Portfolio Constraints and the Fundamental Law of Active Management

Roger Clarke, Harindra de Silva, and Steven Thorley

Active portfolio management is typically conducted within constraints that do not allow managers to fully exploit their ability to forecast returns. Constraints on short positions and turnover, for example, are fairly common and materially restrictive. Other constraints, such as market-capitalization and value–growth neutrality with respect to the benchmark or economic-sector constraints, can further restrict an active portfolio's composition. We derive ex ante *and* ex post *correlation relationships that facilitate the performance analysis of constrained portfolios. The* ex ante *relationship is a generalized version of a previously developed "fundamental law of active management" and provides an important strategic perspective on the potential for active management to add value. The* ex post *correlation relationship represents a practical decomposition of performance into the success of the return-prediction process and the "noise" associated with portfolio constraints. We verify the accuracy of these relationships with a Monte Carlo simulation and illustrate their application with equity portfolio examples based on the S&P 500 Index as the benchmark.*

Most portfolio managers appreciate the fact that value added ultimately depends on their ability to correctly forecast security returns. Managers work hard to create valuable information about future returns but may not pay as much attention to limitations in the portfolio construction process. Constraints such as no short sales, industry limitations, and restrictions on investment style or turnover—all limit a manager's ability to transfer valuable information into portfolio positions. We introduce a conceptual framework, as well as diagnostic tools, for measuring the impact of constraints on value added. The framework provides an important strategic perspective on where and how managers have the potential to add value. The diagnostic tools measure the degree to which realized performance is attributable to return forecasts versus the "noise" induced by portfolio constraints.

Roger Clarke is chairman of Analytic Investors, Los Angeles. Harindra de Silva, CFA, is president of Analytic Investors, Los Angeles. Steven Thorley, CFA, is professor of finance at the Marriott School, Brigham Young University, Provo, Utah.

Terminology and Notation

The discussion will focus on residual security returns, r_i, the portion of security i's total return that is uncorrelated with the benchmark portfolio. Forecasted residual security returns, α_i, are the portfolio manager's forecast of r_i for each of the $i = 1$ to N securities in the portfolio. Our focus on the residual portion of security returns (both forecasted and realized) is motivated by an adjustment for non-unity betas in managed portfolios, as explained in Appendix A, "Framework and Notation." We also refer to the active weight of a security, Δw_i, which is the difference between the weight of security i in the actively managed portfolio and its weight in the benchmark portfolio. A positive active weight indicates that the security is overweighted in the managed portfolio as compared with the benchmark, and a negative active weight indicates that it is underweighted. Active security weights, which sum to 0 (rather than 1), are a compact way to describe a managed portfolio's benchmark-relative positions.

In addition to active weights for securities, we also use the word "active" to describe the difference between the managed portfolio return and the benchmark portfolio return, with adjustment for the managed portfolio's beta with respect to the benchmark. This portfolio return difference is a summary measure of managed portfolio performance commonly referred to as the active return, R_A. The standard deviation of active return, R_A, is the active risk, σ_A, or tracking error of the managed portfolio.

Figure 1 characterizes the relationships between forecasted and realized residual security returns and active security weights. The base of the triangle represents the value added through active management. Value added is measured by the active return, R_A—that is, the difference between the returns on the actively managed and benchmark portfolios. As shown in Appendix A, R_A is algebraically equivalent to the sum of the products of active weights and residual returns for the stocks in the portfolio:

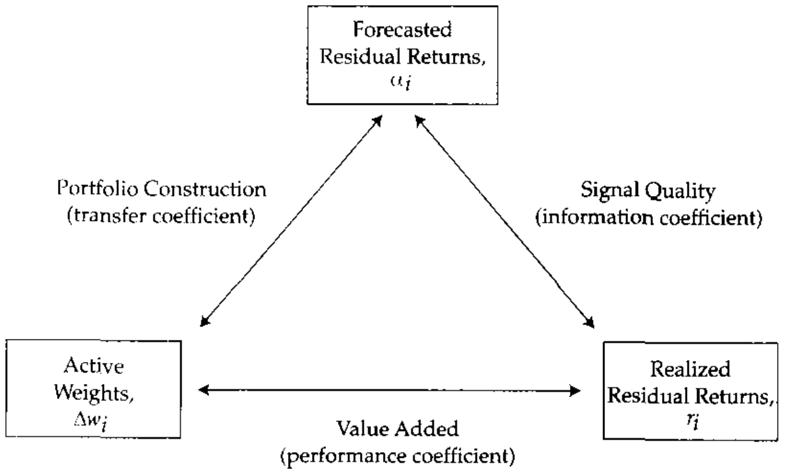
$$R_A = \sum_{i=1}^{N} \Delta w_i r_i. \tag{1}$$

Equation 1 indicates that value is added when positive-active-weight securities have positive residual returns and negative-active-weight securities have negative residual returns. In other words, performance in any given period is related to the cross-sectional correlation between the active security weights and realized residual returns—the security data in the bottom two corners of the correlation triangle in Figure 1.

Although the direct cross-sectional correlation, or "performance coefficient," at the base of the triangle reflects value added, a clearer understanding of the sources and limitations of value added can be obtained by examining the cross-sectional correlations on the two legs. First, there is little hope of value added if the manager's forecasts of returns do not correspond to actual realized returns. Signal quality is measured by the relationship between the forecasted residual returns, or alphas, at the top of the triangle and the realized residual returns at the right corner. This cross-sectional correlation is commonly called "the information coefficient." Managers with a high information coefficient (ability to forecast returns) will add more value over time—but only to the extent that those forecasts are exploited in the construction of the managed portfolio. The second correlation, the relationship between the active weights in the left corner and forecasted residual returns at the top of the triangle, measures the degree to which the manager's forecasts are translated into active weights. We refer to this cross-sectional correlation as "the transfer coefficient."[1]

The transfer coefficient (correlation between active weights and forecasted residual returns) is equal to 1.0 in the absence of constraints in portfolio construction. However, investment managers rarely enjoy the luxury of a completely unconstrained investment portfolio. Portfolio constraints such as no short sales and mandated industry or sector concentrations limit the full transfer of information into active weights and lead to transfer coefficients lower than 1.0. As a result, performance is a function of both signal quality (the right leg of the triangle) and the constraints imposed in the portfolio construction process (the left leg of the triangle).

Figure 1. The Correlation Triangle



The Generalized Fundamental Law

The conceptual framework in Figure 1 can be formalized through a generalized version of Grinold's (1989) fundamental law of active management. We first review the original form of the law; then, we generalize the law to account for portfolio constraints.

The fundamental law describes an *ex ante* relationship between expected performance and the assumed information coefficient, *IC*, of the manager's forecasting process. Expected performance is measured by the *ex ante* information ratio, *IR*, defined as the portfolio's expected active return divided by active risk;[2] that is,

$$IR \equiv \frac{E(R_A)}{\sigma_A}. \tag{2}$$

Appendix A, "The Fundamental Law," contains a proof of the original form of the fundamental law:

$$IR = IC\sqrt{N}, \tag{3}$$

where *IC* is the expected information coefficient and *N* is the breadth, or number of independent "bets" in the actively managed portfolio. In terms of the expected active return, the fundamental law reads

$$E(R_A) = IC\sqrt{N}\sigma_A. \tag{4}$$

As discussed in Appendix A, Equation 4 is based on the operational assumptions that the manager uses mean–variance optimization and an alpha-forecasting process that incorporates individual security risk estimates and the assumed *IC*. For mathematical tractability, however, we make two additional simplifying assumptions. First, we assume a diagonal covariance matrix for residual security returns; that is, we assume that the benchmark portfolio return is the only source of covariance between total security returns. Under this "market model" assumption, *N* is simply the number of securities in the investor's choice set (i.e., the number of securities in the benchmark plus any others with return forecasts). Second, we assume that no budget constraint is imposed in the optimization problem, an assumption that allows the active weight for each security to be perfectly proportional to its risk-adjusted forecasted return. This second assumption does not appear to be a serious limitation for portfolios with a large number of securities and typical residual risk–return parameters. In the next section, we test the materiality of these two simplifying assumptions with data on the S&P 500 Index.

Grinold (1989) readily acknowledged the approximate nature of the fundamental law and presented it as a strategic tool. Thomas (2000) provided intuitive support for the strategic perspectives that come from the fundamental law. The important lesson of the fundamental law is that breadth of application, as well as the quality of the signal, dictates the value expected to be added through active management.

A weakness of the traditional form of the fundamental law in Equation 3 is the assumption that the portfolio manager can take active weights that fully exploit the return-forecasting process. This assumption is explicit in Grinold's (1989) original derivation, where he stated that the law "gives us only an upper bound on the value we can add" because "we presume that we can pursue our information without any limitations" (p. 33). Goodwin (1998) also emphasized that the original fundamental law provides an upper bound on potential information ratios. Unfortunately, portfolio managers sometimes use the law without acknowledging this fact and then wonder why realized information ratios are only a fraction of their predicted value. Indeed, a common rule of thumb in practice is that the theoretical information ratio suggested by the fundamental law should be cut in half. This rule of thumb is sometimes an implicit admission that the true signal quality, or *IC*, is below the manager's assumed value, but much of the reduction in performance is simply the result of constraints in the portfolio construction process.

We now generalize the fundamental law to incorporate a precise measure of how constraints affect value added. Appendix A, "Framework and Notation," shows that under the previously mentioned simplifying assumptions, unconstrained mean–variance portfolio optimization leads to active weights for each security that are proportional to risk-adjusted forecasted residual returns. Specifically, the unconstrained optimal active weight on security *i* is given by

$$\Delta w_i^* = \frac{\alpha_i}{\sigma_i^2}\frac{1}{2\lambda}, \tag{5}$$

where σ_i^2 is the security's residual-return variance and λ is a risk-aversion parameter. Because the risk-aversion parameter is the same for all securities (λ is not subscripted), a perfect cross-sectional correlation exists between the unconstrained optimal active weights, Δw_i^*, and the risk-weighted forecasted residual returns, α_i/σ_i^2. The full information content of the return forecasts is transferred into active weights with no reduction in the potential for value added.

In practice, active managers are often subject to constraints that cause them to deviate from the unconstrained optimal weights given by Equation 5. Searching algorithms are available that can determine optimal weights under constraints, although the results do not generally have a closed-form solution. Define Δw_i (without the asterisk) to be a set of active weights from some portfolio construction process that is subject to one or more constraints. The transfer coefficient, *TC*, is the cross-sectional correlation coefficient between risk-adjusted active weights and risk-adjusted forecasted residual returns for the $i = 1$ to N securities in the portfolio. Based on this definition for the transfer coefficient, Appendix A presents a proof of the generalized fundamental law, which is

$$IR \approx TC\ IC\sqrt{N}, \qquad (6)$$

or in terms of the expected active return,

$$E(R_A) \approx TC\ IC\sqrt{N}\sigma_A. \qquad (7)$$

Like the information coefficient, the transfer coefficient acts as a simple scaling factor in the determination of value added.

The use of the approximate equality notation ≈ in Equation 7 is motivated by an approximation in the proof, as noted in Appendix A. Equation 7 is generalized in the sense that in the original version of the law (Equation 3), *TC* is assumed to be 1.0. In practice, values for *TC* rarely approach 1.0 and, with multiple constraints, can be as low as 0.3.

One can think of the transfer coefficient as an additional adjustment to breadth, *N*, that reflects the reduction in independent bets because of constraints. The best conceptualization of the transfer coefficient, however, is motivated by its mathematical definition; the transfer coefficient measures the degree to which the information in individual security return forecasts is transferred into managed portfolio positions. The transfer coefficient is not the only way to assess the impact of a particular constraint. Given a set of forecasted security returns and an optimization routine, together with an estimated covariance matrix, a manager can calculate the expected information ratio with and without a given constraint and compute the difference.[3] We perform this type of calculation later as a numerical check of Equation 6.

The primary value of the generalized fundamental law is that it provides a strategic framework in which to view the impact of constraints on potential value added. In addition, the transfer coefficient plays a critical role in the *ex post* decomposition of the realized active return that we discuss later.

An alternative formulation of the generalized law in Equation 6 that corresponds to the correlation triangle in Figure 1 is also helpful. Expected value added at the base of the triangle is reflected in the *ex ante* performance coefficient, *PC*. We define *PC* as the expected value of the cross-sectional correlation between risk-adjusted active weights and realized residual returns. Appendix A, "The Fundamental Law," shows that *PC* is equal to the *ex ante* information ratio, *IR*, divided by the square root of *N*. By making this substitution in Equation 6, we can express the generalized fundamental law in correlation coefficient form as

$$PC \approx TC\ IC. \qquad (8)$$

The formulation of the generalized law in Equation 8 is intuitive: The expected correlation of active weights with realized returns, *PC*, is equal to the correlation of active weights with forecasted returns, *TC*, times the expected correlation of forecasted returns with realized returns, *IC*. However, this simple relationship is valid only *ex ante*; the transfer coefficient relates expected performance, *PC*, to expected signal quality, *IC*. Later, we show that realized, or *ex post*, performance is described by a more complicated structure.

Constraint Case Studies

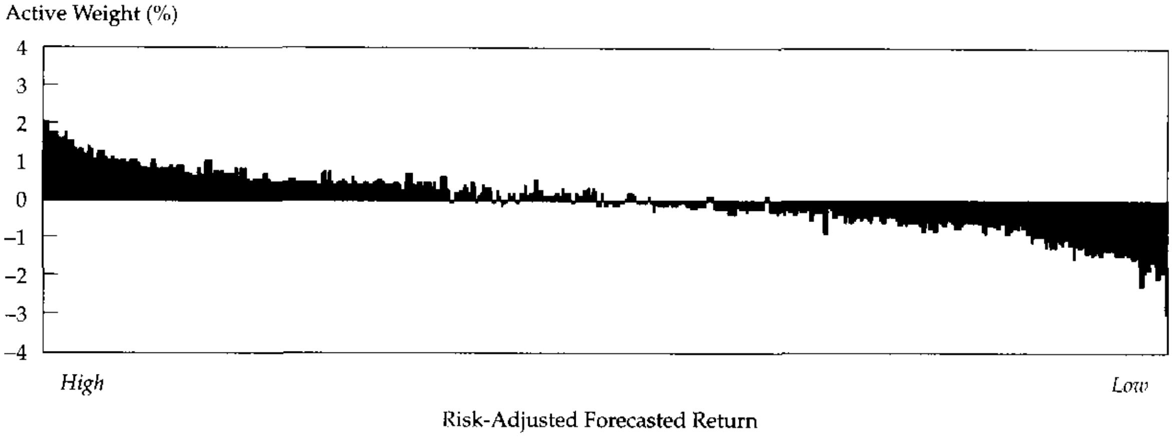
We used the Barra portfolio optimizer and the S&P 500 as the benchmark to create several case studies of the application of the generalized fundamental law.[4] We generated a set of forecasted returns for the 500 stocks in the S&P by using

$$\alpha_i = IC\sigma_i S_i, \qquad (9)$$

where *IC* is an assumed information coefficient, σ_i is the estimated residual-return volatility for each stock, and S_i represents random "scores" drawn from a standard normal distribution. The forecasted returns in this illustration are random, but their relative magnitudes and cross-sectional variations are consistent with actual security volatilities and the assumed quality of the signal.[5] We used an assumed *IC* value of 0.067, so the unconstrained *ex ante* information ratio, according to the fundamental law, has a value of $0.067 \times \sqrt{500} = 1.50$. These expected returns were fed into the optimizer with the condition that active portfolio risk (i.e., tracking error) be limited to 5.0 percent.

As a baseline from which to make comparisons, the first optimization was conducted without any constraints placed on the portfolio except the standard full investment budget constraint. The active weights produced by the optimizer are shown in **Figure 2**, with the active weights sorted from high to low on the basis of each stock's risk-

Figure 2. Unconstrained Optimization
(TC = 0.98)



adjusted forecasted residual return. The active weights in Figure 2 range from about +2.0 percent on the left for the stocks with the highest forecasted risk-adjusted residual returns to less than −2.0 percent on the right for stocks with the lowest forecasted risk-adjusted residual returns.

When the *ex ante* active risk, or tracking error, of the portfolio was calculated, it turned out to be exactly 5.0 percent, as specified to the optimizer. The expected active return, calculated by

$$E(R_A) = \sum_{i=1}^{N} \Delta w_i \alpha_i, \tag{10}$$

is 7.9 percent. Thus, the *ex ante* information ratio is 7.9/5.0 = 1.58, slightly higher than the 1.50 value suggested by the original fundamental law. The discrepancy arises from the simplifying assumptions in the derivation of the law that were not met in the actual optimization. The first and most material cause of the discrepancy is the assumption of a diagonal covariance matrix for residual security returns. Specifically, the proof of the fundamental law assumes that the only source of covariance between total security returns is the benchmark portfolio, an assumption that is violated by the actual risk estimates in the Barra-supplied covariance matrix. Second, no budget constraint was imposed in the derivation of the theoretical results in Equation 5, whereas the optimizer forces the sum of the active weights to be zero.

The weights from the unconstrained optimization routine in Figure 2 provide an important check on the real-world accuracy of the generalized law as well as the original fundamental law. In theory, the active weights should monotonically decline from left to right in perfect correspondence with risk-adjusted expected returns. In fact, small deviations in the pattern occur because of residual covariances among the securities and because the transfer coefficient is 0.98 rather than a perfect 1.00. Small deviations from a perfect transfer of information occur because the optimizer correctly adjusts for the fact that the residual returns of the securities have some correlation. The deviations in a real-world setting are small, however, so the transfer coefficient is almost perfect.

In the sections that follow, we discuss several types of constraints combined with different levels of active portfolio risk. As a preview of the results, **Table 1** provides the transfer coefficient for each case. For example, the results for the unconstrained case just discussed are in the first row.[6] This base case is followed by entries for (1) a long-only constraint with various levels of active risk, (2) long-only and market-cap-neutral constraints with various levels of active risk, (3) turnover limits, and (4) a portfolio with multiple constraints—long-only and market-cap-neutrality constraints, dividend-yield neutrality with respect to the benchmark, and a 50 percent turnover limit—which is similar to what might be found in practice. Notice that the transfer coefficient declines when either the active risk is increased or when additional constraints are added. In either case, the transfer of information into active positions is reduced, lowering the transfer coefficient and decreasing the information ratio. If the decline in the transfer coefficient occurs because of a desired increase in active risk for a given constraint, however, the expected active return may increase. The investor can be rewarded with an increase in expected active return for the additional active risk even though the information ratio declines, which is not the case if the decline in the transfer coefficient is caused by adding additional constraints while trying to maintain the same level of active risk. The additional constraints lower the information ratio as well as the expected active return.

Table 1. Summary of Results for Case Studies

Portfolio Constraint	Active Risk	Transfer Coefficient, *TC*	Information Ratio, *IR*	Expected Active Return
Unconstrained	5.0%	0.98	1.47	7.3%
Long only	2.0	0.73	1.09	2.2
	5.0	0.58	0.87	4.3
	8.0	0.48	0.72	5.8
Long only and market-cap neutral	2.0	0.67	1.00	2.0
	5.0	0.47	0.70	3.5
	8.0	0.37	0.55	4.4
Turnover limit of 50 percent	5.0	0.73	1.09	5.5
Turnover limit of 25 percent	5.0	0.49	0.73	3.7
Multiple constraints[a]	5.0	0.31	0.46	2.3

Note: The information ratios and expected active returns were calculated on the basis of the generalized fundamental law in Equation 6.

[a]Long-only and market-cap-neutrality constraints, dividend-yield neutrality with respect to the benchmark, and turnover limited to 50 percent.

Case Study 1: Long-Only Constraint. The long-only constraint is perhaps the most interesting case study for two reasons. First, the long-only constraint is ubiquitous, so common that it is often not properly acknowledged as a constraint. Second, the long-only constraint is quite material. We found that the transfer coefficient falls more by imposition of the long-only constraint than by any other single restriction, with the possible exception of tight turnover limits.

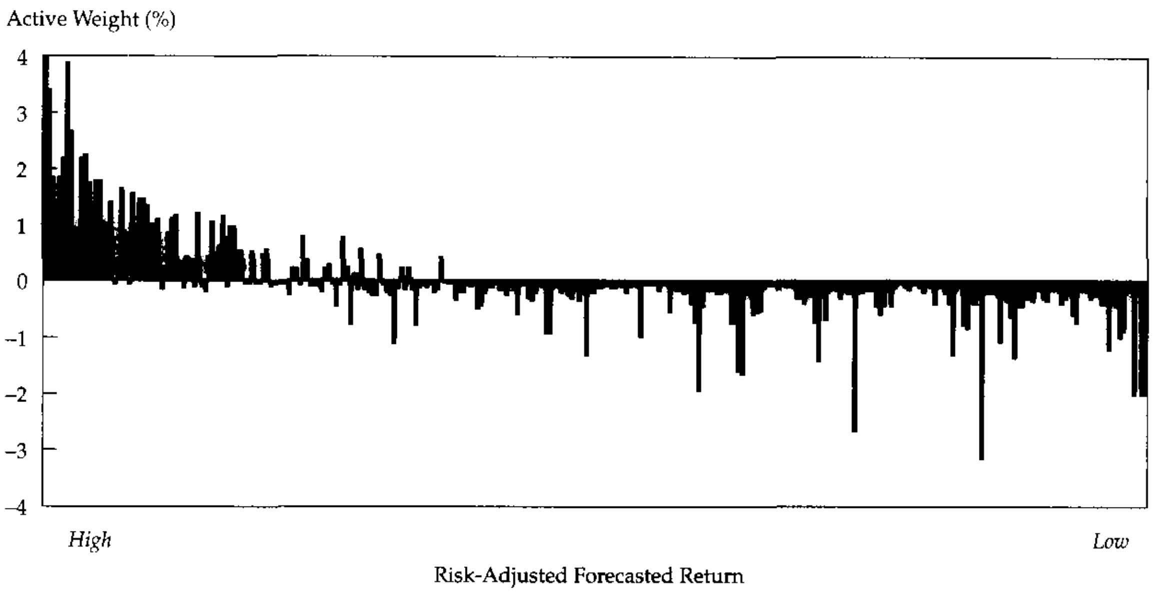
When short sales are prohibited, the manager can reduce the managed portfolio weight of an unattractive security only to zero. Thus, the lowest possible value of a security's active weight is –1 times the benchmark weight. This limit is not binding for securities with large benchmark weights (i.e., large-cap stocks) but is quite restrictive for securities that have smaller benchmark weights. In the extreme, securities with little or no weight in the benchmark cannot receive any measurable negative active weight no matter how pessimistic the manager is about their future returns.

In contrast, when short selling is allowed, securities with large negative forecasted residual returns can be shorted and the short-sale proceeds used to fund long positions in the securities expected to have the highest positive residual returns. In other words, the active weights in a long–short portfolio are primarily determined by the risk-adjusted return forecasts and, therefore, conform more closely than weights in a long-only portfolio to the theoretical optimal active weights in Equation 5.

To analyze the impact of the long-only restriction, we show in **Figure 3** the portfolio of Figure 2 reoptimized with the added constraint that the portfolio weight for each security cannot be less than zero. The same set of forecasted returns was used, and the tracking-error target was again set to 5.0 percent. One notable impact of the long-only constraint is the large number of small negative active weights. In fact, only 89 of the 500 stocks are held in the managed portfolio; the remaining 411 receive negative active weights equal in magnitude to their benchmark weights. The positive active weights are concentrated in the relatively few stocks with the highest forecasted residual returns at the far left of Figure 3. The key point is that the correspondence between forecasted returns and active weights is much weaker in Figure 3 than in Figure 2. As Table 1 indicates, the transfer coefficient calculated by the risk-adjusted cross-sectional correlation between active weights and forecasted residual returns is only 0.58. A *TC* value of 58 percent indicates that 42 percent of the unconstrained potential value added was lost in the portfolio construction process because of the long-only constraint.[7]

The accuracy of the generalized fundamental law can be assessed by calculating the expected active return directly with the use of Equation 10—the sum of the products of active weights taken and forecasted residual returns. The directly calculated expected active return of the long-only portfolio is 4.2 percent, yielding an *IR* of 4.2/5.0 = 0.84, or 56 percent of the total theoretical unconstrained *IR* of 1.50. Thus, the *TC* of 58 percent, together with the

Figure 3. Optimization with Long-Only Constraint (TC = 0.58)



Note: Full height of some bars is not shown.

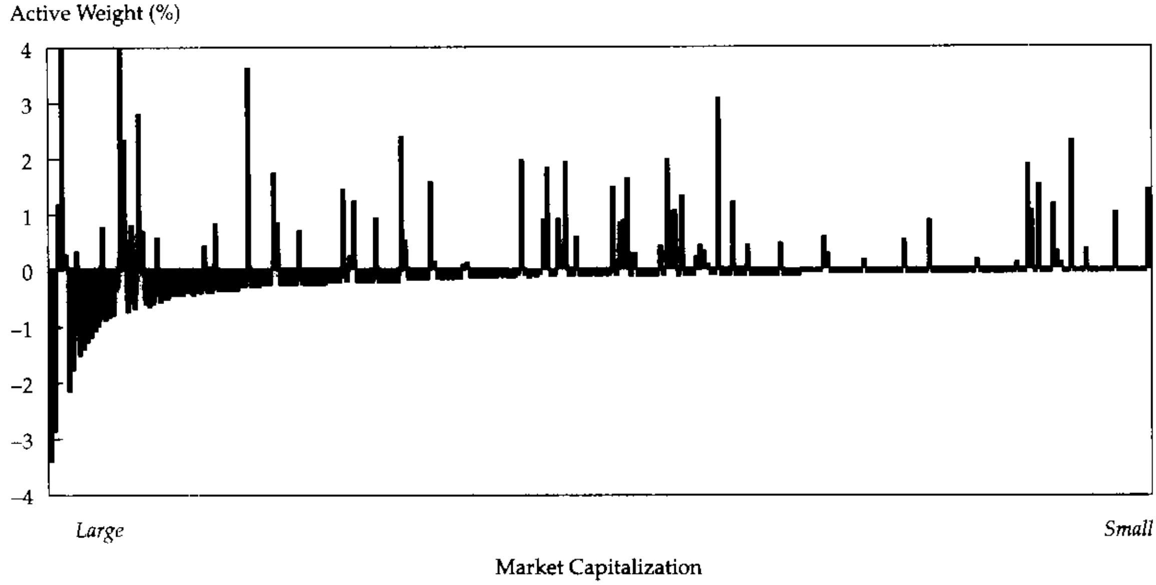
information coefficient and the number of securities, provides a fairly accurate perspective on the potential value added under constraints. The accuracy of the generalized law depends on the degree to which off-diagonal elements in the residual-return covariance matrix are nonzero, but the law appears to be a reasonably accurate description for a large portfolio of securities, such as the S&P 500.

Our findings from using the generalized fundamental law can be compared with the estimated declines in information ratios documented in the examination of long–short versus long-only portfolios in Grinold and Kahn (2000). The relative advantages of long–short portfolios have also been examined by Brush (1997) and by Jacobs, Levy, and Starer (1998, 1999).

Case Study 2: Factor-Neutrality Constraints. Portfolios are often constrained to have characteristics that are similar to the benchmark along one or more dimensions. For example, the managed portfolio may be constrained to have the same style (value–growth) tilt as the benchmark. These constraints can, depending on the investing style and data used to forecast returns, be a material restriction in the portfolio construction process. For example, a growth manager may favor stocks with mid-range P/Es, or "growth at a reasonable price." If the manager is benchmarked against a growth index, however, and is constrained to have the same growth exposure as the index, the positions dictated by the return forecasts based on mid-range P/E will be restricted. Another common neutrality constraint involves the market capitalization of the securities in the managed portfolio compared with the benchmark. Managers, or their clients, might require the managed portfolio to be market-cap neutral—no more sensitive than the benchmark to market-cap exposure.

We used market-cap neutralization as our example of a factor constraint because a small-cap bias happens to be a byproduct of the long-only constraint previously discussed. To illustrate, **Figure 4** displays the active weights in the long-only constrained portfolio in Figure 3 but this time sorted from left to right by market capitalization. With this sorting, the nature of the long-only constraint is readily apparent. All the large negative active weights are associated with the large-cap stocks on the left side of the chart. The magnitude of the negative active weights is very small for the small-cap stocks at the right side of the chart. Yet, positive active weights are as common among the small-cap stocks as the large-cap stocks. The result is a significant small-cap bias in the managed portfolio. The bias toward small-cap stocks is an unintended but natural consequence of the long-only constraint. (Biases in other common risk factors, such as a value–growth tilt, are often a result of data used in the return-forecasting process rather than a byproduct of portfolio constraints.)

Figure 4. Optimization with Long-Only Constraint: Market-Cap Sorting

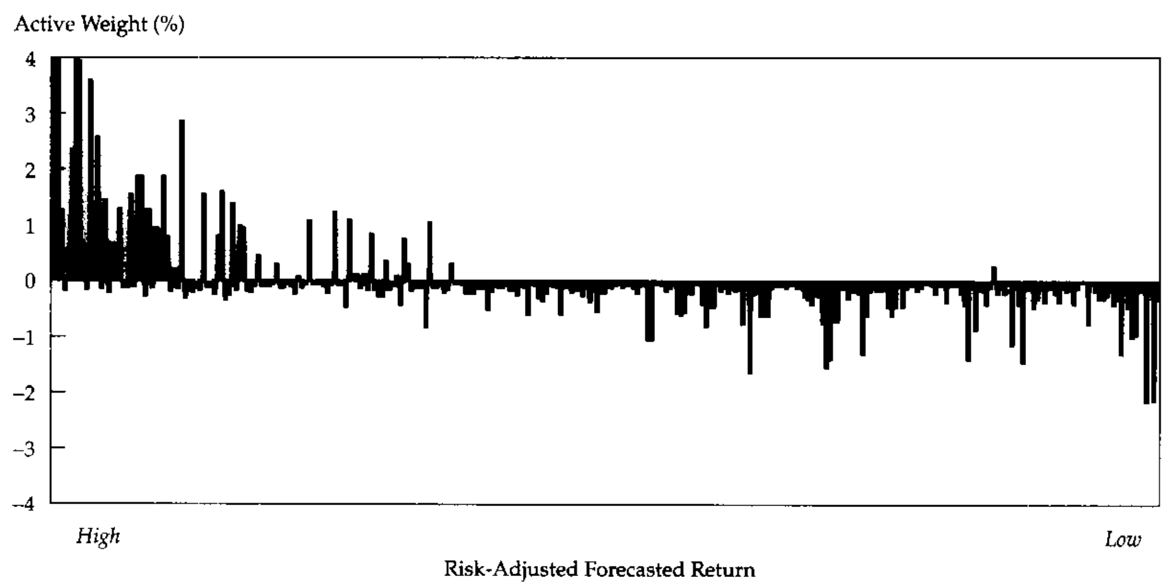


Note: Full height of some bars is not shown.

To eliminate the small-cap bias, **Figure 5** contains a *TC* diagram for an optimization with long-only and market-cap-neutrality constraints.[8] The correlation between active weights and forecasted residual returns is further reduced from the long-only *TC* of 0.58 to a *TC* of 0.47, or 47 percent. In other words, the added constraint leads to an even lower correlation between active weights and forecasted residual returns than in Figure 3, with a correspondingly lower potential for value added.

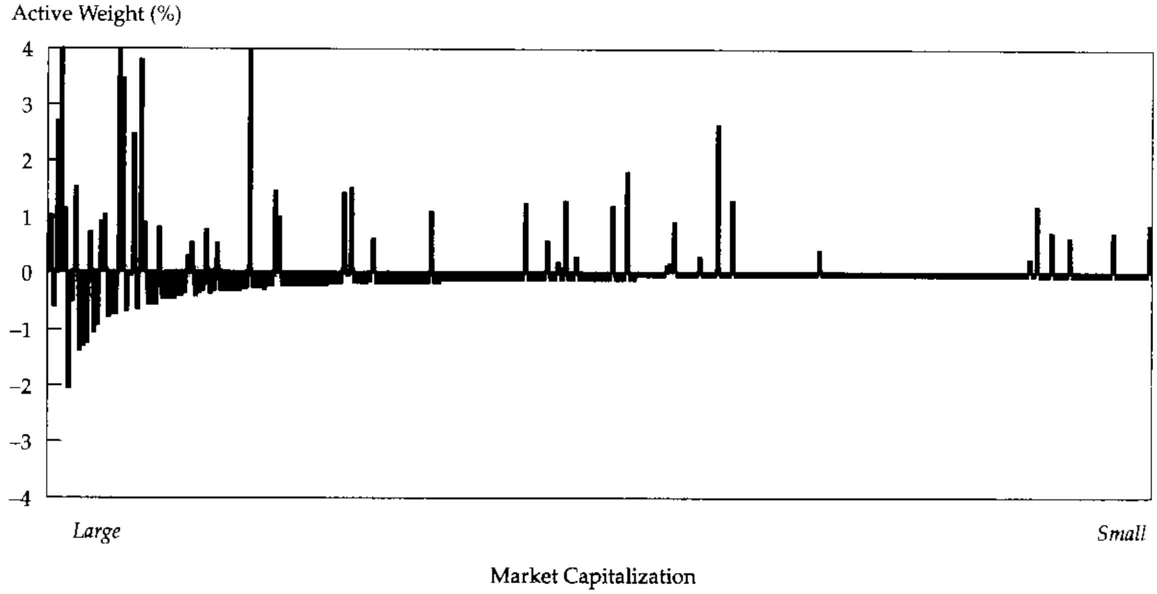
Figure 6 displays the active weights for the long-only and market-cap-neutral portfolio sorted by market capitalization. The small-cap bias induced by the long-only constraint has been corrected; the positive active weights have been reduced to match the negative active weights on the

Figure 5. Optimization with Long-Only and Market-Cap-Neutral Constraints ($TC = 0.47$)



Note: Full height of some bars is not shown.

Figure 6. Optimization with Long-Only and Market-Cap-Neutral Constraints: Market-Cap Sorting



Note: Full height of some bars is not shown.

small-cap end of the graph. The result is a portfolio that is constrained away from having significant active management in anything but the large-cap arena, as indicated by the amount of "ink" on the left end of the *TC* chart. In fact, about half of the active management of the portfolio, as measured by the absolute value of the active weights, is concentrated in the largest 50 of the 500 stocks. Some of the individual active weights exceed 5 percent. Managers may not be comfortable with this much investment riding on a single security, which would motivate additional constraints on the maximum absolute active weight on any single stock. Individual-asset constraints, in conjunction with the constraints already imposed, would lead to even lower *TC* values.

The drop in the information ratio because of the long-only constraint depends on several factors, including the acceptable level of tracking error. Lower tracking error is equivalent to higher risk aversion, λ, in Equation 5. On the one hand, with higher risk aversion, unconstrained optimal active weights have lower absolute values and are less likely to run up against the long-only restriction. We calculated a transfer coefficient of 58 percent in Figure 3 for the long-only constraint at the 5.0 percent tracking-error level. With the addition of the market-cap-neutrality constraint (Figure 5), the transfer coefficient dropped to 47 percent. For the lower tracking errors associated with enhanced index (i.e., low-risk) active strategies, the *TC* values are less affected by the long-only constraint. As Table 1 shows, with the same forecasted security returns as before and a tracking error of only 2.0 percent, the *TC* under the long-only constraint is 73 percent and under both the long-only and market-cap-neutrality constraints, is 67 percent. On the other hand, the long-only constraint becomes a substantial impediment to portfolio construction at higher levels of active risk. At a portfolio tracking error of 8.0 percent (see Table 1), the *TC* under the long-only constraint is 48 percent and is 37 percent under the combined constraints of long only and market-cap neutrality.[9] The long-only manager must decide between more aggressive management (higher tracking error) with a lower transfer coefficient or less aggressive management (lower tracking error) with a higher transfer coefficient. These results are consistent with the findings in Grinold and Kahn's study of the efficiency gains in long–short portfolios.

Case Study 3: Turnover Constraints. In many portfolios, turnover is constrained to reduce transaction costs and, for taxable accounts, to defer the realization of capital gains. In some instances, turnover may also be constrained by mandate or to avoid the appearance of churning. Even when transaction costs are estimated and turnover limits are determined by an optimal trade-off with the higher expected returns, the result is less transfer of the return-forecasting information into active weights. The impact on the transfer coefficient will depend on the degree to which the forecasted returns are correlated with past forecasts and, consequently, with existing portfolio positions. Rapidly changing forecasts, or longer periods between portfolio adjustments, will result in lower turnover-constrained transfer coefficients. The transfer coefficient will also depend on the degree to which the portfolio was allowed to adjust in the past (i.e., based on prior turnover limits).

To avoid the complexities induced by assumptions about the decay in return forecasts and prior limits on turnover, we present the simple example of revising portfolio positions starting from benchmark holdings. This portfolio was, in fact, the starting portfolio in the previous examples of portfolio construction.[10] For example, the unconstrained long–short optimization in Figure 2 led to a turnover of 129 percent from the benchmark starting point. With turnover defined as the percentage of total portfolio value exchanged for new positions, turnover exceeded 100 percent in the long–short optimization because long and short positions are established starting with a benchmark, which has only long positions. The long-only optimization in Figure 3 led to a turnover of 73 percent from the benchmark starting point.

Figure 7 contains a *TC* diagram for a portfolio that is unconstrained (i.e., long–short) except for a turnover limit of 50 percent from an initial benchmark position. The tracking error was set at 5 percent, as in the previous examples. Note that active weights, or deviations from the benchmark weight, are zero for many of the stocks in the middle section of Figure 7. The reason is that turnover cannot be "wasted" on securities with forecasted residual returns that are close to zero. The result (see Table 1) is a *TC* of 73 percent. Tighter turnover limits naturally result in lower transfer coefficients. For example, when turnover is limited to 25 percent, the resulting *TC* is only 49 percent.

Case Study 4: Multiple Constraints. The various types of constraints we have examined have the combined effect of significantly altering actual active weights taken from the unconstrained optimal weights in Equation 5. As a final example of constrained portfolio construction, **Figure 8** contains a *TC* chart for a portfolio with multiple constraints, similar to what might be found in practice. The constraints in this optimization are long only, market-cap neutrality, and dividend-yield neutrality with respect to the benchmark, and turnover limited to 50 percent from a prior long-only portfolio optimized to an unrelated set of forecasted returns. The benchmark is the S&P 500, and active portfolio risk was set at 5 percent. Figure 8 shows only a loose correspondence between active weights and forecasted residual returns. There are negative active weights in some of the highest-forecasted-return stocks and large positive active weights in some of the lowest-forecasted-return stocks. As Table 1 reports, the *TC* is 31 percent. More than two-thirds, or 69 percent, of the value added that was predicted by the information coefficient alone has been lost in the portfolio construction process. *TC* values as low as 30 percent may be common among long-only U.S. equity managers.

Figure 7. Optimization with Turnover Constraint
(*TC* = 0.73)

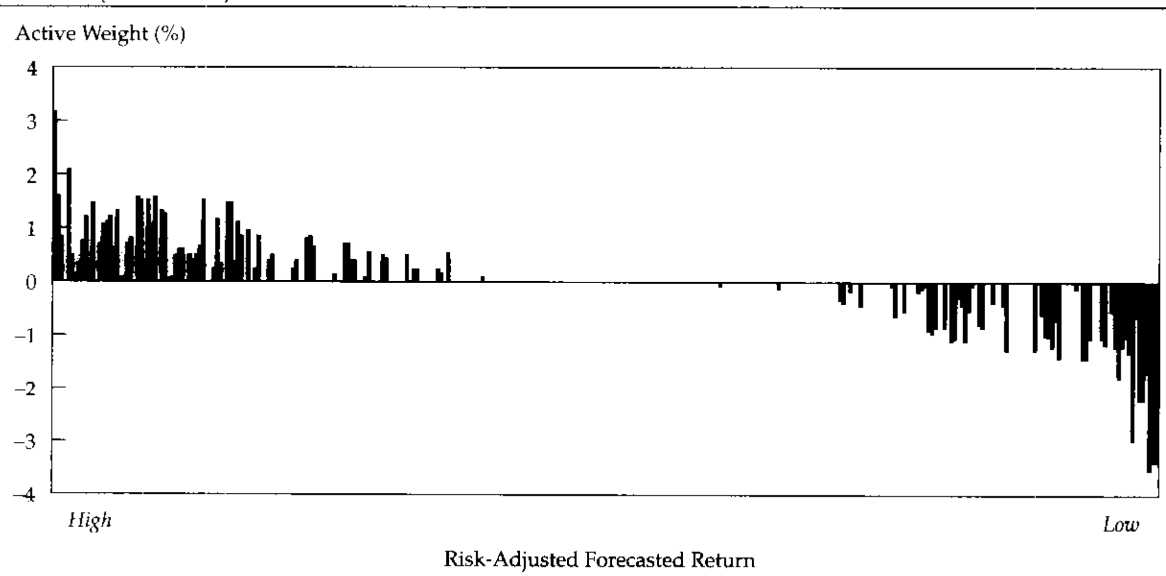
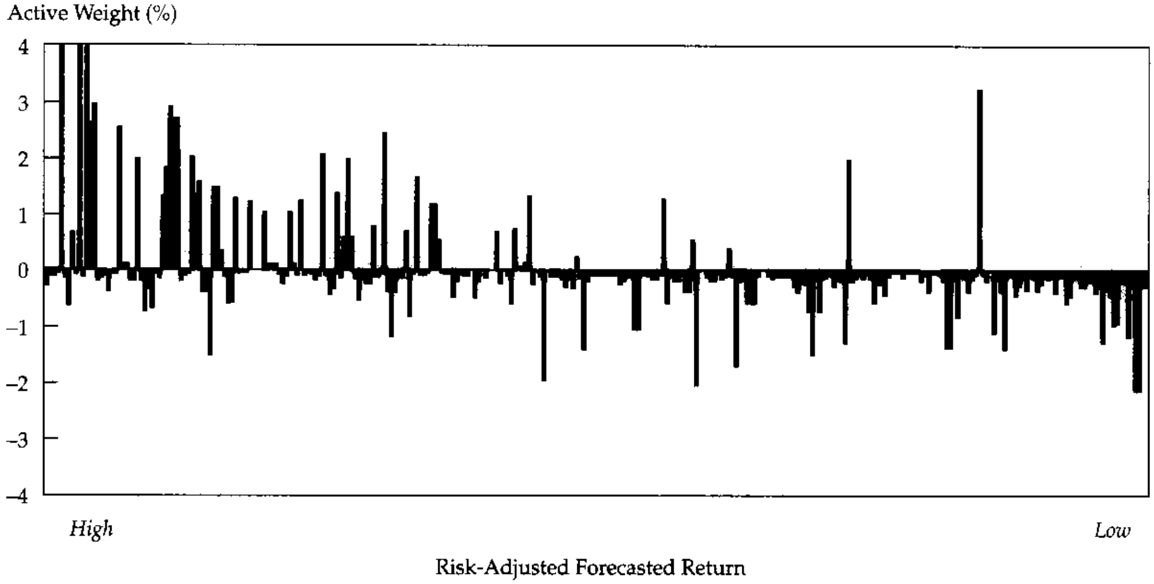


Figure 8. Optimization with Multiple Constraints (TC = 0.31)



Note: Full height of some bars is not shown.

The sets of constraints these managers work under have an unmeasured and perhaps underappreciated impact on performance.[11]

Ex Post Correlation Diagnostics

The perspectives discussed to this point relate to the *potential* value added through portfolio management, as measured by the expected information ratio. Actual performance in any given period will vary from its expected value on the basis of the *ex post* quality of the information, in a range determined by the tracking error. We now turn our attention to diagnosing actual performance—the *ex post* analysis of the realized active returns.

A key determinant of the sign and the magnitude of the realized active portfolio return, R_A, is the degree to which the manager has positive active weights on securities that realize positive residual returns and negative active weights on securities that realize negative residual returns. In other words, actual performance is determined by the relationship between active weights and realized residual returns, a cross-sectional correlation we refer to as the "*ex post* (or realized) performance coefficient."

We use rho with a subscript for realized correlations to distinguish them from expected correlations, which are denoted by capitalized acronyms. Specifically, the realized performance coefficient, $\rho_{\Delta w,r}$, is the cross-sectional correlation between risk-adjusted active weights and realized residual returns. The notation for the expected performance coefficient is PC, which the generalized fundamental law in Equation 8 states is approximately TC times IC. In Appendix A, "*Ex Post* Performance Decomposition," we derive the following important decomposition of the realized performance coefficient:

$$\rho_{\Delta w,r} \approx TC\rho_{\alpha,r} + \sqrt{1-TC^2}\rho_{c,r}, \qquad (11)$$

where $\rho_{\alpha,r}$ is the realized information coefficient and $\rho_{c,r}$ is a realized cross-sectional correlation coefficient that measures the noise associated with portfolio constraints. Note that no realized value is associated with the transfer coefficient, TC, because it measures the relationship between forecasted residual returns and active weights, both of which are established *ex ante*. As in the generalized law, we used the approximate equality notation ≈ in the *ex ante* relationship in Equation 11 because of its dependence on a zero-mean approximation, as specified in Appendix A.

Equation 11 has important implications. In the ideal world of completely unconstrained portfolio construction, TC has an approximate value of 1.0. With a TC value of 1.0 in Equation 11, realized performance, $\rho_{\Delta w,r}$, is determined solely by the actual success of the return-forecasting process,

$\rho_{\alpha,r}$. In other words, if there is a perfect transfer of forecasted residual returns into active weights, then the realized performance coefficient is identical to the realized information coefficient. For *TC* values less than 1.0, however, Equation 11 indicates that realized performance is only partly a function of the success of the signal. For lower values of *TC*, realized performance is increasingly dependent on a second realized correlation coefficient, $\rho_{c,r}$, the risk-adjusted cross-sectional correlation between a new variable, c_i, and realized returns, r_i. As explained in Appendix A, c_i for each security is the difference between the actual active weight taken, Δw_i, and the expected constrained active weight, $TC\Delta w_i^*$.[12] When $\rho_{c,r}$ is positive, performance is increased because, at the margin, the optimization forced higher weights than expected on positively rewarded securities and lower weights than expected on negatively rewarded securities in order to satisfy the portfolio constraints. Of course, the constraint noise coefficient, $\rho_{c,r}$, might just as easily turn out to be negative, which would decrease realized performance.

We can be more specific about the sources of variation in constrained portfolio performance. Under the simplifying assumption we have used of a diagonal residual-return covariance matrix, the variance of each of the three realized correlation coefficients in Equation 11, $\rho_{\Delta w,r}$, $\rho_{\alpha,r}$, and $\rho_{c,r}$, is $1/N$, the number of observations in the cross-sectional correlation calculations. In addition, the two right-hand correlations, $\rho_{\alpha,r}$ and $\rho_{c,r}$, are independent. As a result, the variance of the realized performance coefficient is decomposed into signal and noise variances by

$$\mathrm{var}(\rho_{\Delta w,r}) \approx TC^2\mathrm{var}(\rho_{\alpha,r}) + (1 - TC^2)\mathrm{var}(\rho_{c,r}). \quad (12)$$

For example, if $TC = 0.80$, then 64 percent of the variation in expected performance is attributable to the success of the signal and the remaining 36 percent is attributable to the noise induced by portfolio constraints. For a low *TC* value, 0.30, only 9 percent of the variation in performance is attributable to the success of the signal; the remaining 81 percent is attributable to constraint-related noise. The practical implication is that heavily constrained portfolios will experience frequent periods when, although the forecasting process works, as indicated by a positive realized information coefficient, performance is poor. Alternatively, the realized active portfolio return may be positive even though the realized information coefficient is negative. Without the perspective of Equations 11 and 12, low-*TC* managers might wonder why realized performance seems unrelated to their ability to forecast returns.

Implementation of an *Ex Post* Diagnostic System

The *ex post* relationship in Equation 11 suggests a diagnostic system for analyzing active management performance under constraints. As explained in Appendix A, "The Fundamental Law," the realized active return, R_A, is related to the realized performance coefficient, $\rho_{\Delta w,r}$, by

$$R_A \approx \rho_{\Delta w,r}\sigma_A\sqrt{N}\mathrm{std}\left(\frac{r_i}{\sigma_i}\right). \quad (13)$$

We can think of the term $\sigma_A\sqrt{N}$ as a measure of portfolio aggressiveness. More active risk, σ_A, equates to larger absolute active weights—that is, a more aggressively managed portfolio. The last term in Equation 13, $\mathrm{std}(r_i/\sigma_i)$, is the cross-sectional standard deviation of risk-adjusted realized residual returns, or return dispersion.[13]

Equation 13 can be read as "active return equals the realized performance coefficient times portfolio aggressiveness times realized return dispersion." The sign of the realized performance coefficient determines the sign of the active return. Portfolio aggressiveness and realized-return dispersion simply act as scaling factors. The realized performance coefficient is further decomposed into the realized information coefficient and the constraint noise coefficient according to Equation 11. Combining Equations 11 and 13, we have

$$R_A \approx \left(TC\rho_{\alpha,r} + \sqrt{1 - TC^2}\rho_{c,r}\right) \times \sigma_A\sqrt{N}\mathrm{std}\left(\frac{r_i}{\sigma_i}\right), \quad (14)$$

which forms a diagnostic system in which the realized active return is the product of the decomposed realized performance coefficient (the first term in parentheses) times portfolio aggressiveness times realized return dispersion. Because of the simplifying assumptions used to derive the generalized fundamental law, actual performance as calculated directly from Equation 1 will vary slightly from the performance explained in Equation 14.

We verified the *ex ante* and *ex post* mathematical relationships derived in this article with a Monte Carlo simulation, which will also illustrate the implementation of the diagnostic system. In the simulation, we used the forecasted residual returns and active weights from the long-only constrained optimization shown in Figure 3. Recall that the weights in the case of the long-only constraint have a *TC* of 58 percent (0.578 to be exact). We generated 10,000 sets of 500 realized residual returns by using estimates of individual security residual risks from Barra. The realized-return sets were generated with an IC value of 0.067, the same parameter value used to scale the forecasted returns.[14] We then compiled summary statistics for the 10,000 simulated annual observations of realized portfolio performance. Five observations from the simulation and selected summary statistics are shown in **Table 2**.

As shown in Table 2, the simulation's 10,000 realized information coefficients, $\rho_{\alpha,r}$, had an average value of 0.067, the IC parameter value used in the simulation. The 10,000 realized constraint coefficients, $\rho_{c,r}$, had an average value of –0.002 (effectively zero). Both realized correlations had a standard deviation of about 0.045, approximately $1/\sqrt{500}$. The transfer coefficient and portfolio aggressiveness were constants by design in the simulation.

For the long-only case, the generalized fundamental law predicts an average realized active return of

$$\begin{aligned} E(R_A) &\approx TC\, IC \sqrt{N} \sigma_A \\ &= 0.578 \times 0.067 \times \sqrt{500}\ 5\% \\ &= 4.3\%. \end{aligned}$$

Table 2 shows that the mean active return in the Monte Carlo simulation was 4.2 percent, slightly less than predicted. The standard deviation of the active return in the simulation was 4.9 percent, slightly less than the *ex ante* tracking error of 5.0 percent implicit in the active weights and security residual risks used in the simulation. The IR in the simulation was 4.2/4.9 = 0.86, or 57.3 percent of the theoretical unconstrained IR of 1.50, close to the 57.8 percent predicted by the transfer coefficient. These results verify the approximate accuracy of the *ex ante* relationship described by the generalized fundamental law.

The simulation verifies two critical *ex post* results. First, the explained active return—calculated from the realized information coefficient, constraint noise coefficient, and return dispersion, as per Equation 14—is very similar to the actual realized active return calculated directly from Equation 1. The correlation between explained and actual realized active returns across all 10,000 simulated periods exceeds 99 percent. Second, a regression of the realized performance coefficients, $\rho_{\Delta w,r}$, on realized information coefficients, $\rho_{\alpha,r}$, had an R^2 of 0.332. In other words, 33.2 percent of the performance variation was explained by variation in the realized signal, as suggested by Equation 12. This percentage is close to the predicted value given by TC squared of 0.578^2 = 33.4 percent.

The simulation also provides observations that can be used to illustrate how an *ex post* correlation diagnostic system operates. Table 2 lists parameter values for 5 of the 10,000 simulated annual periods. The first period shown is noteworthy in that the realized information coefficient is 0.075—above the manager's average IC of 0.067. Given the success of the forecasting process in this period, the manager might expect good realized performance even while acknowledging the decline in expected performance because of constraints.

Table 2. Monte Carlo Simulation and *Ex Post* Correlation Diagnostics: S&P 500, Long-Only Constraint

Period	Transfer Coefficient, TC	Realized Information Coefficient, $\rho_{\alpha,r}$	Portfolio Aggressiveness, $\sigma_A\sqrt{N}$	Conditional Active Return	Realized Constraint Coefficient, $\rho_{c,r}$	Explained Performance Coefficient	Realized Return Dispersion, std(r_i/σ_i)	Explained Active Return	Realized Active Return
1	0.578	0.075	1.118	4.8%	–0.057	–0.003	1.008	–0.4%	–0.4%
2	0.578	0.136	1.118	8.8	0.001	0.079	1.026	9.1	9.2
3	0.578	–0.020	1.118	–1.3	0.052	0.031	0.944	3.3	3.2
4	0.578	0.046	1.118	3.0	–0.007	0.021	0.967	2.3	2.3
5	0.578	0.010	1.118	0.6	0.011	0.015	1.000	1.6	1.6
Average		0.067			–0.002		1.000		4.2
Standard deviation		0.044			0.045		0.032		4.9

Notes: IC is 0.067; TC is constant at 0.578; portfolio aggressiveness is constant at $\sigma_A\sqrt{N}$ = 0.05 × 500 = 1.118. The average and standard deviation summary statistics for realized values are for all 10,000 simulated periods. The conditional active return was calculated by Equation 15; the explained performance coefficient was calculated by Equation 11; the explained active return was calculated by Equation 14; the realized active return was calculated by Equation 1.

Taking the expectation of active return in Equation 14, conditional on the realized information coefficient, gives

$$E(R_A|\rho_{\alpha,r}) = TC\rho_{\alpha,r}\sigma_A\sqrt{N}. \quad (15)$$

The conditional active return based on Equation 15 is shown in Table 2. In the first period, the conditional active return is 4.8 percent, slightly higher than the average value of 4.2 percent because the realized information coefficient is slightly higher than average. The *realized* active return in this period, however, is slightly negative (–0.4 percent) because of an unusually large and negative realized constraint coefficient (–0.057). Thus, the portfolio constraints precluded the investor from taking positions in particular securities that were attractive after the fact.

In the third period, the realized information coefficient is negative (–0.020). So, the manager might have expected an active return of –1.3 percent. But despite the poor success of the return-forecasting process in this period, the active return is a positive 3.2 percent. In this case, the portfolio constraints precluded the investor from taking positions in securities that were, after the fact, particularly unattractive (indicated by a positive realized constraint coefficient of 0.052).

The loose correspondence between signal success and actual performance in this simulation is indicative of the mid-range *TC* value of 0.578. The disparity between the success of the return-forecasting process and actual portfolio performance can be even greater and more frequent in portfolios with lower transfer coefficients.

The active return explained by the diagnostic system in this example is highly correlated with the actual active return, but because of simplifying assumptions in the mathematics, the diagnostic system is not exact, as shown in the last two columns of Table 2. The active return explained by the diagnostic system is close to but not an exact match for the actual performance of the portfolio.

The most critical simplifying assumption in the mathematical derivation of the generalized law is the assumption of a diagonal residual covariance matrix, which seems to be an acceptable approximation. We are encouraged by our simulation results for the S&P 500, a commonly used domestic equity benchmark, but we are unable to state any general bounds on the accuracy of the diagnostic system for other benchmarks.

We have not addressed a number of other issues in implementing an *ex post* performance diagnostic. They include nonannual performance periods, the impact of estimation error in security risk parameters, and procedures for estimating the inaccuracy of the diagnostic system because of simplifying assumptions. The simulation does, however, illustrate the key *ex post* concept that the realized information coefficient by itself explains only a portion of actual performance when the portfolio is subject to material constraints. Without an *ex post* diagnostic system that measures constraint-related correlations, managers may be puzzled that actual performance is only loosely related to their ability to predict returns.

Conclusion

We have suggested that the fundamental law of active management, an *ex ante* relationship, can be generalized to include a transfer coefficient as well as an information coefficient. The information coefficient measures the strength of the return-forecasting process, or signal. The transfer coefficient measures the degree to which the signal is transferred into active weights.

The transfer coefficient is a simple scaling factor in the generalized fundamental law and is a quick way to measure the extent to which constraints reduce the expected value of the investor's forecasting ability. In an ideal world with no constraints, the *TC* is approximately 1.0 and the original form of the fundamental law captures the expected value added. In practice, managers must often work within constraints that produce *TC* values ranging from 0.3 to 0.8. The lower transfer coefficient suggests why average performance in practice is only a fraction (0.3 to 0.8) of what is predicted by the original form of the fundamental law.

We also derived a decomposition of *ex post* performance based on the transfer coefficient and the realized information coefficient. The *ex post* performance decomposition indicates that only a fraction (TC^2) of the variation in realized performance, or tracking error, is attributable to variation in the realized information coefficient. If $TC = 1.0$, then variation in performance is wholly attributable to the success of the return-prediction process. If $TC = 0.3$, however, only 9 percent of performance variation is attributable to the success of the signal; the remaining 91 percent is attributable to constraint-induced noise. Managers with low transfer coefficients will experience frequent periods when the signal works but performance is poor and periods when performance is good even though the return-forecasting process failed.

Appendix A. Proofs and Explanations

This appendix explains the motivation for the framework and notation we use, provides proofs of the basic and generalized fundamental law, and examines a decomposition of *ex post* performance.

Framework and Notation

Given a benchmark portfolio, the total excess return (i.e., return in excess of the risk-free rate) on any stock i can be decomposed into a systematic portion that is correlated with the benchmark excess return and a residual return that is not by

$$r_i^{Total} = \beta_i R_B + r_i, \tag{A1}$$

where

R_B = benchmark excess return

β_i = beta of stock i with respect to the benchmark

r_i = security residual return with mean zero and standard deviation σ_i

The benchmark portfolio is defined by the weights, $w_{B,i}$, assigned to each of the N stocks in the investable universe. The benchmark excess return is

$$R_B = \sum_{i=1}^{N} w_{B,i} r_i^{Total}. \tag{A2}$$

If a given stock i in the investable universe is not in the benchmark portfolio, then $w_{B,i}$ is zero.

Like the benchmark, the excess return on an actively managed portfolio, R_P, is determined by the weights, $w_{P,i}$, on each stock:

$$R_P = \sum_{i=1}^{N} w_{P,i} r_i^{Total}. \tag{A3}$$

Define the active return as the managed portfolio excess return minus the benchmark excess return, adjusted for the managed portfolio's beta with respect to the benchmark; that is

$$R_A \equiv R_P - \beta_P R_B. \tag{A4}$$

The managed portfolio's beta, β_P, is simply the weighted-average beta of the stocks in the managed portfolio:

$$\beta_P \equiv \sum_{i=1}^{N} w_{P,i} \beta_i. \tag{A5}$$

With some algebra, and the fact that the beta of the benchmark must be exactly 1, it can be shown that the active return is

$$R_A = \sum_{i=1}^{N} \Delta w_i r_i, \tag{A6}$$

where the active weight for each stock is defined as the difference between the managed portfolio weight and benchmark weight; that is,

$$\Delta w_i = w_{P,i} - w_{B,i}. \tag{A7}$$

The formulation for the active return in Equation A6 (which is also Equation 1 in the body of the article) is the focus of our analysis. Note that the active weights, Äw_i, sum to 0 because they are differences in two sets of weights that each sum to 1.0. Also note that the stock returns, r_i, in Equation A6 are residual, not total, excess returns. Residuals are the relevant component of security returns when performance is measured against a benchmark on a beta-adjusted basis.[15]

Assume that portfolio optimization is based on choosing active weights, Δw_i, that maximize the mean–variance utility function:

$$U = E(R_A) - \lambda \sigma_A^2, \tag{A8}$$

where

$E(R_A)$ = expected active return

σ_A^2 = active return variance

λ = a risk-aversion parameter

Given forecasts for the individual residual stock returns, α_i, the expected active return for the portfolio is

$$E(R_A) = \sum_{i=1}^{N} \Delta w_i \alpha_i. \tag{A9}$$

Under the important simplifying assumption that the residual stock returns are uncorrelated (i.e., the residual covariance matrix is diagonal), the active return variance is

$$\sigma_A^2 = \sum_{i=1}^{N} \Delta w_i^2 \sigma_i^2, \tag{A10}$$

where σ_A^2 is the residual return variance for stock i.

Substituting Equations A9 and A10 into the optimization problem in Equation A8 leads to optimal weights given by the formula

$$\Delta w_i^* = \frac{\alpha_i}{\sigma_i} \frac{1}{2\lambda}. \tag{A11}$$

The simple closed-form solution to optimal weights in Equation A11 (Equation 5 in the body of the article) is based on two simplifying assumptions. First, we assume a diagonal residual-return covariance matrix; that is, we assume that residual returns are perfectly uncorrelated with each other.[16] Second, the formal optimization problem has a budget constraint (the active weights must sum to zero), but we assume that no such condition is imposed in the maximization of Equation A8.

This second assumption will generally not be a problem for portfolios with many securities and typical risk–return parameters. One security can be adjusted to absorb the balancing weight without distorting the analysis to any great extent.

Despite these two simplifying assumptions, the optimization result in Equation A11 will prove useful in understanding the impact of constraints. An intuitive property of Equation A11 is that the mean–variance optimal active weight for each stock is proportional to the stock's forecasted residual return over residual return variance. The constant of proportionality, common to all securities, is related to the inverse of the risk-aversion factor, λ. Lower values of λ lead to more aggressive portfolios with proportionately larger absolute active weights, higher expected active return, and higher active risk. We can insert the optimal weights from Equation A11 into the definition of active return variance in Equation A10 and solve for λ. Using this solution for λ, we find the optimal active weights in terms of active portfolio risk, σ_A, to be

$$\Delta w_i^* = \frac{\alpha_i}{\sigma_i^2}\frac{\sigma_A}{\sqrt{\sum_{i=1}^{N}(\alpha_i/\sigma_i)^2}}. \tag{A12}$$

We make the operational assumption that the security alphas, α_i, are generated from scores, S_i, that have cross-sectional zero mean and unit standard deviation. Specifically, based on Grinold's (1994) prescription, alphas are the product of IC, residual security volatility, and score:

$$\alpha_i = IC\sigma_i S_i. \tag{A13}$$

By not subscripting IC, we are making the explicit assumption that the information coefficient is the same for all securities.

The Fundamental Law

A proof of Grinold's (1989) fundamental law of active management follows directly from the optimization and alpha-generation mathematics. Given a cross-sectional set of scores, S_i, that is constructed to have zero mean and unit standard deviation, Equation A13 dictates that the ratio α_i/σ_i has a cross-sectional mean of zero and standard deviation of IC. Based on the zero-mean property, the denominator in the last term in Equation A12 is a standard deviation calculation for α_i/σ_i but without the requisite N divisor. We use the term "zero-mean property" to justify variance and covariance calculations that are sums of squares and products.[17] Making this substitution and then multiplying each side of Equation A12 by σ_i gives the risk-adjusted optimal weights:

$$\Delta w_i^*\sigma_i = \frac{\alpha_i}{\sigma_i}\left(\frac{\sigma_A}{IC\sqrt{N}}\right). \tag{A14}$$

Kahn (2000) referred to these risk-adjusted weights as "optimal risk allocations." Because the ratio α_i/σ_i has a zero mean, the set of risk-adjusted optimal active weights, $\Delta w_i^*\sigma_i$, in Equation A14 has a zero mean and cross-sectional standard deviation (std) of

$$\begin{aligned}\text{std}(\Delta w_i^*\sigma_i) &= \text{std}\left(\frac{\alpha_i}{\sigma_i}\right)\left(\frac{\sigma_A}{IC\sqrt{N}}\right)\\ &= \frac{\sigma_A}{\sqrt{N}}.\end{aligned} \tag{A15}$$

The zero-mean property of $\Delta w_i^*\sigma_i$ and α_i/σ_i allows the expected active return in Equation A9 to be recast in a correlation formulation as follows:

$$\begin{aligned}E(R_A) &= \sum_{i=1}^{N}\Delta w_i^*\alpha_i\\ &= \sum_{i=1}^{N}(\Delta w_i^*\sigma_i)\left(\frac{\alpha_i}{\sigma_i}\right)\\ &= N\text{cov}\left(\Delta w_i^*\sigma_i, \frac{\alpha_i}{\sigma_i}\right)\\ &= N\rho_{\Delta w^*,\alpha}\text{std}(\Delta w_i^*\sigma_i)\text{std}\left(\frac{\alpha_i}{\sigma_i}\right)\\ &= \rho_{\Delta w^*,\alpha}\sigma_A\sqrt{N}IC,\end{aligned} \tag{A16}$$

where $\rho_{\Delta w^*,\alpha}$ is the correlation coefficient between risk-adjusted optimal weights and alphas—that is, the correlation between $\Delta w_i^*\sigma_i$ and α_i/σ_i. This correlation coefficient is 1.0 because the two variables are proportional, as indicated by Equation A14. Substituting a correlation coefficient value of $\rho_{\Delta w^*,\alpha} = 1.0$ and dividing both ends of Equation A16 by active risk σ_A gives the original form of the fundamental law:

$$IR = IC\sqrt{N}, \tag{A17}$$

where IR is the information ratio of expected active return to active risk; that is, $IR \equiv E(R_A)/\sigma_A$.

In practice, active managers are usually subject to constraints that cause them to deviate from the unconstrained optimal active weights, Δw_i^*. Let Δw_i be a set of constrained active weights generated by an optimizer, where the active risk is set equal to the unconstrained active risk in Equation A10. The covariance and correlation methodology used in Equation A16 for unconstrained optimal active weights can be applied to constrained active weights as follows:

$$\begin{aligned} E(R_A) &= \sum_{i=1}^{N} \Delta w_i \alpha_i \\ &= \sum_{i=1}^{N} (\Delta w_i \sigma_i)\left(\frac{\alpha_i}{\sigma_i}\right) \\ &= N\text{cov}\left(\Delta w_i \sigma_i, \frac{\alpha_i}{\sigma_i}\right) \\ &= (N)(TC)\text{std}(\Delta w_i \sigma_i)\text{std}\left(\frac{\alpha_i}{\sigma_i}\right) \\ &= (N)(TC)IC\sqrt{\sigma_A^2/N - \text{mean}(\Delta w_i \sigma_i)^2}, \end{aligned} \tag{A18}$$

where $TC = \rho_{\Delta w, \alpha}$ is the transfer coefficient and mean($\Delta w_i \sigma_i$) is the cross-sectional mean of $\Delta w_i \sigma_i$. The complexity of the final result arises from the fact that there is nothing to force the zero-mean property on $\Delta w_i \sigma_i$, although the mean is likely to be small because the active weights alone, Δw_i, have a zero mean. In addition, this term is squared in Equation A18. Using zero as an approximation and dividing by active risk, σ_A, as before, produces the generalized fundamental law:

$$IR \approx TC\ IC\sqrt{N}, \tag{A19}$$

which is given as Equation 6 in the body of the article.

The transfer coefficient, TC, is calculated as the cross-sectional correlation coefficient between risk-adjusted forecasted residual returns and actual weights—α_i/σ_i and $\Delta w_i \sigma_i$. Note that the transfer coefficient can be calculated *ex ante* (i.e., before returns are realized) based on the set of weights from an optimizer. We will continue to use the approximation mean($\Delta w_i \sigma_i$) = 0 in the *ex post* analysis that follows and use the $\approx$ notation wherever the approximation affects the results.

The realized active return in any time period is the sum of the product of weights and returns, as defined in Equation A6. The realized active return can be decomposed into cross-sectional statistics by

$$\begin{aligned} R_A &= \sum_{i=1}^{N} \Delta w_i r_i \\ &= \sum_{i=1}^{N} (\Delta w_i \sigma_i)\left(\frac{r_i}{\sigma_i}\right) \\ &= N\text{cov}\left(\Delta w_i \sigma_i, \frac{r_i}{\sigma_i}\right) \\ &= \rho_{\Delta w, r}\sigma_A \sqrt{N}\text{std}\left(\frac{r_i}{\sigma_i}\right). \end{aligned} \tag{A20}$$

The realized correlation coefficient between risk-adjusted weights and returns, $\rho_{\Delta w, r}$, is an important summary measure of performance that will be referred to as the realized "performance coefficient." The notation for the expected value of the performance coefficient is PC. Note that the cross-sectional standard deviation of realized risk-adjusted returns, std(r_i/σ_i), has an expected value of 1.0. Taking expectations on both ends of Equation A20 and rearranging indicates that the expected performance coefficient is simply the expected information ratio, $E(R_A)/\sigma_A$, over the square root of N. With this substitution, the generalized fundamental law in Equation A19 becomes Equation 8 in the body of the article.

Ex Post Performance Decomposition

Given a set of actual active weights, Δw_i, from an optimizer and hypothetical optimal active weights, Δw_i^*, derived analytically from expected returns, we define a new variable,

$$c_i \equiv \Delta w_i - TC\Delta w_i^*, \tag{A21}$$

which can be thought of as the "optimal weight not taken" on each stock because of constraints.[18] The multiplier TC in front of Δw_i^* in Equation A21 is a shrinkage factor related to the expected value of Δw_i given the value of Δw_i^*. Note that TC and thus the c_i values can be calculated *ex ante* (i.e., after the portfolio has been optimized but before returns have been realized). The cross-sectional variance of the risk-adjusted constraint values, $c_i \sigma_i$, is

$$\begin{aligned} \text{var}(c_i \sigma_i) &= TC^2 \text{var}(\Delta w_i^* \sigma_i) + \text{var}(\Delta w_i \sigma_i) \\ &\quad - 2TC\text{cov}(\Delta w_i^* \sigma_i, \Delta w_i \sigma_i) \\ &\approx TC^2 \frac{\sigma_A^2}{N} + \frac{\sigma_A^2}{N} - 2TC\rho_{\Delta w^*, \Delta w}\frac{\sigma_A^2}{N} \\ &= \frac{\sigma_A^2}{N}\left(1 - TC^2\right), \end{aligned} \tag{A22}$$

where the substitution of the transfer coefficient, TC, for the correlation $\rho_{\Delta w^*, \Delta w}$ in the final result is justified by the proportionality between risk-adjusted optimal weights and alphas.

Based on the definitional relationship in Equation A21 and the result in Equation A22, we can decompose the realized active return in Equation A6 into two correlation structures:

$$\begin{aligned} R_A &= \sum_{i=1}^{N}(TC\Delta w_i^* r_i + c_i r_i) \\ &= TC\sum_{i=1}^{N}(\Delta w_i^*\sigma_i)\left(\frac{r_i}{\sigma_i}\right) + \sum_{i=1}^{N}(c_i\sigma_i)\left(\frac{r_i}{\sigma_i}\right) \\ &\approx \left(TC\rho_{\alpha,r} + \sqrt{1-TC^2}\rho_{c,r}\right) \\ &\quad \times \sigma_A\sqrt{N}\text{std}\left(\frac{r_i}{\sigma_i}\right). \end{aligned} \tag{A23}$$

Equating the final expressions in Equations A20 and A23 and dividing out common terms yields the correlation relationship (Equation 11 in the body of the article)

$$\rho_{\Delta w,r} \approx TC\rho_{\alpha,r} + \sqrt{1-TC^2}\rho_{c,r}. \tag{A24}$$

Taking expectations of both sides of Equation A24, and using the relationship in Equation 8, indicates that the expected value of the noise coefficient, $\rho_{c,r}$, is zero. In addition, the two *ex ante* correlation coefficients on the right side of the correlation structure in Equation A24 can be shown to be independent, as follows: The realized values of cross-sectional correlation coefficients have, by definition, a standard deviation of $1/\sqrt{N}$ around their means. Thus, the variance analysis of Equation A24 is

$$\begin{aligned} \frac{1}{N} &\approx TC^2\frac{1}{N} + (1-TC^2)\frac{1}{N} \\ &\quad - 2\text{cor}(\rho_{\alpha,r}, \rho_{c,r})TC\sqrt{1-TC^2}\frac{1}{N}. \end{aligned} \tag{A25}$$

Rearranging Equation A25 indicates that $\text{cor}(\rho_{\alpha,r}, \rho_{c,r})$ is approximately zero.

The value of this independence property is variance decomposition. Because the two right-hand-side realized correlation coefficients in Equation A24 are approximately independent, we have

$$\begin{aligned} \text{var}(\rho_{\Delta w,r}) &\approx TC^2\text{var}(\rho_{\alpha,r}) \\ &\quad + (1-TC^2)\text{var}(\rho_{c,r}). \end{aligned} \tag{A26}$$

Thus, TC^2 percent of the variation in realized performance is the result of the success of the signal as measured by the realized information coefficient, $\rho_{\alpha,r}$. The remaining $1 - TC^2$ percent is attributable to constraint-induced noise, $\tilde{n}_{c,r}$.

Notes

1. The information coefficient and transfer coefficient are not necessarily simple correlation calculations. When securities have different individual residual risks, risk-adjusted variables should be used in calculating the correlations, as shown later.
2. Goodwin (1998) included a useful discussion of procedures for calculating the information ratio.
3. The transfer coefficient plays the same role in our formulation as the notion of "implementation efficiency" in Kahn (2000), although Kahn's measurements were derived through simulation rather than an explicit formulation.
4. Specifically, we used the Barra E3 model as of the end of November 2000. The results are for illustration purposes only and, in all practical respects, are invariant to the time period chosen. Barra is a leading supplier of security covariance matrix estimates and optimization software that structures portfolio weights.
5. As discussed in Appendix A, "Framework and Notation," this process for generating forecasted residual returns uses the rule "alpha is volatility times *IC* times score" as prescribed in Grinold (1994). Each stock's residual risk was calculated from the total security risk, security beta, and benchmark (i.e., S&P 500) risk—all estimated by Barra. The formula for calculating 500 residual risk is $\sigma_i^2 = \sigma_{Total,i}^2 - \beta_i^2\sigma_{S\&P\ 500}^2$.
6. The information ratios and expected active returns shown in Table 1 were based on the generalized fundamental law of Equation 6, not the exact value that Equation 10 would produce. Specifically, the *IR* in the first line was calculated as $0.98 \times 0.067 \times \sqrt{500} = 1.47$, and the expected active return was calculated as $0.98 \times 0.067 \times \sqrt{500} \times 5$ percent = 7.3 percent.
7. The value-added portion of a strategy can be separated from its net market exposure. For example, long–short market-neutral strategies are now commonly overlaid with a long position in equity index futures or an equity index swap. The long–short portfolio makes efficient use of the investor's information, while the derivatives overlay adds market exposure. The combination can have the high transfer coefficient and expected information ratio of a long–short portfolio but also the full equity market exposure of a long-only portfolio.
8. Market-cap neutrality is based on the Barra E3 "size" risk factor, which is defined as the log of market capitalization (price per share times number of outstanding shares). Thus, the constraint imposed was that the average log market capitalization of the stocks in the managed portfolio be equal to the average log market capitalization of the stocks in the S&P 500 benchmark.
9. The *TC* for the unconstrained optimization remains at 98 percent for all three tracking-error values (i.e., 2, 5, and 8 percent). The unconstrained *TC* value varies slightly with the period chosen for the estimated covariance matrix but is always just slightly less than the theoretical value of 100 percent.
10. When turnover is unconstrained, the weights in an optimized portfolio are invariant to the initial portfolio. In other words, the *TC* values in the prior portfolio construction examples did not depend on the assumed starting position.
11. Although it may appear that constraints are always undesirable, some constraints guard against the effect of estimation error in the forecasting process. So, a manager may be willing to tolerate a lower transfer coefficient to protect against sizable underperformance if the forecasting process fails.

12. An alternative, and perhaps more intuitive, definition of active weight not taken, $b_i \equiv \Delta w - \Delta w_i^*$, results in a different *ex post* correlation structure from Equation 11. However, the definition $c_i \equiv \Delta w - TC\Delta w_i^*$ has nicer mathematical properties.
13. Because the security returns are risk adjusted, the expected value of the return dispersion term is 1.0. Risk-adjusted statistics are technically more precise, but the cross-sectional correlations and standard deviations can also be calculated without the risk adjustment. A simple non-risk-weighted system can be implemented with the cross-sectional correlation and standard deviation calculations based on Δw_i, α_i, and r_i instead of the risk-adjusted data, $\Delta w_i \sigma_i$, α_i/σ_i, and r_i/σ_i. Although the characterization of the realized information and transfer coefficients will be incorrect when the securities have different residual risks, the *ex post* system will still "add up." However, the components will not correspond to the conceptual decomposition suggested by the generalized fundamental law.
14. Although the realized residual returns were generated to have a 0.067 correlation with their respective forecasted residual returns and have magnitudes consistent with their respective residual risks, they were cross-sectionally random within each set (i.e., the returns simulate a perfectly diagonal covariance matrix). Thus, the simulation was "real world" in the generation of constrained active weights with an actual nondiagonal covariance matrix, but the simulation controlled the generation of the realized residual returns to conform to assumed statistical parameters.
15. Throughout the article, we took the individual security risk estimates (β_i and σ_i) as given, although in reality, they may be estimated with error.
16. The correlation of "market model" residual returns as defined in Equation A1 cannot be exactly zero, even in a theoretical sense. For example, if the benchmark contains only two stocks, then the residual return on the first has to be perfectly negatively correlated with the residual return on the second. For benchmark portfolios with a large number of stocks (e.g., 500), the correlation matrix is populated with off-diagonal elements that are generally small but, on average, tend to be slightly negative.
17. Specifically, if a variable X has a mean of zero, then the variance of X is $\text{var}(X) = 1/N\sum_{i=1}^{N} X_i^2$. Similarly, if either X or Y has a mean of zero, then the covariance of X and Y is $\text{cov}(X,Y) = 1/N\sum_{i=1}^{N} X_i Y_i$. Also note that the statistical definition of the correlation coefficient between two variables X and Y is $\text{cov}(X,Y) \equiv \rho_{X,Y}\,\text{std}(X)\text{std}(Y)$. This definitional decomposition of covariance into correlation and standard deviation (std) is used repeatedly throughout Appendix A.
18. A more straightforward weight-not-taken variable definition, $b_i \equiv \Delta w_i - \Delta w_i^*$, has less desirable statistical properties. Specifically, it can be shown that the expected value of the correlation of b_i with realized returns is not zero.

References

Brush, John S. 1997. "Comparisons and Combinations of Long and Long–Short Strategies." *Financial Analysts Journal*, vol. 53, no. 3 (May/June):81–89.

Goodwin, Thomas H. 1998. "The Information Ratio." *Financial Analysts Journal*, vol. 54, no. 4 (July/August):34–43.

Grinold, Richard C. 1989. "The Fundamental Law of Active Management." *Journal of Portfolio Management*, vol. 15, no. 3 (Spring):30–37.

———. 1994. "Alpha Is Volatility Times IC Times Score, or Real Alphas Don't Get Eaten." *Journal of Portfolio Management*, vol. 20, no. 4 (Summer):9–16.

Grinold, Richard C., and Ronald N. Kahn. 2000. "The Efficiency Gains of Long–Short Investing." *Financial Analysts Journal*, vol. 56, no. 6 (November/December):40–53.

Jacobs, Bruce I., Kenneth N. Levy, and David Starer. 1998. "On the Optimality of Long–Short Strategies." *Financial Analysts Journal*, vol. 54, no. 2 (March/April):40–51.

———. 1999. "Long–Short Portfolio Management: An Integrated Approach." *Journal of Portfolio Management*, vol. 25, no. 2 (Winter):23–32.

Kahn, Ronald N. 2000. "Most Pension Plans Need More Enhanced Indexing." In *Enhanced Indexing: New Strategies and Technologies for Plan Sponsors*. Edited by Brian R. Bruce. London: Institutional Investor.

Thomas, Lee R. 2000. "Active Management." *Journal of Portfolio Management*, vol. 26, no. 2 (Winter):25–32.