The Journal of Portfolio Management

Risk Reduction in Style Rotation

M. Rodrigo Dupleich Ulloa, Daniel Giamouridis and Chris Montagu

JPM 2012, 38 (2) 44-55
doi: https://doi.org/10.3905/jpm.2012.38.2.044
http://jpm.iijournals.com/content/38/2/44

Risk Reduction in Style Rotation

M. Rodrigo Dupleich Ulloa, Daniel Giamouridis, and Chris Montagu

M. Rodrigo Dupleich Ulloa
is a vice president in Equities Quantitative Analysis, Europe, at Citi Global Markets in London, UK.
rodrigo.dupleich@citi.com

Daniel Giamouridis
is an assistant professor in the Department of Accounting and Finance at Athens University of Economics and Business in Athens, Greece; a senior visiting fellow in the Faculty of Finance at Cass Business School, City University in London, UK; and a research associate at EDHEC-Risk Institute in Nice, France.
dgiamour@aueb.gr

Chris Montagu
is a managing director and head of Global Quantitative Research, Europe, at Citi Investment Research in London, UK.
chris.montagu@citi.com

The market conditions of recent times have challenged a lot of quantitative or systematic investment strategies. The growth of quantitative investing through the early 2000s has meant that many strategies were optimally developed over a period in the markets in which style volatility and correlation levels were low and autocorrelations were high. These benign market conditions meant that many strategies were almost "set and forget" in terms of factor/style weightings. August 2007 changed this, however. The product of cross-asset delevering/sell-off, overcrowded investment strategies, and a difficult macroeconomic environment was abnormal factor/style correlations and volatility, and sudden factor/style reversals. This new environment has made it crucial for quantitative investors to consider more dynamic approaches to their style/factor selection and/or weighting.

Dynamic style selection comes with its own separate set of problems. First, in periods of high volatility, style rotation strategies are at the mercy of frequent turning points in style performance. More recently, we have witnessed increased style volatility and a breakdown in typical correlation structures. In these conditions, a static approach to style weighting is potentially suboptimal, depending on how dynamic or reactive the rotation strategy is, and missing turning points can severely impact portfolio performance. Second, and related, is that the very dynamic nature of the strategy increases portfolio turnover and therefore transaction costs. Too-frequent style re-weighting will erode portfolio performance and a high noise-to-signal ratio will generate unnecessary style rotation. Thus, ideally portfolio managers should seek style rotation strategies that could be dynamic, but are less vulnerable to these risks.

Determining the optimal mix of assets in the context of portfolio construction involves "smart" forecasts of stock returns as well as good estimates of stock return variances and covariances. Typically, sample moments are used as best estimates of the population moments. For example, it is common practice to estimate a covariance matrix (sample covariance matrix) using historical stock returns. Several researchers have pointed out, however, that this generally acceptable practice introduces uncertainty, which can degrade the desirable properties of the constructed portfolio; see, for example, Best and Grauer [1991] and Chopra and Ziemba [1993]. The problem is termed *estimation risk* and is commonly addressed by employing advanced statistical techniques. Giamouridis [2010] discussed the problem of estimation error in the context of portfolio construction and provided a review of this literature.

Style rotation is not, in principle, different. "Smart" forecasts of factor returns are necessary as well as good estimates of factor return variances and covariances. Although researchers have been concerned with the impact of estimation risk in portfolio construction, none to our knowledge has so far been concerned with its impact on decisions involving style portfolio allocations. The vast majority of papers that we have come across are mainly concerned with the development of successful factor return prediction models and are less interested in the optimal mix of styles. Integrating sophisticated portfolio construction approaches that are designed to reduce the risks involved in the portfolio construction process has, in our view, become extremely important.

In this article, we investigate if techniques that address estimation risk, optimal prediction, and variable selection improve equity style-rotation strategies. The general style-rotation framework that we use is similar to that used by Qian, Hua, and Sorensen [2007] in that we focus on style portfolios with the maximum Information Ratio (IR). However, we integrate two new approaches that have recently been shown to work well in the presence of estimation risk and structural breaks in data.

The first approach is motivated by recent research in the portfolio construction literature (see, for example, Brodie et al. [2007], DeMiguel et al. [2009], Lobo, Fazel, and Boyd [2007], Welsch and Zhou [2007], and Giamouridis and Paterlini [2010]) and involves the use of regularization procedures, namely, the ridge regression approach (Hoerl and Kennard [1970]). In this setting, we treat the weights rather than the factor return means, variances, and covariances as the objects of interest to be estimated. The second approach, developed by Pesaran and Timmermann [2007], involves pooling forecasts that are obtained with the same model, but across different observation windows. This approach aims to minimize the impact of structural breaks on moment estimates and can be extended to deal with different types of model uncertainty.

METHODOLOGY

In this section, we describe our methodology for constructing optimal style portfolios using two approaches: ridge regression and multiple estimation windows.

Optimal Style Portfolios

Style rotation portfolios are typically based on the relative scores of stocks with respect to certain factors and equal-weighting schemes for aggregating the score information. Our analysis aims to suggest an alternative weighting scheme that is based on the optimization of the factor portfolio IR. In particular, the optimal weightings for the factor portfolio are obtained through the following generic optimization:

$$\max_{\mathrm{w}} \left(\frac{\bar{R}_{factors}}{\sigma^2_{factors}} \right) \tag{1}$$

where the weights over the K factors are defined as $\mathrm{w} = (w_1, \ldots, w_K)$, $\bar{R}_{factors}$ is the expected return of the factor portfolio, and $\sigma_{factors}$ is the volatility of the factor portfolio. We define $\bar{R}_{factors}$ as $\bar{R}_{factors} = \mathrm{w}\bar{R}$, where $\bar{R}$ is the $K \times 1$ vector of factor expected returns, that is, $\bar{R} = [E(R_1), \ldots, E(R_K)]'$. We also define $\sigma^2_{factors}$ as $\sigma^2_{factors} = \mathrm{w}\Sigma_{factors}\mathrm{w}'$, where $\Sigma_{factors}$ is the $K \times K$ covariance matrix of factor returns. The key inputs for this approach, as with the mean-variance approach, are the means and covariances of factor returns. Using sample moments as best estimates of the population moments involves estimation risk. Additional problems arise when the factor return covariance matrix is singular or nearly singular. We discuss this problem later in the article.

Our approach for determining optimal style weights and, ultimately, constructing the equity portfolio involves two steps. First, we employ IR optimization to determine the optimal factor weightings for a given point in time through variants of Equation (1). We compute the expected returns, variances, and covariances for the styles using historical style return data. We constrain the portfolio optimization problem such that the weights sum to one and the weights are positive. Second, using the optimal style weights from the procedure just outlined, we calculate composite scores for each firm by weighting factor scores with the optimal weights. The composite scores can be used to sort stocks and buy (sell) those with the highest (lowest) composite scores. Composite scores can also be used in an additional layer of portfolio optimization, in which, given the composite scores, the optimal allocation to each firm is sought. Our analysis focuses on the former approach, but we also present some empirical results for the latter.

Ridge Regression

The first approach we study is the ridge regression. In the context of regression analysis, the ridge approach shrinks estimates of the regression coefficients by imposing a penalty on their size in the sum of square errors minimization. In the context of portfolio optimization, we impose a constraint on the sum of square weights in Equation (1). The ridge regression approach aims to address weight sensitivity to possible broad errors in the data. We refer to this as *weights estimation error.*

Consider again Equation (1), which we can rewrite as

$$\max_{w}\left(\frac{\bar{R}_{factors}}{\sigma^2_{factors}}\right) \approx \max_{w}\left[w\bar{R}_{factors} - \xi\, w\Sigma_{factors} w'\right] \quad (2)$$

where ξ is a non-negative constant. The solution to Equation (2) requires computation of the inverse of $\Sigma_{factors}$ and, hence, largely depends on the condition number of the matrix—the ratio of the matrix's largest to smallest singular values—of factor returns. If the condition number of factor returns is large, that is, when factor returns are correlated, $\Sigma_{factors}$ is singular or nearly singular. A standard numerical procedure for the optimization of Equation (2) is then likely to lead to an unstable and unreliable estimate of the vector **w** (Brodie et al. [2007]). This situation is likely to arise in instances where equity portfolio managers consider style rotation across a large universe of factors, subsets of which may be highly correlated.

We tackle these issues with the introduction of a penalty term in the objective function. We consider the following modification of Equation (1), which we augment with a set of basic portfolio construction constraints,

$$\max_{w}\left(\frac{\bar{R}_{factors}}{\sigma^2_{factors}} - \lambda\sum_{i=1}^{K} w_i^2\right) \quad (3)$$

$$\text{subject to } w_{\min} \le w_i \le w_{\max} \text{ and } \sum_{i=1}^{K} w_i = 1$$

where λ is a non-negative constant that determines the relative importance of the penalization of $\sum_{i=1}^{K} w_i^2$ in the optimization. As we did earlier, we rewrite Equation (3) as

$$\begin{aligned}
&\max_{w}\left(\frac{\bar{R}_{factors}}{\sigma^2_{factors}} - \lambda\sum_{i=1}^{K} w_i^2\right)\\
&\approx \max_{\bar{w}}\left[w\bar{R}_{factors} - \xi\, w\Sigma_{factors} w' - \lambda\sum_{i=1}^{K} w_i^2\right]\\
&= \max_{w}\left[w\bar{R}_{factors} - \xi\, w\Sigma_{factors} w' - \lambda w I w'\right]\\
&= \max_{w}\left[w\bar{R}_{factors} - w(\xi\Sigma_{factors} + \lambda I)w'\right] \quad (4)
\end{aligned}$$

$$\text{subject to } w_{\min} \le w_i \le w_{\max} \text{ and } \sum_{i=1}^{K} w_i = 1$$

where I is the identity matrix. The penalty term effectively adds a positive constant to the diagonals of $\Sigma_{factors}$. This ensures that the matrix $(\xi\Sigma_{factors} + \lambda I)$ is nonsingular. Computing its inverse thus does not affect the stability of the optimization problem, and the optimization results in more stable and reliable estimates of the vector **w**.

To gain more intuition about how the ridge regression approach works and also about the optimal portfolios that are obtained through Equations (3) or (4), we turn to Equation (1). Hastie, Tibshirani, and Friedman [2009] suggested that when there are correlated factors, their weights can become poorly determined through Equation (1) and exhibit high variance. A wildly large, positive coefficient on one variable can be cancelled by a similarly large, negative coefficient on its correlated cousin. Adding a penalty term to Equation (1), that is, obtaining optimal weights through Equation (2) instead of Equation (1), is equivalent to adding the constraint $\sum_{i=1}^{K} w_i^2 \le t$ for some positive t in Equation (1). Therefore, by imposing a size constraint on the weights, the problem of poorly determined, highly variable weights is alleviated. Hastie, Tibshirani, and Friedman [2009] provided additional insight into the nature of ridge regression by linking it to principal component analysis (PCA). They suggested that the ridge regression approach shrinks the most principal components' directions—of the factor return matrix in our setup—with small variance.

Turning to the nature of the optimal portfolios that are obtained through Equations (3) or (4), given $\sum_{i=1}^{K} w_i = 1$, we can show that $\sum_{i=1}^{K} w_i^2 = \sum_{i=1}^{K}\left(w_i - \frac{1}{K}\right)^2 + \frac{1}{K}$.[1] This observation suggests that the ridge regression approach penalizes allocations that deviate from $1/K$, or in other words, shrinks portfolio weights toward $1/K$. The shrinkage is more intense, the larger the value of λ. Hence, we expect that the portfolios obtained from Equations (3) or (4) will, in general, remain relatively close to the $1/K$ portfolio. As highlighted in DeMiguel et al. [2009], the ridge regression approach is a relevant estimation error reduction technique for investors who believe that the optimal portfolio is close to the well-diversified $1/K$ portfolio.

Collectively, the ridge regression approach constructs factor portfolios that result from shrinking the portfolio weight vectors instead of shrinking the moments of factor returns. The portfolios obtained through the ridge regression approach can be given a Bayesian interpretation (see Hastie, Tibshirani, and Friedman [2009] and Bishop [2006]) and possess interesting moment-shrinkage interpretations (DeMiguel et al. [2009]). Our earlier discussion and the empirical evidence, so far suggest that the ridge regression approach or regularization methods, in general, can 1) reduce the sensitivity of the optimization to possible collinearities between assets, 2) control transaction costs by promoting stability and sparsity, and 3) improve the out-of-sample performance relative to the classical Markowitz mean–variance approach.

Multiple Estimation Windows

Addressing weight sensitivity to estimation error through regularization has benefits and some attractive interpretations. Dealing explicitly with the inefficiencies of moment estimation (i.e., moment estimation errors) with contemporary techniques can also be beneficial. Recently, Pesaran and Timmermann [2007] and Pesaran and Pick [2011] proposed a method of constructing out-of-sample forecasts when the data are subject to structural breaks.

Pesaran and Timmermann [2007] argued that a forecast averaging procedure can be extended to deal with different types of model uncertainty, such as the uncertainty that arises with the generally arbitrary choice of the size of the estimation window. When a large estimation window is considered, forecasts might be unreliable in the presence of structural breaks in the data. When rolling window estimators are used, it is still possible that forecasting performance may not be, on average, as good as when the Pesaran and Timmermann [2007] pooled forecasts approach is used. The latter simply pools forecasts obtained using the same model, but estimated across different estimation windows. This approach can be seen as a risk diversification strategy in the presence of uncertainty regarding possible structural breaks in the data. An interesting feature of these methods is that no exact information about the structural break is necessary.

Like Pesaran and Timmermann [2007], we compute the average block (window) return for the i^{th} style factor for the most recent $(T-m)$ block data, and denote by $\overline{R^m}$ the vector of factor expected returns, that is, $\overline{R^m} = [E(R_1^m), \ldots, E(R_K^m)]'$. We also denote by $\Sigma_{factors}^m$ the $K \times K$ covariance matrix of factor returns, which we obtain for the most recent $(T-m)$ block data. We use sample estimators for both $\overline{R^m}$ and $\Sigma_{factors}^m$. We implement this approach in two stages, which we depict in Exhibit 1.

In the first stage, which is shown in Panel A of Exhibit 1, each model calculates expected returns, variances, and covariances based on different estimation windows. The individual model's expected return vectors, $\overline{R^m}$, and covariance matrices, $\Sigma_{factors}^m$, are then used

EXHIBIT 1
Pooled Forecasts Approach

Panel A: Stage 1

Time	t – 1	t – 2	t – 3		Expected Return Vector	Covar Matrix	Optimization (3x)	Weight Vector
Model 1					R M1	Cov M1		V M1
Model 2					R M2	Cov M2		V M2
Model 3				Calculate	R M3	Cov M3		V M3

Panel B: Stage 2

	Model Weight Vectors				Average Factor Weights
	V M1	V M2	V M3		
Style/Factor 1	25%	50%	33%		36%
Style/Factor 2	25%	30%	20%		25%
Style/Factor 3	25%	10%	40%	Average across	25%
Style/Factor 4	25%	10%	7%	Models	14%

to perform the optimization scheme of Equation (1) for M different rolling windows, that is,

$$\max_{\mathrm{w}^m}\left[\frac{\overline{R^m_{factors}}}{(\sigma^m_{factors})^2}\right] \tag{5}$$

$$\text{subject to } w_{\min} \leq w_i \leq w_{\max} \ \ and \ \sum_{i=1}^{K} w_i = 1$$

where $\mathrm{w}^m = (w_1^m, \ldots, w_K^m)$ denotes the optimal factor mix for a given block defined through m, $\overline{R^m_{factor}} = \mathrm{w}^m \overline{R^m}$ is the expected return of the factor portfolio, and $(\sigma^m_{factors})^2 = \mathrm{w}^m \Sigma^m_{factors} \mathrm{w}^{m\prime}$ is its variance. The end result from this stage of implementation is M different weight vectors. In the second stage, which is shown in Panel B of Exhibit 1, we calculate optimal factor weights by averaging across the M different weight vectors.

The intuition of this approach is that we remove model uncertainty by averaging all possible models, that is, the different rolling windows in our set up, for mean and covariance matrices. As discussed earlier, averaging across windows can be useful for forecasting in the presence of structural breaks in the data. This case is particularly important in volatile market environments in which sudden factor performance reversals can be very common. Technically, when the data are subject to break(s), a full sample estimator (i.e., an expanding window) produces low variance, but biased forecasts. Forecasts based on a subsample (i.e., recent rolling windows) tend to reduce bias, but inefficiencies arise due to higher variance. Therefore, in the absence of full information regarding the point and size of break(s), an optimal approach would be to use different subwindows to produce forecasts and then average the outcomes, either by simply using equal weights or by cross-validated weights as used by Pesaran and Timmermann [2007]. Using equal weights over forecasts can be seen as setting a uniform prior (uninformative prior) distribution over the mean and covariance estimates and forecasts.

The multiple estimation windows approach can be linked to a model that relies on exponential weighting of more recent observations; see, Pesaran and Pick [2011]. The multiple estimation windows approach can also be related to a stationary bootstrap approach in which factor returns are sampled from the empirical joint distribution of factor returns.

Collectively, the approach advocated by Pesaran and Timmermann [2007] allows us to use short-, medium- and long-term information about factors' performances and their correlation structure. We (optimally) equally weight this information given our uncertainty about the out-of-sample possible outcomes. In addition, given that factors can be subject to sudden reversals (breaks), this approach allows us to control for possible reversals during the out-of-sample period.

EMPIRICAL ANALYSIS

In this section, we empirically investigate whether controlling for estimation risk with either of the ridge or the multiple estimation windows approaches improves the performance of style rotation portfolios. We first discuss the data and the details of our empirical investigation.

Data and Empirical Setup

The focus of our analysis is not on the particular style/factors that we use for our style rotation strategy; rather, we select styles that typical systematic investment managers would possibly use in their investment process. We are also careful to include a blend of style/factors with different alpha decay (or volatilities).

Our base universe is the MSCI Europe. Stocks are assigned scores with respect to 13 standard style factors and fall within eight broad investment themes: GARP, value, growth, low risk, quality, reversals, momentum, and estimates. Style returns are calculated using factor-mimicking portfolios, that is, the (daily) returns between the top and low baskets based on a style factor. Our analysis extends from 1996 to 2009. Although we consider all 13 factors when constructing portfolios, optimal style weights are determined by aggregating the individual factors into their style composite groupings.

The backtest methodology consists of ranking all stocks in the universe by their respective aggregate style score at the end of each month. We then form equal-weighted quintile portfolios and focus on the top- and bottom-quintile portfolios. Subsequently, we calculate the next month's total return for each portfolio and rebalance monthly. We address survivorship bias by using the index constituents at the time of rebalancing. Statistical significance for the mean and median are tested with a t-test and a sign test, respectively. The statistical significance of the IRs is tested on the basis of 10,000 bootstrapped samples from the strategies' original return sample.

Optimization Process

To solve the optimization problem, we use the NUOPT (SPLUS) numerical optimizer. We constrain the factor weightings between 0% and 15%. The lower bound provides a short constraint, but the upper bound is selected on an ad hoc basis—ad hoc, in so far as we use a rough rule of thumb (i.e., we do not want exposure that was double the *1/K* factor weighting).

In our analysis, this constraint plays a significant role in the performance and risk statistics of factor dynamic approaches. Removing this constraint tends to generate slightly higher factor exposures (even under the ridge approach), which tend to undermine the performance-and-risk profile of our strategy. But allowing "factor shorting" is debatable,[2] and in some cases, it seems that it could conflict with the intuition behind finance and quant models (e.g., the logic behind combined value and momentum signals, such as buying expensive stocks with poor price and earnings momentum).[3] From our analysis, we find that, based on monthly rebalances, removing the factor short constraints does not greatly contribute to the overall performance of a given factor model, and it seems that the improvements are observed only during specific time periods.

RESULTS

We compare our results with four style portfolio construction strategies. First, an equal-weighting (EW) scheme of all factors. Second, a weighting scheme that involves weighting factors with respect to their most recent IRs (NOIR). The NOIR approach resembles the characteristics of factor-level momentum strategies. It does not take into account the correlation structure across factors. Qian, Hua, and Sorensen [2007] advocated that style-rotation portfolio strategies can benefit by incorporating Information Coefficient (IC) covariances/correlations. Hence, the third approach we use utilizes cross-sectional factor covariance (IC COV), and the fourth relies on cross-sectional factor correlation (IC COR). The set of statistics we use for comparison involves descriptive measures of the empirical return distribution, extreme risk, drawdown, and stock turnover.[4]

Ridge Approach

The results of the backtesting analysis for the ridge approach are tabulated in Exhibit 2. The first group of statistics shows the empirical distribution of the monthly returns (i.e., the difference in returns between the top- and bottom-quintile baskets). It is interesting to see in Exhibit 2 that the ridge approach tends to improve the return distribution. If we set the ridge parameter λ to 8.0, the average monthly return equals 1.15%, which is approximately 50% higher than the EW and IC COR (both at 0.77%) and the NOIR (0.74%) strategies and also well above the IC COV (0.84%). In the second group of performance and risk statistics, we show that volatility increases to 14.1% from 13.6% and 12.8% for NOIR and EW, respectively, and from 9.5% and 7.1% for IC COV and IC COR, respectively. However, the improvement in IR remains economically significant, that is, from 0.65 and 0.72 for NOIR and EW, respectively, and from 0.79 and 0.82 for IC COV and IC COR, respectively, to a range of 0.87 to 0.98 for the ridge approach depending on the value of λ.

The semi-variance and Conditional Value-at-Risk (CVaR) statistics show mixed results. For $\lambda = 8$, for example, the volatility of negative returns moves from 12.7% (NOIR) and 9.0% (EW) to 9.9%. The CVaR at 5% is −8.06%, which is lower than −9.83% (NOIR) but higher than −7.41% (EW). The drawdown statistics indicate that losses from the high water mark–level reduce with the ridge approach to −20.5%, lower than the −41.5% and −25.0% of the NOIR and EW schemes, respectively. Therefore, semi-variance, CVaR at 5%, and turnover statistics do not favor the ridge, but maximum drawdown does.

It is also important to evaluate the stock-level turnover. Obviously, a more dynamic or flexible approach tends to generate higher stock/style rotation. The ridge approach shows turnover of around 85%, which is higher than 75% (NOIR), 80% (IC COV), and 76% (IC COR), but marginally lower than 86% (EW).

The turnover statistic also shows the direct impact of the ridge hard constraint. As the ridge parameter, λ, increases, the ridge constraint becomes tighter, and factor weights tend to shrink faster. As a result, we allow for a lower flexibility or dispersion in our factor allocation, which directly translates into better performance. On the flipside, as λ approaches zero, the penalty becomes less binding. Setting the $\lambda = 0.02$ can be considered as the benchmark when conducting the analysis in the absence

Exhibit 2

Performance and Risk Statistics (Ridge Approach)

	NOIR	EW	IC COV	IC COR	Ridge λ = 0.02	λ = 0.95	λ = 2.00	λ = 4.00	λ = 8.00
Empirical Return Distribution (monthly)									
Min.	–17.22%	–14.46%	–13.43%	–9.17%	–12.11%	–11.80%	–11.85%	–12.23%	–11.76%
1st Quintile	–0.35%	–0.94%	–0.99%	–1.04%	–0.43%	–0.59%	–0.63%	–0.50%	–0.56%
Median	1.07%**	0.53%**	1.05%**	0.68%**	0.91%**	0.99%**	0.96%**	0.87%**	0.99%**
Mean	0.74%*	0.77%**	0.84%**	0.77%**	1.00%**	1.03%**	1.02%**	1.03%**	1.15%**
Mean (Ann.)	8.85%	9.24%	10.10%	9.21%	12.03%	12.39%	12.29%	12.35%	13.83%
5th Quintile	2.60%	2.54%	2.66%	2.16%	2.61%	2.56%	2.59%	2.73%	2.48%
Max.	13.32%	13.67%	16.35%	12.06%	15.84%	15.72%	16.06%	15.84%	17.27%
Volatility, IR, and Skew Statistics									
Ann. Volatility	13.6%	12.8%	9.5%	7.1%	13.9%	13.9%	13.8%	13.7%	14.1%
Ann. IR	0.65**	0.72**	0.79**	0.82**	0.87**	0.89**	0.89**	0.90**	0.98**
Skew	–0.81	0.13	–0.03	0.38	0.07	0.06	0.11	0.14	0.52
Risk Statistics									
Ann. Semi–Volatility	12.7%	9.0%	9.5%	7.1%	11.0%	10.9%	10.9%	10.7%	9.9%
CVaR 10%	–7.34%	–5.65%	–5.86%	–4.84%	–6.36%	–6.33%	–6.23%	–5.95%	–5.75%
CVaR 5%	–9.83%	–7.41%	–7.90%	–6.12%	–8.83%	–8.80%	–8.62%	–8.18%	–8.06%
Drawdown Statistics									
Length DD (max–days)	17.0	15.0	26.0	11.0	9.0	12.0	12.0	9.0	7.0
DD size (max %)	–41.5	–25.0	–33.3	–23.5	–22.6	–22.1	–22.1	–22.9	–20.5
Turnover									
Length recov (max–days)	14.0	21.0	12.0	11.0	27.0	29.0	26.0	33.0	28.0
Average Turnover	75%	86%	80%	76%	84%	84%	83%	84%	91%

Note: NOIR is a portfolio construction rule that involves weighting factors with respect to their most recent Information Ratio. EW involves weighting factors equally. IC COV is a portfolio construction rule that involves weighting factors based on their Information Coefficient covariance matrix. IC COR is a portfolio construction rule that involves weighting factors based on their Information Coefficient correlation matrix. λ is a non-negative constant that determines the relative importance of the penalization of the sum of square weights in the ridge optimization.
*** and * indicate statistical significance at the 1% and 5% significance levels, respectively.*

of the penalty term in Equation (4). In additional analysis, we investigate lasso, an alternative regularization technique (see Tibshirani [1996]) that is being widely used in portfolio construction studies. Given our constraints on the weights (i.e., $0 \leq w_i \leq 15\%$), we obtain similar results in terms of factor exposures, but slightly less compelling results in terms of performance-and-risk profiles.

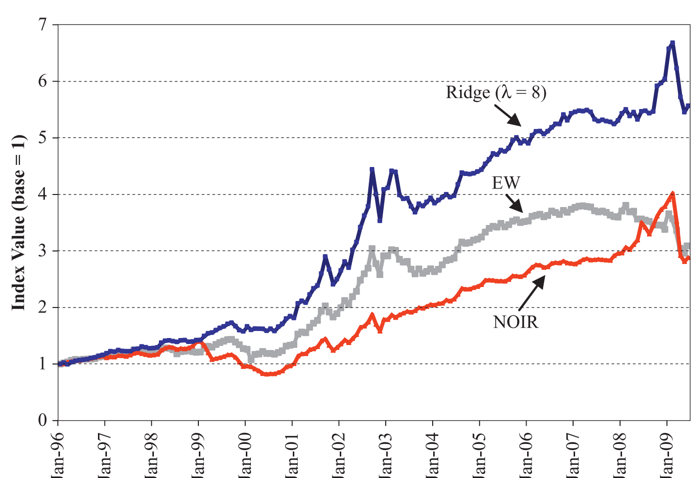
In Exhibit 3, we plot the cumulative performance of the strategy using $\lambda = 8$. It is evident that the ridge approach outperforms EW and NOIR in a relatively consistent way over the sample period. The patterns are similar for IC COV and IC COR.

Multiple Estimation Windows

As before, we conduct the same investigation and show the results of our empirical analysis in Exhibit 4.

Exhibit 3

Wealth Curves of Ridge Approach vs. Benchmark, 1996–2009



For this analysis, the maximum estimation window is 130 days regardless of the window block size. The number of window blocks that we use determines the number of estimation windows that we consider in the context of the overall estimation window. For example, given the overall estimation window size of 130 days, a three-block analysis refers to an average calculated across windows of 43 days, 87 days, and 130 days.

Specifically, in Exhibit 4 we observe that the monthly mean (median) returns increase is 1.16% (1.17%) for the three-block case, which is higher than 0.74% (1.07%) and 0.77% (0.53%) obtained through NOIR and EW, respectively. The volatility tends to increase slightly to 14.1% from 13.6% (NOIR) and 12.8% (EW). These two components have a significant impact on the IR. In particular, there is improvement from 0.65 and 0.72 for NOIR and EW, respectively, to a range of 0.89 to 1.00 for the pooled forecasts approach depending on the number of blocks used. The comparison with IC COV and IC COR also favors the multiple estimation windows approach.

Furthermore, in terms of the extreme risk profile, the semi-volatility measure and CVaR at 5% show mixed results at 10.2% and −8.23%, respectively, while the NOIR and EW benchmarks are 12.7% and 9.0%, respectively, for the semi-volatility measure, and −9.83% and −7.41% for CVaR at 5% statistic, respectively. In addition, the depth of drawdowns tends to be smaller with drops of −21.6% from the high-water mark compared to similar drops of −41.5% and −25.0% for the NOIR and EW

EXHIBIT 4

Performance and Risk Statistics (Multiple Estimation Windows Approach)

					MEW				
	NOIR	**EW**	**IC COV**	**IC COR**	**1 block**	**3 blocks**	**5 blocks**	**7 blocks**	**10 blocks**
Empirical Return Distribution (monthly)									
Min.	–17.22%	–14.46%	–13.43%	–9.17%	–12.37%	–12.38%	–11.47%	–13.09%	–11.97%
1st Quintile	–0.35%	–0.94%	–0.99%	–1.04%	–0.62%	–0.70%	–0.68%	–0.74%	–0.69%
Median	1.07%**	0.53%**	1.05%**	0.68%**	1.00%**	1.17%**	1.13%**	1.15%**	1.11%**
Mean	0.74%*	0.77%**	0.84%**	0.77%**	1.05%**	1.16%**	1.17%**	1.18%**	1.18%**
Mean (Ann.)	8.85%	9.24%	10.10%	9.21%	12.55%	13.94%	14.05%	14.15%	14.21%
5th Quintile	2.60%	2.54%	2.66%	2.16%	2.63%	2.64%	2.75%	2.79%	2.66%
Max.	13.32%	13.67%	16.35%	12.06%	16.05%	15.47%	15.14%	15.73%	15.20%
Volatility, IR, and Skew Statistics									
Ann. Volatility	13.6%	12.8%	9.5%	7.1%	14.1%	14.1%	14.2%	14.6%	14.2%
Ann. IR	0.65**	0.72**	0.79**	0.82**	0.89**	0.99**	0.99**	0.97**	1.00**
Skew	–0.81	0.13	–0.03	0.38	0.04	0.26	0.30	0.19	0.33
Risk Statistics									
Ann. Semi–Volatility	12.7%	9.0%	9.5%	7.1%	11.2%	10.2%	9.9%	10.8%	9.7%
CVaR 10%	–7.34%	–5.65%	–5.86%	–4.84%	–6.44%	–5.95%	–5.92%	–6.22%	–5.81%
CVaR 5%	–9.83%	–7.41%	–7.90%	–6.12%	–8.86%	–8.23%	–8.31%	–8.91%	–7.99%
Drawdown Statistics									
Length DD (max–days)	17.0	15.0	26.0	11.0	12.0	11.0	9.0	9.0	9.0
DD size (max %)	–41.5	–25.0	–33.3	–23.5	–22.8	–21.6	–21.2	–23.6	–22.2
Turnover									
Length recov (max–days)	14.0	21.0	12.0	11.0	27.0	29.0	26.0	33.0	28.0
Average Turnover	75%	86%	80%	76%	84%	87%	87%	87%	88%

Note: NOIR is a portfolio construction rule that involves weighting factors with respect to their most recent Information Ratio. EW involves weighting factors equally. IC COV is a portfolio construction rule that involves weighting factors based on their Information Coefficient covariance matrix. IC COR is a portfolio construction rule that involves weighting factors based on their Information Coefficient correlation matrix. MEW is the portfolio construction rule that relies on Multiple Estimation Windows. 1, 3, 5, 7, and 10 blocks indicate the number of window blocks. Given a maximum window size of 130 days, for example, a 3-block analysis refers to the average calculated across windows of 43 days, 87 days, and 130 days.

*** and * indicate statistical significance at the 1% and 5% significance levels, respectively.*

schemes, respectively. The stock turnover remains relatively stable around 0.87 from the 0.75 and 0.86 showed by the NOIR and EW benchmarks, respectively.

Averaging over different rolling window blocks (i.e., 1 to 10) varies the performance and risk profile of the strategy. Our analysis suggests that averaging optimized factor weights over 3 different blocks will result in an optimal balance between performance and risk. Despite the fact that 10 blocks gives the highest IR and average return, using 3 blocks seems to provide slightly better drawdown and volatility statistics. Therefore, the use of 3 blocks generates an IR of 0.99 combined with a drawdown of –21.6%, which is an improvement to the base case when only one rolling window of 130 days is taken into account, and IR and drawdown are 0.89 and –22.8%, respectively.

Exhibit 5 depicts the cumulative performance of the backtesting analysis using three rolling blocks. This approach also outperforms NOIR and EW in the study period. The patterns are similar for IC COV and IC COR.[5]

Stock-Level Optimization

The multiple estimation windows approach and the bootstrap procedure that we discussed earlier both use the average model weighting and disregard the information that is included in the distribution (or the covariance) of model weighting. In this section, we investigate the benefit of incorporating this information into the portfolio construction process.

To incorporate the information on the distribution of weights, we consider the following optimization problem:

$$\max_{z}\left[z\cdot\alpha-\gamma z(\Sigma_{stock}+\beta\,\Sigma_{weights}\beta')z'\right] \tag{6}$$

$$\text{subject to } z_{\min}\le z_i\le z_{\max} \text{ and } \sum_{i=1}^{K} z_i=1$$

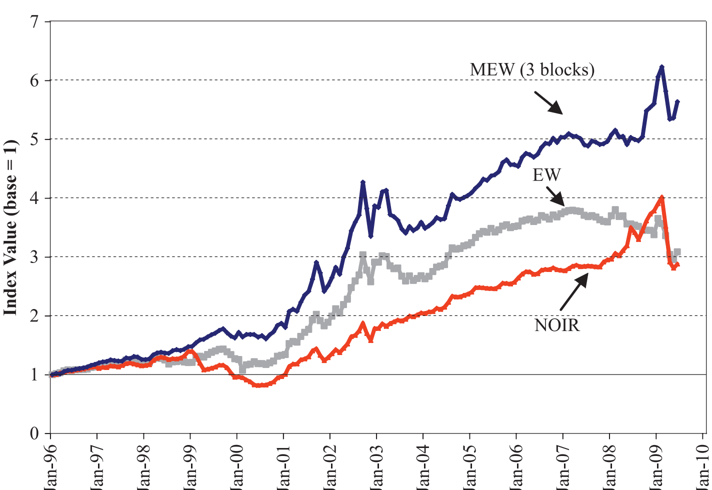
where $z=(z_1,\ldots,z_N)$ denotes the weights over the N stocks, $\alpha=(\alpha_1,\ldots,\alpha_N)'$ is the $N\times 1$ vector of stock alphas, γ is a risk aversion parameter, Σ_{stock} is the $N\times N$ sample covariance matrix of stock returns, β is an $N\times K$ matrix that includes the N stock exposures to the K factors, and $\Sigma_{weights}$ is a $K\times K$ weight covariance matrix.

The optimization in Equation (6) incorporates the first two moments/co-moments of the distribution of model weighting. In particular, the vector of stock alphas is defined as the product of the $N\times K$ matrix of style factor scores for the stocks in the universe and of the mean weights from the distribution of factor model weights, that is, $\bar{w}'=(\bar{w}_1,\ldots,\bar{w}_K)'$. Similarly, $\Sigma_{weights}$, the $K\times K$ covariance matrix of model weights, is computed from the distribution of factor model weights.

The optimization in Equation (6) involves an additional layer of optimization relative to that in Equation (5). First, we obtain the distribution of factor model weights through factor-level optimization as in Equation (5). Next, we pursue stock-level optimization through Equation (6). Numerical considerations involved in conducting both layers of optimization (as we discussed earlier) lead us, however, to choose to pursue a naive weighting for factors that is similar to NOIR. We believe factor-level optimization is less critical in this exercise. Our objective is to derive the distribution of factor model weights, and we believe that we get sufficient information for the distribution, in particular, for the covariance of model weights.

To benchmark the multiple estimation windows approach, we consider a stationary bootstrap procedure. For this procedure, instead of estimating weights through different blocks of data, we sample factor returns from the empirical joint distribution of factor returns.

EXHIBIT 5
Wealth Curves of Multiple Estimation Windows Approach vs. Benchmark, 1996–2009



In Exhibit 6, we tabulate the results from the risk and performance analysis. Exhibit 7 presents the wealth curves of the two strategies. The portfolios obtained with the two models exhibit very similar performance from January 1996 to December 2001. For the rest of the sample period, optimal portfolios obtained through the multiple estimation windows model outperformed those obtained through bootstrapping. Overall, the analysis suggests that when the full information from the model weight distribution is used, the multiple estimation windows approach is economically more attractive than the bootstrap procedure.

EXHIBIT 6

Performance and Risk Statistics (Stock-Level Optimization)

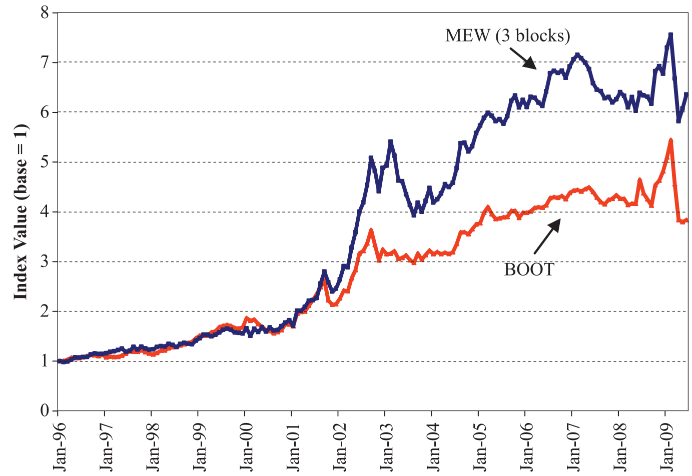
	MEW 3 blocks	BOOT
Empirical Return Distribution (monthly)		
Min.	–12.82%	–16.31%
1st Quintile	–1.56%	–1.05%
Median	1.14%**	0.75%**
Mean	1.27%**	0.93%**
Mean (Ann.)	16.35%	11.75%
5th Quintile	4.27%	2.83%
Max.	17.53%	14.79%
Volatility, IR, and Skew Statistics		
Ann. Volatility	16.9%	14.9%
Ann. IR	0.90**	0.75**
Skew	0.19	–0.55
Risk Statistics		
Ann. Semi-Volatility	9.9%	12.3%
CVaR 10%	–7.03%	–6.66%
CVaR 5%	–8.80%	–9.41%
Drawdown Statistics		
Length DD (max–days)	15.0	11.0
DD size (max %)	–27.4	–30.2
Length recov (max–days)	16.0	15.0
Turnover		
Average Turnover	83%	86%

Note: MEW is the portfolio construction rule that relies on Multiple Estimation Windows. 3 blocks indicates the number of window blocks. Given a maximum window size of 130 days, a 3-blocks analysis refers to the average calculated across windows of 43 days, 87 days, and 130 days. BOOT is the portfolio construction rule that relies on stationary bootstrap of factor returns. Both MEW and BOOT use the full information from the model weight distribution.

*** indicates statistical significance at the 1% significance level.*

EXHIBIT 7

Wealth Curves of Multiple Estimation Windows and Bootstrap Approaches That Use Information from the Model Weight Distribution



CONCLUSION

In this article, we investigate if techniques that address estimation risk, optimal prediction, and variable selection improve equity style-rotation strategies. We employ two approaches that have recently been shown to work well in the presence of estimation risk and structural breaks in data. The results from our analysis indicate that performance and risk gains can be obtained by dynamically adjusting exposures to style factors. Departing from naive approaches generates benefits that in our dataset can be as significant as an improvement in the IR of about 54% (i.e., from 0.65 (naive) to about 1 (dynamic)). We show, however, that the speed of adjustment should be limited in order to improve equal-weighted schemes (i.e., naive diversification). A fast reaction model will tend to generate unnecessary turnover and the model factor reweighting can be triggered by "factor noise."

Although we believe our results overall are compelling, we are aware that using a dynamic approach is not a panacea. Forecasting style returns, and therefore identifying style turning points, is not a trivial task and the ability to do this successfully is the key. As with any quantitative investment strategy, there is not much point in having the world's best portfolio construction technique or in having an accurate market impact cost forecasting model if the ability to successfully forecast return is lacking. Therefore, we believe that we have a good framework for determining optimal style weights, but recognize that the key input is the style return.

ENDNOTES

The core part of this research was conducted while M. Rodrigo Dupleich Ulloa was a vice president in Global Quantitative Research, Europe, at Citi Investment Research in London, UK. We are grateful for the detailed and constructive comments of an anonymous referee that have substantially helped improve our article. We also appreciate the comments and suggestions of Michael Wolf.

[1]This relationship follows from (DeMiguel, Garlappi, and Uppal [2009]),

$$\sum_{i=1}^{K}(w_i - 1/K)^2 = \sum_{i=1}^{K} w_i^2 + \sum_{i=1}^{K}(1/K)^2 - 2\sum_{i=1}^{K} w_i(1/K)$$
$$= \sum_{i=1}^{K} w_i^2 + K/K^2 - 2/K = \sum_{i=1}^{K} w_i^2 - 1/K$$

[2]This is especially true when using the ridge approach as was argued by Jagannathan and Ma [2003] that, with the short-sale constraint in place, the sample covariance matrix performs as well as covariance matrix estimates based on, for example, shrinkage estimators.

[3]This point is also related to how we should interpret a negative IC/IR. On the one hand, a negative IC/IR could be seen as a short signal if the aim of the investment process is to track factor synthetic assets (e.g., a value ETF); from this point of view, see, for example, Qian et al. [2007, p. 205]) Section 7.2.4 where factor shorting is allowed. On the other hand, a negative IC/IR for a signal could be interpreted as the signal has low forecasting power, and therefore, its exposure to this factor/signal should be decreased.

[4]We measure stock turnover as the percentage of in/out stocks for each of the top and low baskets. We divide by two in order to obtain one statistic. This approach resembles $\frac{1}{2}\sum |w_{t,i} - w_{t-1,i}|$.

[5]In additional analysis, we follow a stationary bootstrap approach in which we sample factor returns from the empirical joint distribution of factor returns for different window sizes. We produce subsamples for each window size. We estimate mean vectors and covariance matrices for each bootstrapped sample. We then calculate optimal weights for each bootstrapped sample, and then calculate optimal portfolio weights by averaging the bootstrapped sample-specific optimal weights. This procedure results in portfolios that are superior to NOIR and EW portfolios, but overall are inferior to those produced with the multiple estimation windows approach.

REFERENCES

Best, M., and R. Grauer. "On the Sensitivity of Mean-Variance Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results." *Review of Financial Studies*, 4 (1991), pp. 315-342.

Bishop, C. *Pattern Recognition and Machine Learning*. Springer, 2006.

Brodie, J., I. Daubechies, C. De Mol, and D. Giannone. "Sparse and Stable Markowitz Portfolios." ECORE Discussion Paper 2007/61, 2007.

Chopra, V., and W. Ziemba. "The Effect of Errors in Mean and Covariance Estimates on Optimal Portfolio Choice." *The Journal of Portfolio Management*, 20 (1993), pp. 6-11.

Daubechies, I., M. Defrise, and C. De Mol. "An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint." *Communications on Pure and Applied Mathematics*, 57 (2004), pp. 1416-1457.

DeMiguel, V., L. Garlappi, and R. Uppal. "Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" *Review of Financial Studies*, Vol. 22, No. 5 (2009), pp. 1915-1953.

DeMiguel, V., L. Garlappi, J. Nogales, and R. Uppal. "A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms." *Management Science*, 55 (2009), pp. 798-812.

Giamouridis, D. "Bayesian Approaches for Portfolio Construction: A Review." In *Re-Thinking Risk Management and Reporting, Vol. I: Uncertainty, Bayesian Analysis and Expert Judgment*, edited by K. Böcker, pp. 361-384. Risk Books, 2010.

Giamouridis, D., and S. Paterlini. "Regular(ized) Hedge Fund Clones." *Journal of Financial Research*, Vol. 33, No. 3 (2010), pp. 223-247.

Hastie, T., R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, 2009.

Hoerl, A., and R. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics*, 8 (1970), pp. 27-51.

Jagannathan, R., and T. Ma. "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps." *Journal of Finance*, 58 (2003), pp. 1651-1684.

Lobo, M.S., M. Fazel, and S. Boyd. "Portfolio Optimization with Linear and Fixed Transaction Costs." *Annals of Operation Research*, 152 (2007), pp. 341-365.

Pesaran, M., and A. Pick. "Forecast Combination across Estimation Windows." *Journal of Business Economics and Statistics*, Vol. 29, No. 2 (2011), pp. 307-318.

Pesaran, M., and A. Timmermann. "Selection of Estimation Window in the Presence of Breaks." *Journal of Econometrics*, 137 (2007), pp. 134-161.

Qian, E., R. Hua, and E. Sorensen. *Quantitative Equity Portfolio Management: Modern Techniques and Applications*. Chapman & Hall/CRC, 2007.

Tibshirani, R. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society* Series B (Methodological), 58 (1996), pp. 267-288.

Welsch, R., and X. Zhou. "Application of Robust Statistics to Asset Allocation Models." *REVSTAT- Statistical Journal*, Vol. 5, No. 1 (2007), pp. 97-114.