The Impact of Constraints on Value-Added

A non-cynical approach.

Bernd Scherer and Xiaodong Xu

Are investment constraints a way to manage risks, or rather an outdated approach to risk management? Or a worst case safeguard against a breakdown in risk management practices, and a practical reality of institutional investing?

We want to provide a closer look at the costs of constraints and suggest a new methodology to measure the impact of individual constraints on an investor's value-added. To our knowledge, no such approach has been taken yet.

We first discuss the concept of information ratio shrinkage, as introduced by Grinold and Kahn [2000] and later given a fresh interpretation by Clarke, de Silva, and Thorley [2002] under the name transfer coefficient. Their estimates of the effect of constraints on optimal portfolios involve the calculation of a constrained and unconstrained optimization together with a comparison of the objective functions. This is helpful but not complete.

We therefore consider the use of Lagrange multipliers as discussed in Grinold and Easton [1998] to calculate the distorting effect of various constraints on the forecasts made from an alpha-generating model. We believe our approach provides much more information than a simple distribution of portfolio weights. It also allows us to relate distortions in the portfolio construction process to stock characteristics. We confirm the intuition that the long-only constraint is less damaging to large stocks than to small stocks, while the reverse is true for the size constraint.

Our primary innovation is that we can decompose the effects of portfolio constraints on total value-added. After all, it is value-added that is important, not individual weight distortions. Constraints might create distortions in implied alphas, but could still have little impact on value-added.

BERND SCHERER
is a managing director and global head of Quantitative Structured Products at Morgan Stanley Investment Management in London.
bernd.scherer@morganstanley.com

XIAODONG XU
is a vice president and senior researcher at Deutsche Asset Management in New York City.
xiadong.xu@db.com

In order to account realistically for leverage, we need to calculate the sum of absolute positive and negative weights. For this we introduce mixed integer variables. As this makes the resulting Lagrange multipliers difficult to interpret, we introduce a two-step procedure to arrive at the required multipliers.

INFORMATION RATIO SHRINKAGE

Active quantitative managers usually start with a long/short portfolio that best describes their alpha forecasts, α_i, that arise from their specific valuation model across $i = 1, \ldots, n$ stocks. This portfolio is generally free of institutional constraints (client-specific or regulatory constraints), as the view on a given company is clearly independent of the often arbitrary and conflicting constraints arising from each individual client portfolio.

Views inherent in this portfolio can therefore be summarized as global alpha, i.e., an alpha that is unique for each stock and identical across investment universes. In a world without long-only constraints (and in the absence of any other constraint on factor, industry, and beta exposure), the optimal position for asset i relative to asset j is given by the solution to a standard active investment management problem.

We want to maximize portfolio utility:[1]

$$\sum_i w_i\alpha_i - \frac{\lambda}{2}\sum_i \sum_j w_i w_j \sigma_{ij} \tag{1}$$

that arises from active positions w_i, where σ_{ij} denotes the covariance between asset i and j, and λ captures risk aversion.

The solution to this problem is given by:[2]

$$\frac{w_i^*}{w_j^*} = \left(\frac{\alpha_i - \sum_{j\neq i}\beta_{ij}\alpha_j}{\alpha_j - \sum_{i\neq j}\beta_{ji}\alpha_i}\right)\cdot\left(\frac{\sigma_j^2(1-R_j^2)}{\sigma_i^2(1-R_i^2)}\right) \tag{2}$$

The first term in parentheses reflects relative alphas adjusted for their uniqueness. Note that $\alpha_i - \sum_{j\neq i}\beta_{ij}\alpha_j$ denotes the alpha of asset i that is not already explained by the remaining $n-1$ alpha, where β_{ij} can be interpreted as the optimal position in asset j for a portfolio that attempts to track asset i as closely as possible using *all* the remaining assets.

The second term in the parentheses in Equation (2) calculates the differences in risk that cannot be hedged through long/short positions, where $\sigma_i\sqrt{1-R_i^2}$ denotes the risk of the variance-minimizing hedge.

Essentially we can think of the β_{ij} as the optimal positions of a funding portfolio. Note that we do not use β_{ij} and β_{ji} interchangeably, as hedging positions are not generally symmetric.

Without long-only constraints, the optimal relative weights w_i^* / w_j^* do not depend on tracking error requirements, or risk aversion, or benchmarks. In fact, optimal portfolios differ only by a scaling factor, which leaves the information ratio constant. This is exactly what one would expect from a performance measure.

The efficient frontier becomes a straight line, and all active portfolios along this line have unit correlation, as again they are clones that differ only in leverage. Imposing constraints will result in inefficient portfolios, as the optimal relative allocation changes.

In short, the relative weights in a constrained portfolio will differ from the relative weights in an unconstrained portfolio:

$$\frac{w_i^*}{w_j^*} \neq \frac{w_i^c}{w_j^c} \tag{3}$$

Removing or even relaxing the long-only constraints allows a more efficient portfolio construction process. Alternative explanations are likely to be overoptimistic as to their explanatory power. This powerful insight was provided by Grinold and Kahn [2000] and remains the main reason behind the appeal of unconstrained investing.

To exemplify their argument, we compare an unconstrained benchmark relative optimization versus the S&P 500 (benchmark weights are as of June 30, 2006) with optimizations that show varying levels of constraints. To simulate the effect of constraints, we maximize expression (1) subject to a cash-neutrality, a beta-neutrality, a size-neutrality, and a typical no-short constraint:

$$\sum_{i=1}^n w_i = 0 \tag{4}$$

$$\sum_{i=1}^n w_i\beta_i = 0 \tag{5}$$

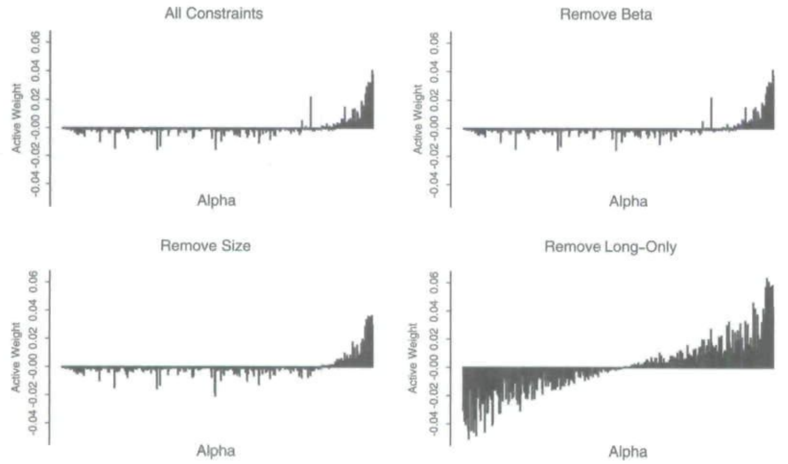
$$\sum_{i=1}^n w_i s_i = 0 \tag{6}$$

$$-w_{b,i} \leq w_i \tag{7}$$

where $-w_{b,i}$ denotes the benchmark stock weight. In other words, the maximum underweight (negative active position) of a given stock equals its benchmark allocation. Note that β_i and s_i are the exposures to a given characteristic, i.e., the stock size or benchmark beta.

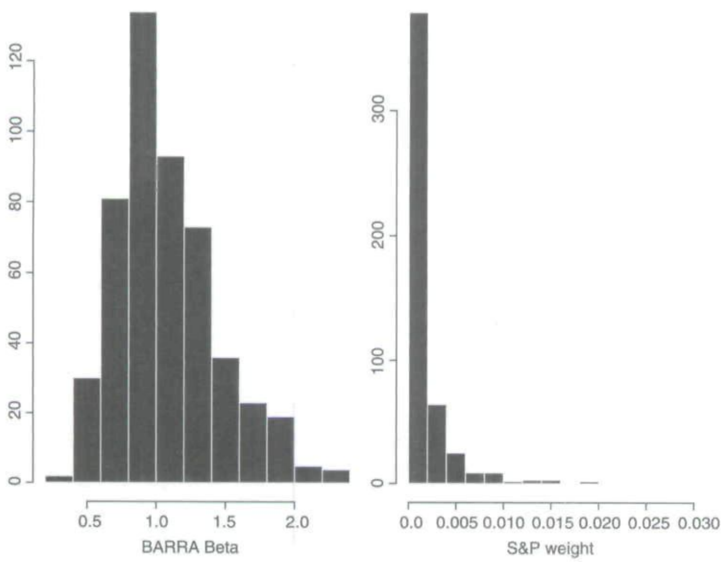
The effect of Equations (4) through (7) can be seen in Exhibit 1, which plots active weights relative to sorted alphas. The highest alphas are found at the right end of the x-axis. They represent the top-ranked overweights.

EXHIBIT 1
Active Weight versus Sorted Alpha



Each graph shows the optimal active weights for an S&P relative investor with risk aversion of $\lambda = 4$ *versus the sorted security alphas. Covariance matrix for S&P 500 from BARRA as of 06/30/2006. Alphas are lowest at the left-hand side and increase moving toward the right-hand side. Roughly 50% of alphas relative to the S&P 500 are positive, while the other half are negative. All alphas are drawn from a normal distribution with mean and standard deviation equal to 0 and 0.006, respectively.*

EXHIBIT 2
Distribution of Benchmark Weights and Stock Beta



Data from BARRA as of 06/30/2006. Note that benchmark weights can differ by a factor of 500, while the maximum factor is 7 for betas.

In an optimization including all constraints (top left-hand graph), where we include a long-only as well as cash-, beta-, and size-neutrality constraints, the optimal weights seem to bear little relation to the relative attractiveness of a given stock as measured by its alpha. In fact stocks with high alphas get underweighted by the same amount as stocks with very negative alphas. The associated *IR* equals 0.40.

Removing beta-neutrality will leave this effect virtually unchanged. The first noticeable change arises when the size constraint is removed. Comparing the upper right-hand with the lower left-hand barplot, we can see that a removal of the size constraint will lead to a smoother transition from overweights to underweights. This is intuitive, as betas vary to a much smaller degree than benchmark weightings.

It is easy to find stocks that differ in their weightings by a factor of 200 or more, but literally impossible to find two stocks that differ in beta by the same factor, as can be seen in Exhibit 2. The reason is that underweighting large stocks despite their positive alpha in order to generate size-neutrality is now no longer required.

Finally, removing the long-only constraint altogether seems to have the greatest impact on implementation efficiency. We observe a smooth transition from the largest negative to the largest positive weight positions, where over- and underweights are closely aligned with the size of the corresponding alpha. Consequently, the *IR* is highest (0.65).

Note that the size constraint impacts optimal portfolio choice in three different ways. First, the long constraint leads to a reduction in breadth of active decision-making. For stocks with low index weights, we can implement only positive positions. Theoretically, the IR on those stocks drops by 1 over the square root of 2, as only half of the views can be implemented.

Second, the long-only constraint can lead to an unwanted size bias that can entirely swamp the performance of a stock-picker. Trivial overweights need to be financed by underweights, but as small stocks offer little opportunity for underweighting, we eventually need to use large-caps. This will lead to a positive small-cap bias.

This bias is the real reason it is sometimes so difficult and sometimes so easy to beat the S&P 500. If large-caps outperform, the portfolio construction-induced positive small-cap bias will make it literally impossible to beat the S&P 500. This has nothing to do with any claimed efficiency of the S&P 500.

Finally, imposing a long-only constraint removes the ability to hedge stock price risks; i.e., it makes risk management more difficult. For example, relative value trades between two small-cap stocks become virtually impossible to implement.

The Clarke, de Silva, and Thorley [2002] interpretation of Grinold and Kahn [2000] has proven to be highly valuable for portfolio practitioners. The so-called transfer coefficient τ effectively represents the ratio of two information ratios $\tau - IR^c/IR$, where IR and IR^c denote the information ratios of both an unconstrained and a constrained optimization.[3] By definition, the transfer coefficient will always be smaller than or equal to one ($\tau \leq 1$), as a constrained optimization will ex ante (i.e., in-sample) always do more poorly than an unconstrained optimization. In the example of Exhibit 1, the transfer coefficient from an unconstrained to a completely constrained portfolio is $\tau = 0.40/0.65 = 0.62$.

Both Grinold and Kahn [2000] and Clarke, de Silva, and Thorley [2002] mainly target to provide the total impact (rather than a decomposition) on the cost of constraints. We would like instead to measure the impact of constraints down to individual restrictions at the security holding level or else be able to exactly specify the cost of alternative sets of constraints.

USE OF LAGRANGE MULTIPLIERS

Rather than providing a top line number, we would like to be able to specify the costs of constrained investing (especially the long-only constraint) at the individual security holding level. If we can achieve this, we can relate the costs of constraints to individual security characteristics like benchmark weighting or stock market beta.

In order to do this we must return briefly to first principles. We first rewrite (1) in vector notation, so the problem can be stated as:

$$\max \mathbf{w}^T \boldsymbol{\alpha} - \frac{\lambda}{2} \mathbf{w}^T \boldsymbol{\Omega} \mathbf{w} \tag{8}$$

where $\mathbf{w} = (w_1, \ldots, w_n)^T$ represents the $n \times 1$ vector of active weights for all $i = 1, \ldots, n$ stocks; $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)^T$ denotes the $n \times 1$ vector of raw alphas (original forecasts); $\lambda = (\alpha / \sigma^2)$ represents investors' risk aversion; and $\boldsymbol{\Omega}$ stands for the $n \times 1$ covariance matrix of asset returns. We also allow a set of linear equality constraints, which can be summarized as:

$$\mathbf{Aw} = 0 \tag{9}$$

where $\mathbf{A}$ stands for a $k \times n$ matrix of stock characteristics. Note that k represents the number of equality constraints, in our case, $k = 3$, and the exact form is given by:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ s_1 & s_2 & \cdots & s_n \\ \beta_1 & \beta_2 & \cdots & \beta_n \end{bmatrix} \quad \mathbf{Aw} = \begin{bmatrix} \sum_i w_i \\ \sum_i w_i s_i \\ \sum_i w_i \beta_i \end{bmatrix}$$

The final constraint is a long-only constraint in disguise, as we require:

$$\mathbf{w} \geq -\mathbf{w}_b \tag{10}$$

and again $\mathbf{w}_b$ is an $n \times 1$ vector including the individual benchmark weightings.

We can summarize all this in a Lagrange function:

$$L = \mathbf{w}^T \alpha - \tfrac{\lambda}{2} \mathbf{w}^T \Omega \mathbf{w} \\ + \boldsymbol{\gamma}^T (\mathbf{w} + \mathbf{w}_b - \mathbf{s}^2) + \boldsymbol{\theta}^T (\mathbf{Aw}) \tag{11}$$

where $\mathbf{s}^2$ is used for a vector of slack variables.

Note that $\boldsymbol{\gamma}$ represents an $n \times 1$ vector of Lagrange multipliers on the individual no-short constraint. These multipliers are always positive, as a negative value would actually reverse the constraint. The vector $\boldsymbol{\theta}$ is $k \times 1$ and represents the Lagrange multipliers on our equality constraints. In contrast to the Lagrange multipliers on the long-only constraint, their signs are undetermined.

The first-order condition of (11) tells us that we can write the optimal constrained solution in closed form as a function of the Lagrange multipliers:

$$\mathbf{w}^c = \lambda^{-1} \boldsymbol{\Omega}^{-1} (\boldsymbol{\alpha} + \boldsymbol{\gamma} + \mathbf{A}^T \boldsymbol{\theta}) \tag{12}$$

This is important, as these multipliers are a natural way to describe the costs associated with a given constraint. For any binding constraint, they describe the shadow price of relaxing these constraints in terms of units of the objective function won. In fact (12) is only one among many Karush-Kuhn-Tucker (KKT) conditions for solving a quadratic program.[4]

At this stage we do not intend to ask ourselves how to determine the Lagrange multipliers optimally. Virtually all optimization packages routinely provide the Lagrange multipliers of a constrained optimization problem. There is no need for the reader to replicate the calculations performed by sophisticated packages like CPLEX or NUOPT. All one needs to do is to store, use, and interpret the Lagrange multipliers in the way we suggest in this article.

Note that we can use this framework to express liquidity costs. *The inability to enter a position works like a constraint, so we can relate the lack of liquidity to its costs.*

We start with the well-known expression for implied returns. For a given vector of constrained portfolio weights, $\mathbf{w}^c$, the implied alphas $\boldsymbol{\alpha}^c$ can be defined as $\lambda\boldsymbol{\Omega}\mathbf{w}^c$. In other words: The required alpha for a given asset equals its marginal contribution to risk times risk aversion.

Multiplying both sides of (12) by $\lambda\boldsymbol{\Omega}$, and using that identity, we can rewrite the first-order condition to arrive at:[5]

$$\boldsymbol{\alpha}^c = \boldsymbol{\alpha} + \boldsymbol{\gamma} + \mathbf{A}^T\boldsymbol{\theta} \tag{13}$$

where $\boldsymbol{\alpha}^c$ denotes the vector of implied returns of the constrained solution.

Economically this means that we now have a way to decompose the alphas that have been distorted through the introduction of constraints into the origins of these distortions. For the *i*-th alpha, (13) can be expressed more transparently as:

$$\alpha_i^c = \alpha_i + \gamma_i + \theta_c \cdot 1 + s_i\theta_s + \theta_\beta\beta_i \tag{14}$$

where θ_c, θ_s, and θ_β reflect the multipliers on cash-, size-, and beta-neutrality constraints.

It is worthwhile to spend some time on (14), as it essentially provides valuable insight into how raw alphas (α_i) are transformed during a process of constrained optimization into implied alphas (α_i^c). The first adjustment to raw alphas is made for hitting the lower barrier on individual position limits. If a lower limit for the *i*-th stock is hit (for example, a minus 3.5% benchmark weight), the Lagrange multiplier will be positive, $\gamma_i \geq 0$, and thus the implied alpha (α_i^c) of this stock is higher than the original (α_i). This correction is needed to make a larger short position unattractive.

Constraints implicitly change the forecast our valuation models make and therefore create inefficiencies. Suppose we face only a no-short constraint. In this case, we can rewrite (2) as (15):

$$\frac{w_i^c}{w_j^c} = \left(\frac{(\alpha_i + \gamma_i) - \sum_{j\neq i}\beta_{ij}(\alpha_j + \gamma_j)}{(\alpha_j + \gamma_j) - \sum_{i\neq j}\beta_{ji}(\alpha_i + \gamma_i)}\right)\cdot\left(\frac{\sigma_j^2(1-R_j^2)}{\sigma_i^2(1-R_i^2)}\right) \tag{15}$$

We can see in (15) that even if two assets end up hitting no constraints ($\gamma_i = \gamma_j = 0$), their relative weights under short constraints are still likely to differ from the unconstrained case as the multipliers on the remaining stocks might be non-zero, indicating a change in relative attractiveness. This is what we expect (and observe) in a constrained optimization, and it reflects the strange way constraints can work through the system.

The remaining equality constraints are broken down to the individual security level by using the security exposure to a given constraint. How does this work? Suppose we have an exceptionally high alpha for a large-cap stock, but the size constraint forces us to remain size-neutral. In this case, we are most likely forced to underweight another large-cap stock (even though it might have decent positive returns), as the overweight needs to be funded by an underweight.

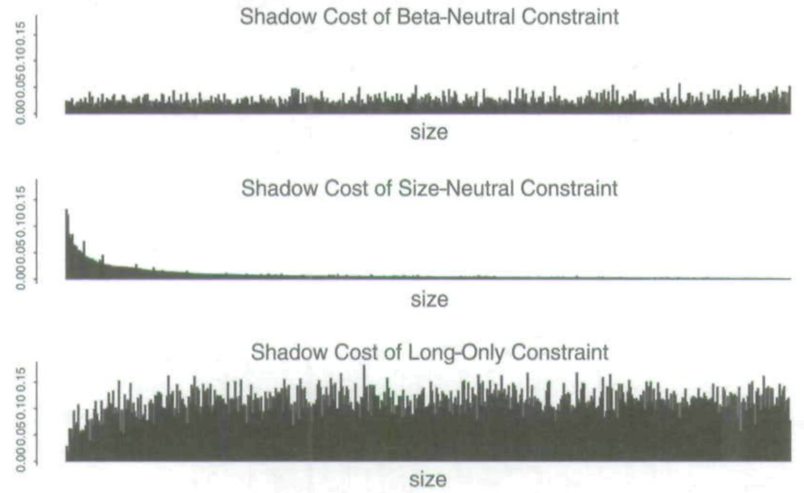
In short, while a given alpha makes the same contribution to active returns, irrespective of size, it will have a much greater (distorting effect) on constraints for a large-cap stock.

This approach allows us to identify the distortions in implicit forecasts created by each single constraint, rather than merely looking at optimized weights or calculating a top-line figure like the transfer coefficient. We are also now able to relate these distortions (shadow costs of a particular constraint) to stock characteristics or benchmark allocations.

Exhibit 3 shows the decomposition of Equation (14) for the example. The most interesting effects of constraints on implied alphas arise from the size and long-only constraints. We can see that the shadow costs (equivalently, the Lagrange multiplier) are greatest for high-capitalization stocks. Any active weight in a large-capitalization stock will lead to more distortions than an active weight in a small-cap stock, as the same weight will break the size constraint to a much greater degree. Twenty positive active 1% positions in stocks with a market cap of $1 billion are

EXHIBIT 3

Costs of Constraints—Disruption of Raw Alphas



Each graph shows the Lagrange multiplier (more precisely, the average across 10 simulations) for an S&P relative investor with a high degree of tracking risk aversion ($\lambda = 10$), versus the sorted security benchmark weighting. Along the x-axis stocks rank from large stocks (left) to small stocks (right). Covariance matrix for S&P 500 from BARRA as of 06/30/2006.

required to neutralize a single negative active 1% position in a $20 billion stock.

In simple math terms, we write this as:

$$1\% \cdot 1 + \cdots + 1\% \cdot 1 - 1\% \cdot 20 = 0$$

Compared to the size constraint, there is little impact from beta-neutrality. The greatest impact comes from the long-only constraint. Here we see a negative relation between size and shadow costs. A long-only constraint on large stocks creates little distortion (for small tracking errors) in implied alphas as the short side becomes rarely binding, while the same constraint causes much more disruption for small-stocks, due to the inability to leverage negative alphas.

CONSTRAINTS, ALPHA, AND VALUE-ADDED

Follwing Grinold and Easton [1998], we use the relation expressed in Equation (13) to explain the distorted alphas as a function of constraints. This allows us to answer questions such as: What harm can constraints do to your forecasts? The more interesting question, however, is: What harm can constraints do to value-added, i.e., investor utility? Remember that constraints might create distortions in implied alphas, but could still have little impact on value-added.

Let us focus on the alpha differential first. We suggest using the first-order KKT conditions in (12) to directly break down the difference in portfolio alpha for the constrained and unconstrained portfolios. In other words, we want to decompose the difference between the returns of an unconstrained portfolio (α) and a constrained portfolio (α^c) into a function of Lagrange multipliers.

We start with definition of the alpha differential:

$$\alpha - \alpha^c = (\mathbf{w} - \mathbf{w}^c)^T \boldsymbol{\alpha} \qquad (16)$$

Then we substitute (12) into (16) and use $\mathbf{w} = \lambda^{-1}\Omega^{-1}\boldsymbol{\alpha}$ to arrive at

$$\begin{aligned}
\alpha - \alpha^c &= (\lambda^{-1}\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha} - \lambda^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{\alpha} + \gamma + \mathbf{A}^T\boldsymbol{\theta}))^T \boldsymbol{\alpha} \\
&= (\lambda^{-1}\boldsymbol{\Omega}^{-1}(-\gamma - \mathbf{A}^T\boldsymbol{\theta}))^T \boldsymbol{\alpha} \\
&= (-\gamma^T - \boldsymbol{\theta}^T\mathbf{A})\lambda^{-1}\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha} \\
&= -(\gamma^T\mathbf{w} + \boldsymbol{\theta}^T\mathbf{A}\mathbf{w})
\end{aligned}$$

In other words, the (positive) difference between unconstrained and constrained return is a function of the Lagrange multipliers and thus can be decomposed into its origins.

Expanding this expression, we get the breakdown:

$$\alpha - \alpha^c = -\sum_i w_i \gamma_i - \theta_c \sum_i w_i - \theta_s \sum_i w_i s_i - \theta_\beta \sum_i w_i \beta_i$$

The first term, $-\Sigma_i w_i \gamma_i$, represents the contribution of the long-only constraint. It is clearly positive as γ_i is positive only for those stocks where the long-only constraint is binding, i.e., where w_i is negative.

The ultimate objective of any analysis regarding the costs of constraints is to quantify the impact of constraints on the loss in utility relative to an unconstrained solution. This has not been addressed in the literature.

Using the same framework as above, we will show how to decompose the difference in value-added into its contributors. Let us first substitute $\boldsymbol{\alpha} = \lambda\Omega\mathbf{w}$ into the difference in value-added given in (17):

$$U - U^c = \left(\mathbf{w}^T\boldsymbol{\alpha} - \frac{\lambda}{2}\mathbf{w}^T\Omega\mathbf{w}\right) - \left(\mathbf{w}^{cT}\boldsymbol{\alpha} - \frac{\lambda}{2}\mathbf{w}^{cT}\Omega\mathbf{w}^c\right) \quad (17)$$

where unconstrained and constrained utility are denoted by U and U^c, respectively.

The difference in utility is then more compactly written as

$$\begin{aligned} U - U^c &= \frac{\lambda}{2}(\mathbf{w} - \mathbf{w}^c)^T\,\Omega(\mathbf{w} - \mathbf{w}^c) \\ &= \frac{\lambda}{2}(\lambda^{-1}\Omega^{-1}(\gamma + \mathbf{A}^T\boldsymbol{\theta}))^T\;\Omega(\lambda^{-1}\Omega^{-1}(\gamma + \mathbf{A}^T\boldsymbol{\theta})) \\ &= \frac{\lambda}{2\lambda}(\gamma + \mathbf{A}^T\boldsymbol{\theta})^T\;\Omega^{-1}(\gamma + \mathbf{A}^T\boldsymbol{\theta}) \end{aligned}$$

We can view this as the risk term $\psi^2 \equiv 2\lambda(U - U^c)$.[6] The reader can recognize this is linear-homogeneous and hence decompose it into its marginal contributions:

$$\begin{aligned} \psi &= \gamma^T\frac{d\psi}{d\gamma} + \theta^T\frac{d\psi}{d\theta} \\ &= \gamma^T \cdot \frac{\Omega^{-1}(\gamma + \mathbf{A}^T\theta)}{\psi} + \theta^T\frac{\mathbf{A}\Omega^{-1}(\gamma + \mathbf{A}^T\theta)}{\psi} \end{aligned}$$

Finally, multiplying both sides by ψ again, we arrive at:

$$U - U^c = \frac{1}{2\lambda}\left(\underbrace{\gamma^T \cdot \Omega^{-1}(\gamma + \mathbf{A}^T\theta)}_{\substack{\text{contribution of}\\ \text{no short constraint}}} + \underbrace{\theta^T \cdot \mathbf{A}\Omega^{-1}(\gamma + \mathbf{A}^T\theta)}_{\substack{\text{contribution of}\\ \text{equality constraint}}}\right) \quad (18)$$

This gives us a compact expression for the difference in utility as a function of the respective Lagrange multipliers. We can now use (18) to continue our S&P example and distinguish the relative contributions as shown in Exhibit 4.

We start with the base case, where we assume that alphas and stock size are unrelated, i.e., their correlation is zero. This is the case of true alphas reflected in the first two rows for lambdas of 4 and 10, respectively. The long-only constraint is responsible for almost 100% of the utility loss generated by introducing constraints.

What might be surprising is the zero impact of the size-neutrality constraint. Where does this come from?

EXHIBIT 4

Percentage Contribution to Loss in Utility

Correlation of alpha and size	Lambda	Utility loss	Percentage contribution to loss in utility			
			Cash neutrality	Beta Neutrality	Size neutrality	Long only
$\rho(\alpha_i, s_i) = 0$	10	0.0133	2%	0%	0%	98%
	4	0.0469	4%	0%	0%	96%
$\rho(\alpha_i, s_i) = -0.3$	10	0.0146	2%	0%	13%	85%
	4	0.0445	3%	0%	4%	93%
$\rho(\alpha_i, s_i) = -0.6$	10	0.0183	2%	0%	47%	51%
	4	0.0517	4%	0%	43%	53%

Calculations performed using BARRA covariance matrix as of 06/30/2006. All alphas are drawn from a normal distribution with mean and standard deviation equal to 0 and 0.006, respectively. The dominance of the long-only constraint becomes considerably less if the focus of an investment process shifts away from stock selection.

As long as the universe is large and alphas are unrelated to size, the optimizer will very likely find another large-cap stock with negative alpha for every large-cap stock with positive alpha. Maintaining size-neutrality causes minimum disruption in the portfolio construction process.

We could say alternatively that we don't need a size constraint if our signals (alphas) are almost size-neutral. After all, a size constraint is a safeguard against sloppy analysis that confuses the premium for small-cap stocks with genuine alpha. A common example is firms that rank stocks on the basis of their dividend discount rate (discount rate that equates stock price with projected cash flow to equity owners). Small-cap stocks naturally have higher dividend discount rates because they tend to represent higher risks.

Let us instead introduce a negative correlation between alpha forecasts and firm size. This can be either the result of sloppy analysis as mentioned earlier or because small-cap stocks tend to be underresearched and hence offer higher alphas than large-cap stocks. Assuming a negative correlation of –0.3 between alpha signal and firm size (more precisely benchmark weighting), we arrive at rows three and four in Exhibit 4. The size constraint starts to have an impact on total value-added.

Increasing this number to -0.6 changes the picture entirely. The size constraint now has the same negative impact as a long-only constraint.

The reason lies in the inability of the optimizer to compensate active weights in a large-cap stock by another large-cap stock, because this group of stocks has alphas that tend to be similar in direction. As a result, the optimizer cannot make use of the information that large-cap stocks as a group outperform or underperform small-cap stocks. If this information becomes dominant in the alpha signal (raising correlation), the corresponding in-sample loss due to the long-only constraint becomes very damaging.

This raises a more philosophical point. Size constraints have traditionally arisen from both clients' unwillingness to pay portfolio managers for collecting a risk premium and asset managers' unwillingness to swamp stock-picking performance by large timing bets.

While our decomposition in (18) will always show that it is in-sample a good idea to relax the size constraint, the same is not necessarily true out-of-sample. If one has no skill in timing the size premium (overweight size if the size premium is higher than its equilibrium premium), taking size bets will only crowd out alpha bets for any given tracking risk and depress the information ratio. Ex post, the size constraint might very well increase value-added.

An analysis like that in Exhibit 4 also helps a diagnosis. If the size constraint becomes a major drag on value-added, the investment process needs to be reviewed. Either the constraint needs to be relaxed, or the analysts need to refocus on alpha instead of collecting a risk premium.

LEVERAGE AND DIMINISHING SHADOW COSTS

To analyze the benefit of *partially* relaxing the long-only constraint, we first lay out the long/short optimization formulation used in our example. For convenience of formulation, we denote securities' active holding positions as

$$w_i = w_{l,i} - w_{s,i} - w_{b,i} \quad w_{l,i} \geq 0, \quad w_{s,i} \geq 0 \tag{19}$$

where w_{ij} represents a security's long position while $w_{s,i}$ denotes its short position, and w_{bi} is the benchmark weight.

As the partially relaxed long-only portfolio, e.g., 120/20, allows short positions as long as the portfolio's total leverage exposure is within a prespecified range, we need to replace the long-only constraint by a leverage constraint that specifies the portfolio's maximum leverage exposure L:

$$\sum_i w_{l,i} + \sum_i w_{s,i} \leq 1 + L \tag{20}$$

For the long side of 120, we need to short (borrow against a lending fee and sell) stock worth 100. We also might want to introduce integer variables to prevent simultaneous long and short of the same securities. We use the constraints to achieve this:

$$b_i m \geq w_{l,i}, \; (1 - b_i)m \geq w_{s,i} \tag{21}$$

where m is a high number.

Note that Equation (21) operates like a switch. If $b_i = 1$, then the short position of security i will be forced to equal zero; otherwise, the long position will equal zero. The size-neutrality and beta-neutrality constraints are the same as before.

Although the formulation in (21) allows us to find the optimal active holding position of a relaxed problem, it also makes explanation of the shadow costs unclear, because of the separation of the long and short position. Another downside is that the Lagrange problem becomes non-differentiable exactly because of the integer variable.

To make our utility decomposition methodology work for the partially relaxed long-only problem, we propose a two-step optimization approach to calculate the shadow costs of constraints. In the first step, we solve the optimization problem with the help of the binary variables to achieve a portfolio satisfying the leverage and other equality constraints. We then prespecify the sign of the holding positions of all the securities according to the optimal value of the binary variables and resolve an equivalent problem.

The equality constraints can be summarized as:

$$\begin{bmatrix} l_1 & l_2 & \dots & l_n \\ 1 & 1 & \dots & 1 \\ s_1 & s_2 & \dots & s_n \\ \beta_1 & \beta_2 & \dots & \beta_n \end{bmatrix} \quad Aw = \begin{bmatrix} \sum_i w_i l_i \\ \sum_i w_i \\ \sum_i w_i s_i \\ \sum_i w_i \beta_i \end{bmatrix}$$

$$B = \begin{bmatrix} 1 + L - \sum_i w_{b,i} l_i \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The first constraint is the leverage constraint, which requires the portfolio's leverage to equal L. The coefficients of the leverage constraint are defined as

$$l_i = \begin{cases} 1 & if \quad b_i = 1 \\ -1 & if \quad b_i = 0 \end{cases} \tag{22}$$

where b_i is the value of the integer solution of the model with integer decision variables.

In addition to the equality constraints, the inequality constraints specify the sign of the active positions based on the value of b_i:

$$w_i + w_{b,i} \geq 0 \quad \text{if} \quad b_i = 1$$
$$w_i + w_{b,i} \leq 0 \quad \text{if} \quad b_i = 0$$

We can now find the dual variable for each constraint and analyze the effects of constraints on alpha distortion and utility loss. In fact, this two-step optimization approach assumes that the optimizer uses the branch and bound approach to tackle the integer problem, and the second step simply takes the optimal integer solutions to find the meaningful shadow costs for each constraint. Equipped with this theoretical insight, we can now recalculate Equation (18) under the additional use of leverage.

Exhibit 5 summarizes the results, where we explicitly assume that $\rho(\alpha_i, s_i) = 0$. Thanks to our methodology, we can observe two characteristics of leveraged solutions.

First, we see that the loss in value-added is lower after allowing leverage if we compare the numbers for utility loss in both examples. Note that leverage does not improve performance per se; i.e., the result is not due to taking greater risks, but rather to taking risks more intelligently and efficiently.

Second, we observe that as leverage increases, the long-only constraint becomes less and less binding; i.e., we see a decline in its contribution to the difference in value-added. This should come not as a surprise but as a confirmation. After all, we used leverage to partially make the long-only constraint obsolete.

EXHIBIT 5
Impact of Leverage

Leverage	Utility Loss	Percentage contribution to loss in utility			
		Leverage	Cash Neutrality	Size Neutrality	Long-Only
120/20	0.0125	75%	0%	0%	25%
140/40	0.0112	79%	0%	0%	21%
160/60	0.0100	81%	0%	0%	19%
180/80	0.0098	83%	0%	0%	17%

Calculations performed using BARRA covariance matrix as of 06/30/2006. All alphas are drawn from a normal distribution with mean and standard deviation equal to 0 and 0.006, respectively. Correlation between alpha and size has been kept at zero: $\rho(\alpha_i, s_i) = 0$. An increase in leverage reduces the slippage in utility across all levels of risk aversion. The more leverage allowed, the less binding the long-only constraint becomes. Note that L 5 0.4 relates to a 120/20 product.

SUMMARY

We have provided a closer look at the costs of constraints and suggested a new methodology to measure the impact of individual constraints on an investor's value-added. Rather than calculating the decay in information ratio, we show how to use Lagrange multipliers in order to decompose the difference in value-added into the contributions coming from alternative constraints.

This differs considerably from current methods that either focus on a headline measure of information ratio shrinkage (also called transfer coefficient) or express the impact of constraints as distortions in implied alphas. While the latter approach can answer questions like: "What harm can constraints do to your forecasts?" the more interesting question to us is to ask: "What harm can constraints do to value-added, i.e., investor utility?" Constraints might create distortions in implied alphas, but could still have little impact on value-added, as we have seen from the size constraint in our setting.

We also suggest using a decomposition of value-added into its constraint-related parts as a diagnostic measure. As real alphas show no correlation with firm-specific characteristics (or they would be called beta), constraints on various betas will have little impact on the optimality of a solution.

Size-, beta-, or value-neutrality constraints become disruptive only when an investment process becomes dominated by taking factor bets. In this case, it is not clear that one should remove the constraints. In fact, this has been the reason why constraints are there in the first place, i.e., to align portfolio construction with skill.

Finally, we should note that this kind of analysis can be performed only for an asset management process that provides explicit signals (rather than directly choosing weights) and uses portfolio optimization. In other words, quantitative managers will benefit by far the most.

ENDNOTES

At the time this article was written, Bernd Scherer was head of research and head of portfolio engineering at Deutsche Asset Management in New York. The authors thank Richard Grinold and Colm O-Cinneide for many useful suggestions.

[1]See Fabozzi, Focardi, and Kolm [2006] for an excellent review of portfolio and utility optimization in an equity context.

[2]See Stevens [1998] and Lee [2000].

[3]Grinold and Kahn [2000] call this the information ratio shrinkage.

[4]See Fletcher [1987] for an excellent resource on solving optimization problems.

[5]See Grinold and Easton [1998].

[6]See Grinold [2006], Equations (5)–(9) for an independent derivation with a similar focus.

REFERENCES

Clarke, Roger, Harindra de Silva, and Steven Thorley. "Portfolio Constraints and the Fundamental Law of Active Management." *Financial Analysts Journal*, September/October 2002, Vol. 58, pp. 48–66.

Fabozzi, Frank J., Sergio M. Focardi, and Peter N. Kolm. *Financial Modeling of the Equity Market*. Hoboken, NJ: Wiley, 2006.

Fletcher, R. *Practical Methods of Optimization*. New York: Wiley, 1987.

Grinold, Richard C. "Implementation Efficiency." *Financial Analysts Journal*, September/October 2006, Vol. 62, pp. 52–63.

Grinold, Richard C., and Kelly A. Easton. "Attribution of Performance and Holdings." In Willliam T. Ziemba and John M. Mulvey, eds., *Worldwide Asset Liability Modeling*. Cambridge: Cambridge University Press, 1998.

Grinold, Richard C., and Ronald N. Kahn. *Active Portfolio Management*. New York: McGraw-Hill, 2000.

Lee, Wai. *Theory and Methodology of Tactical Asset Allocation*. New Hope, PA: Fabozzi Associates, 2000.

Stevens, Guy V. "On the Inverse of the Covariance Matrix in Portfolio Analysis." *Journal of Finance* (1998), Vol. 53, pp. 1821–1827.