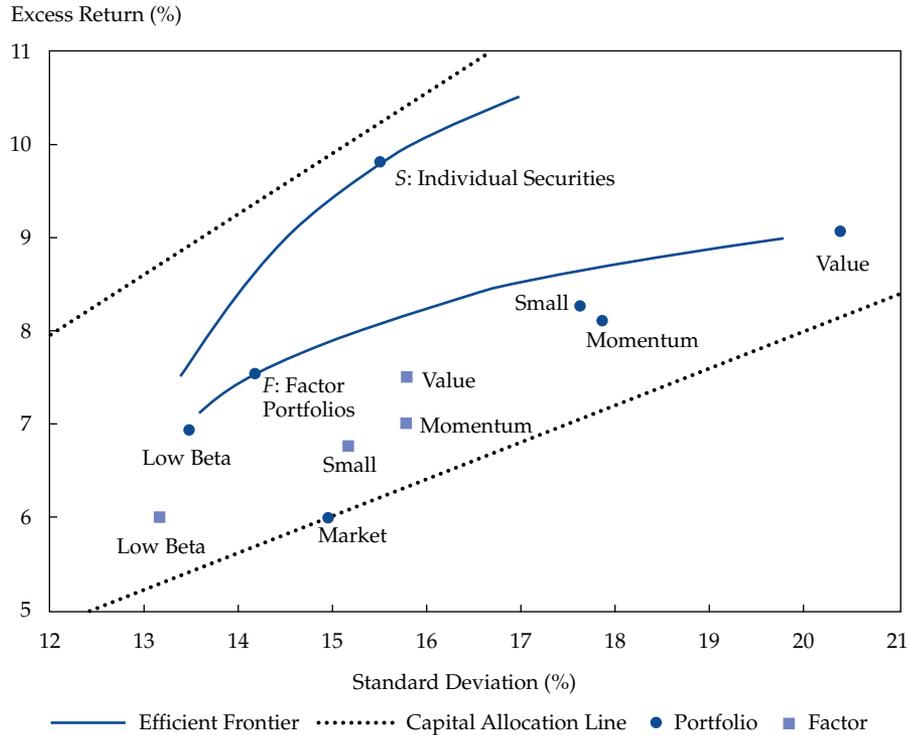
	Low Beta	Small	Value	Momentum	Portfolio F
<i>A. Maximum Sharpe ratio portfolios</i>					
Return	6.92%	8.27%	9.06%	8.11%	7.54%
Risk	13.51%	17.66%	20.41%	17.88%	14.21%
Sharpe ratio	0.512	0.468	0.444	0.453	0.531
Factor exposures					
Market	1.000	1.000	1.000	1.000	1.000
Low beta	1.373	-0.440	-0.705	-0.228	0.701
Small	1.450	2.740	1.810	1.799	1.578
Value	-0.038	0.243	1.802	-0.447	0.255
Momentum	-0.107	-0.152	-0.996	1.431	-0.027
Weight in F	63.9%	0.0%	19.6%	16.5%	100.0%
N (securities)	302	260	243	432	
Effective N	136.7	185.1	64.8	192.9	
<i>B. Equal-weighted quintile portfolios</i>					
Return	6.95%	7.62%	8.66%	8.10%	7.48%
Risk	13.67%	16.78%	18.39%	17.39%	14.40%
Sharpe ratio	0.508	0.454	0.471	0.466	0.519
Factor exposures					
Market	1.000	1.000	1.000	1.000	1.000
Low beta	1.072	-0.351	-0.492	-0.235	0.564
Small	1.464	2.062	1.794	1.784	1.575
Value	-0.097	0.121	1.280	-0.298	0.214
Momentum	-0.004	-0.112	-0.600	1.207	-0.026
Weight in F	65.9%	0.0%	24.1%	10.1%	100.0%
N (securities)	200	800	200	200	
Effective N	200.0	800.0	200.0	200.0	
<i>C. Cap-weighted quintile portfolios</i>					
Return	5.51%	7.27%	7.54%	6.58%	6.81%
Risk	13.23%	16.30%	17.83%	16.50%	15.04%
Sharpe ratio	0.416	0.446	0.423	0.399	0.453
Factor exposures					
Market	1.000	1.000	1.000	1.000	1.000
Low beta	1.070	-0.271	-0.516	0.008	0.047
Small	-0.056	1.699	0.148	0.106	0.716
Value	-0.251	0.074	1.287	-0.520	0.222
Momentum	-0.076	-0.120	-0.497	1.282	-0.062
Weight in F	26.0%	40.3%	23.9%	9.8%	100.0%
N (securities)	200	800	200	200	
Effective N	48.7	547.4	35.1	30.8	

Note: Nonmarket portfolio exposures to the primary factor of interest are in boldface.

Panel C of Table 5 reports on a fourth example of factor subportfolio construction methodology: cap-weighted quintile sorts. Cap-weighted factor portfolios are commonly used in practice because of automatic rebalancing and general equity market representation. Cap-weighted factor portfolios also come closer to pure long–short factor representations, as shown by the relatively large values for

the bolded exposures in Panel C of Table 5, and generally have low off-diagonal or secondary factor exposures. As a result, the long-only-constrained combination for portfolio F has positive weights for all four factor subportfolios, in contrast to the maximum Sharpe ratio portfolios and the equal-weighted quintile portfolios. But the combined portfolio F in Panel C of Table 5 captures only

Figure 3. Factor Portfolios and Pure Factors



$(0.453 - 0.400) / (0.658 - 0.400) = 21\%$ of the potential Sharpe ratio improvement.

We now turn to the issue of constructing portfolio S from individual securities. As with subportfolio construction, there are different methodologies for selecting and weighting the securities. Table 6 reports on several alternatives, along with the market portfolio in the first column. The second column shows the long-only implementation of Equation 2, selecting from all 1,000 securities, as seen in the earlier illustration.³ The maximum Sharpe ratio portfolio S contains 117 securities and has substantial exposures to all the nonmarket factors—and, despite being long only, obtains an

ex ante Sharpe ratio of 0.632, capturing $(0.632 - 0.400) / (0.658 - 0.400) = 90\%$ of the maximum potential. The two other security portfolio construction methodologies in Table 6 are the equal and capitalization weighting of the 200 out of 1,000 securities (i.e., top quintile) with the highest weights in Equation 2. Both have positive exposures to all the nonmarket factors, although not as high as the maximum Sharpe ratio security portfolio, and slightly lower *ex ante* Sharpe ratios of 0.583 (71% capture) and 0.590 (74% capture), respectively. The final security portfolio in Table 6 uses the long-short weights specified in Equation 2 and thus achieves the maximum potential Sharpe ratio of

Table 6. Security Portfolios in 2016

	Market	Maximum Sharpe	Top Quintile Equal Weight	Top Quintile Cap Weight	Long-Short Optimal
Return	6.00%	9.81%	9.02%	9.14%	12.19%
Risk	15.00%	15.52%	15.48%	15.48%	18.51%
Sharpe ratio	0.400	0.632	0.583	0.590	0.658
<i>Factor exposures</i>					
Market	1.000	1.000	1.000	1.000	1.000
Low beta	0.000	1.014	0.494	0.554	1.500
Small	0.000	2.261	2.271	2.302	3.375
Value	0.000	0.972	0.614	0.637	1.688
Momentum	0.000	0.653	0.394	0.456	1.125
N (securities)	1,000	117	200	200	1,000
Effective N	171.7	40.7	200.0	188.7	

0.658, as predicted by the generalized Treynor–Black result in Equation 4.

To summarize, under the assumed investor views in Table 3, the maximum possible *ex ante* Sharpe ratio is 0.658, compared with a passive 0.400 Sharpe ratio for the market portfolio. Given the factor exposures built into US stocks at the beginning of 2016, the cap-weighted top-quintile security portfolio (next-to-last column of Table 6) has an *ex ante* Sharpe ratio of 0.590, capturing $(0.590 - 0.400) / (0.658 - 0.400) = 74\%$ of the potential improvement in the Sharpe ratio over the market portfolio. In contrast, a long-only portfolio built from cap-weighted top-quintile factor subportfolios (Panel C of Table 5) has an *ex ante* Sharpe ratio of 0.453, capturing only $(0.453 - 0.400) / (0.658 - 0.400) = 21\%$ of the potential improvement over the market portfolio. Our *ex ante* analysis for 2016 assumes that the long-only investor is aware of any secondary exposures of the factor subportfolios and takes them into account in maximizing the Sharpe ratio of portfolio *F*. Under more *ad hoc* methodologies for combining the factor subportfolios, the capture of potential Sharpe ratio improvement would be even lower.

Considerations for Idiosyncratic Risk. The theoretical and empirical results up to this point have ignored the impact of security-specific or idiosyncratic risk, focusing solely on Sharpe ratios driven by factor exposures. The focus on expected return to factor exposures rather than the alphas of individual securities is consistent with the philosophy of factor-based investing, but even well-diversified security portfolios retain some idiosyncratic risk. To understand the potential impact of idiosyncratic risk, consider the hypothetical example of an equally weighted portfolio of 100 securities, each with idiosyncratic risk of 20%. Because idiosyncratic risks are by definition uncorrelated, the calculation of the portfolio's idiosyncratic risk in this case is simply $0.50 / (100)^{1/2} = 2.00\%$. Given a relatively low assumed factor risk of 15.00%, total portfolio risk

would be $(0.15^2 + 0.02^2)^{1/2} = 15.13\%$. Coupled with an expected return of 6.00%, the total risk Sharpe ratio is $6.00 / 15.13 = 0.397$, compared with a factor risk Sharpe ratio of $6.00 / 15.00 = 0.400$, verifying that idiosyncratic risk has little practical impact on well-diversified portfolios.

The realized versus *ex ante* impact of idiosyncratic risk is a complex combination of the general level of idiosyncratic risk, the heterogeneity or relative magnitudes of idiosyncratic risk across securities, and the correlations between those magnitudes and the various factor exposures. In addition, the *ex post* or realized performance of optimized portfolios depends on the persistence (i.e., predictability) of the idiosyncratic risks. For example, given homogeneous idiosyncratic risks, the number of securities in a long-only-constrained optimization will monotonically increase with the magnitudes of the idiosyncratic risks. Alternatively, the magnitudes of heterogeneous idiosyncratic risk estimates may be correlated with one of the nonmarket factors, increasing or decreasing the total risk optimal exposure to that factor.

Table 7 provides a brief examination of the 2016 *ex ante* impact of considering idiosyncratic risks in the cap-weighted quintile subportfolio and total portfolio methodologies. The third row of Table 7 reports the estimated idiosyncratic risk for each of the portfolios, under the assumption that idiosyncratic risk is homogeneous at 20% for each of the 1,000 investable securities. The 20% estimate is based on the median observed five-factor idiosyncratic risk from the prior 60 months of returns and the set of factor exposures in 2016. For instance, the small subportfolio's idiosyncratic risk estimate is 0.85% and the value subportfolio's idiosyncratic risk estimate is 3.37%. At these levels, the impact of idiosyncratic risk is generally immaterial to the analysis of factor-based investing. For example, idiosyncratic risk increases the small subportfolio's total risk from 16.30% to only 16.32% and the value subportfolio's total risk from 17.83% to only 18.15%.

Table 7. The Impact of Idiosyncratic Risks

	Market	Low Beta	Small	Value	Momentum	Portfolio F	Portfolio S
Return	6.00%	5.51%	7.27%	7.54%	6.58%	6.81%	9.81%
Factor risk	15.00%	13.23%	16.30%	17.83%	16.50%	15.04%	15.52%
Homogeneous idiosyncratic risk	1.53%	2.87%	0.85%	3.37%	3.60%	1.27%	3.13%
Total risk	15.08%	13.54%	16.32%	18.15%	16.89%	15.09%	15.83%
Total Sharpe ratio	0.398	0.407	0.445	0.416	0.390	0.451	0.619
<i>N</i> (securities)	1,000	200	800	200	200	904	117
Effective <i>N</i>	171.7	48.7	547.4	35.1	30.8	248.9	40.7
Heterogeneous idiosyncratic risk	1.41%	1.93%	1.00%	2.43%	4.96%	1.08%	4.61%

The lower idiosyncratic risk estimate for the small subportfolio in Table 7, compared with the other subportfolios, is largely driven by the fact that 800 rather than 200 individual securities are included in that portfolio, but it is also a function of the distribution of the weights. Under the homogeneous idiosyncratic risk assumption, the impact on the portfolio’s idiosyncratic risk is scaled by the inverse square root of the portfolio’s effective N :

$$\sigma_{\varepsilon,P} = \frac{\sigma_{\varepsilon}}{\sqrt{N_P}}. \tag{17}$$

For instance, Table 7 reports that the idiosyncratic risk estimate for the market portfolio is $20.00/(171.7)^{1/2} = 1.53\%$ and the estimate for portfolio S is $20.00/(40.7)^{1/2} = 3.13\%$, even though the market portfolio has five times as many securities as portfolio S .

If idiosyncratic risk estimates are heterogeneous across securities, an important issue for factor-based investors is how the magnitude of such risks correlates with the various factor exposures. As in most prior years, a regression of the log realized idiosyncratic risk from the prior 60 months on the 2016 factor exposures shows statistically significant tendencies for stocks with greater exposures to the small and momentum factors to have higher idiosyncratic risk and for stocks with greater exposures to low beta and value to have lower idiosyncratic risk. For example, the estimate in Table 7 for the low-beta subportfolio drops from 2.87% for homogeneous idiosyncratic risk to 1.93% for heterogeneous idiosyncratic risk. But the estimate for the small subportfolio increases from 0.85% for homogeneous idiosyncratic risk to

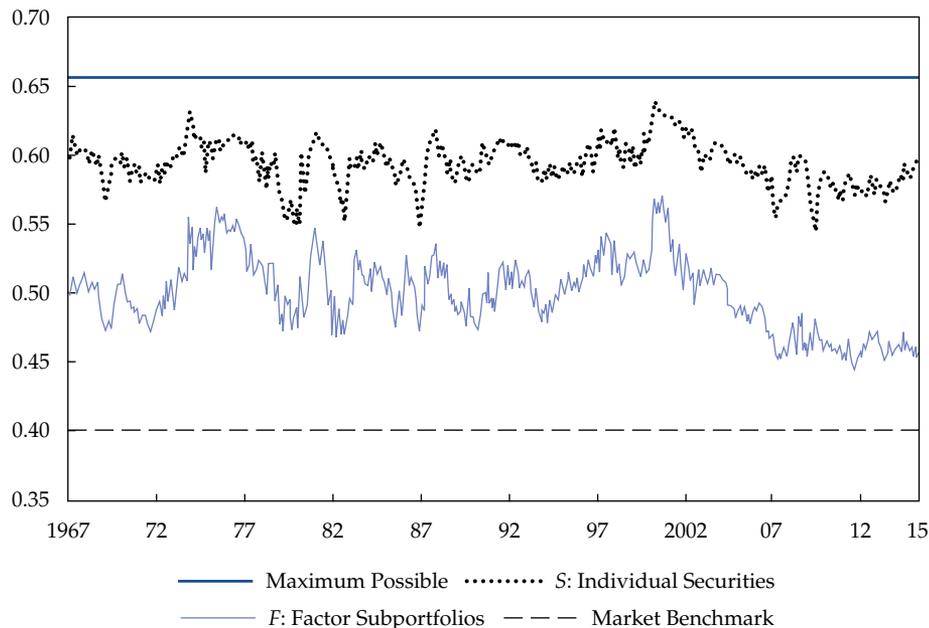
1.00% for heterogeneous idiosyncratic risk. Although the impacts are small, such patterns of heterogeneous idiosyncratic risk would tilt optimal weights slightly away from the small and momentum factors and slightly toward the low-beta and value factors.

Ex Ante and Ex Post Historical Performance, 1968–2015

The 1968–2015 (48-year) track record of security versus factor portfolio investing can be examined in two ways: (1) the evolution of the *ex ante* or expected factor Sharpe ratios at each point in time and (2) simulated *ex post* or realized return performance over the entire history. First, Figure 4 plots the expected factor Sharpe ratio for portfolios S and F at the beginning of each month from January 1968 to December 2015, using the top-quintile cap-weighted specification reported in Panel C of Table 5 and the fourth column of Table 6. Figure 4 illustrates the comparative advantage of constructing long-only portfolios directly from individual securities versus factor subportfolios. The *ex ante* factor Sharpe ratio for portfolio S plots much closer than portfolio F ’s ratio to the unconstrained long–short or maximum possible factor Sharpe ratio of 0.658. On average, the additional factor Sharpe ratio capture of portfolio S over the market Sharpe ratio is about 80%, compared with about 40% for portfolio F .

Note that as an actual historical record of factor-based investing, Figure 4 makes the implausible assumption that starting back in the 1960s, investors knew about the equity factors that are now widely

Figure 4. Ex Ante Sharpe Ratios, 1968–2015



followed and had some perspective on the magnitude of their information ratios. We used constant information parameter values, as given in Table 3, so changes in the *ex ante* factor Sharpe ratios for *S* and *F* are due solely to changes in the factor exposures in the underlying securities over time. Although some of the changes may also be related to the cross-sectional correspondence of factor exposures to the capitalization weights in the benchmark portfolio, the key drivers are changes in the cross-sectional correlations in the factor exposure matrix.

Table 2 shows that at the beginning of calendar year 2016, large-cap US stocks with high value exposure tended to have low momentum exposure, as measured by the cross-sectional correlation value of -0.313 . Similarly, stocks with low beta tended to be large rather than small, as measured by the correlation value of -0.173 . As discussed earlier, negative correlations between factor exposures make long-only combinations of factor subportfolios less able to preserve factor premiums, because the constraint against shorting becomes more binding. Indeed, Figure 4 suggests that the gap in expected factor Sharpe ratios between portfolios *S* and *F* in the US equity market is currently (2016) as large as at any point in the 48-year history. As in Table 5, versions of Figure 4 for the maximum Sharpe ratio and equally weighted quintile portfolios over 1968–2015 also show higher factor Sharpe ratio capture for portfolios *F* and *S*, with a significant and persistent gap between them.

Second, the long-term historical data from the CRSP and Compustat also allow for a backtest or simulation of the *ex post* performance of various portfolio construction methodologies, with the same caveat about what investors would have known back in the 1960s. Although a host of details in the simulation can affect the results, we focused on constant factor return expectations over time. **Table 8** reports the performance of portfolios *F* and *S*, rebalanced monthly using the three long-only construction methodologies in Tables 5 and 6. Specifically, the last two columns of Table 8 compare a version of portfolio *F* that combines cap-weighted top-quintile factor

subportfolios and a version of portfolio *S* constructed by cap weighting the top quintile of stocks that score high on all four nonmarket factors simultaneously.

As Table 8 shows, the market (i.e., the Russell 1000 proxy) excess return over 1968–2015 was 5.73%, with a standard deviation of 15.56% and a realized Sharpe ratio of 0.368, compared with the expected 0.400 market Sharpe ratio incorporated into the simulation. The cap-weighted quintile portfolio *S* (last column of Table 8) has an excess return of 9.51%, compared with portfolio *F*'s 7.61%, with slightly lower realized risk. Although the focus in Table 8 is on the cap-weighted quintile portfolios, the difference in realized Sharpe ratios between portfolios *F* and *S* is not quite as pronounced in the maximum Sharpe ratio and equal-weighted quintile portfolios. The two other subportfolio construction alternatives thus provide *ex post* support for equal-weighted methodologies, at least over 1968–2015, when the larger implicit exposure to the small factor paid off.

Returning to the cap-weighted quintile portfolios in Table 8, we see that the realized active risk (“tracking error”) is higher for portfolio *S* than for portfolio *F*, but the realized information ratio, which adjusts for active risk, is 0.669, much higher than the realized information ratio of 0.417 for portfolio *F*. Alternatively, a $5.66/6.98 = 81\%$ weighted combination of portfolio *S* and the market portfolio would yield a $0.81 \times 4.67 = 3.78\%$ alpha, compared with a 2.36% alpha for portfolio *F*, on the basis of an equal realized tracking error. But as discussed earlier, measuring risk-adjusted performance in this way ignores the optimal level of active risk in portfolio construction. Specifically, a risk-averse investor could simply relever portfolio *S* to match the total risk of portfolio *F*, which would *increase* the realized excess return of portfolio *S*.

Sensitivity analysis of the simulations reported in Table 8 could adjust for a wide variety of specifications, including investor views on factor information ratios, dynamically changing factor and market risks, and the estimation process for idiosyncratic risks. Indeed, backtest parameters are often adjusted in practice to control turnover and to balance the best Sharpe ratio against other characteristics, such

Table 8. Ex Post Performance, 1968–2015

	Market	Maximum Sharpe Ratio		Equal-Weighted Quintile		Cap-Weighted Quintile	
		<i>F</i>	<i>S</i>	<i>F</i>	<i>S</i>	<i>F</i>	<i>S</i>
Average	5.73%	7.40%	9.80%	7.63%	10.26%	7.61%	9.51%
Std. dev.	15.56%	13.40%	15.02%	13.22%	15.27%	15.35%	14.89%
Sharpe ratio	0.368	0.552	0.652	0.577	0.672	0.496	0.639
Market beta	1.000	0.741	0.788	0.760	0.877	0.917	0.845
Market alpha	0.00%	3.15%	5.29%	3.28%	5.24%	2.36%	4.67%
Tracking error	0.00%	6.83%	8.69%	5.93%	6.85%	5.66%	6.98%
Information ratio	0.000	0.462	0.608	0.554	0.764	0.417	0.669

as drawdown risk and benchmark tracking error. Because the focus of our study was on *ex ante* portfolio construction and the *relative* performance of two general portfolio structures, we did not explore the broad set of specifications that could be backtested to produce the best historical results.

Conclusion

Does a long-only-constrained investor with views about equity market factor returns and risks but no views about individual security returns still need a portfolio constructed directly from individual stocks? In this article, we have shown that the answer is yes, both theoretically (*ex ante*) and empirically (*ex post*). The mathematics of mean–variance portfolio theory indicates that the long-only constraint applied to a large portfolio of securities only slightly reduces the potential efficiency from exploiting nonmarket factors. Even when the investor has no views on security alphas, a well-constructed security portfolio has the flexibility, or “degrees of freedom,” needed for a near-optimal simultaneous exposure to the underlying factors. In contrast, the additional layer of constraints in combining factor-replicating subportfolios materially reduces mean–variance efficiency. Ironically, the reduction in *ex ante* Sharpe ratios increases with the number of nonmarket factors and the magnitude of their *ex ante* information ratios, and a belief in several nonmarket factors with positive active returns is probably the reason for deviating from the passive market portfolio in the first place.

The implications for the evolution of prepackaged portfolios, such as multifactor exchange-traded funds, depend on the assumed correlation structure of factor exposures across the securities. In the equity market, a large set of securities with differential exposures to various factors allows the investor to construct security portfolios that capture almost all the potential gains, so the loss from using prepackaged factor portfolios can be substantial. In other asset classes, where individual securities fall into categories with similar factor exposures, the loss from using prepackaged portfolios would be less material. For example, a fixed-income investor may face a set of securities in which exposure to one factor, such as duration, is generally associated with exposure to another factor, such as optionality or credit. If the set of investable securities has high cross-sectional correlations with factor exposures, combining them into long-only subportfolios does not impose as much of a reduction in *ex ante* performance.

Multifactor products that have been introduced by various providers in the institutional investment management industry may avoid some of the suboptimality associated with predefined factor portfolio construction. Specifically, a version of Equation 2

can be used to establish security weights that optimize across multiple factors simultaneously. The end investor, however, would need to be aware of the secondary as well as primary factor exposures in the product and the assumed structure of the idiosyncratic risks. For instance, a product optimized to the small and value factors could not be combined with a product optimized to the momentum and recently popularized quality (Novy-Marx 2013) factors without facing the same suboptimality issues associated with combining single-factor products. In other words, the prepackaged portfolios would need to measure *all* the factors that the end investor views as relevant. In addition, the weighting of the underlying factors in a multifactor product would need to change over time with changes in the investor’s view on the magnitude of the various factor information ratios.

The use of factor subportfolios may still have advantages for some investors. Factor portfolios come prepackaged, and the process of subsequently constructing a multifactor portfolio from subportfolios is generally less complex and easier to explain than constructing a multifactor portfolio directly from individual securities. The level of sophistication in portfolio theory needed to isolate factor returns and move them from one portfolio construct to another is typically associated with institutional investors that directly manage portfolios of securities, rather than individual investors. The performance of the separate factor subportfolios is easier to track and facilitates performance attribution in multifactor combinations. Separate factor subportfolios allow the end investor to easily select the factors thought to be most relevant and also allow for a simple reweighting of prepackaged subportfolios over time based on changes in the investor’s views on expected factors. In the end, the results of our study suggest that such simplicity comes at a high cost when the factor subportfolios are combined in a long-only setting.

Because the dominant factor in equity portfolios is market risk, equity index futures may provide an alternative way to reconfigure the factor exposures in predefined portfolio products. Although many investors have policy or other constraints that restrict shorting of individual securities or nonmarket subportfolios, equity index futures that track the market portfolio are highly liquid and a short futures position could be used to “squeeze out” enough market exposure that the nonmarket factors would have closer-to-optimal exposures. But if some security shorting is allowed (e.g., long–short 120/20 portfolios), the results in Table 4 suggest that pure factor subportfolios allow for the conceptual simplicity of weights on factor subportfolio combinations being equal to the desired factor exposures. Thus, replacing

the long-only constraint with a long–short 120/20 constraint could capture most of the remaining factor Sharpe ratio potential in multifactor security portfolios.



Appendix A. Multifactor Portfolio Theory

When an investor can use risk-free cash to lever or delever the final result, utility theory does not affect the construction of optimal risky asset portfolios. The objective is to find the unique portfolio along the efficient frontier that maximizes the Sharpe ratio. The well-known solution to the Sharpe ratio–maximizing portfolio is

$$\mathbf{w}_P = \left(\frac{\sigma_P^2}{\mu_P} \right) \boldsymbol{\Omega}^{-1} \boldsymbol{\mu}, \quad (\text{A1})$$

where \mathbf{w}_P is an N -by-1 vector of optimal asset weights, $\boldsymbol{\Omega}$ is the forecasted N -by- N asset return covariance matrix, and $\boldsymbol{\mu}$ is the N -by-1 vector of forecasted asset returns in excess of the risk-free rate. The intuition behind Equation A1 is that optimal asset weights increase with forecasted returns via the asset return vector and decrease with risk via the inverse covariance matrix. The scalar multiplier (in parentheses in Equation A1) is given in terms of the optimal portfolio’s risk and expected return, although the asset weights can be calculated without those values. The raw weights specified by $\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}$ can be divided by the scalar value $\boldsymbol{\mu}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}$ to ensure that the weights sum to 100%.

Additional structure for the general mean–variance solution in Equation A1 requires a specification of the asset return–generating process. The well-known linear factor model assumes that the realized excess return on each asset i is a linear function of K common risk factors:

$$r_i = \alpha_i + \beta_{i,1} R_1 + \dots + \beta_{i,K} R_K + \varepsilon_i, \quad (\text{A2})$$

where the R_j are realized returns on factors $j = 1$ to K , the $\beta_{i,j}$ are exposures of asset i to factor j , and ε_i is the idiosyncratic risk of asset i . We assume that α_i is zero for all N assets (i.e., no security-specific alphas) in order to focus on factor returns. Under that assumption, the vector of expected asset returns is simply

$$\boldsymbol{\mu} = \mathbf{B} \mathbf{U}, \quad (\text{A3})$$

where \mathbf{B} is the N -by- K matrix of asset factor exposures and \mathbf{U} is a K -by-1 vector of forecasted factor

returns. Given the linear factor model in Equation A2, the asset return covariance matrix is

$$\boldsymbol{\Omega} = \mathbf{B} \mathbf{V} \mathbf{B}' + \boldsymbol{\Delta}, \quad (\text{A4})$$

where \mathbf{V} is a K -by- K factor return covariance matrix, and $\boldsymbol{\Delta}$ is a diagonal matrix of idiosyncratic return variances. Without loss of generality, we assume that all but the first (i.e., market exposure) column vectors in \mathbf{B} have been standardized to have a cross-sectional mean of 0 and a standard deviation of 1 (note that the first column of \mathbf{B} is filled with 1s). In other words, the variance of each nonmarket factor exposure has been subsumed by the factor return covariance matrix, \mathbf{V} .

Inserting Equations A3 and A4 into Equation A1 gives the optimal asset weight vector as

$$\mathbf{w}_P = \left(\frac{\sigma_P}{\text{SR}_P} \right) (\mathbf{B}\mathbf{V}\mathbf{B}' + \boldsymbol{\Delta})^{-1} \mathbf{B} \mathbf{U}. \quad (\text{A5})$$

The matrix inverse in Equation A5 can be analytically derived and used for unconstrained long–short security portfolios, although numerical optimization is required when a long-only constraint is imposed.⁴ In our study, we focused on *factor* Sharpe ratio maximization because the source of expected returns is factor exposures rather than security-specific alphas. Thus, eliminating the diagonal matrix $\boldsymbol{\Delta}$ in Equation A5 gives asset weights that maximize the expected return from factor exposures divided by the risk from factor exposures.

Without $\boldsymbol{\Delta}$, the remaining N -by- N matrix $\mathbf{B}\mathbf{V}\mathbf{B}'$ in Equation A5 has rank $K < N$ and is thus not directly invertible. However, an inverse can be derived by using the Penrose (1955) left-side and right-side inverse of a nonsquare matrix: $\mathbf{B}^{-1} = (\mathbf{B}'\mathbf{B})^{-1} \mathbf{B}'$ and $\mathbf{B}'^{-1} = \mathbf{B}(\mathbf{B}\mathbf{B}')^{-1}$, commonly called “Moore–Penrose pseudoinverses.” Applying these matrix identities to Equation A5 and given that the inverse of a matrix product is the product of the inverses in reverse order, we obtain the factor Sharpe ratio–maximizing portfolio

$$\mathbf{w}_P = \left(\frac{\sigma_P}{\text{SR}_P} \right) \mathbf{B}(\mathbf{B}\mathbf{B}')^{-1} \mathbf{V}^{-1} \mathbf{U}, \quad (\text{A6})$$

where $\mathbf{B}\mathbf{B}'$ is the K -by- K factor exposure covariance matrix. Although exact, the analytic solution in Equation A6 is not unique, so a numerical search requires additional restrictions, such as the long-only constraint or the reintroduction of homogeneous idiosyncratic risks. Although almost all large-scale optimizations in practice require a numerical search, the analytic solution in Equation A6 provides an intuitive structure for factor-optimal weights and is a key step in establishing the Treynor–Black result (which follows).

We use the subscript F to designate an optimal combination of K factor subportfolios, as opposed to an optimal combination of N securities. Relabeling Equation A6, we obtain the K -by-1 vector of Sharpe ratio-maximizing weights for portfolio F :

$$\mathbf{w}_F = \begin{pmatrix} \frac{\sigma_F^2}{\mu_F} \\ \mu_F \end{pmatrix} \mathbf{B}_F (\mathbf{B}'_F \mathbf{B}_F)^{-1} \mathbf{V}^{-1} \mathbf{U}, \quad (\text{A7})$$

where \mathbf{B}_F is a K -by- K matrix of the subportfolio factor exposures. In the absence of constraints, the portfolio of individual securities (Equation A6) and portfolio F (constructed from subportfolios according to Equation A7) have the same squared factor Sharpe ratio of

$$\text{SR}_P^2 = \mathbf{R}' \mathbf{\Pi}^{-1} \mathbf{R}, \quad (\text{A8})$$

where \mathbf{R} is a K -by-1 vector comprising the market Sharpe ratio and the other factor information ratios.⁵ An important principle, first identified by Treynor and Black (1973), is that under the assumption of uncorrelated factor returns (i.e., $\mathbf{\Pi} = \mathbf{\Pi}^{-1} = \mathbf{I}$), the optimal portfolio's squared Sharpe ratio in Equation A8 is the squared market Sharpe ratio plus the sum of squared information ratios of the nonmarket factors:

$$\text{SR}_P^2 = \text{SR}_M^2 + \sum_{j=2}^K \text{IR}_j^2. \quad (\text{A9})$$

Factor subportfolios can be constructed by using a variety of techniques, whereby some subportfolios are optimal by different criteria and other subportfolios are constructed in an *ad hoc* fashion. For example, Clarke, de Silva, and Thorley (2014) showed that the active (i.e., benchmark-relative) asset weights of "pure" factor subportfolios are given by the columns of

$$\Delta \mathbf{w} = (\mathbf{B} \circ \mathbf{w}_M) [(\mathbf{B} \circ \mathbf{w}_M)' \mathbf{B}]^{-1}, \quad (\text{A10})$$

where \mathbf{w}_M is the vector of market (benchmark) weights and \circ is the matrix dot product. Equation A10 provides factor-replicating subportfolios that are pure in that the matrix product $\Delta \mathbf{w}' \mathbf{B}$, absent the first column and row, forms the $K - 1$ by $K - 1$ identity matrix. That is, the factor portfolios described in Equation A10 have an exposure of exactly 1 to the nonmarket factor of interest and an exposure of exactly 0 to the other factors.

The optimal amount of active risk in a Sharpe ratio-maximizing portfolio can be derived by defining the total return as the portfolio's market exposure (e.g., market beta) times the expected market return plus the active return. The expected active return is in turn defined by the fixed information ratio (collectively from one or more nonmarket factors) times a variable level of active risk:

$$\mu_P = \beta_P \mu_M + \text{IR} \sigma_A. \quad (\text{A11})$$

Similarly, a portfolio's total risk squared in terms of market exposure, market risk, and uncorrelated active risk is

$$\sigma_P^2 = \beta_P^2 \sigma_M^2 + \sigma_A^2. \quad (\text{A12})$$

Combining Equations A11 and A12 gives the squared portfolio Sharpe ratio as

$$\text{SR}_P^2 = \frac{(\beta_P \mu_M + \text{IR} \sigma_A)^2}{\beta_P^2 \sigma_M^2 + \sigma_A^2}. \quad (\text{A13})$$

Maximizing Equation A13—by setting the first derivative with respect to active risk equal to zero and using some algebra in which terms cancel out—gives the optimal level of active risk as

$$\sigma_A = \frac{\text{IR}}{\text{SR}_M} \beta_P \sigma_M. \quad (\text{A14})$$

The optimal level of active risk in Equation A14 increases with the portfolio's *ex ante* information ratio. Intuitively, if the portfolio's *ex ante* information ratio is zero, the optimal level of active risk is zero and the passive market portfolio is optimal. Substituting the optimal level of active risk from Equation A14 back into Equation A13 (and using some algebra in which market exposure drops out) gives a simple version of the Treynor-Black result:

$$\text{SR}_P^2 = \text{SR}_M^2 + \text{IR}^2, \quad (\text{A15})$$

where, again, the portfolio's *ex ante* information ratio may be derived from several factors.

For long-only or other constraints imposed on an optimized portfolio, Clarke, de Silva, and Thorley (2002) introduced the transfer coefficient (TC), which affects the *constrained* portfolio Sharpe ratio (the notation SR with no subscript) according to the formula

$$\text{SR}^2 = \text{SR}_M^2 + (\text{TC} \times \text{IR})^2. \quad (\text{A16})$$

If $\text{TC} = 1$ (i.e., constraints are not binding), then Equation A16 reduces to Equation A15. Combining Equations A15 and A16 and solving for TC gives a transfer coefficient formula based on squared Sharpe ratios:

$$\text{TC} = \left(\frac{\text{SR}^2 - \text{SR}_M^2}{\text{SR}_P^2 - \text{SR}_M^2} \right)^{1/2}. \quad (\text{A17})$$

In the main text of this article, we use the slightly less complicated "percentage capture" of maximum potential Sharpe ratio, $(\text{SR} - \text{SR}_M) / (\text{SR}_P - \text{SR}_M)$, rather than the transfer coefficient in Equation A17, as the *ex ante* measure of a constrained portfolio's mean-variance efficiency.

Notes

1. For a review of information ratios, factor-based portfolios, and the Treynor–Black rule, see Bodie, Kane, and Marcus (2013), especially Chapters 8 and 10.
2. Specifically, the factor returns are estimated from a series of monthly cap-weighted cross-sectional regressions of security returns on standardized factor exposures.
3. As explained in Appendix A, Equation 2 provides exact closed-form optimal security weights for an unconstrained portfolio but does not work well as the basis for long-only numerical optimization because the solution is not unique absent idiosyncratic risk.
4. Based on the matrix inversion lemma, the general form of the inverse covariance matrix in Equation A5 is $(\mathbf{B}\mathbf{V}\mathbf{B}' + \mathbf{\Delta})^{-1} = \mathbf{\Delta}^{-1} - \mathbf{\Delta}^{-1}\mathbf{B}(\mathbf{V}^{-1} + \mathbf{B}'\mathbf{\Delta}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{\Delta}^{-1}$, which can be

used to provide an analytic but fairly complex matrix algebra solution for optimal security weights.

5. The proof uses the matrix definitions of a portfolio's expected return, $\mu_p = \mathbf{w}_p' \mathbf{B}\mathbf{U}$, and variance, $\sigma_p^2 = \mathbf{w}_p' \mathbf{B}\mathbf{V}\mathbf{B}' \mathbf{w}_p$, and the weights in Equation A6. With these substitutions, the terms involving \mathbf{B} cancel out, giving a squared Sharpe ratio of $\mu_p^2 / \sigma_p^2 = \mathbf{U}'\mathbf{V}^{-1}\mathbf{U}$. Equation A8 uses the additional matrix relationship $\mathbf{V} = (\boldsymbol{\sigma}\boldsymbol{\sigma}') \circ \mathbf{\Pi}$, where \circ is the matrix dot product, to isolate the factor correlation matrix $\mathbf{\Pi}$ instead of the factor covariance matrix \mathbf{V} .

References

- Black, Fischer, and Robert Litterman. 1992. "Global Portfolio Optimization." *Financial Analysts Journal*, vol. 48, no. 5 (September/October): 28–43.
- Blitz, David. 2015. "Factor Investing Revisited." *Journal of Index Investing*, vol. 6, no. 2 (Fall): 7–17.
- Bodie, Zvi, Alex Kane, and Alan J. Marcus. 2013. *Investments*, 10th ed. New York: McGraw-Hill.
- Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance*, vol. 52, no. 1 (March): 57–82.
- Chan, Louis K.C., Jason Karceski, and Joseph Lakonishok. 1998. "The Risk and Return from Factors." *Journal of Financial and Quantitative Analysis*, vol. 33, no. 2 (June): 159–188.
- Clarke, Roger, Harindra de Silva, and Steven Thorley. 2002. "Portfolio Constraints and the Fundamental Law of Active Management." *Financial Analysts Journal*, vol. 58, no. 5 (September/October): 48–66.
- . 2010. "Know Your VMS Exposure." *Journal of Portfolio Management*, vol. 36, no. 2 (Winter): 52–59.
- . 2014. "The Not-So-Well-Known Three-and-One-Half-Factor Model." *Financial Analysts Journal*, vol. 70, no. 5 (September/October): 13–23.
- Fama, Eugene F., and Kenneth R. French. 1996. "Multifactor Explanations of Asset Pricing Anomalies." *Journal of Finance*, vol. 51, no. 1 (March): 55–84.
- Fan, Jianqing, Yingying Fan, and Jinchi Lv. 2008. "High Dimensional Covariance Matrix Estimation Using a Factor Model." *Journal of Econometrics*, vol. 147, no. 1 (November): 186–197.
- Frazzini, Andrea, and Lasse Pedersen. 2014. "Betting against Beta." *Journal of Financial Economics*, vol. 111, no. 1 (January): 1–25.
- Grinold, Richard. 1989. "The Fundamental Law of Active Management." *Journal of Portfolio Management*, vol. 15, no. 3 (Spring): 30–37.
- Jegadeesh, Narasimhan, and Sheridan Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance*, vol. 48, no. 1 (March): 65–91.
- Kahn, Ronald N., and Michael Lemmon. 2016. "The Asset Manager's Dilemma." *Financial Analysts Journal*, vol. 72, no. 1 (January/February): 15–20.
- Kritzman, Mark, and Sébastien Page. 2003. "The Hierarchy of Investment Choice." *Journal of Portfolio Management*, vol. 29, no. 4 (Summer): 11–23.
- Ledoit, Oliver, and Michael Wolf. 2003. "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection." *Journal of Empirical Finance*, vol. 10, no. 5 (December): 603–621.
- Markowitz, Harry. 1952. "Portfolio Selection." *Journal of Finance*, vol. 7, no. 1 (March): 77–91.
- Novy-Marx, Robert. 2013. "The Other Side of Value: The Gross Profitability Premium." *Journal of Financial Economics*, vol. 108, no. 1 (April): 1–28.
- Penrose, Roger. 1955. "A Generalized Inverse for Matrices." *Proceedings of the Cambridge Philosophical Society*, vol. 51, no. 3 (July): 406–413.
- Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance*, vol. 19, no. 3 (September): 259–263.
- Treynor, Jack, and Fischer Black. 1973. "How to Use Security Analysis to Improve Portfolio Selection." *Journal of Business*, vol. 46, no. 1 (January): 66–88.