INFORMATION HORIZON, PORTFOLIO TURNOVER, 
AND OPTIMAL ALPHA MODELS

Edward Qian, PhD, CFA
Director, Head of Research, Macro Strategies
PanAgora Asset Management
260 Franklin Street
Boston, MA 02109
Phone: 617-439-6327
Fax: 617-790-2642
Email: eqian@panagora.com

Eric H. Sorensen, PhD
President & Chief Executive Officer
PanAgora Asset Management
260 Franklin Street
Boston, MA 02109
Phone: 617-439-6315
Fax: 617-790-2649
Email: esorensen@panagora.com

Ronald Hua, CFA
Director, Head of Research, Equity Group
PanAgora Asset Management
260 Franklin Street
Boston, MA 02109
Phone: 617-439-6171
Fax: 617-790-2635
Email: rhua@panagora.com

Draft Date: 02/26/2007

(Forthcoming, Journal of Portfolio Management, Fall 2007)
INFORMATION HORIZON, PORTFOLIO TURNOVER, AND OPTIMAL ALPHA MODELS

INTRODUCTION

The bottom-line value of an active investment process comprises two parts: the theoretical value of the alpha skill (the gross paper profit) minus the cost of implementation. The larger is the former and the smaller is the latter, the happier is the investor. Clearly, the total asset under management (AUM) influences the latter. A strategy might be profitable with small assets under management and unprofitable with larger assets under management - as asset amounts grow, transaction costs grow. Recently, Kahn and Shaffer (2005) pointed out that one remedy to the “size” problem is to reduce portfolio turnover. This is a sensible suggestion. However, their work is based on a hypothetical relationship between turnover and expected alpha that might be too general to be applicable. In reality, any relationship between turnover and expected alpha is not exogenous. It depends on alpha factors, their weights in an alpha model, and rebalance horizon.

In this paper, we propose an analytic framework for integrating alpha models with portfolio turnover. In practice, many alpha models are not constructed in such an integrated framework. Typically, managers first develop an alpha model (with little consideration given to turnover), and then throw the alpha model into an optimizer, setting turnover constraints to handle the transactions costs. There are 2 drawbacks to this two-step process: 1) it creates difficulty knowing the true effectiveness of the alpha model; and 2) it does not allow managers to adjust the alpha model along the way as AUM grows.

To integrate alpha model with portfolio turnover, we extend the previous work by Qian
models with the objective of maximum information ratio. This work relies heavily on the time series correlation of information coefficients (IC’s) as well as contemporaneous correlations between factor signals. In this paper, we explore these critical correlation analyses to include serial correlations of factors and the concept of factor information horizon. This allows us to formally evaluate implementation costs in finding the factor weights that optimize information ratio (IR) with net returns.\textsuperscript{1} We first present a general discussion about information horizon, and then derive an analytic formula for portfolio turnover conditioned on changes in forecasts. This solution allows us to estimate portfolio turnover for different quantitative alpha factors and related investment strategies. We find that portfolio turnover can be endogenous in a complete system, and that factor autocorrelation is the key exogenous ingredient. Finally, we build multi-factor model by maximizing IR under portfolio turnover constraint. A numerical example employing a value and a momentum factor is used to illustrate our framework.

**INFORMATION HORIZON**

The concept of information horizon is crucial to portfolio turnover. If, on the one hand, a factor has relatively short information horizon – it predicts security returns only for the very near term. The signal decays quickly. Momentum factors, especially the one-month reversal factor, behave this way. Such factors cause high turnover since their exposures must be replenished constantly. On the other hand, if a factor has a relatively long information horizon – it predicts long after its information becomes available. The signal decays slowly. Valuation factors tend to behave this way. These factors will have low turnover – portfolios constructed based on lagged factors can still generate excess returns.

In reality, no two factors (models) have the same profile in terms of information...
long as several years. It seems reasonable that, depending on the predictive power of different factors, the turnover frequency\(^2\) will vary with the particular alpha model. The two should be in balance.

**Some IC terms**

We study the general concept of information horizon through two specific expressions related to the information coefficient (IC): *lagged IC* and *horizon IC*. We denote IC as the cross sectional correlation coefficient between the factors value at the start of time \(t\) and the security returns over time period \(t\), i.e., \(\text{corr}_{tt}\). Consider this the typical one-period IC measure, for a month or a quarter. An example is the first quarter return IC. The factor values are observed December 31, and the return period is January to March.

The *lagged IC* is the correlation coefficient between time \(t\) factor values and a later period (by 1, 2 or more quarters) return vector, i.e., \(\text{corr}_{tlt}\), with lag \(l\). Compared to the return, the factor is lagged by \(l\) periods. For example, using factor readings on December 31, we can correlate it with returns for later periods (second quarter \([l = 1]\), third quarter \([l = 2]\)), and so on. The IC will typically decay in power as the lag increases. However, the rate of decrease differs across different factors such as momentum and value. For simplicity, we shall assume the ICs are generated by “stationary processes” implying that

Table 1 shows the average quarterly ICs and standard deviation of ICs of a specific price momentum factor based on 9-month returns (referred hereafter as PM). The average IC is high with no lag but it declines steadily with greater lags. With a lag of 3, the average IC is close to zero, indicating no predictability for the return. Table 2 shows the average quarterly
ICs and standard deviation of ICs for a specific value factor based on trailing earning yield (referred hereafter as E2P). The average IC is lower compared to that of momentum, but shows less decay in power as the lag increases.

In both cases, the IC standard deviations are stable with respect to the change in lags. The standard deviation of IC is much higher for the momentum factor than that of the value factor. As a result, the IR (annualized by multiplying 2) of the value factor is higher than that of the momentum factor at all lags. With lag of three quarters, the IR of momentum factor is nearly zero while that of the value factor is still close to 1!

**Table 1 Quarterly average and standard deviation of IC and lagged ICs for the momentum factor - PM. The IR is the annualized ratio of the average IC to the standard deviation of IC.**

<table>
<thead>
<tr>
<th>Lag</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg</td>
<td>0.073</td>
<td>0.047</td>
<td>0.026</td>
<td>0.003</td>
</tr>
<tr>
<td>Stdev</td>
<td>0.115</td>
<td>0.112</td>
<td>0.103</td>
<td>0.085</td>
</tr>
<tr>
<td>IR</td>
<td>1.269</td>
<td>0.847</td>
<td>0.505</td>
<td>0.078</td>
</tr>
</tbody>
</table>
Table 2 Quarterly average and standard deviation of IC and lagged ICs for the value factor – E2P. The IR is the annualized ratio of the average IC to the standard deviation of IC.

<table>
<thead>
<tr>
<th>Lag</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg</td>
<td>0.043</td>
<td>0.029</td>
<td>0.030</td>
<td>0.028</td>
</tr>
<tr>
<td>Stdev</td>
<td>0.066</td>
<td>0.061</td>
<td>0.060</td>
<td>0.057</td>
</tr>
<tr>
<td>IR</td>
<td>1.304</td>
<td>0.939</td>
<td>1.002</td>
<td>0.992</td>
</tr>
</tbody>
</table>

Another important variant on the standard IC helps us understand the cumulative effect of multiple lagged ICs. We define horizon IC as IC of a factor at a given time, $t$, for subsequent returns over multi-period horizons. For example, if we have a factor available at December 31, we are interested in its correlations with cumulative returns of next quarter, next two quarters, or next three quarters, etc. We denote $\hat{R}_{tt}$ as the risk-adjusted cumulative returns from period $t$ to period $t+h$, and denote $\mathbb{I}_{tt+h}$ as the horizon IC. For example, $\hat{R}_{t}$ is the standard IC for the return in period $t$, and $\mathbb{I}_{tt}$ is correlation between the factor and the return vectors over the next six months, periods 1 and 2.

The relationship between lagged IC and horizon IC

Although the lagged IC typically decays with the lag, horizon IC often increases with the horizon, at least initially. By definition, the cumulative multi-period return in the horizon IC relates to the single period return by $\mathbb{I}_{tt+h} = \hat{R}_{tt+h}$. When the periods return are small, we can approximate with $\mathbb{I}_{tt+h} \approx \hat{R}_{tt}$. Using it in the horizon IC yields
If we further assume that the risk-adjusted returns from different periods are uncorrelated\(^3\), then the IC simplifies to:

\[
\text{(1)}
\]

The horizon IC is an average of lagged ICs times the square root of the horizon length (\(L\)). Note that the horizon IC covers returns of multiple periods, and the lagged ICs cover forecasts for single intervals for future periods. Suppose there is no information decay in the lagged forecasts, i.e., \(1 \approx L\). Then from (2) we have \(1 \approx \sqrt{L}\). In this case the horizon IC is the IC times the square root of the horizon length, and it therefore increases as the horizon lengthens.

Even when there is information decay, the horizon IC can still initially increases with the horizon length. It then declines as the horizon lengthens further and/or as the lagged IC’s decline more rapidly. Figure 1 plots one such case, in which the initial period IC is 0.10. The lagged IC is 0.08 with lag 1, 0.06 with lag 2, and so on. It reaches zero with lag 5, and turns negative thereafter. The horizon IC increases at first. For example, the IC is 0.128 for returns over the next two periods (with 1 on the horizontal axis), and 0.139 for returns over the next three periods (with 2 on the horizontal axis). However, the horizon IC is eventually dragged down by the declining lagged ICs.
Table 3 shows the horizon IC of the two factors from Tables 1 and 2, based on equation (2). The horizon IC for the PM factor starts to decline at three quarters while that of E2P keeps increasing.

Table 3 Horizon ICs for the two factors

<table>
<thead>
<tr>
<th>Horizon</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>0.073</td>
<td>0.085</td>
<td>0.085</td>
<td>0.075</td>
</tr>
<tr>
<td>E2P</td>
<td>0.043</td>
<td>0.051</td>
<td>0.059</td>
<td>0.065</td>
</tr>
</tbody>
</table>

Horizon IC and trading horizon

The propensity for the horizon IC to increase initially with the horizon does not necessarily mean that we can increase the total IC for a longer trading horizon. Longer trading horizons allow fewer opportunities to rebalance, or fewer chances along the time dimension. Basically, the longer horizon reduces the *breadth* of the process, which adversely impacts IR.
The reduced breadth leads to higher variability of horizon IC. As a result, when we measure IR in terms of ratio of average IC to the standard deviation of IC, the longer trading horizon may not have any advantage. Consider the following example in which both forecasts and returns are of quarterly frequency. The quarterly IC has a mean of 0.1 and standard deviation of 0.2. Then the quarterly IR is $0.5^4$ and the annualized IR is $\sqrt[4]{0.5}$. Let’s assume that all subsequent one-period lagged ICs have the identical distribution, and are uncorrelated. Then according to equation (2), the horizon IC of one year, or 4 quarters, will have a mean of $\sqrt[4]{0.1} \cdot \sqrt[4]{0.2}$, and a standard deviation of $\sqrt[4]{0.1} \cdot \sqrt[4]{0.2}$. Hence, the annual IR is also $1 - \sqrt[4]{0.5}$ the same as annualized IR of quarterly trading. There is no difference in terms of the performance, gross of any trading costs.

However, there could be difference in trading cost that arises from the implication of quarterly trading (rebalancing), as opposed to annual trading. In one case, we trade once a year. In the other case, we trade four times per year. The question is whether the total trading costs in the latter case exceeds the cost in the former case. The answer to this question depends on the nature of the turnover induced by changes in alpha factors over each quarter versus the one-year holding period. Thus, the alpha model profile jointly with trading costs become endogenous with respect to each other, and interact in determining the maximum deliverable net (after cost) IR.

**TURNOVER CAUSED BY CHANGE IN FORECASTS**

Consider turnover over a single trading period, in which the active weights change from $w_i^t$ to $w_i^{t+1}$, for each security i. We define one-way turnover as one half times the sum of the absolute value of weight changes$^5$
We assume the active weights for each security result from an unconstrained mean-variance optimization based on residual return and residual risk, respectively at time $t$ and $t+1$:

$$\sum_{i} w_i \mu_i = \sum_{i} w_i \lambda \sigma_i = -\lambda \sum_{i} w_i \lambda \sigma_i,$$  

(4)

are risk-adjusted forecasts at $t$ and $t+1$; are risk aversion parameters. For simplicity, we assume that all stock specific risks remain unchanged and that number of stocks remains unchanged. If we hold constant the targeted tracking error for the portfolio, then the risk aversion parameter is given by (Qian & Hua 2004)

$$\sqrt{\frac{\sum_{i} w_i \lambda \sigma_i}{\sqrt{\sum_{i} w_i \lambda \sigma_i}}}.$$  

(5)

Substituting (5) into (4) gives

$$\sqrt{\frac{\sum_{i} w_i \lambda \sigma_i}{\sqrt{\sum_{i} w_i \lambda \sigma_i}}}.$$  

(6)

in which are now standardized with . In other words, they reduce to simple z scores. Equation (6) states the active weight of a stock is directly proportional to the portfolio target tracking error and its z score, but is inversely proportional to the specific risk and the square root of the number of stocks. Using equation (6) in (3) gives

$$\sqrt{\frac{\sum_{i} w_i \lambda \sigma_i}{\sqrt{\sum_{i} w_i \lambda \sigma_i}}}.$$  

(7)

The most difficult aspect of analyzing turnover is dealing with the absolute value function above. Our solution to this problem is to approximate the turnover in equation (3) as the expectation of the absolute difference of two continuous variables underlying two sets of
forecasts. When we do this, we can rely on standard statistical theory to evaluate various expectations. In the appendix, we show the portfolio turnover is given by

\[ \sqrt{\frac{1}{N}} \sqrt{1 - \rho^2} \cdot \sigma \pi \]

Equation (8) represents our solution for forecast-induced turnover for an unconstrained portfolio\(^6\). It depends on four elements. The turnover is higher:

- The higher the tracking error
- The larger the number of stocks - proportional to the square root of \(N\)
- The lower the forecast autocorrelation - cross sectional correlation between the consecutive forecasts,

\[ \rho_{\text{corr}} \]

- The lower the average stock specific risk

The forecast autocorrelation becomes the most relevant for our analysis of turnover. There is considerable intuition behind this. For example, consider two extremes. On the one hand, if the correlation between the consecutive forecasts is equal to one, then the weights are identical and there is no turnover. When the correlation is less than unity, it will be advantageous to incur turnover. And, on the other hand, at another extreme, with a correlation of -1, turnover will be at the maximum. In deed, all weights flip signs and the portfolio “reverses” itself.

For example, consider the case where stock specific risks are the same for all stocks; in this case the turnover is reduced to

\[ \sqrt{\frac{1}{N}} \sqrt{1 - \rho^2} \cdot \sigma \pi \]

(9)
For an active portfolio (which could either be long-short market neutral or active versus a benchmark) with $N = 500$, model $5\% \sigma = 0.30\%$, and $0.9\rho = \sqrt{\cdot}$, the one-time turnover would be 

\[
T = \sqrt{\cdot}.
\]

Figure 2 plots the function $\sqrt{\cdot}$. This is the dependence of turnover on the forecast autocorrelation. Turnover is a decreasing function of forecast autocorrelation. The function behaves close to a linear function for most of the range, but it drops more precipitously when $\rho$ is greater than 0.8.

**Figure 2 The dependence function of turnover on the forecast autocorrelation**

- **Forecast autocorrelations of quantitative factors**

  Among the common quantitative factors employed by practitioners, value factors as a category generally have the highest forecast autocorrelation, and thus the lowest turnover. They also have the slowest information decay characterized by high lagged ICs. Their forecast autocorrelation can be as high as 0.95. Among value factors, those based on cash flow have slightly lower forecast autocorrelation. In contrast, the momentum factors have the lowest
forecast autocorrelation, and thus the highest turnover. Momentum factors’ forecast autocorrelations generally lies between 0.6 and 0.7. Also, for price momentum factors, the autocorrelation typically increases as the time window used for return calculation lengthens up to 12 months. Therefore, one should use a longer time window to measure price momentum in order to reduce turnover.

TURN OVER OF MULTI-FACTOR MODELS

Multi-factor models offer diversification among factors and can increase information ratio of the overall model [see Sorensen et al 2004] drawing on results of the previous section, we analyze the turnover of multi-factor models by studying their forecast autocorrelation. Multi-factor models can depend on both cross sectional and time series average of multiple factors. We can reduce the portfolio turnover if the composite model has the higher forecast autocorrelation than individual factors.

Cross sectional average

In the cross sectional dimension, the average consists different factors whose performance are evaluated over the same time interval. To illustrate, consider a two-factor case, in which the composite forecasts are linear combinations $c = F_1 + F_2$, both $F_1$ and $F_2$ are standardized factors, and $v_1$, $v_2$ are their weights. The autocorrelation of the composite factor is

$$\rho_c = \rho_{F_1, F_2} + \rho_{v_1, v_2} + \rho_{F_1, v_2} + \rho_{F_1, v_1} + \rho_{v_1, v_2} + \rho_{F_2, v_2} + \rho_{F_2, v_1} + \rho_{v_1, v_1} + \rho_{F_1, F_2}.$$  

In the formula, $\rho_{F_1, F_2}$ is the contemporaneous correlation between the two factors, and $\rho_{v_1, v_2}$ is the correlation between $v_1$ and $v_2$. For $\rho_{F_1, v_2}$, $\rho_{v_1, v_2}$ is the serial autocorrelation of a single factor, and for $\rho_{F_1, v_1}$, $\rho_{v_1, v_2}$ is the serial cross-correlation between two different factors.
The autocorrelation of the composite factor depends on this correlation structure and on factor weights. Other things equal, the autocorrelation of the composite factor will be high if the two factors have high serial auto and cross- correlation, and the composite factor has low volatility. The low composite volatility depends, in turn, on a low contemporaneous factor correlation. Generally, low contemporaneous correlation also means better factor score diversification and therefore may imply lower volatility of IC.

For a numerical example, suppose the serial autocorrelations of two factors - PM and E2P, are \( \rho_{11} = 0.11 \) and \( \rho_{22} = 0.94 \), the serial cross-correlations are \( \rho_{12} = -0.09 \) and \( \rho_{21} = 0.00 \), and the contemporaneous correlation \( \rho_{12} \). Then

\[
\begin{bmatrix}
0.68 & 0.94 \\
0.09 & 0.00 \\
\end{bmatrix}
\]

Figure 3 plots the autocorrelation as a function of \( v_1 \) and we have \( \rho_{12} \). When \( v_1 = 0 \), it is equal to the autocorrelation of the factor 2 while \( v_1 = 1 \), it is equal to the autocorrelation of factor 1. When \( v_1 \) is small, there is a slight increase in the composite autocorrelation.
Figure 3 The autocorrelation of composite forecast as a function of the weight in factor 1.

Mathematically, all the correlation coefficients discussed above can be neatly placed into a single symmetric non-singular matrix - the correlation matrix for the stacked vector $\mathbf{\psi}$:

$$
\begin{bmatrix}
\rho & \rho & \rho & \rho \\
\rho & \rho & \rho & \rho \\
\rho & \rho & \rho & \rho \\
\rho & \rho & \rho & \rho \\
\end{bmatrix}
$$

(11)

The calculation of autocorrelation of composite factors can be carried out using this correlation matrix (see below).

**Forecast autocorrelation of moving averages**

When signals are volatile, we can smooth them using moving averages. In the framework of multi-factor models, moving averages are also considered composite factors - a
linear combination of new and past information. A natural question is why we would use “outdated” information in the forecasts since one tends to think that forecast based on the most recent information is better than the lagged forecast, in terms of more predictive power for subsequent returns, i.e., better IR. This may be true, but one should verify this empirically.

The primary reason to use lagged forecasts for many alpha factors, is moving averages increase the autocorrelation thus lowering turnover. Despite possible information decay of lagged forecasts, the tradeoff between potential profit reduction and improved transaction costs may favor the inclusion of lagged factors in a multi-factor model. Our next step is to provide a mathematical framework for analyzing this important tradeoff.

Autocorrelations of moving averages are calculated similarly to cross sectional averages. Given forecast series \( \hat{F}_t \), we form a moving average of order \( L \) as

\[
\bar{F}_t = \frac{1}{L} \sum_{i=1}^{L} \hat{F}_{t-i}
\]

For \( 2L = 1 \). The serial autocorrelation is given by

\[
\rho(1) = \frac{1}{1} \frac{\sum_{i=1}^{L} \hat{F}_{t-i} \hat{F}_t}{\sum_{i=1}^{L} \hat{F}_{t-i}^2}
\]

In the equation, \( \hat{F}_t \) is the serial autocorrelation function of \( \hat{F}_t \), with \( \hat{F}_t = \hat{F}_t - \mu \).

For given serial autocorrelations \( \rho(1), \rho(2) \), the correlation (12) is a function of the weights, \( \omega_1, \omega_2 \). Since the correlation is invariant to a scalar, we can require \( \omega_1 + \omega_2 = 1 \). Figure 4 plots equation (12) as a function - the weight of the lagged forecast for the E2P factor, with \( \rho(1), \rho(2) \). When \( \omega_1 = 1 \), the moving average is identical to the original factor, so the serial autocorrelation is 0.94. As \( \omega_1 \) increases, the lagged forecast is added to the moving average, the serial autocorrelation of \( \rho(1) \) increases; it reaches a maximum close to
0.96 at when the two terms are equally weighted. As changes from 0.5 to 1, the autocorrelation declines from the maximum to 0.94.

**Figure 4** Serial autocorrelation of forecast moving average with , and

![Figure 4 Serial autocorrelation of forecast moving average with](image)

Figure 5 is equation (12) for the PM factor, with . Similarly, as increases, the lagged forecast is added to the moving average, the serial autocorrelation of increases; it reaches a maximum close to 0.82 at when the two terms are equally weighted. For both factors, the composite forecast is made more stable by including the lagged factors leading to reduced portfolio turnover. This is generally the case in practice.
Cross sectional and time series averages

An all-encompassing multi-factor model should have both cross sectional and time series averages. We write such models as \( \sum_{v_1} \), in which there are \( M \) factors and each of which have a moving average of order \( L \). For expository clarity, we consider here the case of two factors and one lag, all standardized with standard deviation one, i.e.,

\[
\begin{align*}
\sum_{v_1} &= \sum_{v_1}  \\
&= 0.68, 20.40 \\
\end{align*}
\]

Although it is still possible to calculate the serial autocorrelation of (13) algebraically as before, the expression becomes cumbersome and intractable with more factors and more lags. It is much more succinct to derive the autocorrelation by matrix notation instead. To this end,
we denote the weights in (13) as a vector, . We consider the stacked vector and denote its correlation matrix by

\[
\begin{pmatrix}
v_1 & v_2 & \ldots & v_n
\end{pmatrix}
\end{pmatrix}.
\]

In the matrix, we have . The autocorrelation of the composite (13) is the ratio of the covariance between the composite and its lagged value to its variance. To calculate the covariance and variance, we denote the matrix in the upper left corner of as and the matrix in the upper right corner of as

\[
(14)
\]

Then the variance of is

\[
(15)
\]

And the covariance is

\[
(16)
\]

Combining (16) and (17) yields the serial autocorrelation of

\[
(18)
\]

For a given factor correlation matrix \( \textbf{C} \), the autocorrelation is an analytic function of the
OPTIMAL ALPHA MODEL WITH INCLUSIVE TURNOVER CONSTRAINTS

The previous sections lay the groundwork for building optimal alpha models that explicitly consider explicit turnover constraints. We consider multi-factor linear models consisting of both current and lagged factors. The autocorrelation of the model sets the constraint on the portfolio turnover while the model’s IR is optimized according to the average IC’s and covariances of ICs, based on the framework developed first in Qian & Hua (2004) and extended in Sorensen at al (2004).

The key insight from this constrained optimization is the optimal use of lagged forecasts as part of the composite alpha model even if the lagged ICs are sometimes weaker than the contemporaneous ICs. Including lagged forecasts increases composite forecast autocorrelation, and thus lowers the portfolio turnover, giving rise to saving in transaction costs. The equilibrium tradeoff between the lagged ICs and the forecast autocorrelations determines the optimal model weights in the current as well as the lagged factors.

Constrained optimization

Continuing with the case of two factors and one lag, equation (13) describes an alpha model; and we are interested in the model weights: that maximize IR while controlling the turnover. The autocorrelation of the composite is given by equation (18) and the matrices and are sub-matrices of , given by equation (15). Since the turnover is a function of , we thus constrain the autocorrelation to be at a targeted value.

The IR of the alpha model is approximately the ratio of average IC to the standard deviation of IC (Qian & Hua 2004). Denote the average IC of by and the IC
covariance matrix by \( \overline{\Sigma} \), the average IC of the model is \( \overline{\mu} \) and the standard deviation of IC is \( \sqrt{\overline{\Sigma}} \). The constrained optimization problem is
\[
\text{Maximize: } \overline{IR} \quad \text{subject to: }
\overline{cma} = \overline{vIC} \quad \overline{Dv}.
\]

(19)

The target autocorrelation is denoted by \( \overline{\rho} \). The autocorrelation constraint is quadratic in nature. This implies that (19) is a nonlinear optimization with a quadratic constraint, for which no analytic solution is readily available. However, it is easy to solve the problem numerically, as in the following example. We also note that the problem can be easily extended to include more factors and multiple lags.

**An example – the inputs**

We present a numerical example using the two previous factors – PM and E2P. To make the example more realistic we consider models with three lags. In this section we describe all the inputs to the constrained optimization and in the next section we discuss the results.

First, we consider the IC inputs associated with the factors from Tables 1 and 2 above. These are quarterly data from the Russell 3000 universe from 1987 to 2004, and include the average ICs and standard deviations of IC for the two factors and their lagged factors. Thus, with two factors and three lags, we have eight different sources of alpha. Next, we compute the IR of the composite model using factor IC correlations, presented in Table 3. The subscripted numbers denote lags. The most notable feature of Table 3 is the negative IC correlation between the two factors. For instance, the ICs of PM_0 and E2P_0 have a
correlation of 0.42, indicating significant diversification benefit. That is, when the PM factor has a higher (lower) IC adding more alpha, the E2P factor tends to have a lower (higher) IC adding less alpha. The diversification carries out to ICs of the lagged forecasts. For example, the ICs of PM_1 and E2P_1 have a correlation of 0.45 and the ICs of PM_0 and E2P_1 have a correlation of 0.37. We also note that the IC correlations among the same factors but of different lags tend to be high, indicating less diversification of information. However, the correlation drops with increasing time span between the forecasts. For instance, for the PM factor, the correlation is 0.86 between PM_0 and PM_1, 0.78 between PM_0 and PM_2, and 0.61 between PM_0 and PM_3. For the value factor, the correlations are even higher, 0.92 between E2P_0 and E2P_1, 0.84 between E2P_0 and E2P_2, and 0.78 between E2P_0 and E2P_3.

Table 3 IC correlation matrix of current and lagged values for the price momentum factor and the earning yield factor

<table>
<thead>
<tr>
<th></th>
<th>PM_0</th>
<th>E2P_0</th>
<th>PM_1</th>
<th>E2P_1</th>
<th>PM_2</th>
<th>E2P_2</th>
<th>PM_3</th>
<th>E2P_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM_0</td>
<td>1.00</td>
<td>-0.42</td>
<td>0.86</td>
<td>-0.37</td>
<td>0.78</td>
<td>-0.26</td>
<td>0.61</td>
<td>-0.19</td>
</tr>
<tr>
<td>E2P_0</td>
<td>-0.42</td>
<td>1.00</td>
<td>-0.44</td>
<td>0.92</td>
<td>-0.31</td>
<td>0.84</td>
<td>-0.29</td>
<td>0.78</td>
</tr>
<tr>
<td>PM_1</td>
<td>0.86</td>
<td>-0.44</td>
<td>1.00</td>
<td>-0.45</td>
<td>0.88</td>
<td>-0.36</td>
<td>0.71</td>
<td>-0.30</td>
</tr>
<tr>
<td>E2P_1</td>
<td>-0.37</td>
<td>0.92</td>
<td>-0.45</td>
<td>1.00</td>
<td>-0.33</td>
<td>0.94</td>
<td>-0.30</td>
<td>0.86</td>
</tr>
<tr>
<td>PM_2</td>
<td>0.78</td>
<td>-0.31</td>
<td>0.88</td>
<td>-0.33</td>
<td>1.00</td>
<td>-0.28</td>
<td>0.83</td>
<td>-0.22</td>
</tr>
<tr>
<td>E2P_2</td>
<td>-0.26</td>
<td>0.84</td>
<td>-0.36</td>
<td>0.94</td>
<td>-0.28</td>
<td>1.00</td>
<td>-0.28</td>
<td>0.94</td>
</tr>
<tr>
<td>PM_3</td>
<td>0.61</td>
<td>-0.29</td>
<td>0.71</td>
<td>-0.30</td>
<td>0.83</td>
<td>-0.28</td>
<td>1.00</td>
<td>-0.30</td>
</tr>
<tr>
<td>E2P_3</td>
<td>-0.19</td>
<td>0.78</td>
<td>-0.30</td>
<td>0.86</td>
<td>-0.22</td>
<td>0.94</td>
<td>-0.30</td>
<td>1.00</td>
</tr>
</tbody>
</table>

These IC correlations determine the composite alpha model’s strategy risk. The factor correlations determine the alpha model autocorrelation. Table 4 presents the average factor correlations between factors of different lags, which are needed in determining the forecast autocorrelation. Notice there are four lags in Table 4, instead of 3 lags in Table 3. This is due
made of factors of three lags. We note correlations among the same factor are high, for E2P in particular. This is not surprising since earning yields or PE multiples changes slowly.

Essentially, high serial autocorrelation of value factors is consistent with their minimal information decay. The factor correlations are much smaller for the PM factor: the lag 1 correlation is 0.68; the lag 2 correlation is 0.40; and the lag 3 and lag 4 correlations drop nearly to zero. This implies the winners and losers defined by the price momentum change drastically over time - winners (losers) today have little resemblance to the winners (losers) nine months earlier. Thus, the construction process will incur more turnover in maintaining momentum exposure as the PM factor updates frequently. Lastly, we note the correlations between PM and E2P of different lags are small and not as quite negative as the IC correlations.

**Table 4 Factor correlation matrix of current and lagged values for the price momentum factor and earning yield factor**

<table>
<thead>
<tr>
<th></th>
<th>PM 0</th>
<th>E2P 0</th>
<th>PM 1</th>
<th>E2P 1</th>
<th>PM 2</th>
<th>E2P 2</th>
<th>PM 3</th>
<th>E2P 3</th>
<th>PM 4</th>
<th>E2P 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM 0</td>
<td>1.00</td>
<td>-0.08</td>
<td>0.68</td>
<td>0.00</td>
<td>0.40</td>
<td>0.05</td>
<td>0.09</td>
<td>0.08</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>E2P 0</td>
<td>-0.08</td>
<td>1.00</td>
<td>-0.09</td>
<td>0.94</td>
<td>-0.06</td>
<td>0.84</td>
<td>0.01</td>
<td>0.73</td>
<td>0.03</td>
<td>0.61</td>
</tr>
<tr>
<td>PM 1</td>
<td>0.68</td>
<td>-0.09</td>
<td>1.00</td>
<td>-0.08</td>
<td>0.68</td>
<td>0.00</td>
<td>0.40</td>
<td>0.05</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>E2P 1</td>
<td>0.00</td>
<td>0.94</td>
<td>-0.08</td>
<td>1.00</td>
<td>-0.09</td>
<td>0.94</td>
<td>-0.06</td>
<td>0.84</td>
<td>0.01</td>
<td>0.73</td>
</tr>
<tr>
<td>PM 2</td>
<td>0.40</td>
<td>-0.06</td>
<td>0.68</td>
<td>-0.09</td>
<td>1.00</td>
<td>-0.08</td>
<td>0.68</td>
<td>0.00</td>
<td>0.40</td>
<td>0.05</td>
</tr>
<tr>
<td>E2P 2</td>
<td>0.05</td>
<td>0.84</td>
<td>0.00</td>
<td>0.94</td>
<td>-0.08</td>
<td>1.00</td>
<td>-0.09</td>
<td>0.94</td>
<td>-0.06</td>
<td>0.84</td>
</tr>
<tr>
<td>PM 3</td>
<td>0.09</td>
<td>0.01</td>
<td>0.40</td>
<td>-0.06</td>
<td>0.68</td>
<td>-0.09</td>
<td>1.00</td>
<td>-0.08</td>
<td>0.68</td>
<td>0.00</td>
</tr>
<tr>
<td>E2P 3</td>
<td>0.08</td>
<td>0.73</td>
<td>0.05</td>
<td>0.84</td>
<td>0.00</td>
<td>0.94</td>
<td>-0.08</td>
<td>1.00</td>
<td>-0.09</td>
<td>0.94</td>
</tr>
<tr>
<td>PM 4</td>
<td>0.07</td>
<td>0.03</td>
<td>0.09</td>
<td>0.01</td>
<td>0.40</td>
<td>-0.06</td>
<td>0.68</td>
<td>-0.09</td>
<td>1.00</td>
<td>-0.08</td>
</tr>
<tr>
<td>E2P 4</td>
<td>0.09</td>
<td>0.61</td>
<td>0.08</td>
<td>0.73</td>
<td>0.05</td>
<td>0.84</td>
<td>0.00</td>
<td>0.94</td>
<td>-0.08</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Optimal alpha models – the results and insights**

Given the inputs, we solve the optimization problem (equation 19) for a series of forecast autocorrelation targets, ranging from 0.85 to 0.97. The optimal model weights for
notice as \( \rho \) goes from 0.85 to 0.97, the optimal IR first increases from 2.30 to 2.39 when \( \rho \) is 0.89 and it then decreases to 1.88 when \( \rho \) reaches 0.97. The highest IR is when the optimal weights are 36% for PM_0 and 64% for E2P_0, with no lagged factors. This represents the unconstrained optimal alpha model using the current factors only. Second, we can see that for targeted autocorrelation above .89, the optimal model begins to add weight to lagged factors, which will slow turnover – but this is at the expense of IR. PM_1 is the first to appear and it is followed by E2P_1, E2P_2, and E2P_3. And the other two lagged momentum factors, PM_2 and PM_3, never obtain any significant weight in the model. The reason for their absence is both PM_2 and PM_3 have very low average ICs. In contrast, all E2P factors have consistent IC and high autocorrelation. Their optimal weights get near 10% each as the target autocorrelation gets very high. Finally, we notice that the weight increase in the lagged factors comes at the expense of the current momentum factor PM_0 first and then the current value factor E2P_0.

**Table 5 Optimal weights of the composite model for different levels of autocorrelation and their optimal IR**

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>IR</th>
<th>PM_0</th>
<th>E2P_0</th>
<th>PM_1</th>
<th>E2P_1</th>
<th>PM_2</th>
<th>E2P_2</th>
<th>PM_3</th>
<th>E2P_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.85</td>
<td>2.30</td>
<td>45%</td>
<td>55%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.86</td>
<td>2.33</td>
<td>43%</td>
<td>57%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.87</td>
<td>2.36</td>
<td>41%</td>
<td>59%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.88</td>
<td>2.38</td>
<td>39%</td>
<td>61%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.89</td>
<td><strong>2.39</strong></td>
<td>36%</td>
<td>64%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.90</td>
<td>2.38</td>
<td>34%</td>
<td>65%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.91</td>
<td>2.37</td>
<td>31%</td>
<td>65%</td>
<td>4%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.92</td>
<td>2.36</td>
<td>28%</td>
<td>65%</td>
<td>7%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.93</td>
<td>2.33</td>
<td>24%</td>
<td>65%</td>
<td>10%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td>0.94</td>
<td>2.28</td>
<td>21%</td>
<td>58%</td>
<td>12%</td>
<td>4%</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td>0.95</td>
<td>2.21</td>
<td>18%</td>
<td>50%</td>
<td>12%</td>
<td>8%</td>
<td>0%</td>
<td>4%</td>
<td>0%</td>
<td>8%</td>
</tr>
<tr>
<td>0.96</td>
<td>2.09</td>
<td>15%</td>
<td>43%</td>
<td>14%</td>
<td>10%</td>
<td>3%</td>
<td>7%</td>
<td>3%</td>
<td>10%</td>
</tr>
</tbody>
</table>
To understand the quantitative tradeoff between IR and turnover, consider the following comparison. The maximum IR 2.39 occurs when the forecast autocorrelation $\rho$ is at 0.89, and the model IR drops to 2.33 when $\rho$ is at 0.93. Hence the IR drops by roughly 2.5%. At the same time, the turnover changes according to the function $\sqrt{\cdot}$, dropping from $\sqrt{\cdot}$ to $\sqrt{\cdot}$, or 20%!

In Table 6 we aggregate optimal weights into weights for PM and E2P, and into weights for factors of lags 0, 1, 2, and 3 in Table 6. As $\rho$ increases from 0.85 to 0.97, the PM weight decreases from 45% to 28% while the E2P weight increases from 55% to 72%. At the same time, the weight with lagged factors increases from 0% to 58%, offset by reductions in the weights for zero-lagged factors.

**Table 6 Aggregated optimal weights of the composite model with autocorrelation targets and associated IRs**

<table>
<thead>
<tr>
<th>IR</th>
<th>PM</th>
<th>E2P</th>
<th>PM</th>
<th>E2P</th>
<th>PM</th>
<th>E2P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.85</td>
<td>2.30</td>
<td>45%</td>
<td>55%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.86</td>
<td>2.33</td>
<td>43%</td>
<td>57%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.87</td>
<td>2.36</td>
<td>41%</td>
<td>59%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.88</td>
<td>2.38</td>
<td>39%</td>
<td>61%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.89</td>
<td>2.39</td>
<td>36%</td>
<td>64%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>0.90</td>
<td>2.38</td>
<td>35%</td>
<td>65%</td>
<td>98%</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>0.91</td>
<td>2.37</td>
<td>35%</td>
<td>65%</td>
<td>96%</td>
<td>4%</td>
<td>0%</td>
</tr>
<tr>
<td>0.92</td>
<td>2.36</td>
<td>35%</td>
<td>65%</td>
<td>93%</td>
<td>7%</td>
<td>0%</td>
</tr>
<tr>
<td>0.93</td>
<td>2.33</td>
<td>34%</td>
<td>66%</td>
<td>88%</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>0.94</td>
<td>2.28</td>
<td>33%</td>
<td>67%</td>
<td>79%</td>
<td>15%</td>
<td>1%</td>
</tr>
<tr>
<td>0.95</td>
<td>2.21</td>
<td>30%</td>
<td>70%</td>
<td>68%</td>
<td>20%</td>
<td>4%</td>
</tr>
<tr>
<td>0.96</td>
<td>2.09</td>
<td>30%</td>
<td>70%</td>
<td>57%</td>
<td>21%</td>
<td>9%</td>
</tr>
<tr>
<td>0.97</td>
<td>1.88</td>
<td>28%</td>
<td>72%</td>
<td>42%</td>
<td>23%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Transaction cost and net returns**

One last piece to the puzzle is an assumption for trading costs. First, we measure the effect of autocorrelation on turnover, by calculating the turnover according to equation (9), on an annual basis, for a long-short portfolio. The inputs are target risk and stock specific risk. The turnover results and the IR are graphed in Figure 6. First note the extremely high turnover when the autocorrelation is low; it is nearly 550% when \( \tau \) is 0.89. But the most important feature of the graph is that the rate of decrease is markedly different for the IR and the turnover, as the autocorrelation \( \tau \) increases. While the turnover drops consistently, the IR changes rather slowly except when the autocorrelation reaches very high level. Since the turnover drops more rapidly than the IR over a large range, it is entirely feasible that net expected return - expected return less transaction cost, is higher for alpha models with higher autocorrelations with lagged factors.

**Figure 6 The IR and portfolio turnover of optimal alpha models with give forecast autocorrelation. The IR scale is on the left axis and the turnover scale is on the right axis.**
Continuing with the same factors, we compute net expected return by imposing different levels of transaction costs, simply with a linear proportion of the portfolio turnover. For example, at 50 basis points or 0.5%, a turnover of 100% would cost 0.5% and a turnover of 200% would cost 1%, etc. With increasing asset under management, one can use a higher multiple of turnover as transaction costs. Table 7 lists the gross return given by IR times the target tracking error, turnover, and net returns with different transaction assumptions. In addition to 0.5%, we also include transaction cost as 1% and 1.5%.

Table 7 The gross excess return and net excess returns under different transaction cost assumptions for portfolios with target risk, and stock specific risk.

<table>
<thead>
<tr>
<th>IR</th>
<th>Gross Return</th>
<th>Turnover</th>
<th>Net Return (0.5%)</th>
<th>Net Return (1.0%)</th>
<th>Net Return (1.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.85</td>
<td>2.30</td>
<td>9.19%</td>
<td>638%</td>
<td>6.00%</td>
<td>2.81%</td>
</tr>
<tr>
<td>0.86</td>
<td>2.33</td>
<td>9.32%</td>
<td>617%</td>
<td>6.24%</td>
<td>3.15%</td>
</tr>
<tr>
<td>0.87</td>
<td>2.36</td>
<td>9.43%</td>
<td>594%</td>
<td>6.46%</td>
<td>3.49%</td>
</tr>
<tr>
<td>0.88</td>
<td>2.38</td>
<td>9.51%</td>
<td>571%</td>
<td>6.66%</td>
<td>3.80%</td>
</tr>
<tr>
<td>0.89</td>
<td><strong>2.39</strong></td>
<td><strong>9.55%</strong></td>
<td>547%</td>
<td>6.81%</td>
<td>4.08%</td>
</tr>
<tr>
<td>0.90</td>
<td>2.38</td>
<td>9.53%</td>
<td>521%</td>
<td>6.93%</td>
<td>4.32%</td>
</tr>
<tr>
<td>0.91</td>
<td>2.37</td>
<td>9.50%</td>
<td>494%</td>
<td>7.03%</td>
<td>4.56%</td>
</tr>
<tr>
<td>0.92</td>
<td>2.36</td>
<td>9.44%</td>
<td>466%</td>
<td>7.11%</td>
<td>4.78%</td>
</tr>
<tr>
<td>0.93</td>
<td>2.33</td>
<td>9.33%</td>
<td>436%</td>
<td><strong>7.15%</strong></td>
<td>4.97%</td>
</tr>
<tr>
<td>0.94</td>
<td>2.28</td>
<td>9.13%</td>
<td>404%</td>
<td>7.11%</td>
<td>5.09%</td>
</tr>
<tr>
<td>0.95</td>
<td>2.21</td>
<td>8.83%</td>
<td>369%</td>
<td>6.98%</td>
<td><strong>5.14%</strong></td>
</tr>
<tr>
<td>0.96</td>
<td>2.09</td>
<td>8.35%</td>
<td>330%</td>
<td>6.70%</td>
<td>5.06%</td>
</tr>
<tr>
<td>0.97</td>
<td>1.88</td>
<td>7.53%</td>
<td>285%</td>
<td>6.10%</td>
<td>4.68%</td>
</tr>
</tbody>
</table>

As expected, the gross return is maximized at , where the IR is the maximum. But the net return attains its maximum at higher factor model autocorrelation , corresponding to different alpha model with lagged factors. When the cost of 100% turnover is
drops to 436% from 547%. When the transaction cost is higher at 1.0%, the maximum net return is at $0.95 \rho$ where the paper IR is 2.21 and the expected net return is higher than the model with $0.89 \rho$, by over 1%. At even higher transaction cost of 1.5% cost for 100% turnover, the optimal model for net return would be at $0.96 \rho$. Alpha models with these autocorrelation targets include significant weights of lagged factors (see Table 5 and 6).

Figure 7 The gross excess return and net excess returns under different transaction cost assumption for portfolios with $N$, target risk $\sigma$, and stock specific risk.

In Figure 7 we plot the data in Table 6. The square on each curve denotes the model of the maximum net return. As the transaction costs increases, the net return gets lower and lower. This is especially true for the left side of the return curves due to higher turnover. The right side of the curves drops to a less extent because the turnover is lower (higher model autocorrelation). On each cost curve, the point of maximum net return shifts to the right as the
autocorrelation increases. We also note that when transactions costs are high, there is a more rapid increment in net return as autocorrelation increases -- optimal models with high autocorrelation have a better chance to yield positive net returns as trading costs rise.

These results have strong implications for the construction of optimal alpha models. First, using the model with maximum gross IR can be sub-optimal in terms of net return where the transaction costs are taken into account. For example, with 1% cost, the net return of that model is lower than that of the optimal model with $\hat{\rho}$, by more than 1%. Second, the optimal model in terms of highest net return changes as the transaction cost becomes larger. This indicates the need to evolve the alpha model as AUM grows. Our results reveal that one way to do so is to add lagged factors based on their risk/return/turnover tradeoff. Third, both the net return and the optimal model are sensitive to IR assumption. If the IR’s were lower than those in the example, then for a given level of transactions costs, the maximum net return would have been achieved with even higher $\hat{\rho}$ models. In other words, when the information content of the factors is lower, we need to pay even more attention to reduce portfolio turnover to reduce transaction cost. This inevitably leads to more weights in the lagged factors, especially lagged value factors.$^{10}$

**SUMMARY**

Realistic alpha models necessitate the direct, endogenous analysis of implementation costs. This means that we should build optimal alpha models with an integration of transaction cost in the modeling process. This paper provides an analytic framework to do so by maximize IR of alpha models under portfolio turnover constraint.
The solution we provide is quite general. We show the portfolio turnover is an algebraic function of forecast autocorrelation. Hence, factor autocorrelation is a key diagnostic in evaluating single factor efficacy, and in creating optimal composite models. For linear composite forecasts, the autocorrelation depends on auto- and cross- factor correlations. Optimal models by definition must depend on factor correlations for contemporaneous and lagged values as well as average IC’s, standard deviation of IC’s, and IC correlations.

These results should be useful to practitioners in several ways. First, autocorrelation is an important diagnostic for evaluating factors. Second, comprehensive alpha models necessitate the direct integration of transaction costs. Third, the analyses here lead directly to reasonable estimations of strategy capacity associated with increasing AUM levels. They suggest adjusting the weights in favor of lagged factors to reduce turnover by increasing the target autocorrelation, thus maximize the net return.

APPENDIX

Portfolio turnover

In this appendix we derive the theoretical solution of portfolio turnover. We first rewrite (7) as an expectation

\[ E[\sum_{i=1}^{N} (\sqrt{F_{it} T_{it} \sigma_{N}} - T_{it} \sigma_{N})]. \]  

(20)

To evaluate the expectation, we further assume that the change in the risk-adjusted forecast and the stock specific risk are independent. Hence (20) can be written as

\[ \sqrt{E[\sum_{i=1}^{N} (F_{it} T_{it} \sigma_{N})]}. \]  

(21)
Both sets of forecasts have standard deviation of one. We further assume they form a bivariate normal distribution with mean zero and the cross sectional correlation between the two sets of consecutive forecasts is \( \rho_f \). This is simply the lag 1 autocorrelation of the risk-adjusted forecasts. If the correlation is high, then the change in forecasts is minimal and the turnover is low. Conversely, if the correlation is low, then the forecast change is significant and the turnover will be high. Since both forecasts are normally distributed, the change is still a normal distribution with zero mean and standard deviation \( \sqrt{\rho} \). For a random variable \( x \) with distribution \( \mathcal{N}(0, \sigma^2) \), we can show that

\[
\sqrt{\frac{\rho}{\sigma^2}} = \frac{1}{\sqrt{\sigma^2}}.
\]

Therefore,

\[
\sqrt{\frac{\rho}{\sigma^2}} = \frac{1}{\sqrt{\sigma^2}}.
\]

(22)

Substituting (22) into (21) yields

\[
\sqrt{\frac{\rho}{\sigma^2}} = \frac{1}{\sqrt{\sigma^2}}.
\]

(23)

**Leverage and turnover**

Since targeted tracking error is linked to the portfolio leverage, we can derive a relationship between the leverage and forecast-induced turnover, using expectations. We have

\[
\sqrt{\frac{\rho}{\sigma^2}} = \frac{1}{\sqrt{\sigma^2}}.
\]

(24)

It is apparent from the equation that the leverage of the portfolio should remain the same if the targeted tracking error and the specific risks remain constant.
Since $F$ is a standard normal variable, we have $\sqrt{F}$. Using this result in (24) yields

\[ \sqrt{F} \]  

(25)

Combining (25) and (23) yields

\[ \sqrt{F} \]

(26)

Therefore, the turnover is directly proportional to the leverage times the square root of 1 minus forecast autocorrelation.

**REFERENCE**


---

1. We assume that on an aggregated portfolio level the implementation cost is proportional to the amount of trading -- or portfolio turnover. The majority of implementation cost is related to trading. These costs could be commission, and bid/ask spread, and market impact. The trading cost may vary from stock to stock. As an approximation, we simply assume the trading cost is a fixed multiple of portfolio turnover.

2. Turnover can also be caused by flows in and out of portfolios. These forced turnovers are not due to portfolio rebalance and they are easy to analyze. We shall exclude them from our analysis.

3. If there is short-term reversion between consecutive period returns, then the horizon IC will be higher.

4. See Qian & Hua (2004) for a development of the IR expression: $IR = \frac{\text{Expected IC}}{\text{IC}}$.

5. Our definition of turnover measures the percentage change of the portfolio versus portfolio capital, which is most relevant in terms of amount of trading and costs. There are other variations that use total portfolio leverage or notational exposures as denominators, which tend to lower turnover percentage.

6. For constrained portfolios such as long-only portfolios, the turnover can be substantially less, since constraints work to suppress changes in portfolio weights (Qian, Hua, and Tilney 2004). The turnover can also be reduced through other means such as cutting them proportionally or ignoring small trades. For detail, see Grinold and Stuckelman (1993) and Qian, Hua, and Tilney (2004).

7. Our analysis shows that the inclusion of lagged forecast would increase the serial autocorrelation as long as autocorrelation of lag 2 is above certain threshold. The value of threshold is given by two times lag 1 correlation squared minus one. These values are easily exceeded for most factors encountered in practice.

8. While factor correlation has some bearing on the corresponding IC correlation, they are not the same. In building alpha model, it is the IC correlation NOT factor correlation that plays a crucial role in determining the weights of factors.

9. In practice, the actual turnover will be lower when the transactions are incorporated in the portfolio optimization.

10. It is not hard to imagine this situation might apply to market segments that are relatively less inefficient, such as US large cap stocks.