Deutsche Bank

7 June 2011

# Portfolios Under Construction

# Tail risk in optimal signal weighting

**Research summary**
We demonstrate how to manage tail risk in the optimal signal weighting decision process. We develop a series of structured models to better estimate tail distribution. We also design an efficient PGP algorithm to perform the multi-dimensional optimization.

**Team Contacts**

**Yin Luo, CFA**
Strategist
(+1) 212 250-8983
yin.luo@db.com

**Rochester Cahan, CFA**
Strategist
(+1) 212 250-8983
rochester.cahan@db.com

**Miguel-A Alvarez**
Strategist
(+1) 212 250-8983
miguel-a.alvarez@db.com

**Javed Jussa**
Strategist
(+1) 212 250-4117
javed.jussa@db.com

**Zongye Chen**
Strategist
(+1) 212 250-2293
john.chen@db.com

**Innovative, rigorous, and practical quantitative research**

**Why do we need to worry about tail risk?**
Traditional multi-factor stock selection models are built on mean-variance optimization without explicitly accounting for tail risk (i.e., the Grinold & Kahn or Qian, Hua, and Sorensen factor weighting approach). Most common factors have negative skewness/excess kurtosis; therefore, most common multi-factor models also show greater tail risk than what's implied by a normal distribution.
In this research, we demonstrate the benefit of incorporating tail risk in our optimal signal weighting decision process.

**Structured models to better estimate tail distribution**
The biggest challenge in tail risk management is dimensionality, i.e., the number of parameters to be estimated is too large to properly account for higher moments. We develop a set of structured models to better estimate the tail distribution.

**Higher moment optimization – PGP to the rescue**
The second challenge in tail risk management is how to optimize the multiple and conflicting goals of maximizing return (and skewness) and minimizing risk (and kurtosis). We design a PGP algorithm to facilitate this task.

**Conclusion**
In the end, we find our MVSK (mean-variance-skewness-kurtosis) alpha models are intuitive and reasonably easy to implement. Our MVSK models are also aware of the underlying macroeconomic environment. We are able to substantially reduce model downside risk (as measured by drawdown and expected shortfall), while model IR (and Sortino ratio) can be maintained at roughly the same level (or increased moderately).

Source: gettyimages.com

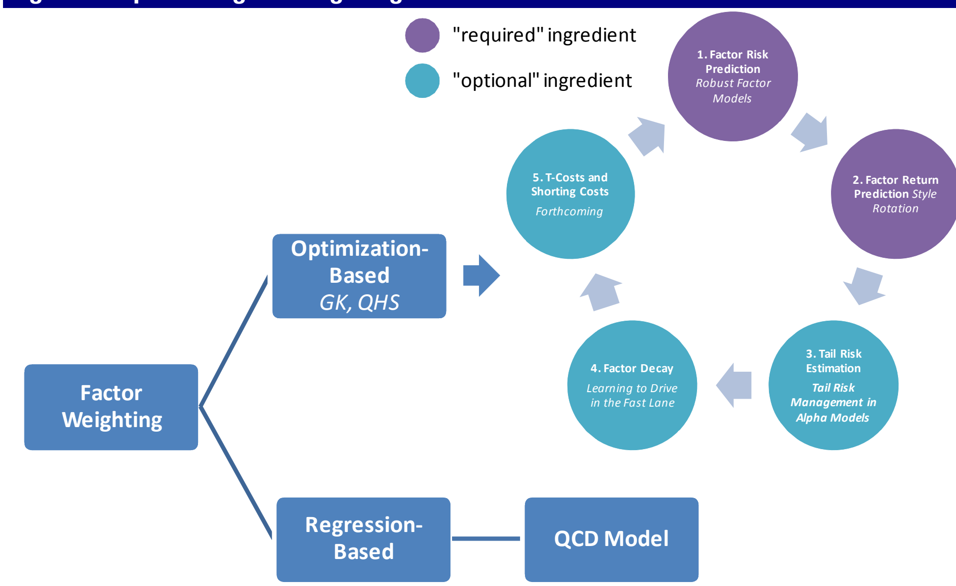
## Table of Contents

# A letter to our readers

## Managing tail risk in the alpha model construction process

As shown in Figure 1, there are two major signal weighting approaches[1]. The most common approach is based on a simple mean-variance optimization, which treats each factor as an "asset class". This is typically called the Grinold & Kahn (GK) or Qian, Hua, and Sorensen (QHS) approach. In the GK/QHS approach, there are two required raw ingredients (expected factor returns and predicted factor risk) and some additional/optional components.

**Figure 1: Optimal signal weighting decision**




*Source: Deutsche Bank Quantitative Strategy*

One of the issues with the GK/QHS approach is that it is based on the assumption that factor returns follow a multivariate normal distribution. However, in reality, we know that this assumption is invalid. Similar to stock returns, factor returns also show skewness and excess kurtosis. Therefore, stock selection models based on the GK/QHS models are not explicitly designed to handle extreme downside risk and are likely to expose managers to larger tail risk than what is being suggested by a normal distribution.

How to manage tail risk is a very challenging task. Managing tail risk directly at the stock level is probably too ambitious. In this research, we attempt to manage tail risk at the factor model level. The most significant benefit is the decrease in dimensionality – we deal with tens of factors rather than thousands of stocks. However, even at the factor level, trying to estimate the tail distribution is still a huge challenge. For an alpha model with 12 factors, we need to estimate 1,819 higher moment parameters, which requires more than 12 years of monthly data or almost two years of daily data. For this many parameters, the estimation error can easily overshadow any added benefit from better tail risk control.

In this research, we further extend the robust factor covariance models developed in Luo, Cahan, Alvarez, Jussa, and Chen [2011] to the higher moment tensors (i.e., coskewness and cokurtosis matrices). We find that structured models can substantially reduce the model

[1] Please refer to Luo, Y., Cahan, R., Alvarez, M., Jussa, J., and Chen, Z. [2011]. "Emerging Issues: A Roadmap for Quantitative Investing", Deutsche Bank Quantitative Strategy, May 17, 2011, for details.

parameter space. For example, based on our single index model, we only need to estimate 63 higher moment parameters (instead of 1,819) for 12 factors, a reduction of over 96%.

Once we have successfully estimated all required and optional raw ingredients (factor returns, factor covariance matrix, factor coskewness and cokurtosis matrices), the next obstacle is how to balance the four conflicting goals of: 1) maximizing expected return; 2) minimizing expected risk; 3) maximizing expected skewness; and 4) minimizing expected kurtosis. We develop an interesting optimization routine called PGP or polynomial goal programming to optimally balance the four goals. The optimal factor weights based on the MVSK (mean-variance-skewness-kurtosis) algorithm is supposed to have the best trade-off of the four goals.

In the end, we find the MVSK models have the following benefits compared to the equivalent GK/QHS models:

- Factor timing or style rotation is no longer essential. MVSK models with and without factor return prediction make little difference in IR, Sortino ratio, drawdown, or conditional VaR. This essentially eliminates the controversy of estimating future factor returns.
- After incorporating tail risk in our signal weighting decision, our model's downside risk (as measured in drawdown or expected shortfall) declines significantly, while IR (and Sortino ratio) can be maintained at roughly the same level (or increased moderately).
- MVSK models tend to have much lower factor turnover (and portfolio turnover) than the comparable MVO models; therefore, the after cost performance is even more attractive.

Yin, Rocky, Miguel, Javed, and John

**Deutsche Bank US/Global Quantitative Equity Strategy Team**

# Optimal signal weighting

## An overview of the optimal factor weighting decision

Discovering interesting and strong stock selection signals is only the first step. The next natural step is how to optimally combine these factors to construct a multi-factor alpha model.

As shown in Figure 2, one of the most common signal weighting approaches is based on a simple mean-variance optimization, by treating each factor as an "asset class". This is typically called the Grinold & Kahn (GK) or Qian, Hua, and Sorensen (QHS) approach[2].

Let us assume that we have $N$ stock-selection factors (i.e., alpha factors or signals). Now, we want to find an optimal combination of the $N$ factors, in order to maximize expected model IR. The problem is the same as the case of an optimal portfolio of $N$ stocks. Therefore, we can translate the objective into an optimization problem.

The objective is to $\underset{\omega}{Max}\left(ModelIC\right) = \underset{\omega}{Max}\left(\omega_t' \overline{R}_t\right)$

Subject to

$$\omega_t' \Omega_t \omega_t \le \xi_t$$

$$\omega_t^{LB} \le \omega_t \le \omega_t^{UB}$$

$$\sum_{i=1}^{N} \omega_{i,t} = 1$$

where,

$\overline{R}_t$ is a vector (of size $N$ ) of expected factor returns at time $t$ ,

$\omega_t$ is a vector (of size $N$ ) of factor weights at time $t$ (this is the decision variable),

$\omega_t^{LB}$ is the minimum factor weight (typically set to $0$ , meaning that we do not typically go against a factor) at time $t$ ,

$\omega_t^{UB}$ is the maximum factor weight at time $t$ ,

---

[2] Please refer to Luo, Y., Cahan, R., Alvarez, M., Jussa, J., and Chen, Z. [2011]. "Emerging Issues: A Roadmap for Quantitative Investing", Deutsche Bank Quantitative Strategy, May 17, 2011, for details.

$\Omega_t$ is the factor variance-covariance matrix at time $t$, and

$\xi_t$ is the targeted variance of the multifactor model at time $t$.

In this setting, we add a subscript $t$ for expected factor returns ($\overline{R}_t$), factor covariance matrix ($\Omega_t$). factor weights ($\omega_t$), and all constraints ($\omega_t^{LB}$, $\omega_t^{UB}$, and $\xi_t$) to assume that they can all be time varying.

**Figure 2: Optimal signal weighting decision**


Source: Deutsche Bank Quantitative Strategy

In the GK/QHS approach, there are two required raw ingredients (expected factor returns and predicted factor risk) and some additional/optional components. In our style rotation research[3], we focus on predicting factor returns, i.e., $\overline{R}_t$, while using a sample covariance matrix in the optimization. Factor timing or return prediction is one of the most controversial topics in quantitative investing. One camp argues that factor timing is the future of quantitative investing, as most traditional factors now behave like risk factors. When expected returns are close to zero, timing the factors is the only way to add alpha. The opposite argument is that factor timing is a low breadth strategy – rather than investing in thousand of stocks, now we invest in tens of factors – so it requires too high a skill to be possible in practice. Also, the opponents suggest there are at least two drawbacks to factor timing: increased portfolio turnover[4] and potential data mining bias.

Our second factor weighting research focuses on a less ambitious and less controversial topic – the factor covariance matrix, i.e., $\Omega_t$. Risk is generally considered to be much easier to predict than returns. Because of the small dimensionality in the factor space, the factor covariance matrix is typically estimated using sample data alone. In our Robust Factor Model

[3] Please refer to Luo, Y., Cahan, R., Jussa, J., and Alvarez, M. [2010]. "GTAA/Signal Processing: Style Rotation", Deutsche Bank Quanitative Strategy, September 7, 2010, for details.

[4] In Cahan, R., Luo, Y., Jussa, J., and Alvarez, M. [2010]. "Portfolios under Construction: It's all in the timing", Deutsche Bank Quantitative Strategy, August 19, 2010, we discuss the trade-off of factor timing skills and increased portfolio turnover.

research[5], we illustrate that, even at the factor space, structured covariance matrices often outperform the sample covariance matrix. In the end, we recommend a particular structured model called the single index (SF) model, which suggests that factor correlation is only related to its beta sensitivity to a "hyper" index.

Our third signal weighting research[6] takes into account factor decay/turnover. The simple framework above assumes all factors have the same decay profile and the same level of transaction costs. In reality, we know a fast decay factor like reversal would have a very different after-cost performance profile than a slow decay factor like valuation. Incorporating transaction costs in the signal weighting decision is intuitive, but proves to be very difficult in practice. The framework we develop in the paper seems to be simple and intuitive, while it also gives us a good starting point. It seems to be particularly useful when the portfolio's turnover constraint is tight.

One of the issues with the GK/QHS approach is that it is based on the assumption that factor returns follow a multivariate normal distribution. However, in reality, we know that this assumption is invalid. Similar to stock returns, factor returns often show negative skewness and excess kurtosis. Therefore, stock selection models based on the GK/QHS models are not explicitly designed to handle extreme downside risk. The purpose of this research is to attempt to incorporate tail risk management in the optimal factor weighting decision framework.

[5] Please refer to Luo, Y., Cahan, R., Alvarez, M., Jussa, J., and Chen, Z. [2011]. "Portfolios under Construction: Robust Factor Models", Deutsche Bank Quantitative Strategy, January 24, 2011.

[6] Please refer to Alvarez, M., Luo, Y., Cahan, R., Jussa, J., and Chen, Z. [2011]. "Portfolios under Construction: Learning to Drive in the Fast Lane", Deutsche Bank Quantitative Strategy, April 26, 2011.

# Managing tail risk

It is widely known that financial assets exhibit significant downside risk compared to what is implied by an equivalent normal distribution. Downside risk, also called tail risk, is often associated with extreme events or "black swan" events (see Taleb [2007]). Although it is widely known and accepted, in practice, the mean-variance approach still plays a predominant role in model design and portfolio construction.

Risk model vendors like Barra and Axioma either have specific products (e.g., Barra-BxR – see Goldberg, Hayes, Menchero, and Mitra [2009] for details) or have been doing active research on this topic, especially since the so-called "summer 2007 quant crisis" when certain quant factors showed tremendous intra-month drawdown in August 2007. Risk vendors mostly deal with tail risk at the asset level, by assuming asset returns are not necessarily normally distributed.

Managing tail risk at the portfolio level is extremely difficult for two reasons. The first hurdle is that a tail event, by definition, is rare; and therefore, we have only a limited data history. We need to make some kind of assumptions to account for the tail distribution. Estimating higher moments like skewness and kurtosis at the stock level is almost impossible, due to the large number of parameters. Even if we can somehow estimate the tail distribution, how to build optimal portfolios taking into account the tail distribution is also very challenging.

In this paper, we decide to take a different approach. We try to model and manage tail risk at the factor level. There are many potential benefits – the most important one being dimensionality. Once we are able to estimate the tail distribution, we further develop an interesting optimization technique that allows us to optimally combine factors, controlling for tail risk.

## Why do we need to worry about higher moments?

Mean-variance optimization critically depends on two fundamental assumptions: 1) investors have a quadratic utility function; and 2) asset returns are normally distributed. Both assumptions are problematic in practice. In the presence of non-quadratic preferences and nonnormally distributed asset returns, mean-variance optimized portfolios may yield welfare loss (see Hong, Tu, and Zhou [2007]). Horvath and Scott [1980] and Kimball [1993] show that investors exhibit nontrivial preferences for higher moments, in addition to mean and variance. In particular, it has been shown that investors are willing to accept lower expected return and higher variance (compared to the mean-variance benchmark) in exchange for higher skewness and lower kurtosis (see Harvey and Siddique [2000], Dittmar [2002], and Mitton and Vorkink [2007]).

When asset returns are not normally distributed, higher moments like skewness and kurtosis are needed to properly model the overall distribution.

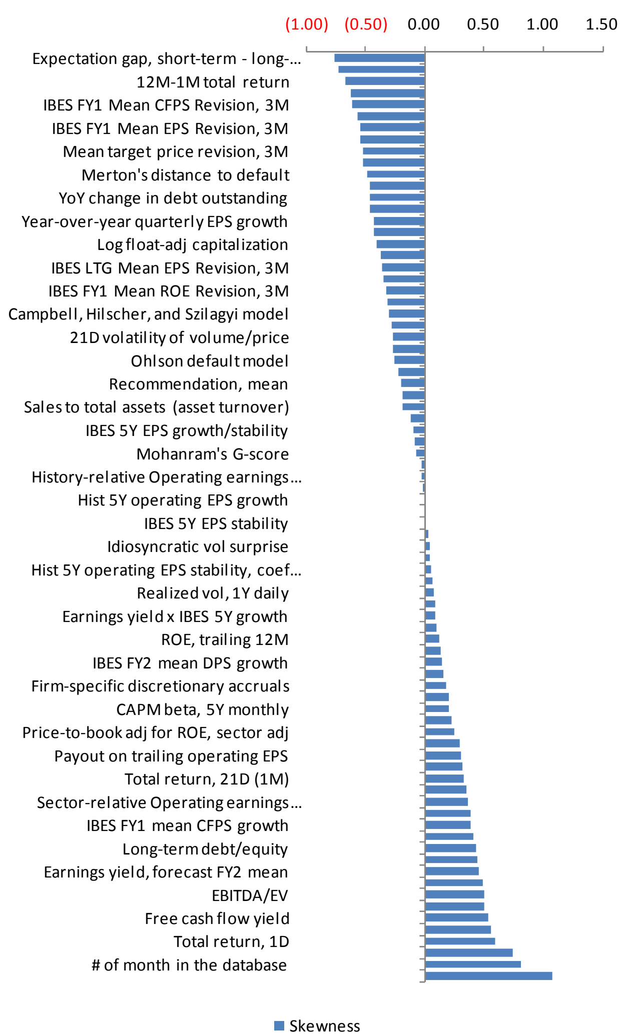
## Most factors have negative skewness and excess kurtosis

The traditional GK/QHS approach assumes that factor returns (or IC's) follow a normal distribution. Therefore, a standard mean-variance optimization gives us the best risk adjusted model. In reality, the Jacque-Bera test rejects the null hypothesis of normal distribution for 47 of the 80 standard factors (or 59% of the factors) at the 10% confidence level. In addition, as shown in Figure 3 and Figure 4, 38 factors (or 48%) show negative skewness and 64 factors (or 80%) show excess kurtosis.

Factor models based on mean-variance optimization will inevitably show more downside risk than suggested by a normal distribution. To test this hypothesis, we calculate the Jarque-Bera statistics for the EQW (equally weighting 12 factors) and 39 models based on mean-variance optimization – the Jacque-Bera test reject the null hypothesis of normal distribution for 50% of these models. More interestingly, about 85% of these models show negative skewness and 100% show excess kurtosis.

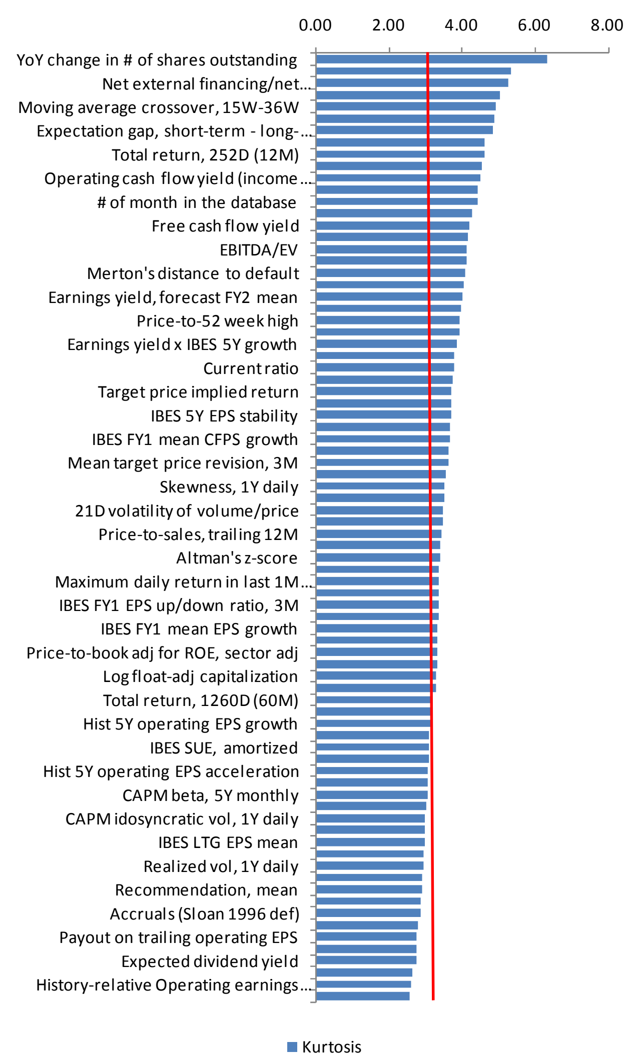
In particular, Figure 5 and Figure 6 show the empirical model performance distribution of the two most commonly used models, EQW and mean-variance optimized model (assuming naïve factor return prediction/naïve factor covariance estimation). Both models clearly show excess left tails compared to the normal distribution.

**Figure 3: Factor skewness**



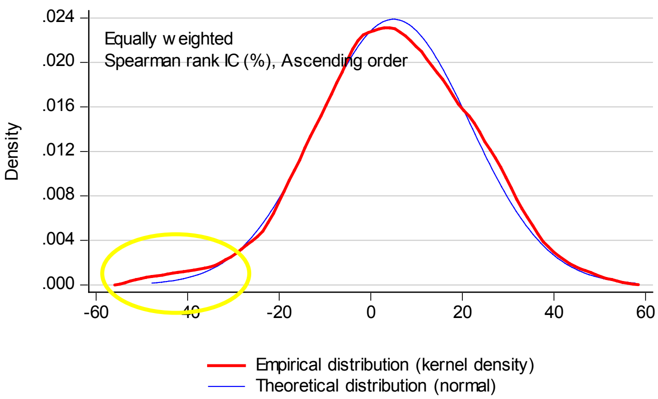

Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 4: Factor kurtosis**



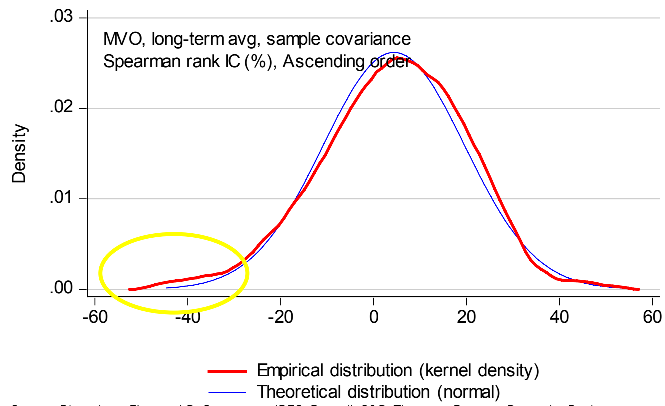

Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 5: Empirical distribution of EQW model**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 6: Empirical distribution of MVO model**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

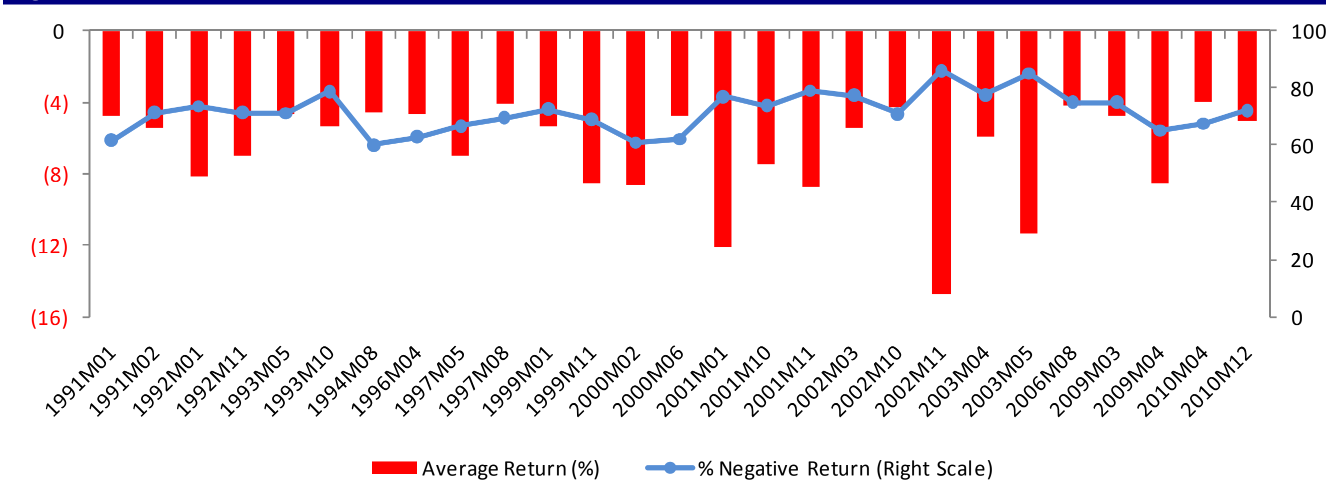
## What's the big deal with rare events

First of all, we need to define tail event. In the factor space, extreme negative events occur when most factors have negative performance and the average performance is severely negative. In the past approximately 25 years (280 months), we can identify 27 such periods or slightly less than 10% of the time. The average monthly factor IC during the stress periods was -6.7%, while the overall monthly average over the past 25 years is about 2.1%.

The multi-factor models we constructed following the GK/QHS approach have an average monthly information coefficient of 4.3% over the entire out-of-sample backtesting period (about 131 months or about 11 years), while the performance drops to -13.7% during those periods of stress.

On paper, one month or even a few months of extreme negative returns do not appear to be that important, so long as the overall performance is good. In reality, the consequences of rare events and extreme downside performance can be catastrophic.

**Figure 7: Tail events**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

## Difficulties with estimating higher moments

The problem of incorporating higher moments is that it increases the number of parameters exponentially. For a universe of 10,000 stocks, for example, we need to estimate 10,000 expected returns, 50 million covariance coefficients, 167 billion coskewness and 417 trillion cokurtosis parameters (see Figure 8). It is impossible to get a precise estimate of this many parameters, as shown in Brandt, Santa-Clara, and Valkanov [2009]. Academic research involving higher momentum optimization either assumes a very small number of assets or assumes asset returns follow certain known forms of distribution, or both (see Harvey, Liechty, Liechty, and Muller [2010].

**Figure 8: Required number of parameters**

| | Mean | Covariance | Coskewness | Cokurtosis | Total |
|---|---|---|---|---|---|
| Sample | N | (N+1)xN/2 | (N+2)x(N+1)xN/6 | (N+3)x(N+2)x(N+1)xN/24 | Dominated by (N+3)x(N+2)x(N+1)xN/24 |
| Constant correlation (CC) | N | N+1 | N+2 | 2xN+4 | 5N+7 |
| Single index (SF) | N | 2xN+1 | N+1 | N+1 | 5xN+3 |
| **Example 1: 10,000 assets** | | | | | |
| Sample | 10,000 | 50,005,000 | 166,716,670,000 | 416,916,712,502,500 | 417,083,479,187,500 |

Source: Deutsche Bank Quantitative Strategy

## Managing risk (especially tail risk) at the alpha model level

In our example, since we are modeling at the factor level, for 12 factors, we only need to estimate 12, 78, 364, and 1,365 mean, covariance, coskewness, and cokurtosis parameters (see Figure 8) – a tiny fraction of what we would need to estimate at the asset level.

In some of the structured models (to be elaborated on later sections), we can further reduce the number of parameters (see Figure 9). For example, in our single index model, we only need to estimate 63 parameters to fully account for the expected factor return, risk, coskewness, and cokurtosis of the 12 factors.

**Figure 9: Required number of parameters**

| | Mean | Covariance | Coskewness | Cokurtosis | Total |
|---|---|---|---|---|---|
| Sample | N | (N+1)xN/2 | (N+2)x(N+1)xN/6 | (N+3)x(N+2)x(N+1)xN/24 | Dominated by (N+3)x(N+2)x(N+1)xN/24 |
| Constant correlation (CC) | N | N+1 | N+2 | 2xN+4 | 5N+7 |
| Single index (SF) | N | 2xN+1 | N+1 | N+1 | 5xN+3 |
| **Example 1: 10,000 assets** | | | | | |
| Sample | 10,000 | 50,005,000 | 166,716,670,000 | 416,916,712,502,500 | 417,083,479,187,500 |
| Constant correlation (CC) | 10,000 | 10,001 | 10,002 | 20,004 | 50,007 |
| Single index (SF) | 10,000 | 20,001 | 10,001 | 10,001 | 50,003 |
| **Example 2: 12 factors** | | | | | |
| Sample | 12 | 78 | 364 | 1,365 | 1,819 |
| Constant correlation (CC) | 12 | 13 | 14 | 28 | 67 |
| Single index (SF) | 12 | 25 | 13 | 13 | 63 |

Source: Deutsche Bank Quantitative Strategy

# Higher moment optimization – preliminary

In this section, we outline the basics of higher moment optimization and mathematics. A more practical implementation will be further elaborated on later sections.

As suggested in Horvath and Scott [1980], investors are assumed to have preferences for higher odd (e.g., mean and skewness) and lower even moments (e.g., variance and kurtosis). Hence, an extension of the simple mean-variance optimization with higher moments can be expressed as:

$$\underset{\omega}{Max}\left[\omega_t'\overline{R}_t + \frac{\lambda(\lambda+1)}{6}\omega_t' M_3(\omega_t \otimes \omega_t) - \frac{\lambda(\lambda+1)(\lambda+2)}{24}\omega_t' M_4(\omega_t \otimes \omega_t \otimes \omega_t)\right]$$

Subject to

$$\omega_t'\Omega_t\omega_t \le \xi_t$$

$$\omega_t^{LB} < \omega_t < \omega_t^{UB}$$

$$\sum_{i=1}^{N}\omega_{i,t} = 1$$

where,

$\lambda$ is the risk aversion parameter to be calibrated with data,

$M_3$ is the coskewness matrix,

$M_4$ is the cokurtosis matrix, and

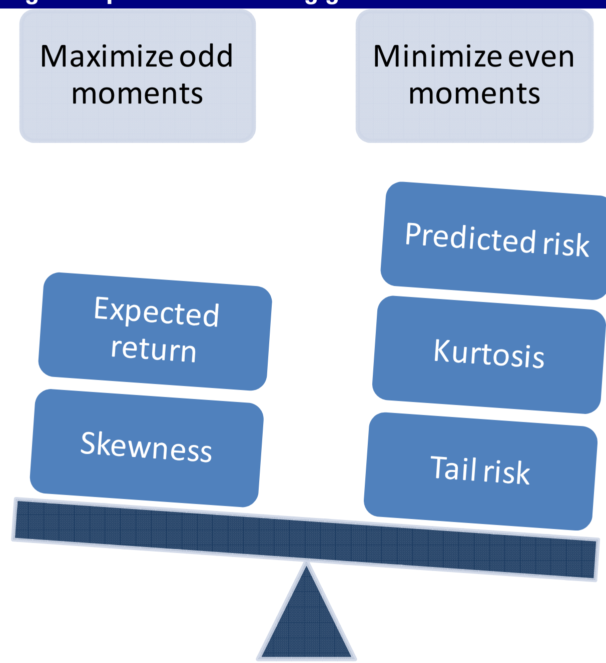
$\otimes$ is the Kronecker product or tensor product.

There are two issues with the above setting. First, we need estimate the coskewness and cokurtosis matrices. As describe in the above section, there are too many parameters to be estimated with sample data. It is simply impossible to estimate coskewness and cokurtosis matrices with any precision once the number of assets exceed more than ten or twenty. In our case, since we are dealing with factors and the number of factors is typically much smaller than the number of assets, using sample data is difficult but achievable.

Second, the objective function in the above setting involves third and fourth order terms. The problem is no longer convex, and therefore exhibits multiple local optima. Standard quadratic optimizers cannot handle such an optimization problem. Therefore, we need to use global optimizers, which tend to be extremely slow. In practice, it is almost impossible to get it done in a reasonable amount of time, once the number of variables exceeds ten or twenty. Again, since we are dealing with factors, it is difficult but not unattainable.

The higher moment optimization problem essentially consists in balancing four conflicting goals simultaneously – maximizing mean and skewness (i.e., odd moments) and minimizing variance and kurtosis (i.e., even moments), as shown in Figure 10. In later section, we will demonstrate how we propose to solve the problem involving multiple and conflicting goals numerically.

**Figure 10: Balancing multiple and conflicting goals**




*Source: Deutsche Bank Quantitative Strategy*

## Higher moment mathematics

The mathematics of higher moments is described in Fabozzi, Kolm, Pachamanova, and Focardi [2007]. Following Fabozzi, Kolm, Pachamanova, and Focardi [2007], we introduce the higher-order moment tensors. The second moment tensor $M_2$ of dimension $N \times N$ is the standard variance/covariance matrix. The third moment tensor (skewness) $M_3$ can be described as a three-dimensional cube of $N \times N \times N$. Similarly, the fourth moment tensor (kurtosis) $M_4$ is a four-dimensional cube of $N \times N \times N \times N$.

Mathematically, it is easier to transform the three-dimensional skewness and four-dimensional kurtosis cubes into two-dimensional matrices using Kronecker product operator:

$$M_3 = (s_{ijk})_{N \times N^2} = E\left[(R-\mu)(R-\mu)' \otimes (R-\mu)'\right]$$

$$M_4 = (s_{ijkl})_{N \times N^3} = E\left[(R-\mu)(R-\mu)' \otimes (R-\mu)' \otimes (R-\mu)'\right]$$

where each element is defined:

$$s_{ijk} = E\left[(R_{it}-\mu_{it})(R_{jt}-\mu_{jt})(R_{kt}-\mu_{kt})\right],\ i,j,k = 1,2,\ldots N,$$

$$s_{ijkl} = E\left[(R_{it}-\mu_{it})(R_{jt}-\mu_{jt})(R_{kt}-\mu_{kt})(R_{lt}-\mu_{lt})\right],\ i,j,k,l = 1,2,\ldots N,$$

$R_{it}$ is the return of factor $i$ at time $t$, and

$\mu_{it}$ is the mean of factor $i$ at time $t$ (mean is assumed to be time varying).

Similar to the covariance matrix, the coskewness and cokurtosis matrices have dimensions of $N \times N^2$ and $N \times N^3$, respectively.

Given a portfolio with a weighting vector $\omega$ (of dimension $N \times 1$), the variance, skewness, and kurtosis of the portfolio are therefore:

$$\mu^{(2)} = \omega' M_2 \omega,$$

$$\mu^{(3)} = \omega' M_3 (\omega \otimes \omega),$$

$$\mu^{(4)} = \omega' M_4 (\omega \otimes \omega \otimes \omega)$$

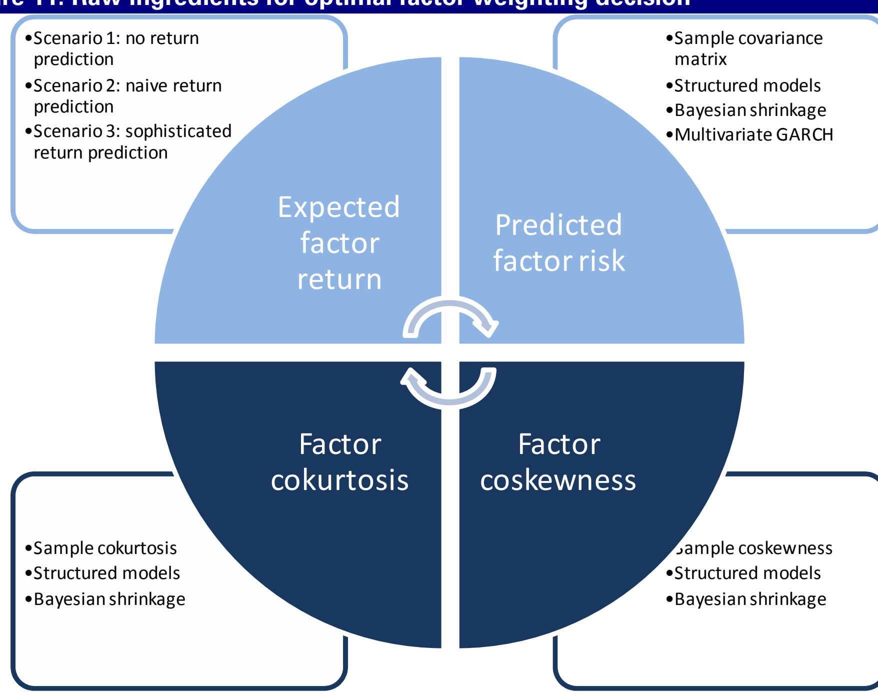
# Estimating higher moments

While a number of methods have been proposed on how to get a more precise estimate of the covariance matrix, little has been done on higher moments. Even the few papers (e.g., Kim and White [2004]) on higher moment estimation are exclusively on univariate skewness and kurtosis parameters, which count for only a small fraction of the coskewness and cokurtosis matrices. One paper of particular interest to us is Martellini and Ziemann [2010], where the authors explicitly deal with Bayesian estimation of coskewness and cokurtosis matrices. Martellini and Ziemann [2010] is an extension of Ledoit and Wolf [2003]. Another paper dealing with higher moment estimation is Malevergne and Sornette [2005]. Malevergne and Sornette [2005] estimate higher moment matrices by making assumptions on the multivariate distribution of the asset returns.

We find the improved covariance, coskewness, and cokurtosis estimators significantly raise model performance (IR and overall distribution) and portfolio performance (IR and overall distribution). Additionally, we find the use of these improved estimators significantly enhances the stability of the multifactor models and therefore produces portfolios with lower model/portfolio turnover.

Figure 11 shows the four required raw ingredients for our MVSK (mean-variance-skewness-kurtosis) optimal signal weighting model. There are a few models for each of the four elements, which will be further described in the following sections[7].

**Figure 11: Raw ingredients for optimal factor weighting decision**




*Source: Deutsche Bank Quantitative Strategy*

[7] Expected factor return is discussed in Luo, Y., Cahan, R., Jussa, J., and Alvarez, M. [2010]. "GTAA/Signal Processing: Style Rotation", Deutsche Bank Quantitative Strategy, September 7, 2010, while predicted factor risk is described in Luo, Y., Cahan, R., Alvarez, M., Jussa, J., and Chen, Z. [2011]. "Portfolios under Construction: Robust Factor Models", Deutsche Bank Quantitative Strategy, January 24, 2011. In this research, we focus on factor coskewness and cokurtosis estimations.

## Seven approaches

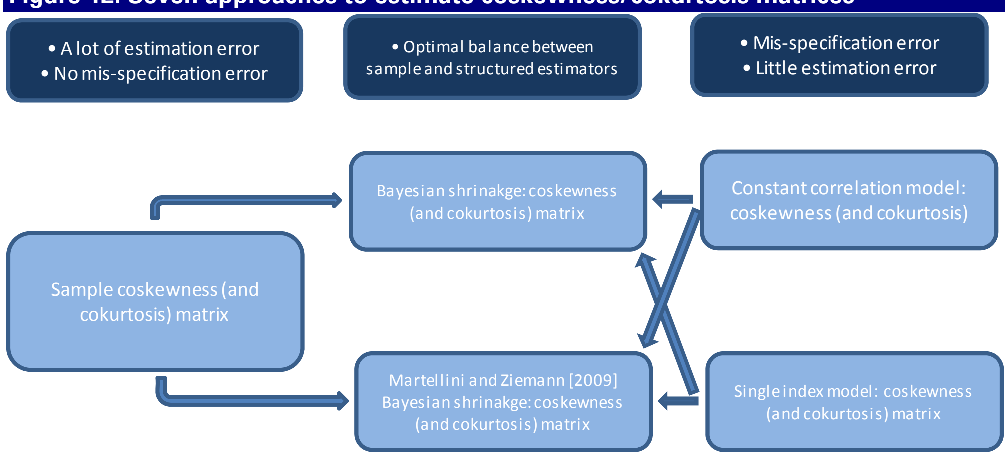
In this research, we test seven approaches to estimate the factor coskewness/cokurtosis matrices (see Figure 12).

On the one extreme, we could use sample data to directly estimate factor coskewness/cokurtosis matrices. This is very simple. However, because we need to estimate too many parameters using sample data, this approach suffers from too much estimation error.

On the other extreme, we test two highly structured models: constant correlation and single index. The structured models make explicit assumptions about the structure of the coskewness/cokurtosis matrices. The structured models simplify and reduce the parameter space significantly, which cuts down estimation error. The downside is that the models might be mis-specified.

Martellini and Ziemann [2010] developed a Bayesian shrinkage framework, which supposedly makes an optimal balance between the sample approach and structured models.

**Figure 12: Seven approaches to estimate coskewness/cokurtosis matrices**




*Source: Deutsche Bank Quantitative Strategy*

## Structured models for higher moments

Following Martellini and Ziemann [2010], we propose two structured coskewness and cokurtosis estimators: constant correlation (CC) and single index (SF). The technical details for factor return prediction are covered in our Style Rotation research[8] and factor covariance matrix are described in our Robust Factor Model research[9], so in this paper, we will focus on the higher moments (i.e., coskewness and cokurtosis matrices).

**Constant correlation (CC) model**

The constant correlation model is an extension of Elton and Gruber [1973] on higher moments.

[8] Please refer to Luo, Y., Cahan, R., Jussa, J., and Alvarez, M. [2010]. "GTAA/Signal Processing: Style Rotation", Deutsche Bank Quanitative Strategy, September 7, 2010, for details.

[9] Please refer to Luo, Y., Cahan, R., Alvarez, M., Jussa, J., and Chen, Z. [2011]. "Portfolios under Construction: Robust Factor Models", Deutsche Bank Quantitative Strategy, January 24, 2011.

Similar to sample average correlation coefficient[10], we now introduce seven extended correlation coefficients for higher moments.

$$\hat{r}^{(1)} = \frac{2}{TN(N-1)} \sum_{i>j} \sum_{t=1}^{T} \frac{\left(\overline{R}_{it}\overline{R}_{jt}\right)}{\sqrt{m_i^{(2)} m_j^{(2)}}}$$

$$\hat{r}^{(2)} = \frac{2}{TN(N-1)} \sum_{i>j} \sum_{t=1}^{T} \frac{\left(\overline{R}_{it}^{2}\overline{R}_{jt}\right)}{\sqrt{m_i^{(4)} m_j^{(2)}}}$$

$$\hat{r}^{(3)} = \frac{2}{TN(N-1)} \sum_{i>j} \sum_{t=1}^{T} \frac{\left(\overline{R}_{it}^{3}\overline{R}_{jt}\right)}{\sqrt{m_i^{(6)} m_j^{(2)}}}$$

$$\hat{r}^{(4)} = \frac{6}{TN(N-1)(N-2)} \sum_{i>j>k} \sum_{t=1}^{T} \frac{\left(\overline{R}_{it}\overline{R}_{jt}\overline{R}_{kt}\right)}{\sqrt{m_k^{(2)} \hat{r}^{(5)} \sqrt{m_i^{(4)} m_j^{(4)}}}}$$

$$\hat{r}^{(5)} = \frac{2}{TN(N-1)} \sum_{i>j} \sum_{t=1}^{T} \frac{\left(\overline{R}_{it}^{2}\overline{R}_{jt}^{2}\right)}{\sqrt{m_i^{(4)} m_j^{(4)}}}$$

$$\hat{r}^{(6)} = \frac{6}{TN(N-1)(N-2)} \sum_{i>j>k} \sum_{t=1}^{T} \frac{\left(\overline{R}_{it}^{2}\overline{R}_{jt}\overline{R}_{kt}\right)}{\sqrt{m_i^{(4)} \hat{r}^{(5)} \sqrt{m_j^{(4)} m_k^{(4)}}}}$$

$$\hat{r}^{(7)} = \frac{24}{TN(N-1)(N-2)(N-3)} \sum_{i>j>k>l} \sum_{t=1}^{T} \frac{\left(\overline{R}_{it}\overline{R}_{jt}\overline{R}_{kt}\overline{R}_{lt}\right)}{\sqrt{\hat{r}^{(5)} \sqrt{m_i^{(4)} m_j^{(4)}}\, \hat{r}^{(5)} \sqrt{m_k^{(4)} m_l^{(4)}}}}$$

where,

$\hat{r}^{(1)}$ is the standard correlation coefficient, and

$m_i^{(n)} = \frac{1}{T} \sum_{t=1}^{T} \left(\overline{R}_{it}\right)^n$ is the $n$ th centered sample moment of factor $i$.

The coskewness and cokurtosis matrices can therefore be estimated:

$\hat{s}_{iik} = \hat{r}^{(2)} \sqrt{m_i^{(4)} m_k^{(4)}}$, where $i = j$

$$\hat{s}_{ijk} = \hat{r}^{(4)} \sqrt{m_k^{(2)} \hat{r}^{(5)} \sqrt{m_i^{(4)} m_j^{(4)}}}$$

[10] The constant correlation model is described in Luo, Y., Cahan, R., Alvarez, M., Jussa, J., and Chen, Z. [2011]. "Portfolios under Construction: Robust Factor Models", Deutsche Bank Quantitative Strategy, January 24, 2011.

$$\hat{s}_{iiil} = \hat{r}^{(3)}\sqrt{m_i^{(6)}m_l^{(2)}}$$, where $i = j = k$

$$\hat{s}_{iikk} = \hat{r}^{(5)}\sqrt{m_i^{(4)}m_k^{(4)}}$$, where $i = j$ and $k = l$

$$\hat{s}_{iikl} = \hat{r}^{(6)}\sqrt{m_i^{(4)}\hat{r}^{(5)}\sqrt{m_k^{(4)}m_l^{(4)}}}$$, where $i = j$

$$\hat{s}_{ijkl} = \hat{r}^{(7)}\sqrt{\hat{r}^{(5)}\sqrt{m_i^{(4)}m_j^{(4)}}\hat{r}^{(5)}\sqrt{m_k^{(4)}m_l^{(4)}}}$$, where $i \neq j \neq k \neq l$

Under the constant correlation (CC) model, for 12 factors, the total number of parameters to be estimated is reduced from 1,819 based on the sample estimation approach to 67 (a reduction of 96%).

**Single index (SF) model**

The industry standard risk models are typically based on multiple factors rather than a single factor. Using multiple factors in our context is, however, problematic. Because we are not estimating stock return distribution, rather, we are estimating factor return distribution, we are essentially looking for a "hyper" index (or indices) that style factors are built upon. The dimension of factor space is typically small – in our case, we have 12 factors. Therefore, rather than using multiple factors, we choose a single "hyper" factor for simplicity.

The single factor model (Sharpe [1963]) assumes a single-factor linear model for $N$ asset returns:

$$R_{it} = c + \beta_i R_{Ft} + \varepsilon_{it}$$

where,

$R_{it}$ is the return of asset $i$ in period $t$,

$R_{Ft}$ is index return in period $t$, and

$\varepsilon_{it}$ is the idiosyncratic error term of asset $i$ in period $t$.

On the stock level, $R_{Ft}$ is typically assumed to be a broad-based index, e.g., S&P 500. In our application, because we are modeling factor returns, there is no intuitive reason why a broad-based equity index should be the driver of factor returns. On the other hand, we can probably argue that market sentiment (as measured by the VIX index) is a common driver behind many factors).

In the SF model, the regression residuals are assumed to be homoscedastic and cross-sectionally uncorrelated, i.e.,

$$\varepsilon \sim (0, \Psi)$$

All off-diagonal elements in $\Psi$ are zero. The idiosyncratic risks of the assets are reflected by the diagonal elements of $\Psi$. Therefore, the covariance matrix $S$ based on SF model can be calculated as:

$$S = \beta\beta' m_0^{(2)} + \Psi$$

where,

$\beta$ is the $N \times 1$ vector of the regression coefficients, and

$m_0^{(2)}$ is the variance of the single factor.

The structure of the $N \times N$ covariance matrix $\Psi$ of residual returns is:

$\psi_{ii} = E(\varepsilon_i^2)$, and

$\psi_{ij} = 0$, when $i \neq j$

A sample estimate of $\psi_{ii}$ is:

$$\hat{\psi}_{ii} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$$

where,

$\hat{\varepsilon}_{it}$ is the regression residual of asset $i$ in period $t$.

Extending the Sharpe [1963] model to higher moments, we can show:

$$M_2 = (\beta\beta') m_0^{(2)} + \Psi,$$

$$M_3 = (\beta\beta' \otimes \beta') m_0^{(3)} + \Phi,$$

$$M_4 = (\beta\beta' \otimes \beta' \otimes \beta') m_0^{(4)} + \Upsilon$$

where, similar to $m_0^{(2)}$ (the variance of the single index, i.e., VIX), $m_0^{(3)}$ and $m_0^{(4)}$ are the sample skewness and kurtosis of the single index. The three residual matrices $\Psi$, $\Phi$, and $\Upsilon$ can then be estimated by the following equations.

The specific return covariance matrix $\Psi$ (of dimension $N \times N$) is defined as:

$\psi_{ii} = E(\varepsilon_i^2)$, and

$\psi_{ij} = 0$, when $i \neq j$

A sample estimate of $\psi_{ii}$ is:

$$\hat{\psi}_{ii} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^{2}$$

The specific return coskewness matrix $\Phi$ (of dimension $N \times N^2$) is similarly defined as:

$\phi_{iii} = E\left(\varepsilon_i^3\right)$, and

$\phi_{iik} = \phi_{ijk} = 0$, when $i \neq j \neq k$

A sample estimate of $\phi_{iii}$ is:

$$\hat{\phi}_{iii} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^{3}$$

The specific return cokurtosis matrix $\Upsilon$ (of dimension $N \times N^3$) is similarly defined as:

$\upsilon_{iiii} = E\left(\varepsilon_i^4\right)$,

$\upsilon_{iiil} = 3\beta_i\beta_l m_0^{(2)}\psi_{ii}$,

$\upsilon_{iikk} = \beta_i^2 m_0^{(2)}\psi_{kk} + \beta_k^2 m_0^{(2)}\psi_{ii} + \psi_{ii}\psi_{kk}$,

$\upsilon_{iikl} = \beta_k\beta_l m_0^{(2)}\psi_{ii}$, and

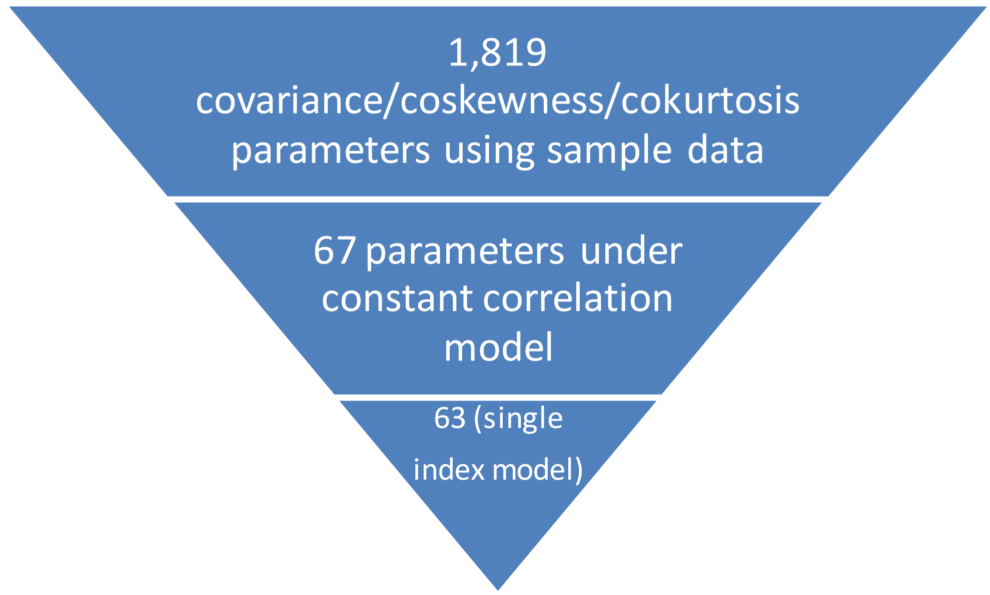
$\upsilon_{ijkl} = 0$, when $i \neq j \neq k \neq l$

A sample estimate of $\upsilon_{iiii}$ is:

$$\hat{\upsilon}_{iiii} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^{4}$$

As shown in Figure 13, under the single index (SF) model, the total number of parameters to be estimated is reduced from 1,819 based on the sample estimation approach to 63.

**Figure 13: The benefit of structured models**




*Source: Deutsche Bank Quantitative Strategy*

## Bayesian shrinkage estimators of higher moments

In order to find the "optimal" trade-off between estimation risk and misspecification error, we follow Ledoit and Wolf [2003] and Martellini and Ziemann [2010] to calculate the Bayesian shrinkage estimators for higher moments (coskewness and cokurtosis matrices) based on the two structured models (CC and SF). The mathematics for higher moment Bayesian shrinkage is quite involved. Readers can skip this section without losing continuity[11].

The essence of Bayesian shrinkage is to find a compromise between a non-structured estimator (i.e., the sample covariance/coskewness/cokurtosis matrix, $S$ ) and a highly structured estimator (i.e., the constant correlation or the single index model), $\Lambda$ , by computing a convex linear combination $\alpha\Lambda + (1-\alpha)S$ . $\alpha$ is a number between 0 and 1, called the shrinkage intensity or shrinkage constant. It measures the weight that is given to the structured estimator.

Any shrinkage estimator has three ingredients: An estimator with no structure, an estimator with a heavy structure, and a shrinkage constant. The estimator without structure is generally the sample covariance/coskewness/cokurtosis matrix. In this paper, we propose two structured estimators: the constant correlation estimator (Elton and Gruber [1973]) and the single factor estimator (Sharpe [1963]).

In Bayesian shrinkage, sampling error is reduced at the cost of specification error. Several authors have studied the optimal trade-off between sampling error and specification error in the context of optimal shrinkage theory, e.g., Jorion [1985, 1986], Ledoit and Wolf [2004a, 2004b].

Compared to sample estimators, the structured estimators (e.g., the constant correlation estimator and the single index estimator) have significantly less parameters to be estimated; therefore, the structured estimators tend to have lower estimation errors. On the other hand, they involve misspecification errors, because the models can be misspecified, i.e., the "true"

[11] We find some inconsistency and missing derivations in Martellini and Ziemann [2010] paper. In this research, we strive to provide a more consistent and complete description of the Bayesian shrinkage estimator.

correlation structure may not be constant correlation or based on the single index. The trick is to find an optimal trade-off between estimation error and specification error.

In the spirit of Ledoit and Wolf [2003, 2004a], we define the posterior misspecification function $L$ of the combined estimator as:

$$L(\alpha) = \left\| \alpha\Lambda + (1-\alpha)S - \Omega \right\|_F$$

where,

$\Omega$ is the true (unobserved) covariance/coskewness/cokurtosis matrix,

$S$ is the sample covariance/coskewness/cokurtosis matrix,

$\Lambda$ is the shrinkage target, i.e., a structured estimator of $\Omega$ (in our case, either the constant correlation estimator or the single index estimator),

$\left\| \bullet \right\|_F$ represents the Frobenius norm of a matrix (i.e., the square root of the sum of its squared entries), and

$\alpha$ is the shrinkage constant or shrinkage intensity.

Similar to the Bayesian shrinkage of the covariance matrix, the higher moment equivalent of $\alpha$ (Bayesian constant), $\pi$ (asymptotic variance of the sample estimator), $\gamma$ (the squared error of the structured estimator), and $\rho$ (the asymptotic covariances between the sample and the structured estimator) will be developed below. $\pi$ are the same for both structured models (CC and SF), while $\rho$ and $\gamma$ (and therefore $\alpha$) are different.

To review, the shrinkage constant $\alpha$ is defined as:

$$\hat{\alpha} = \frac{1}{T}\frac{\pi - \rho}{\gamma}$$

where,

$$\pi = \sum_{i=1}^{N}\sum_{j=1}^{K}\pi_{ij},$$

$$\rho = \sum_{i=1}^{N}\sum_{j=1}^{K}\rho_{ij},$$

$$\gamma = \sum_{i=1}^{N}\sum_{j=1}^{K}\gamma_{ij}$$

where,

$K = N$ for covariance matrix,

$K = N^2$ for coskewness matrix, and

$K = N^3$ for cokurtosis matrix,

The three raw ingredients behind the shrinkage constant, $\pi$ , $\rho$ , $\gamma$ represent the asymptotic variance of the sample estimator, the asymptotic covariance between the sample and structured estimators, and the squared error of the structured estimator.

**Asymptotic variance of the sample estimator**
$\pi$ is the asymptotic variance of the sample estimator, with the consistent estimators for $\pi_{ijk}$ (for coskewness calculation) and $\pi_{ijkl}$ (for cokurtosis calculation) given by:

$$\hat{\pi}_{ijk} = \frac{1}{T}\sum_{t=1}^{T}\left[(R_{it}-m_i)(R_{jt}-m_j)(R_{kt}-m_k)-\hat{\sigma}_{ijk}\right]^2,$$

$$\hat{\pi}_{ijkl} = \frac{1}{T}\sum_{t=1}^{T}\left[(R_{it}-m_i)(R_{jt}-m_j)(R_{kt}-m_k)(R_{lt}-m_l)-\hat{\sigma}_{ijkl}\right]^2$$

where,

$\hat{\sigma}_{ijk}$ and $\hat{\sigma}_{ijkl}$ are the sample estimates of the corresponding tensor entries.

**The squared error of the structured estimator**
$\gamma$ is the squared error of the structured estimator, with the consistent estimators for $\hat{\gamma}_{ijk}$ (for coskewness calculation) and $\hat{\gamma}_{ijkl}$ (for cokurtosis calculation) given below.

$$\hat{\gamma}_{ijk} = (\hat{\lambda}_{ijk} - \hat{\sigma}_{ijk})^2,$$

$$\hat{\gamma}_{ijkl} = (\hat{\lambda}_{ijkl} - \hat{\sigma}_{ijkl})^2,$$

where,

$\hat{\sigma}_{ijk}$ is the sample estimate of the coskewness parameter among factor $i$, $j$, and $k$,

$\hat{\sigma}_{ijkl}$ is the sample estimate of the cokurtosis parameter among factor $i$, $j$, $k$, and $l$,

$\hat{\lambda}_{ijk}$ is the structured estimate of the coskewness parameter among factor $i$, $j$, and $k$, and

$\hat{\lambda}_{ijkl}$ is the sample estimate of the cokurtosis parameter among factor $i$, $j$, $k$, and $l$,

**The asymptotic covariance between the sample and the structured estimators**
$\rho$ represents the asymptotic covariance between the sample and the structured estimators. Ledoit and Wolf [2003 and 2004a] derived closed-form formula of the consistent estimators for the covariance matrix ( $\rho_{ij}$ ). Martellini and Ziemann [2010] extend the analysis to coskewness and cokurtosis. This is the most critical aspect in the Bayesian shrinkage calculation (and the most difficult step). We use the delta-method to derive the asymptotic covariance. This step is fairly technical and involved. Interested readers should follow Greene [2007], Ledoit and Wolf [2003, 2004a, 2004b], and Martellini and Ziemann [2010]. In this paper, we omit the mathematical derivations and show the results directly. Depending on whether we use the constant correlation estimator (CC) or the single-index estimator (SF) as the structured estimator, $\rho$ are different.

**Coskewness**
First, $\rho_{iii}^{CC}$ and $\rho_{iii}^{SF}$ are simple – they are the same as $\pi_{iii}$ .

If we use the constant correlation estimator as the structured model, $\rho_{ijk}^{CC}$ can be estimated with:

$$\hat{\rho}_{ijj}^{CC} = \frac{\hat{r}^{(2)}}{2}\left[\sqrt{\frac{m_j^{(4)}}{m_i^{(2)}}} AsyCov\left(\sqrt{T}m_i^{(2)}, \sqrt{T}s_{ijj}\right) + \sqrt{\frac{m_i^{(2)}}{m_j^{(4)}}} AsyCov\left(\sqrt{T}m_j^{(4)}, \sqrt{T}s_{ijj}\right)\right],$$

$$\hat{\rho}_{ijk}^{CC} = \left(\frac{\hat{r}^{(4)}}{4}\sqrt{m_k^{(2)}\hat{r}^{(5)}\sqrt{\frac{m_j^{(4)}}{\left(m_i^{(4)}\right)^3}}}\right) AsyCov\left(\sqrt{T}m_i^{(4)}, \sqrt{T}s_{ijk}\right),$$

$$+\left(\frac{\hat{r}^{(4)}}{4}\sqrt{m_k^{(2)}\hat{r}^{(5)}\sqrt{\frac{m_i^{(4)}}{\left(m_j^{(4)}\right)^3}}}\right) AsyCov\left(\sqrt{T}m_j^{(4)}, \sqrt{T}s_{ijk}\right)$$

$$+\left(\frac{\hat{r}^{(4)}}{4}\sqrt{\frac{\hat{r}^{(5)}}{m_k^{(2)}}\sqrt{m_i^{(4)}m_j^{(4)}}}\right) AsyCov\left(\sqrt{T}m_k^{(2)}, \sqrt{T}s_{ijk}\right)$$

where,

$\hat{\rho}_{ijk}^{CC}$ is the consistent estimated asymptotic coskewness between the sample and the structured estimator, and

$AsyCov(\bullet)$ refers to the asymptotic covariance.

The asymptotic covariance can then be estimated using:

$$AsyCov\left(\sqrt{T}m_i^{(n)}, \sqrt{T}s_{ijk}\right) =$$

$$\frac{1}{T}\sum_{t=1}^{T}\left[(R_{it}-m_i)^n - m_i^{(n)}\right]\left[(R_{it}-m_i)(R_{jt}-m_j)(R_{kt}-m_k) - s_{ijk}\right]$$

If we use the single index estimator as the structured model, $\rho_{ijk}^{SF}$ can be estimated with:

$$\hat{\rho}_{ijk}^{SF} = \frac{s_{j0}s_{k0}m_0^{(3)}}{\left(m_0^{(2)}\right)^3} AsyCov\left(\sqrt{T}s_{i0}, \sqrt{T}s_{ijk}\right),$$

$$+\frac{s_{i0}s_{k0}m_0^{(3)}}{\left(m_0^{(2)}\right)^3} AsyCov\left(\sqrt{T}s_{j0}, \sqrt{T}s_{ijk}\right)$$

$$+\frac{s_{i0}s_{j0}m_0^{(3)}}{\left(m_0^{(2)}\right)^3} AsyCov\left(\sqrt{T}s_{k0}, \sqrt{T}s_{ijk}\right)$$

$$-\frac{3s_{i0}s_{j0}s_{k0}m_0^{(3)}}{\left(m_0^{(2)}\right)^4} AsyCov\left(\sqrt{T}m_0^{(2)}, \sqrt{T}s_{ijk}\right)$$

$$+\frac{s_{i0}s_{j0}s_{k0}}{\left(m_0^{(2)}\right)^3}AsyCov\left(\sqrt{T}m_0^{(3)},\sqrt{T}s_{ijk}\right)$$

The asymptotic covariance can then be estimated using:

$$AsyCov\left(\sqrt{T}s_{i0},\sqrt{T}s_{ijk}\right)=$$

$$\frac{1}{T}\sum_{t=1}^{T}\left[(R_{it}-m_i)(R_{Ft}-m_0)-s_{i0}\right]\left[(R_{it}-m_i)(R_{jt}-m_j)(R_{kt}-m_k)-s_{ijk}\right]$$

$$AsyCov\left(\sqrt{T}m_0^{(n)},\sqrt{T}s_{ijk}\right)=$$

$$\frac{1}{T}\sum_{t=1}^{T}\left[(R_{Ft}-m_0)^n-m_0^{(n)}\right]\left[(R_{it}-m_i)(R_{jt}-m_j)(R_{kt}-m_k)-s_{ijk}\right]$$

**Cokurtosis**

First, $\rho_{iiii}^{CC}$ and $\rho_{iiii}^{SF}$ are simple – they are the same as $\pi_{iiii}$.

If we use the constant correlation estimator as the structured model, $\rho_{ijkl}^{CC}$ can be estimated with:

$$\hat{\rho}_{ijjj}^{CC}=\frac{\hat{r}^{(3)}}{2}\left[\sqrt{\frac{m_j^{(6)}}{m_i^{(2)}}}AsyCov\left(\sqrt{T}m_i^{(2)},\sqrt{T}s_{ijjj}\right)\right.$$

$$\left.+\sqrt{\frac{m_i^{(2)}}{m_j^{(6)}}}AsyCov\left(\sqrt{T}m_j^{(6)},\sqrt{T}s_{ijjj}\right)\right]$$

$$\hat{\rho}_{iikk}^{CC}=\frac{\hat{r}^{(5)}}{2}\left[\sqrt{\frac{m_k^{(4)}}{m_i^{(4)}}}AsyCov\left(\sqrt{T}m_i^{(4)},\sqrt{T}s_{iikk}\right)\right.$$

$$\left.+\sqrt{\frac{m_i^{(4)}}{m_k^{(4)}}}AsyCov\left(\sqrt{T}m_k^{(4)},\sqrt{T}s_{iikk}\right)\right]$$

$$\hat{\rho}_{iikl}^{CC}=\left(\frac{\hat{r}^{(6)}}{2}\sqrt{\frac{\hat{r}^5}{m_i^{(4)}}\sqrt{m_k^{(4)}m_l^{(4)}}}\right)AsyCov\left(\sqrt{T}m_i^{(4)},\sqrt{T}s_{iikl}\right),$$

$$+\left(\frac{\hat{r}^{(6)}}{4}\sqrt{m_i^{(4)}\hat{r}^{(5)}\sqrt{\frac{m_l^{(4)}}{\left(m_k^{(4)}\right)^3}}}\right)AsyCov\left(\sqrt{T}m_k^{(4)},\sqrt{T}s_{iikl}\right)$$

$$+\left(\frac{\hat{r}^{(6)}}{4}\sqrt{m_i^{(4)}\hat{r}^{(5)}\sqrt{\frac{m_j^{(4)}}{\left(m_l^{(4)}\right)^3}}}\right)AsyCov\left(\sqrt{T}m_l^{(4)},\sqrt{T}s_{iikl}\right)$$

$$\hat{\rho}_{ijkl}^{CC}=\left(\frac{\hat{r}^{(7)}}{4}\sqrt{\hat{r}^{(5)}\sqrt{\frac{m_j^{(4)}}{\left(m_i^{(4)}\right)^3}}\hat{r}^{(5)}\sqrt{m_k^{(4)}m_l^{(4)}}}\right)AsyCov\left(\sqrt{T}m_i^{(4)},\sqrt{T}s_{ijkl}\right),$$

$$+\left(\frac{\hat{r}^{(7)}}{4}\sqrt{\hat{r}^{(5)}\sqrt{\frac{m_i^{(4)}}{\left(m_j^{(4)}\right)^3}}\hat{r}^{(5)}\sqrt{m_k^{(4)}m_l^{(4)}}}\right)AsyCov\left(\sqrt{T}m_j^{(4)},\sqrt{T}s_{ijkl}\right)$$

$$+\left(\frac{\hat{r}^{(7)}}{4}\sqrt{\hat{r}^{(5)}\sqrt{m_i^{(4)}m_j^{(4)}}\hat{r}^{(5)}\sqrt{\frac{m_l^{(4)}}{\left(m_k^{(4)}\right)^3}}}\right)AsyCov\left(\sqrt{T}m_k^{(4)},\sqrt{T}s_{ijkl}\right)$$

$$+\left(\frac{\hat{r}^{(7)}}{4}\sqrt{\hat{r}^{(5)}\sqrt{m_i^{(4)}m_j^{(4)}}\hat{r}^{(5)}\sqrt{\frac{m_k^{(4)}}{\left(m_l^{(4)}\right)^3}}}\right)AsyCov\left(\sqrt{T}m_l^{(4)},\sqrt{T}s_{ijkl}\right)$$

where,

$\hat{\rho}_{ijkl}^{CC}$ is the consistent estimated asymptotic cokurtosis between the sample and the structured estimator, and

$AsyCov(\bullet)$ refers to the asymptotic covariance.

The asymptotic covariance can then be estimated using:

$$AsyCov\left(\sqrt{T}m_i^{(n)},\sqrt{T}s_{ijkl}\right)=$$

$$\frac{1}{T}\sum_{t=1}^{T}\left[\left(R_{it}-m_i\right)^n-m_i^{(n)}\right]\left[\left(R_{it}-m_i\right)\left(R_{jt}-m_j\right)\left(R_{kt}-m_k\right)\left(R_{lt}-m_l\right)-s_{ijkl}\right]$$

If we use the single index estimator as the structured model, $\rho_{ijkl}^{SF}$ can be estimated with:

$$\hat{\rho}_{ijkl}^{SF}=\frac{s_{j0}s_{k0}s_{l0}m_0^{(4)}}{\left(m_0^{(2)}\right)^4}AsyCov\left(\sqrt{T}s_{i0},\sqrt{T}s_{ijkl}\right),$$

$$+\frac{s_{i0}s_{k0}s_{l0}m_0^{(4)}}{\left(m_0^{(2)}\right)^4}AsyCov\left(\sqrt{T}s_{k0},\sqrt{T}s_{ijkl}\right)$$

$$+\frac{s_{i0}s_{j0}s_{l0}m_0^{(4)}}{\left(m_0^{(2)}\right)^4}AsyCov\left(\sqrt{T}s_{k0},\sqrt{T}s_{ijkl}\right)$$

$$+\frac{s_{i0}s_{j0}s_{k0}m_0^{(4)}}{\left(m_0^{(2)}\right)^4}AsyCov\left(\sqrt{T}s_{l0},\sqrt{T}s_{ijkl}\right)$$

$$-\frac{4s_{i0}s_{j0}s_{k0}s_{l0}m_0^{(4)}}{\left(m_0^{(2)}\right)^5}AsyCov\left(\sqrt{T}m_0^{(2)},\sqrt{T}s_{ijkl}\right)$$

$$+\frac{s_{i0}s_{j0}s_{k0}s_{l0}}{\left(m_0^{(2)}\right)^4}AsyCov\left(\sqrt{T}m_0^{(4)},\sqrt{T}s_{ijkl}\right)+r_{ijkl}^{*}$$

where,

$$r^{*}_{iiil} = 3\left[\frac{s_{j0}\psi_{ii}}{m_0^{(2)}} AsyCov\left(\sqrt{T}s_{i0}, \sqrt{T}s_{iiil}\right)\right.$$

$$+\frac{s_{i0}\psi_{ii}}{m_0^{(2)}} AsyCov\left(\sqrt{T}s_{l0}, \sqrt{T}s_{iiil}\right)$$

$$-\frac{s_{i0}s_{j0}\psi_{ii}}{\left(m_0^{(2)}\right)^2} AsyCov\left(\sqrt{T}m_0^{(2)}, \sqrt{T}s_{iiil}\right)$$

$$\left.+\frac{s_{i0}s_{j0}}{m_0^{(2)}} AsyCov\left(\sqrt{T}\psi_{ii}, \sqrt{T}s_{ijjj}\right)\right]$$

$$r^{*}_{iikk} = \frac{2s_{i0}\psi_{kk}}{m_0^{(2)}} AsyCov\left(\sqrt{T}s_{i0}, \sqrt{T}s_{iikk}\right)$$

$$-\frac{s_{i0}^2\psi_{kk}}{m_0^{(2)}} AsyCov\left(\sqrt{T}\psi_{ii}, \sqrt{T}s_{iikk}\right)$$

$$+\frac{s_{i0}^2}{m_0^{(2)}} AsyCov\left(\sqrt{T}\psi_{ii}, \sqrt{T}s_{iikk}\right)$$

$$+\frac{2s_{k0}\psi_{ii}}{m_0^{(2)}} AsyCov\left(\sqrt{T}s_{k0}, \sqrt{T}s_{iikk}\right)$$

$$-\frac{s_{k0}^2\psi_{ii}}{\left(m_0^{(2)}\right)^2} AsyCov\left(\sqrt{T}m_0^{(2)}, \sqrt{T}s_{iikk}\right)$$

$$+\frac{s_{k0}^2\psi_{ii}}{m_0^{(2)}} AsyCov\left(\sqrt{T}\psi_{ii}, \sqrt{T}s_{iikk}\right)$$

$$+\psi_{kk} AsyCov\left(\sqrt{T}\psi_{ii}, \sqrt{T}s_{iikk}\right)$$

$$+\psi_{ii} AsyCov\left(\sqrt{T}\psi_{kk}, \sqrt{T}s_{iikk}\right)$$

$$r^{*}_{iikl} = \frac{s_{l0}\psi_{ii}}{m_0^{(2)}} AsyCov\left(\sqrt{T}s_{k0}, \sqrt{T}s_{iikl}\right)$$

$$+\frac{s_{k0}\psi_{ii}}{m_0^{(2)}} AsyCov\left(\sqrt{T}s_{l0}, \sqrt{T}s_{iikl}\right)$$

$$-\frac{s_{k0}s_{l0}\psi_{ii}}{\left(m_0^{(2)}\right)^2} AsyCov\left(\sqrt{T}m_0^{(2)}, \sqrt{T}s_{iikl}\right)$$

$$+\frac{s_{k0}s_{l0}}{m_0^{(2)}} AsyCov\left(\sqrt{T}\psi_{ii}, \sqrt{T}s_{iikl}\right)$$

$r_{ijkl}^* = 0$, where $i \neq j \neq k \neq l$

The asymptotic covariances in the above equations can be consistently estimated by:

$$AsyCov\left(\sqrt{T}s_{i0}, \sqrt{T}s_{ijkl}\right) = \frac{1}{T}\sum_{t=1}^{T}\left[(R_{it}-m_i)(R_{Ft}-m_0)-s_{i0}\right]$$

$$\times\left[(R_{it}-m_i)(R_{jt}-m_j)(R_{kt}-m_k)(R_{lt}-m_l)-s_{ijkl}\right]$$

$$AsyCov\left(\sqrt{T}m_0^{(n)}, \sqrt{T}s_{ijkl}\right) = \frac{1}{T}\sum_{t=1}^{T}\left[(R_{Ft}-m_0)^n - m_0^{(n)}\right]$$

$$\times\left[(R_{it}-m_i)(R_{jt}-m_j)(R_{kt}-m_k)(R_{lt}-m_l)-s_{ijkl}\right]$$

$$AsyCov\left(\sqrt{T}\psi_{ii}, \sqrt{T}s_{ijkl}\right) = \frac{1}{T}\sum_{t=1}^{T}\left[\varepsilon_{it}^2 - \hat{\psi}_{ii}\right]$$

$$\times\left[(R_{it}-m_i)(R_{jt}-m_j)(R_{kt}-m_k)(R_{lt}-m_l)-s_{ijkl}\right]$$

**Bayesian shrinkage intensity**

The Bayesian shrinkage intensities for the covariance, coskewness, and cokurtosis matrices can then be calculated as follows:

$$\hat{\alpha}_2^{CC} = \frac{1}{T}\frac{\hat{\pi}_2 - \hat{\rho}_2^{CC}}{\gamma_2^{CC}},\ \hat{\alpha}_3^{CC} = \frac{1}{T}\frac{\hat{\pi}_3 - \hat{\rho}_3^{CC}}{\gamma_3^{CC}},\ \hat{\alpha}_4^{CC} = \frac{1}{T}\frac{\hat{\pi}_4 - \hat{\rho}_4^{CC}}{\gamma_4^{CC}}$$

$$\hat{\alpha}_2^{SF} = \frac{1}{T}\frac{\hat{\pi}_2 - \hat{\rho}_2^{SF}}{\gamma_2^{SF}},\ \hat{\alpha}_3^{SF} = \frac{1}{T}\frac{\hat{\pi}_3 - \hat{\rho}_3^{SF}}{\gamma_3^{SF}},\ \hat{\alpha}_4^{SF} = \frac{1}{T}\frac{\hat{\pi}_4 - \hat{\rho}_4^{SF}}{\gamma_4^{SF}}$$

## Martellini and Ziemann [2010] model

Martellini and Ziemann [2010] empirically test their model using 100 randomly chosen baskets of US stocks, annually rebalanced from 1973 to 2006. They report their calculated Bayesian shrinkage intensities for the above six Bayesian shrinkage intensities. For comparison purpose, we use the parameters calculated in Martellini and Ziemann [2010] as two additional estimators.

# PGP optimization

After taking into account the higher moments, investors face multiple and conflicting goals. An ideal portfolio is designed to:

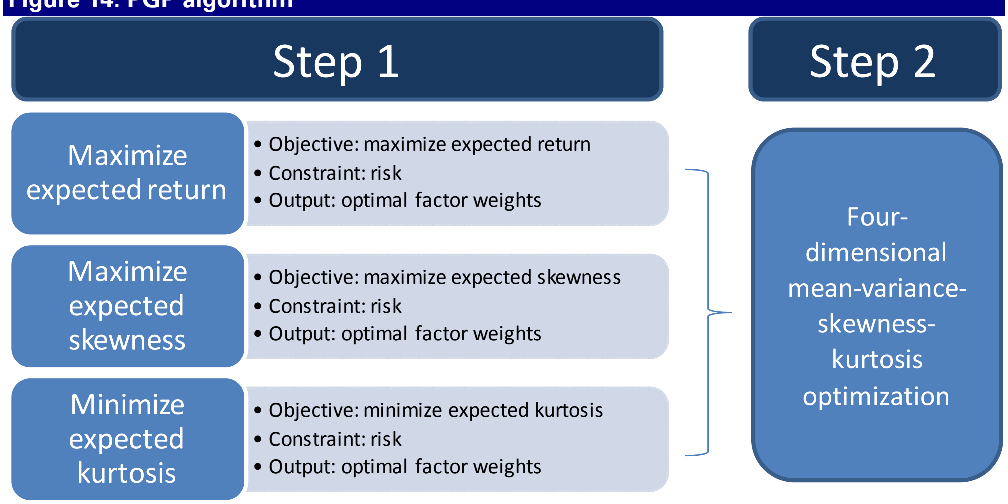
- Maximize expected return
- Minimize expected variance or volatility
- Maximize skewness
- Minimize kurtosis

The four goals are conflicting; therefore, we have to make some trade-offs. One way is to directly optimize a four-dimension utility function as described above. However, it is not only very difficult to compute (and very slow), but also there is no guarantee one can find the global optimal point.

One recently developed computational technique dealing with multiple conflicting goals is called polynomial goal programming or PGP (see Fabozzi, Kolm, Pachamanova, and Focardi [2007], Chunhachinda, Dandapani, Hamid, and Prakash [1997], Sun and Yan [2003], Prakash, Chang, Pactwa [2003], and Davies, Kat, and Lu [2004]). The basic idea behind PGP is simple – break the overall problem into smaller (and solvable) elements and then iteratively attempt to find solutions that preserve, as closely as possible, the individual (conflicting) goals.

The PGP algorithm can be implemented in a two-step approach (see Figure 14). In the first stage, we need to optimize the expected return, skewness, and kurtosis functions separately. Then, in the second step, we use the optimal values from the first step as the initial values for the second (four-dimensional) optimization.

**Figure 14: PGP algorithm**




*Source: Deutsche Bank Quantitative Strategy*

In the first stage, we need to optimize the expected return, skewness, and kurtosis functions separately, subject to the same set of constraints.

$$\underset{\omega}{Max}\, Z_1 = \left[\omega_t' \overline{R}_t\right]$$

$$\underset{\omega}{Max}\, Z_3 = [\omega_t' M_3 (\omega_t \otimes \omega_t)]$$

$$\underset{\omega}{Min}\, Z_4 = [\omega_t' M_4 (\omega_t \otimes \omega_t \otimes \omega_t)]$$

subject to

$$\omega_t' \Omega \omega_t \leq \xi$$

$$\omega_t^{LB} < \omega_t < \omega_t^{UB}$$

$$\sum_{i=1}^{N} \omega_{i,t} = 1$$

where, the covariance, coskewness, and cokurtosis matrices ($\Omega$, $M^3$, and $M^4$) can be estimated using either sample data, the two structural models, or the four Bayesian shrinkage estimators.

Once we solve the optimal values of $Z_1$, $Z_3$, and $Z_4$ (i.e., $Z_1^*$, $Z_3^*$, $Z_4^*$), we can substitute the optimal values into the following second step optimizer.

$$\underset{\omega}{Max}\, Z = (1+d_1)^{\alpha} + (1+d_3)^{\beta} + (1+d_4)^{\gamma}$$

Subject to

$$[\omega_t' \overline{R}_t] + d_1 = Z_1^*$$

$$[\omega_t' M_3 (\omega_t \otimes \omega_t)] + d_3 = Z_3^*$$

$$-[\omega_t' M_4 (\omega_t \otimes \omega_t \otimes \omega_t)] + d_4 = -Z_4^*$$

$$d_1, d_3, d_4 \geq 0$$

$$\omega_t' \Omega \omega_t \leq \xi$$

$$\omega_t^{LB} < \omega_t < \omega_t^{UB}$$

$$\sum_{i=1}^{N} \omega_{i,t} = 1$$

In the above specification, it ensures that it is monotonically increasing in $d_1$, $d_3$, and $d_4$ for all possible asset weights. In this setup, $\alpha$, $\beta$, and $\gamma$ reflect the nonnegative investor-specific preferences for the mean, skewness, and kurtosis, i.e., risk aversion parameters. These three parameters can then be mapped to the marginal rate of substitution (RMS). As shown in Davies, Kat, and Lu [2004], the MRS between expected return and skewness is given by:

$$MRS_{1,3} = \frac{\partial Z / \partial d_1}{\partial Z / \partial d_3} = \frac{\alpha(1+d_1)^{\alpha-1}}{\beta(1+d_3)^{\beta-1}}$$

Similarly, the MRS between expected return and kurtosis is expressed by:

$$MRS_{1,4} = \frac{\partial Z / \partial d_1}{\partial Z / \partial d_4} = \frac{\alpha(1+d_1)^{\alpha-1}}{\gamma(1+d_4)^{\gamma-1}}$$

In summary, the PGP algorithm involves two steps. In the first step, we solve for the optimal weights to maximize portfolio expected return, skewness, and kurtosis (subject to the same set of constraints) separately. Then, the optimal values of the above three objective functions ($Z_1^*$, $Z_3^*$, $Z_4^*$) are substituted into the overall function.

## Computational issues

The gain in computational speed is significant by using the PGP algorithm. A full scale mean-variance-skewness-kurtosis optimization requires a global optimizer with heuristic functionality. In our 12-factor example, it takes more than 12 hours for a single period optimization. More problematic, in many cases we encounter the no-solution problem. With the PGP algorithm, it takes about one second for each of the first-step optimizations and about 30 seconds for the second-step optimization; in total, it takes less than one minute for a single period optimization. More importantly, we have never experienced the no-solution problem with the PGP algorithm.

# Empirical backtesting

## Setting the scene

We somewhat arbitrarily pick 12 factors (see Figure 15) from the six broad style categories (value, growth, momentum/reversal, analyst sentiment, quality, and technicals). We use the same 12 factors as in two of our previous factor weighting research papers[12]. Five of the 12 factors show negative skewness and eight have excess kurtosis.

**Figure 15: 12 factors used in this research**

| # | Factor | Style | Direction* | Mean | Median | Maximum | Minimum | Std. Dev. | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Earnings yield, forecast FY1 mean | Value | Ascending | 4.1 | 3.5 | 27.4 | (13.5) | 6.7 | 0.3 | 3.3 |
| 2 | Price-to-book | Value | Ascending | 1.1 | 0.6 | 32.9 | (20.1) | 8.3 | 0.4 | 3.8 |
| 3 | IBES 5Y EPS growth | Growth | Ascending | 0.6 | 0.9 | 20.2 | (20.8) | 6.3 | (0.1) | 3.5 |
| 4 | Maximum daily return in last 1M (lottery factor) | Momentum/Reversal | Descending | 4.8 | 5.1 | 26.6 | (26.5) | 8.0 | (0.1) | 3.4 |
| 5 | 12M-1M total return | Momentum/Reversal | Ascending | 3.1 | 4.0 | 23.7 | (37.8) | 9.5 | (0.8) | 4.2 |
| 6 | IBES FY1 Mean EPS Revision, 3M | Analyst Sentiment | Ascending | 2.7 | 2.8 | 20.8 | (25.4) | 7.2 | (0.4) | 3.6 |
| 7 | Sales to total assets (asset turnover) | Quality | Ascending | 1.2 | 1.0 | 18.9 | (14.7) | 5.0 | 0.4 | 4.1 |
| 8 | Long-term debt/equity | Quality | Ascending | 0.6 | 0.8 | 17.7 | (15.1) | 4.9 | (0.0) | 3.6 |
| 9 | Net external financing/net operating assets | Quality | Ascending | 2.5 | 2.3 | 25.9 | (12.4) | 5.5 | 0.7 | 4.6 |
| 10 | Accruals (Sloan 1996 def) | Quality | Descending | 0.6 | 0.7 | 10.2 | (9.4) | 3.5 | (0.0) | 3.0 |
| 11 | CAPM idosyncratic vol, 1Y daily | Technical | Descending | 4.4 | 4.4 | 29.8 | (32.0) | 9.5 | (0.2) | 3.5 |
| 12 | Log float-adj capitalization | Technical | Ascending | 2.8 | 3.6 | 29.3 | (30.0) | 10.1 | (0.2) | 3.1 |

* Direction indicates how the factor scores are sorted. Ascending order means higher factor scores are likely to be associated with higher subsequent stock returns, a
vice versa for descending order.

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

## The model composition

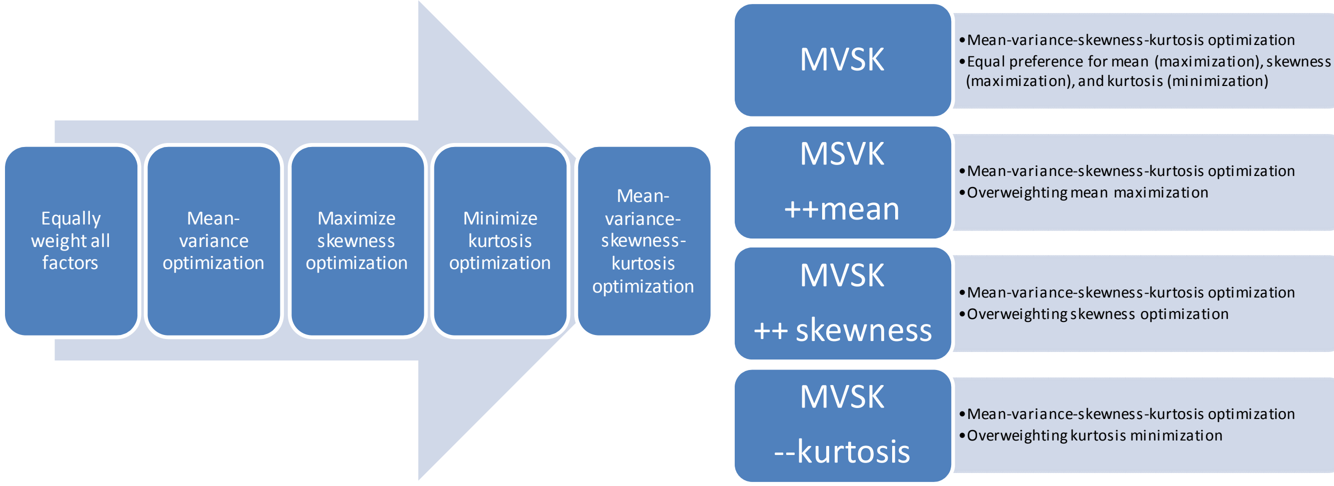
The overall benchmark model is to equally weight all factors (EQW). This is arguably the most robust model of all, as it makes no assumptions about expected factor return, risk, or distribution.

We further build 123 optimal signal weighting models along four dimensions (Figure 16):

- GK/QHS – mean variance optimized models
- Skewness maximization models
- Kurtosis minimization models
- MVSK (mean-variance-skewness-kurtosis) optimization models

[12] The be consistent, we use the same 12 factors as in two of our previous factor weighting research: 1) *Style Rotation*, which focuses on factor return prediction; and 2) *Robust Factor Models*, which is about factor risk estimation.

**Figure 16: The 124 model composition**




*Source: Deutsche Bank Quantitative Strategy*

**GK/QHS – mean-variance optimized models**

The next 39 models are based on the traditional mean-variance optimization, i.e., GK/QHS approach. These 39 models are based on three different return prediction scenarios (no return prediction, naïve return prediction, and sophisticated return prediction) and 13 factor risk models (sample covariance matrix, two structured models, four Bayesian shrinkage models, and six multivariate GARCH models). The details of the 39 models are covered in Luo, Cahan, Alvarez, Jussa, and Chen [2011]; therefore, we won't spend much time here.

**Skewness maximization models**

The third set of models is constructed to maximize expected skewness, following a similar philosophy to minimum variance portfolios. In this case, we try to maximize the expected skewness of our alpha model. For each of the two return prediction scenarios (naïve and sophisticated models), we match with the seven skewness estimation techniques (sample, constant correlation, single index, and the four Bayesian shrinkage approaches). Therefore, in total, we have tested 14 maximum skewness models.

**Kurtosis minimization models**

In a similar fashion as maximum skewness, we also build 14 kurtosis minimization models, again, along the two return prediction scenarios and seven kurtosis estimation techniques.

**MSVK (mean-variance-skewness-kurtosis) optimized models**

Lastly, we construct a set of alpha models to balance the four conflicting goals of mean maximization/risk minimization/skewness maximization/kurtosis minimization.

We understand that different investors have different preferences for mean, risk, skewness, and kurtosis. Therefore, we create three sets of utility functions with:

- Equal preference for mean, skewness, and kurtosis
- More emphasis to mean maximization
- More emphasis to skewness maximization
- More emphasis to kurtosis minimization

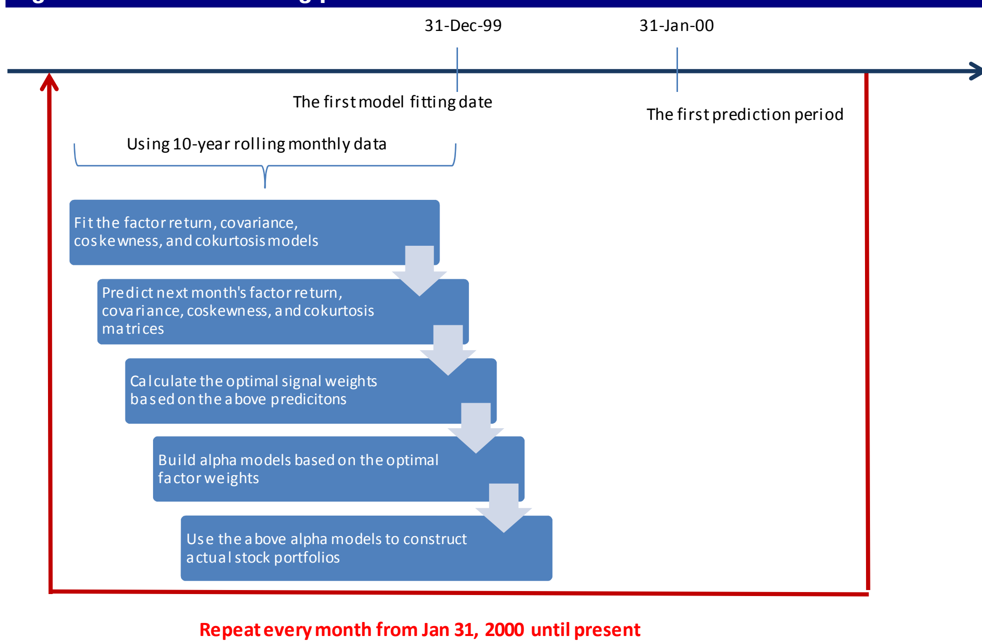
There are in total 56 alpha models that can be constructed using MVSK optimization, along three dimensions:

- Two return prediction scenarios: naïve prediction (using sample mean) and sophisticated prediction (using our style rotation model)
- Seven higher moment estimators: sample covariance/coskewness/cokurtosis, constant correlation, single index, and four Bayesian shrinkage models
- Four different utility functions: equal preference for mean/skewness/kurtosis; more for mean, more for skewness, and more for kurtosis

## The backtesting process

The pure out-of-sample backtesting is conducted from January 2000 to present (about 137 months). On December 31, 1999, we use 10-year rolling monthly data to fit the first set of models for future factor returns, risk, coskewness, and cokurtosis matrices. Then, we use this set of models to predict the factor returns and distributions for January 2000. Based on the predicted distribution statistics, we further calculate the optimal signal weights under different factor weighting models. The stock-selection models are constructed on the respective weights. Finally, we create (or rebalance) a series of stock portfolios using the respective alpha models. The same process is then repeated every month from January 2000 until present; therefore, all portfolios are formed monthly and rebalanced monthly.

**Figure 17: The backtesting process**

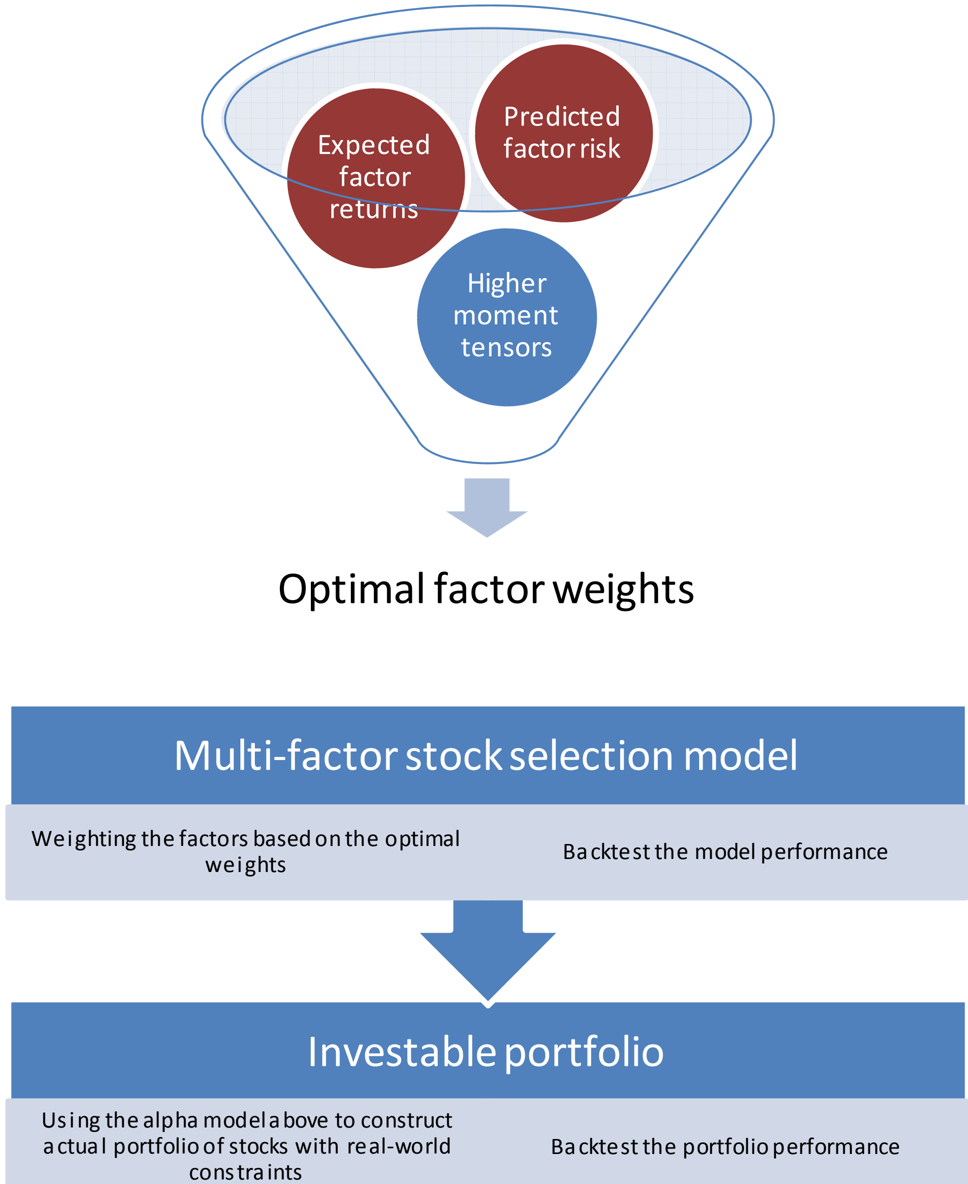



*Source: Deutsche Bank Quantitative Strategy*

## Model versus portfolio performance

In order to measure the effectiveness of these signal weighting schemes, we backtest the performance of: 1) our alpha models (using their rank information coefficient or their stock-selection ability); and 2) our stock portfolios (using Sharpe ratios). Figure 18 describes these two different performance measurement processes.

**Figure 18: Model performance versus portfolio performance**




*Source: Deutsche Bank Quantitative Strategy*

**Model performance**

For each one of the 124 factor weighting models, we further build a multi-factor model using the 12 style factors. The efficacy of the alpha models is measured by their respective rank information coefficient (and information ratio), for their stock selection ability.

**Portfolio performance**

The ultimate test, of course, is what multi-factor models have the highest portfolio IR in real life. For each multi-factor model, we build a long-short market neutral strategy, assuming all typical institutional constraints (e.g., beta neutral, size neutral, sector neutral, holding constraint, turnover constraint) and track the model performance after transaction costs.

## Performance measurement beyond mean-variance

The standard performance measures (e.g., IR and Sharpe ratio) are in the mean-variance framework. A few recently proposed methods try to mitigate and capture the entire distribution or more importantly, the tail distribution of the return series. For all of the 124 models, we can also calculate these unconventional performance measures. For simplicity, in the following discussion we focus on the Sortino ratio and conditional VaR (also called expected shortfall). The full details on other metrics are available upon request.

### Sortino ratio

Sortino and Price [1994] proposed an improvement on the Sharpe ratio to better account for skill and excess return by using only downside semivariance as the measure of risk.

### Upside Potential ratio

The upside potential ratio further improves the Sortino ratio using only upside on the numerator and only downside of the denominator (Plantinga, van der Meer, and Sortino [2001]).

### Omega

Keating and Shadwich [2002] proposed the Omega ratio as a way to capture all of the higher moments of the return distribution. It involves partitioning returns into loss and gain above and below a given threshold. The Omega ratio is then the ratio of the probability of having a gain to the probability of having a loss.

### VaR

VaR or value at risk is the industry standard for measuring downside risk. For a return series, VaR is defined as the high quantile (e.g., a 95% quantile) of the negative value of the returns. This quintile can be estimated using historical data (historical VaR) or assuming certain distribution (e.g., the Cornish-Fisher estimate or "modified" VaR).

### Conditional VaR

Conditional value at risk is also known as expected shortfall (or ES). Conditional VaR is defined as the expected value of the return series when the return is less than its c-quantile. Conditional VaR has more desirable properties than VaR. Similar to VaR, conditional VaR can also be estimated either using historical data (historical ES) or the Cornish-Fisher method (modified ES).

# Model performance

One of the common questions from clients is how we assess the incremental value from each of the optimal factor weighting components: 1) return prediction, 2) risk estimation, and 3) higher moment integration.

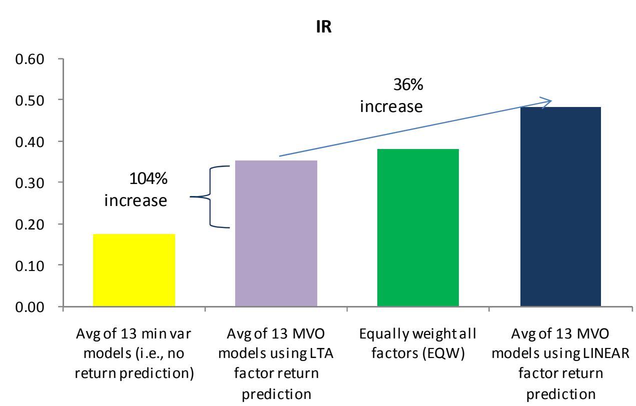
## The incremental value of better factor return predictions

It is not surprising that a better return prediction will lead to better performance. All strategies (using the 13 covariance estimators) using the more sophisticated LINEAR style forecasting model outperform their respective naïve LTA[13] counterparts. As shown in Figure 19 and Figure 20, on average, the LINEAR models outperform LTA models by 36% in model IR and reduce drawdown by 22%. Regardless of which covariance matrix we use, the LINEAR model outperforms the LTA model in every occasion.

If we measure the model performance using Sortino ratio (taking into account of downside risk), LINEAR factor timing models outperform naïve LTA models by 41%, while keeping conditional VaR down by 7% (see Figure 21 and Figure 22).

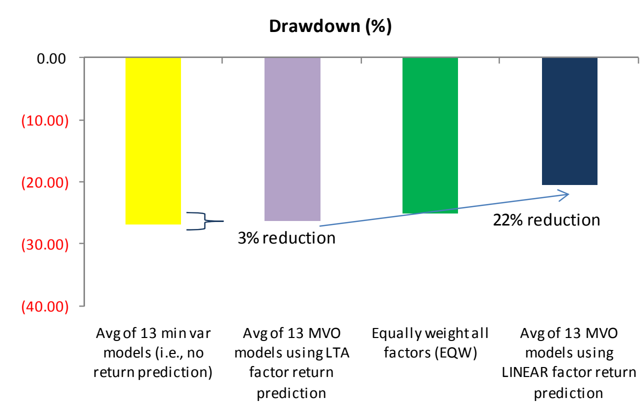
It is also interesting to note that the benchmark model, where we equally weight the 12 factors, is actually very robust, in line with those strategies using naïve factor return forecasting models (i.e., LTA).

**Figure 19: IR**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

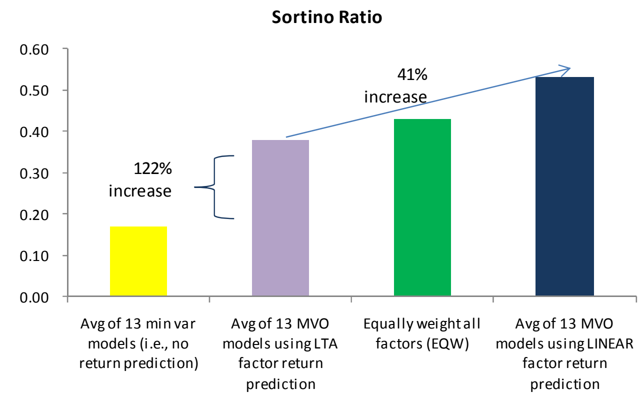
**Figure 20: Drawdown (%)**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

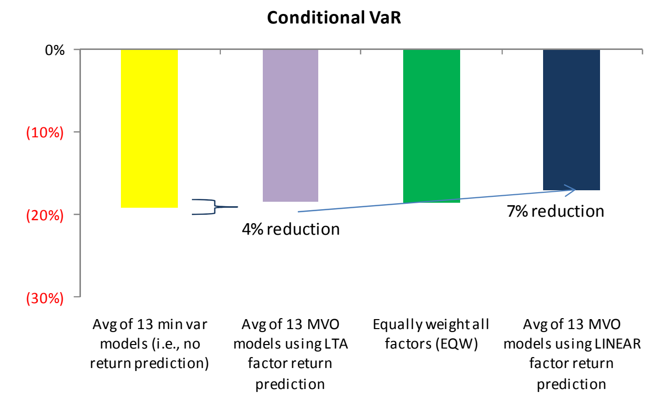
[13] In the LTA models, future factor returns are estimated using sample mean (i.e., the rolling average of 10-year monthly factor returns).

**Figure 21: Sortino ratio**




Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 22: Conditional VaR**

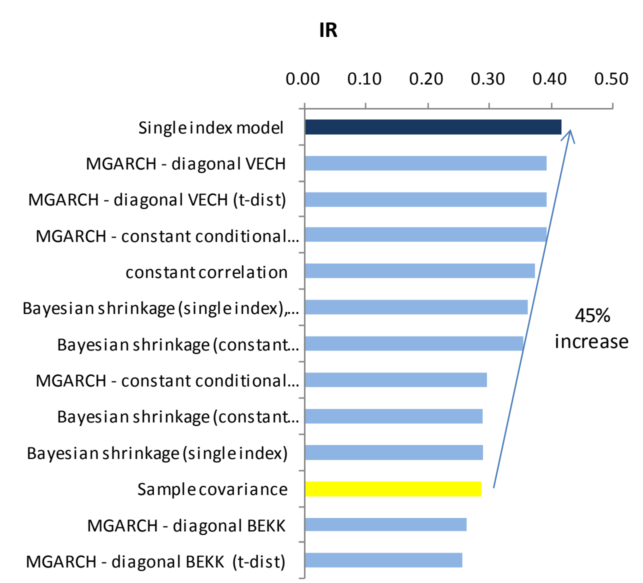



Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

## The incremental value of better risk estimations

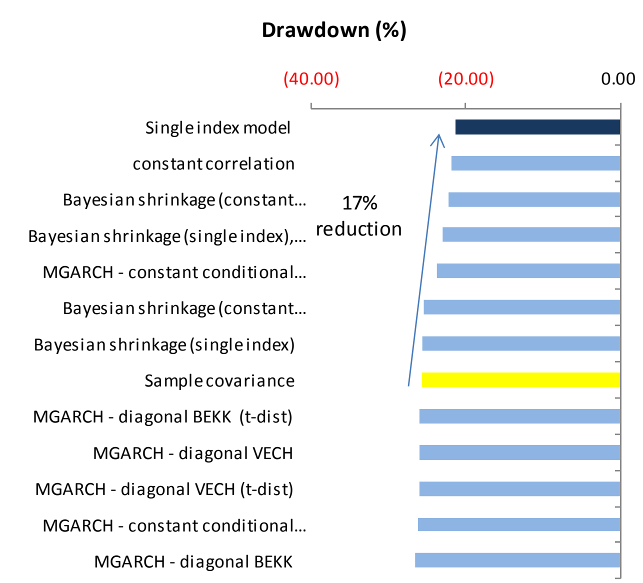
It is more interesting to note that an improvement in risk estimation can significantly lift model performance. Using the single index model to estimate the factor covariance matrix raises model IR by 45% (Figure 23) and reduces drawdown by 17% (Figure 24). Similarly, we find model Sortino ratio is improved by 53%, while conditional VaR is decreased by 13% (see Figure 25 and Figure 26).

**Figure 23: IR**



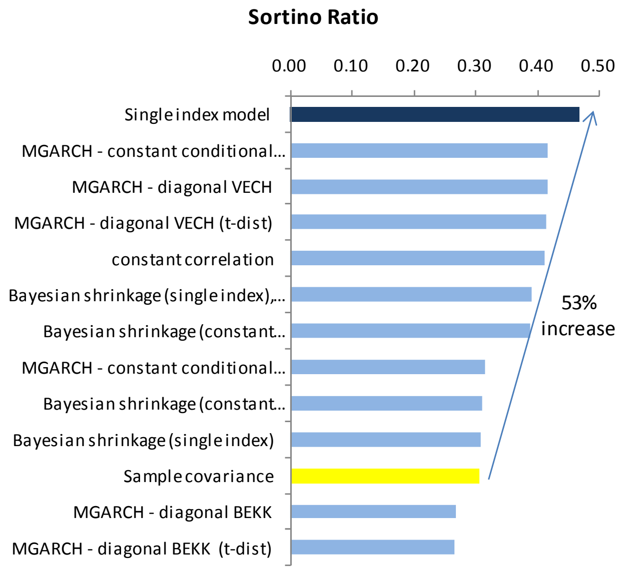

Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 24: Drawdown (%)**



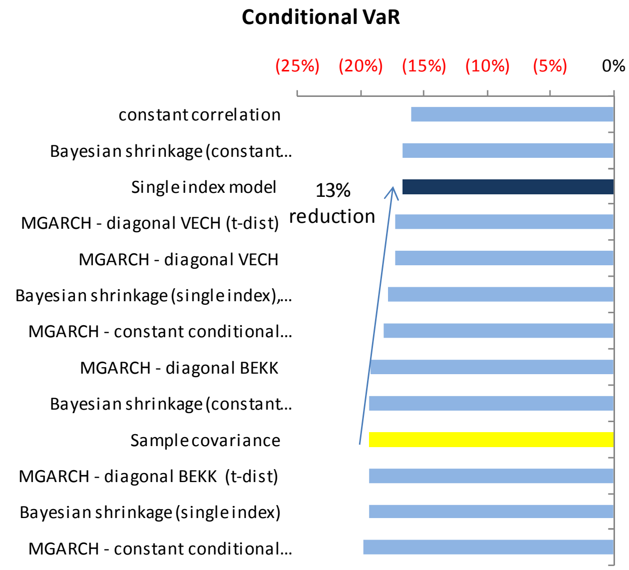

Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 25: Sortino ratio**




Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 26: Conditional VaR**




Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

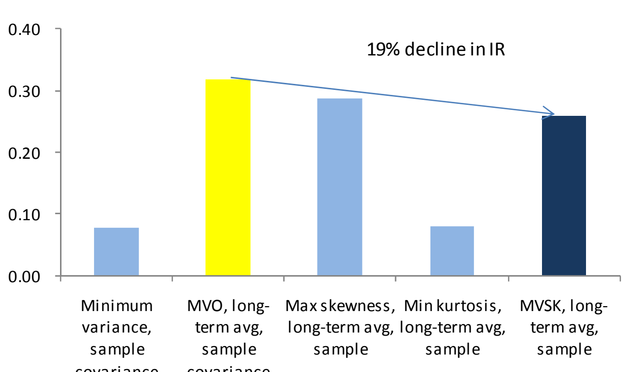
## The incremental benefit from tail risk management

Incorporating tail distribution in the optimal signal weighting decision is essentially shifting our focus from mean-variance to higher moments. In theory, MVSK-type of models should have lower downside risk, but the trade-off is that these models' IR may suffer moderately.

### Naïve factor return prediction + naïve higher moment estimation

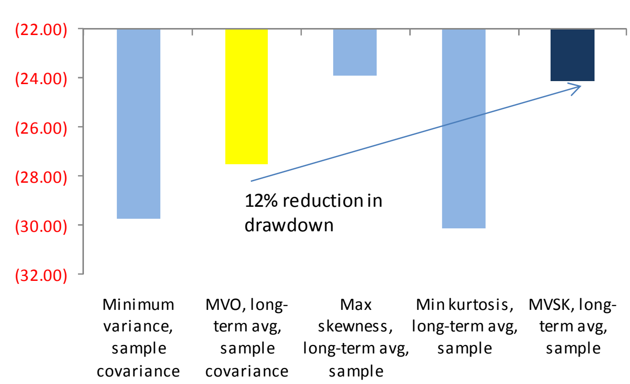
If we use sample mean, covariance, coskewness, and cokurtosis to optimize our factor weights, the MVSK model shows 12% reduction in drawdown – a moderate improvement (Figure 28). On the other hand, model's IR drops 19% - a substantial decline (Figure 27). The intuition is simple – estimation techniques based on sample data suffer estimation error, especially at the higher moments – as we shift our focus from mean-variance (MVO) to skewness, MVSK, and kurtosis, the performance declines monotonically (see Figure 27).

**Figure 27: IR**



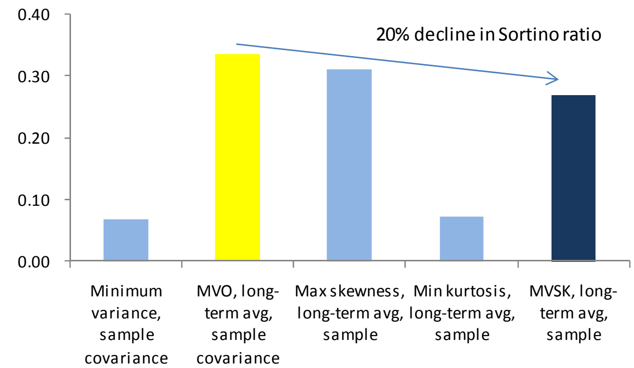

Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 28: Drawdown (%)**



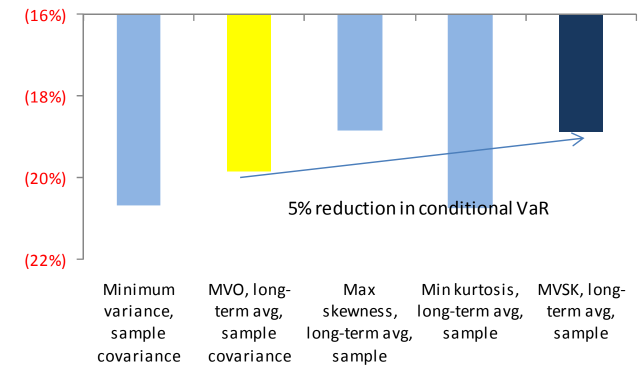

Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 29: Sortino ratio**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 30: Conditional VaR**

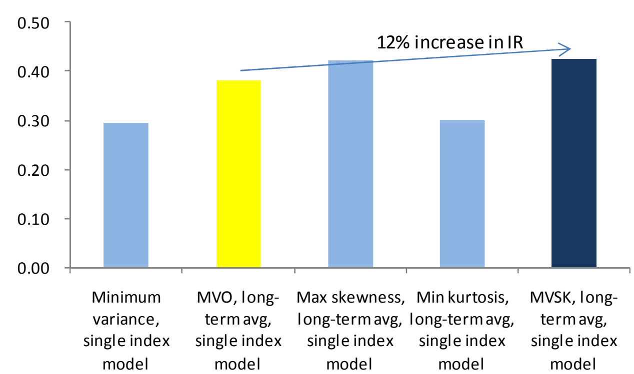



*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Naïve factor return prediction + sophisticated higher moment estimation**

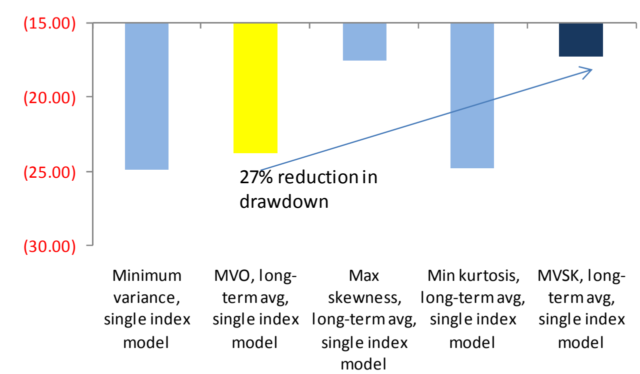
As we move to more sophisticated estimation techniques for higher moments (factor covariance, coskewness, and cokurtosis matrices), the MVSK model shows a significant reduction in downside risk – indeed, the worst month performance drops 27% (Figure 32). Surprisingly, the model's IR actually increases by 12% (Figure 31). The intuition is also simple – now we use sample mean to predict future factor returns, which is highly imprecise. On the other hand, we use our single index model to estimate factor covariance, coskewness, and cokurtosis matrices, which are more precise. Measured by Sortino ratio and conditional VaR, the results are even more encouraging – Sortino ratio increases by 32% while expected shortfall declines by 24% (see Figure 33 and Figure 34).

**Figure 31: IR**



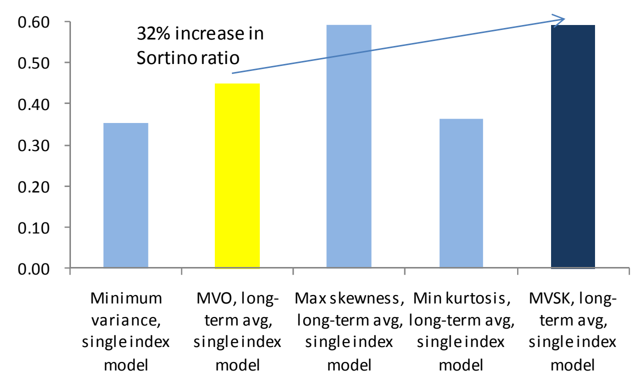

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 32: Drawdown (%)**



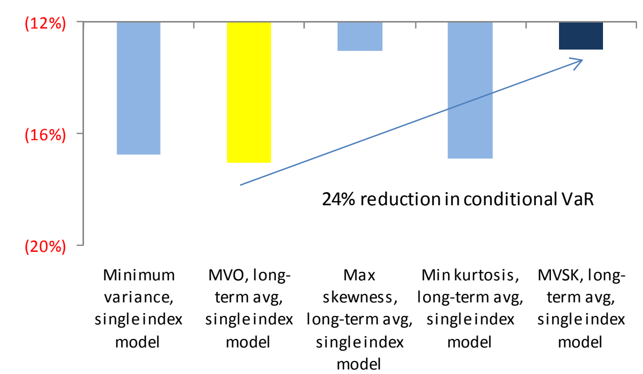

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 33: Sortino ratio**




Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 34: Conditional VaR**

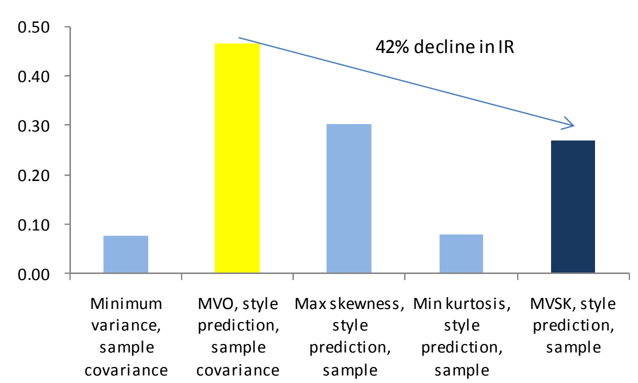



Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Sophisticated factor return prediction + naïve higher moment estimation**

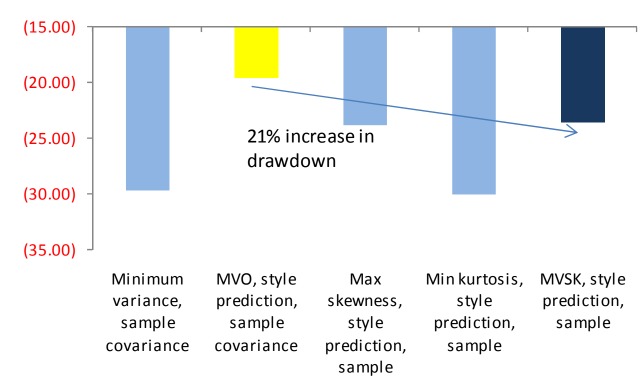
Now, we use our LINEAR model to predict future factor returns. As shown in our previous research, the LINEAR model has strong predictive power of forward factor performance. If we use naïve sample data to estimate all higher moments (i.e., factor covariance/coskewness/cokurtosis matrices), we would expect that we shift our weighting scheme from more reliable factors like expected returns to less reliable sources (i.e., higher moments). Therefore, our expectation is that model performance should suffer (in terms of IR), without much meaningful improvement in downside risk reduction. Our backtesting results seem to prove our hypothesis – model IR drops 42% (Figure 35), while drawdown actually increase by 21% (Figure 36).

**Figure 35: IR**



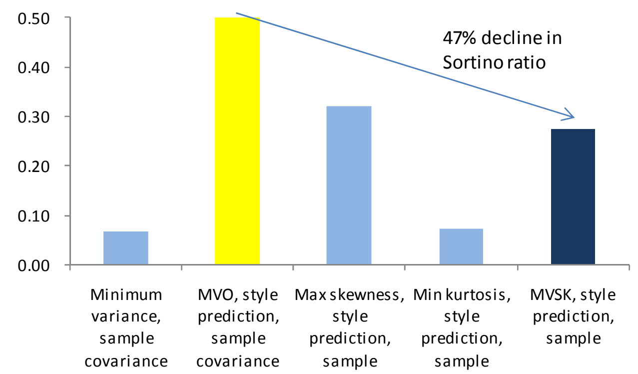

Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 36: Drawdown (%)**



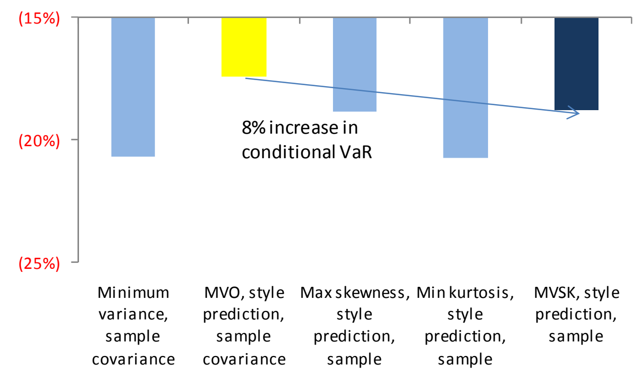

Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 37: Sortino ratio**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 38: Conditional VaR**



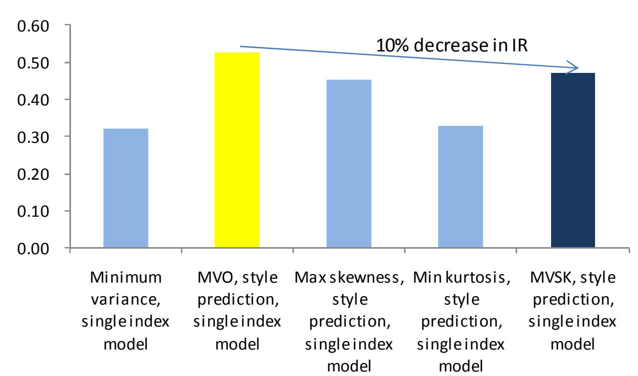

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Sophisticated factor return prediction + sophisticated higher moment estimation**

If we use our LINEAR model to predict future factor returns, while at the same time using single index model to estimate factor covariance, coskewness, and cokurtosis matrices, we would expect a reduction in our model's downside risk. Depending on how good we are at estimating factor returns, the MVSK model may have higher or lower IR. It turns out that our MSVK model has successfully reduced downside risk by 24% (Figure 40), but the model IR also drops 10% (Figure 39).

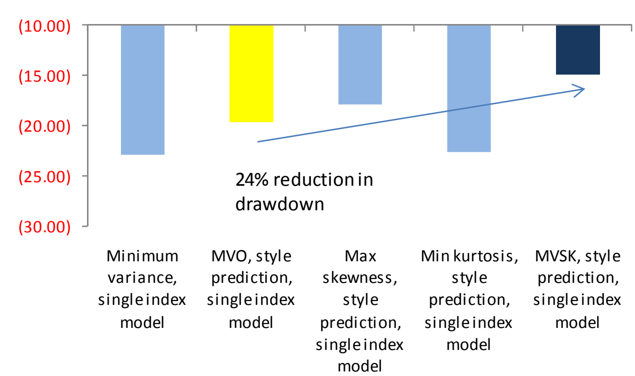
To prove our hypothesis that the decline in IR is due to our shift of weights from mean-variance to skewness/kurtosis, we re-run our optimization but giving higher weights to mean – we call the new model MVSK (++mean). Interestingly, the new model does appear to increase IR moderately.

**Figure 39: IR**



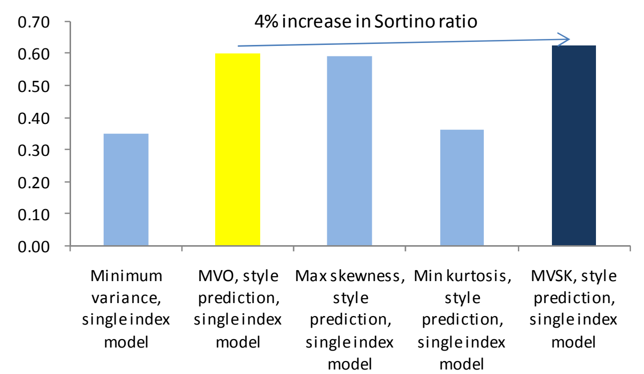

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 40: Drawdown (%)**



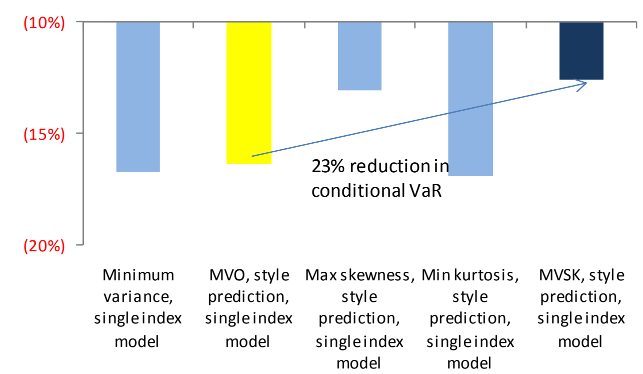

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 41: Sortino ratio**




Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 42: Conditional VaR**




Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

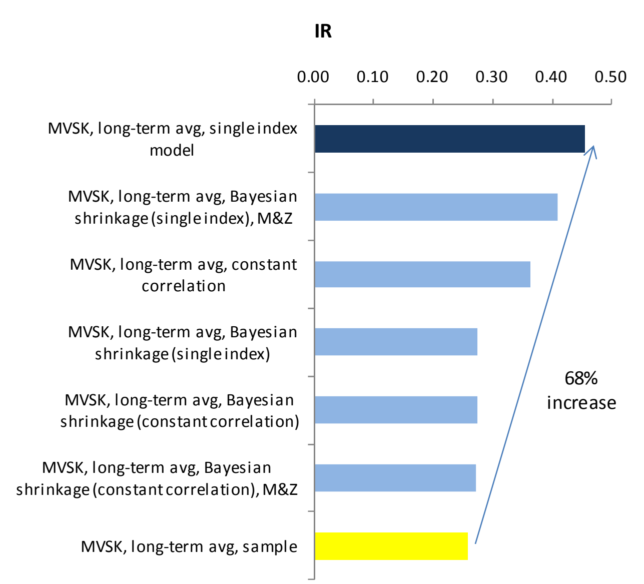
## The impact of different estimation techniques

As shown in previous sections, it appears that structured models for factor covariance/coskewness/cokurtosis are very important. In this section, we compare the various moment estimation techniques.

**Naive factor return prediction/MVSK optimization**

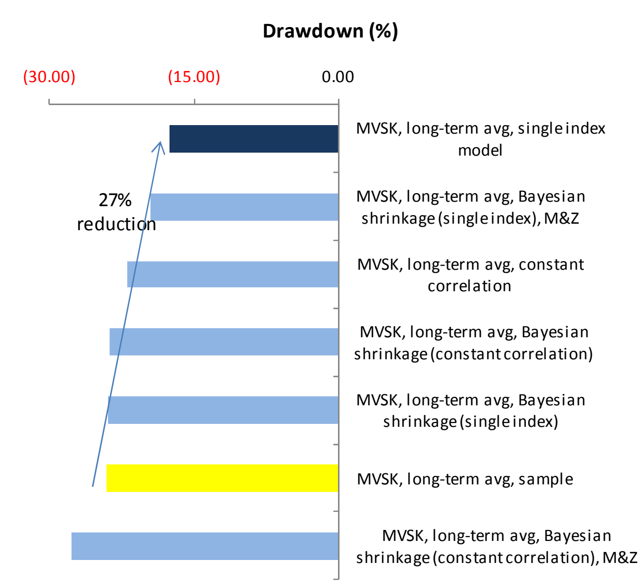
For managers who do not follow factor timing strategies, using past performance (i.e., sample mean) as the prediction of future factor returns is probably a reasonable assumption. Under this scenario, structured models for higher moments can significantly increase model IR/Sortino ratio and reduce downside risk (see Figure 43 to Figure 46).

**Figure 43: IR**



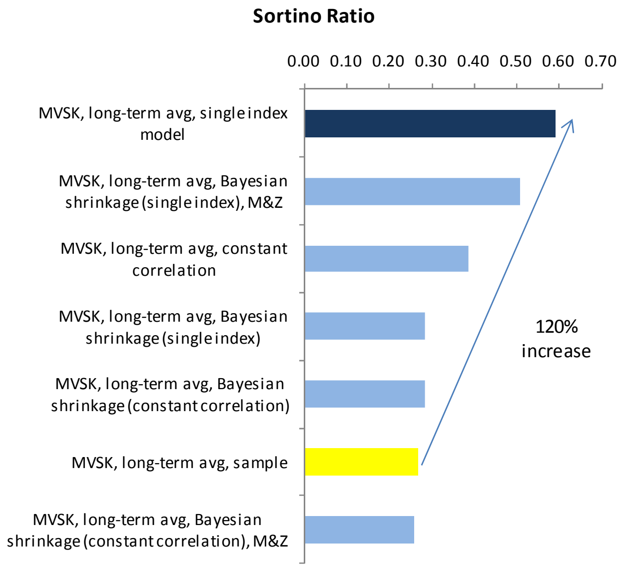

Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 44: Drawdown (%)**



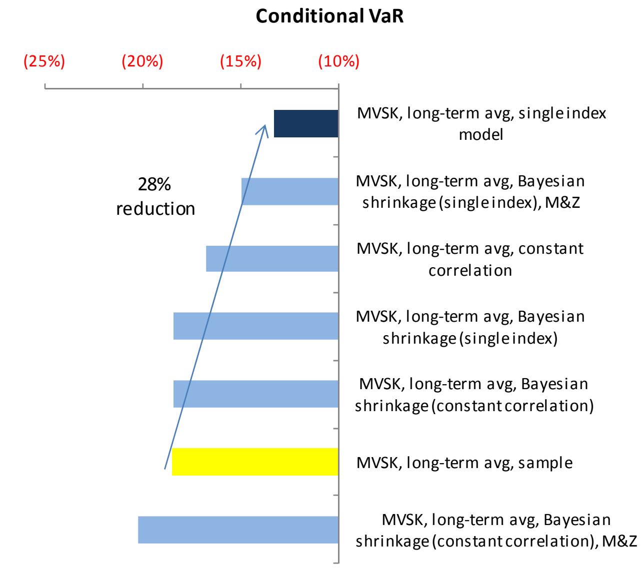

Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy

**Figure 45: Sortino ratio**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 46: Conditional VaR**

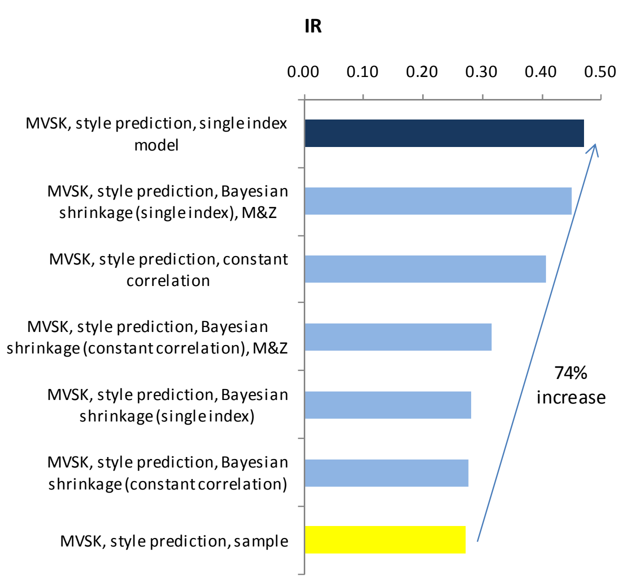



*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

### Sophisticated factor return prediction/MVSK optimization

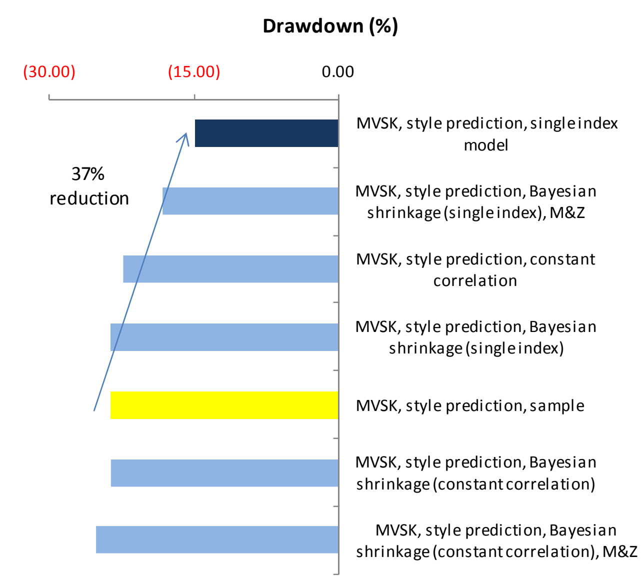
Even when we have a good factor timing model to predict factor returns, we still find a better structured model for the higher moments very useful. Indeed, MVSK strategies based on the single index model significantly outperformed those strategies based on sample data alone (see Figure 47 to Figure 50).

**Figure 47: IR**



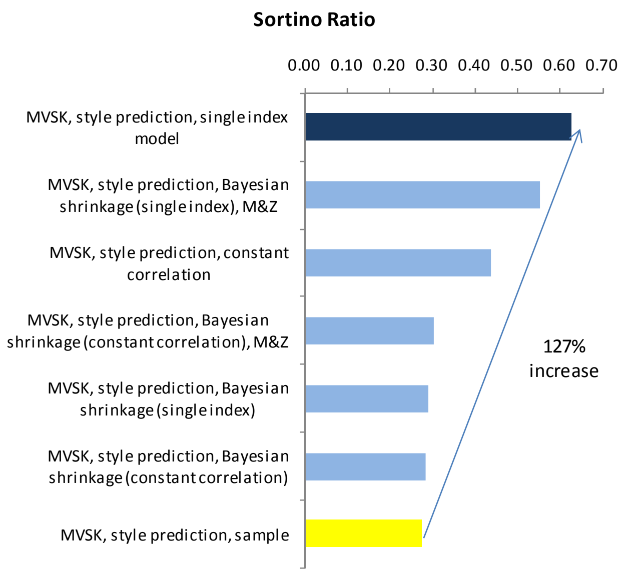

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 48: Drawdown (%)**



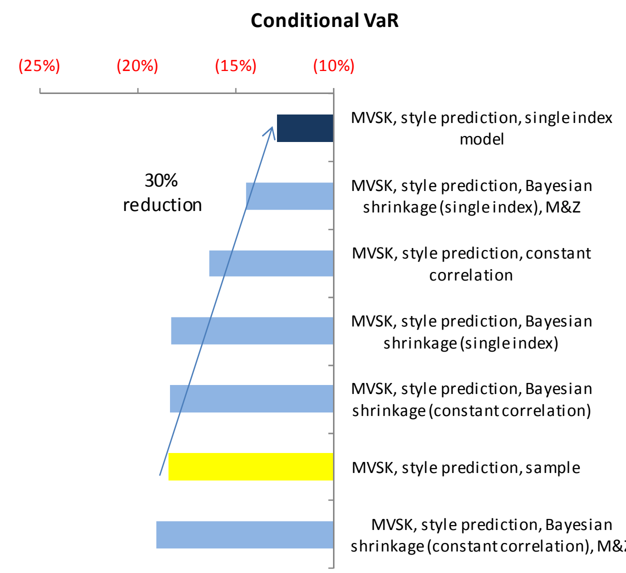

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 49: Sortino ratio**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 50: Conditional VaR**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

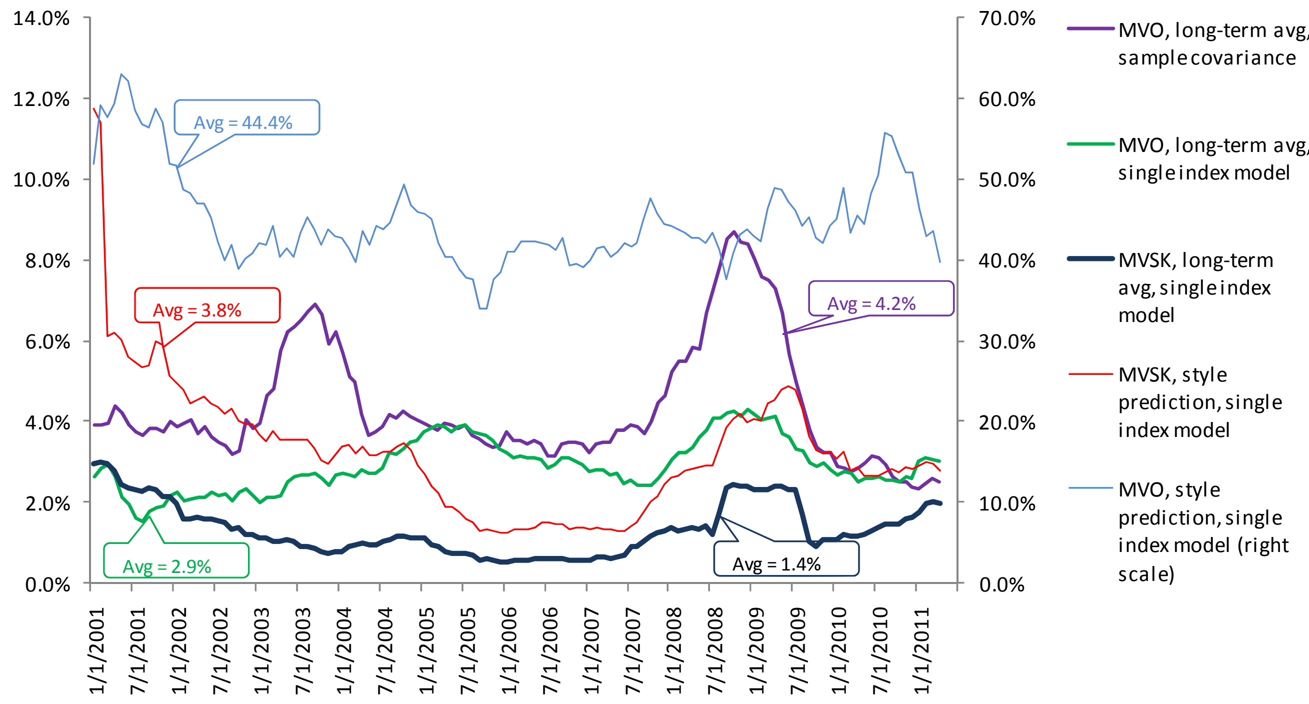
## Model turnover

It is also important to understand the impact of more sophisticated factor covariance/coskewness/cokurtosis estimators on factor turnover, because factor turnover eventually will translate into portfolio turnover.

As shown in Figure 51, it is interesting to note that the MVO signal weighting model with factor timing/single-index covariance matrix has the highest factor turnover (average = 44.4% per month), while MSVK model with factor timing effectively reduces the turnover to 3.8% – very much in line with those models without factor timing (e.g., MVO with LTA model/sample covariance of 4.2%, MVO with LTA/single-index covariance of 2.9%, or MVSK model with LTA of 1.4%). In almost all cases, MVSK models have significantly lowered factor turnover.

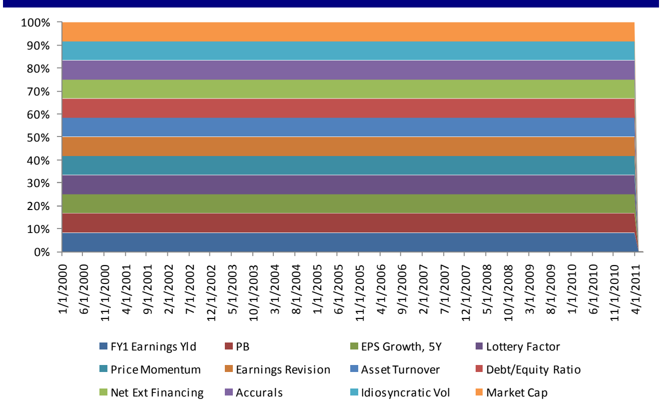
Figure 52 to Figure 57 show the time series of factor weights for the six comparable models. It appears that the two MVSK models have more stable factor weighting profiles, likely due to the reduction in estimation errors.

**Figure 51: Factor turnover**



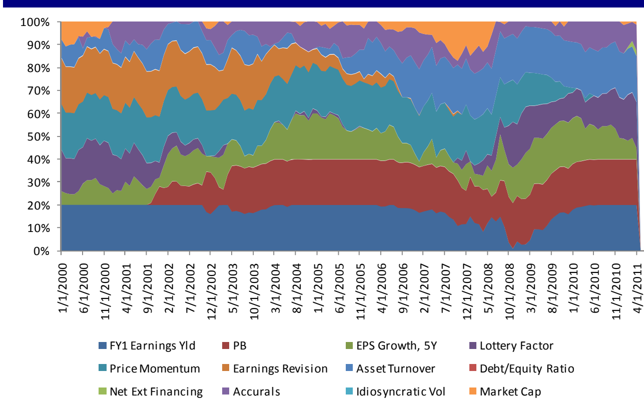

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 52: Equally weight all 12 factors**



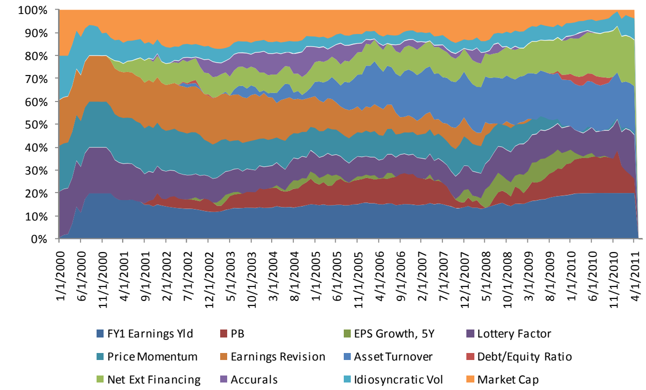

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 53: MVO with naïve return prediction/sample covariance**



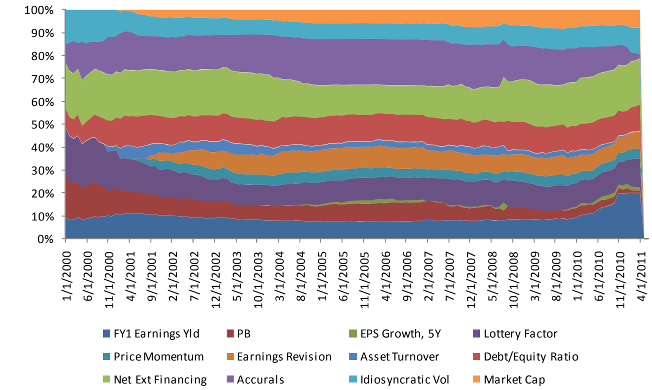

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 54: MVO with naïve return prediction/single-index covariance matrix**



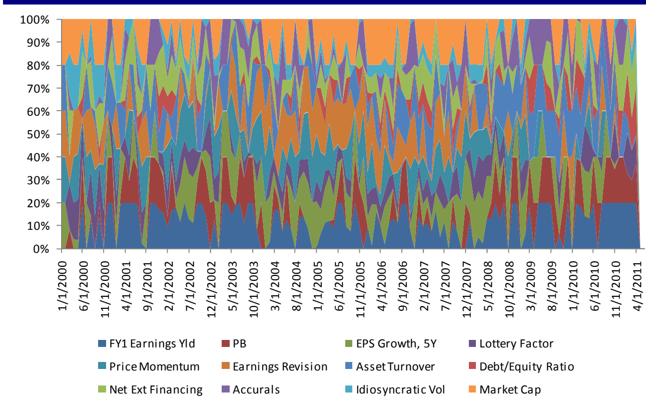

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 55: MVSK with naïve return prediction/single-index covariance, coskewness, cokurtosis matrices**



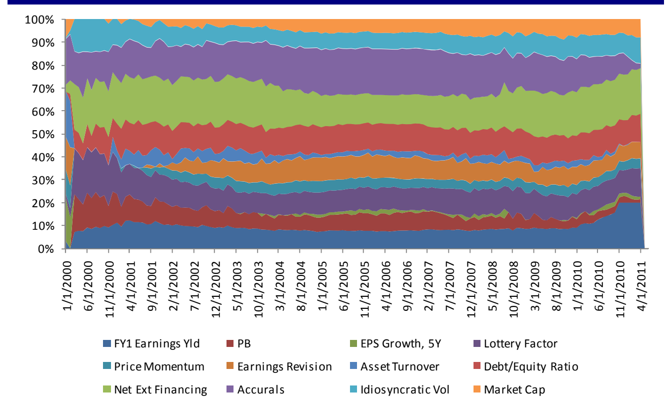

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 56: MVO with sophisticated return prediction/single-index covariance matrix**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 57: MVSK with sophisticated return prediction/single-index covariance, coskewness, cokurtosis matrices**

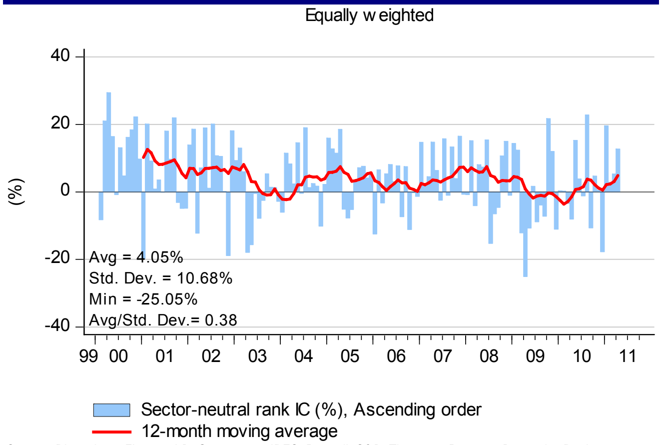



*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

## Time series performance

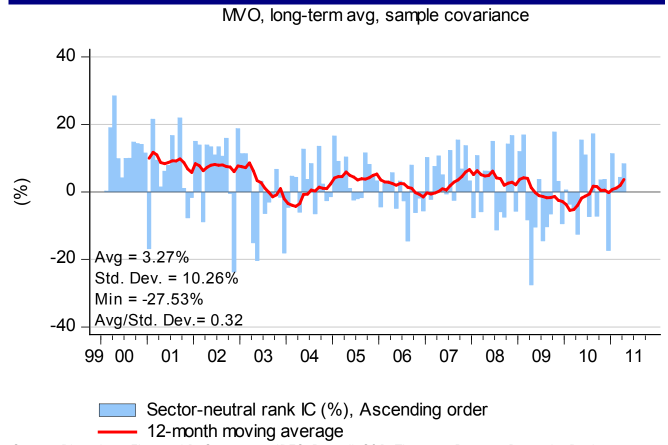
In order to show the performance of each model over time, from Figure 58 to Figure 63, we show the time series rank IC for the six comparable models. The benefits of the MVSK model can be seen even more clearly in the time series graphs.

**Figure 58: Equally weight all 12 factors**



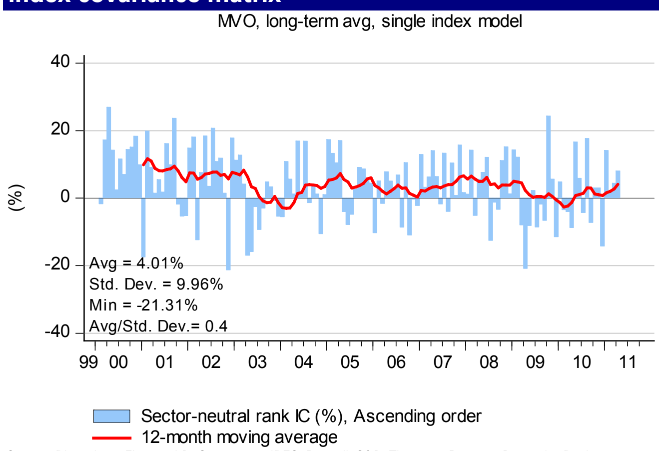

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 59: MVO with naïve return prediction/sample covariance matrix**



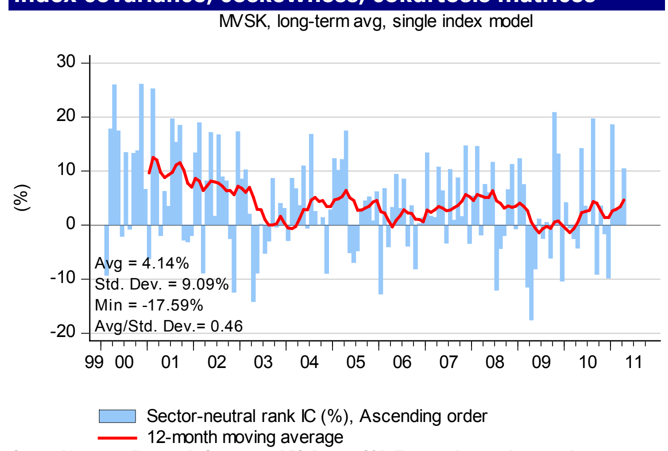

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 60: MVO with naïve return prediction/single-index covariance matrix**



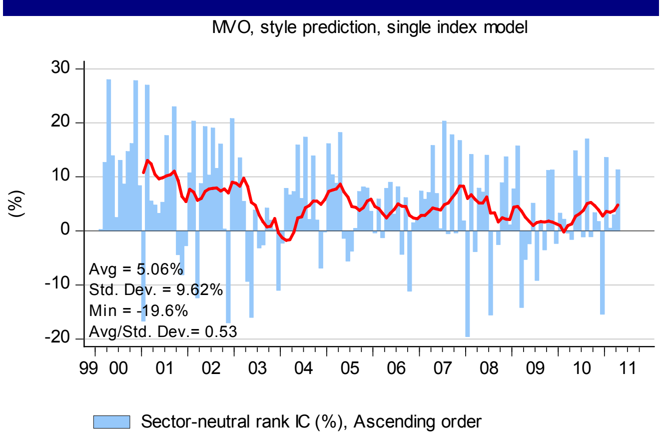

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 61: MVSK with naïve return prediction/single-index covariance, coskewness, cokurtosis matrices**



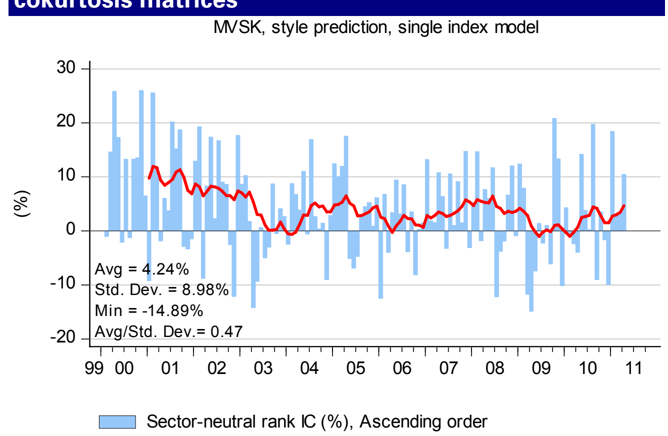

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 62: MVO with sophisticated return prediction/single-index covariance matrix**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 63: MVSK with sophisticated return prediction/single-index covariance, coskewness, cokurtosis matrices**

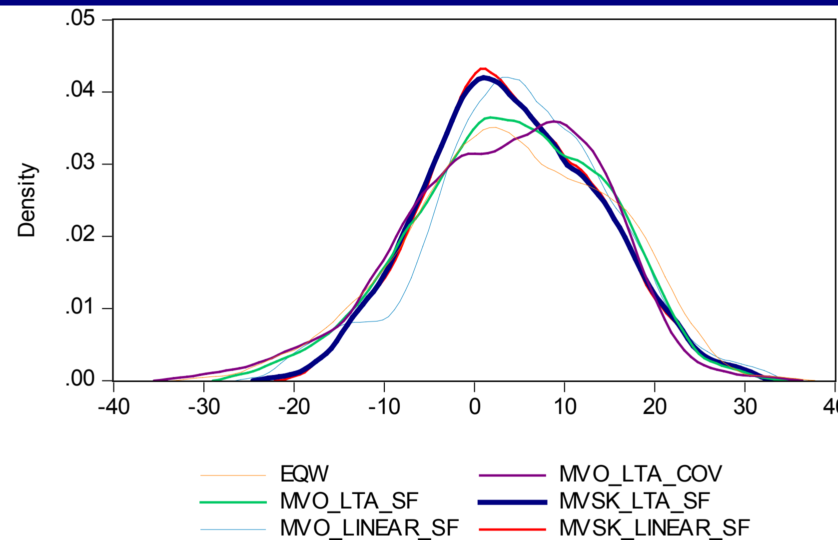



*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

## Model return distribution

All the above statistics try to summarize the overall distribution in a more concise way, but the best way is probably still to look at the empirical distribution. As shown in Figure 64, the two MVSK models clearly have much lower left tails than the other four models. The two most traditional models, EWQ (equally weight all factors) and MVO with naïve return prediction/sample covariance matrix have the largest downside risks.

**Figure 64: Model return distribution**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

## An example – Spring 2009 "junk" rally

In order to demonstrate the benefit of tail risk management, let's use the Spring 2009 "junk" rally as an example. In the Spring of 2009 (from March to May 2009), investors' risk appetite suddenly improved, in a sign that we had reached the bottom of the financial crisis. As a result, investors shifted their focus toward low quality/high risk stocks. Most of the traditional quant factors underperformed significantly, as most of them tend to bet on higher quality/lower risk stocks.

As shown in Figure 65, many risk-related (e.g., the lottery factor, idiosyncratic vol, and market capitalization), price momentum, earnings revision, and quality (e.g., debt/equity ratio[14]) factors had a terrible month of performance. The MVO model using naïve factor return prediction and sample covariance matrix overweighted the lottery factor and price momentum; and therefore had a miserable performance of -27.5%.

The MVO model with factor timing put less weights toward the lottery factor and price momentum, while adding more weight toward net external financing. Our LINEAR style timing model worked very well in April 2009 by predicting a deep negative value for the price momentum, earnings revision, and size factors.

Interestingly, the MVSK model without style rotation achieved the same goal of reducing weight toward the lottery factor/price momentum and adding weight toward net external financing – at the same time, it also successfully allocated more weight toward debt/equity

[14] Please note that the long-term performance suggests that companies with higher financial leverage (i.e., higher debt/equity ratio) are likely to produce higher return on equity; therefore, are more likely to outperform. In our debt/equity ratio factor, we typically bet on stocks with higher leverage. Therefore, in April 2009, the very strong performance of debt/equity ratio indicates that higher leverage (i.e., lower quality) companies have outperformed.

ratio and accruals factors (see Figure 66). The benefit of this model is that it does not rely on factor return prediction or factor timing. Factor weights are fully determined by factor risk and distribution, which are perceived as easier to estimate (than factor returns) without controversy.

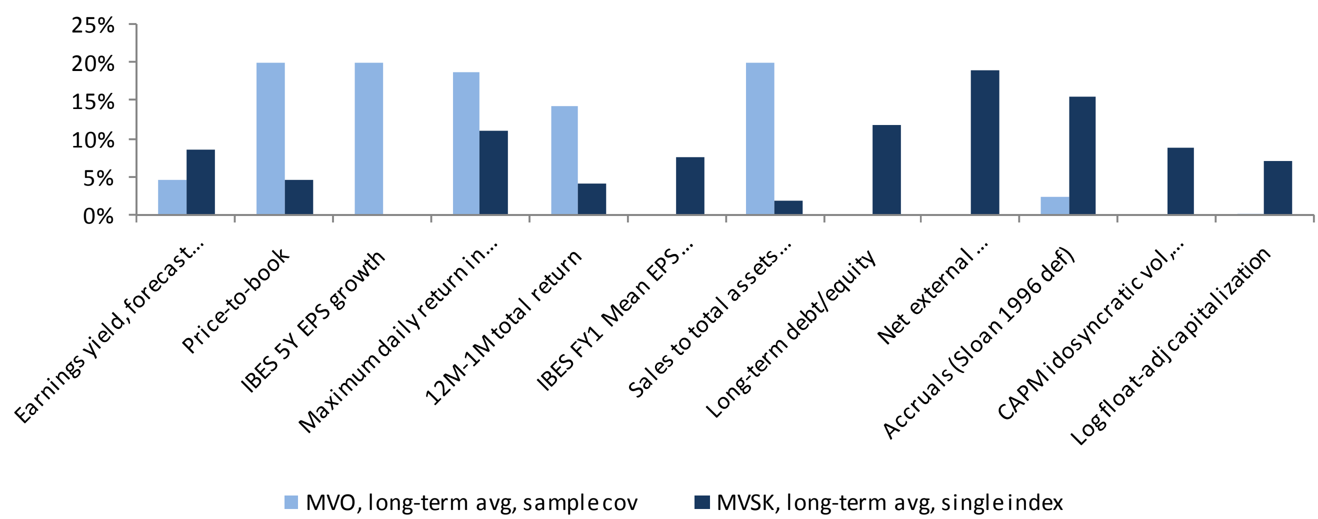
**Figure 65: April 2009 with a magnifier**

| | Earnings Yield | PB | EPS Growth 5Y | Lottery Factor | Price Momentum | Earnings Revision | Asset Turnover | Debt/Equity Ratio | Net Ext Financing | Accurals | Idiosyncratic Vol | Market Cap | 2009M4 Rank IC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Factor performance statistics*** | | | | | | | | | | | | | |
| Factor performance, rank IC (%) | 0.4 | 31.3 | (0.8) | (26.5) | (37.8) | (25.4) | 1.7 | 17.7 | 2.3 | 5.7 | (25.0) | (16.6) | (6.1) |
| Long-term average rank IC (%) | 4.1 | 1.1 | 0.6 | 4.8 | 3.1 | 2.7 | 1.2 | 0.6 | 2.5 | 0.6 | 4.4 | 2.8 | |
| Skewness | 0.3 | 0.4 | (0.1) | (0.1) | (0.8) | (0.4) | 0.4 | (0.0) | 0.7 | (0.0) | (0.2) | (0.2) | |
| Kurtosis | 3.3 | 3.8 | 3.5 | 3.4 | 4.2 | 3.6 | 4.1 | 3.6 | 4.6 | 3.0 | 3.5 | 3.1 | |
| Direction* | Asc | Asc | Asc | Des | Asc | Asc | Asc | Asc | Asc | Des | Des | Asc | |
| Predicted by the LINEAR model | 14.1 | 22.5 | (3.5) | (0.0) | (18.1) | (17.7) | 10.4 | 9.4 | 11.3 | 4.1 | 1.4 | (11.3) | |
| ***Factor weights for April 2009*** | | | | | | | | | | | | | |
| Equally weighted | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | (25.0) |
| MVO, long-term avg, sample cov | 5% | 20% | 20% | 19% | 14% | 0% | 20% | 0% | 0% | 2% | 0% | 0% | (27.5) |
| MVO, long-term avg, single index | 16% | 6% | 8% | 13% | 8% | 1% | 20% | 0% | 13% | 0% | 7% | 8% | (20.9) |
| MVSK, long-term avg, single index | 8% | 5% | 0% | 11% | 4% | 8% | 2% | 12% | 19% | 15% | 9% | 7% | (17.6) |
| MVO, style prediction, single index | 18% | 20% | 0% | 0% | 0% | 0% | 20% | 16% | 19% | 8% | 0% | 0% | (5.3) |
| MVSK, style prediction, single index | 9% | 6% | 0% | 11% | 3% | 4% | 3% | 13% | 20% | 17% | 9% | 6% | (14.9) |

* Direction indicates how the factor scores are sorted. Ascending order means higher factor scores are likely to be associated w ith higher subsequent stock returns, and vice versa for descending order.

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 66: Factor weight – MVO versus MVSK**



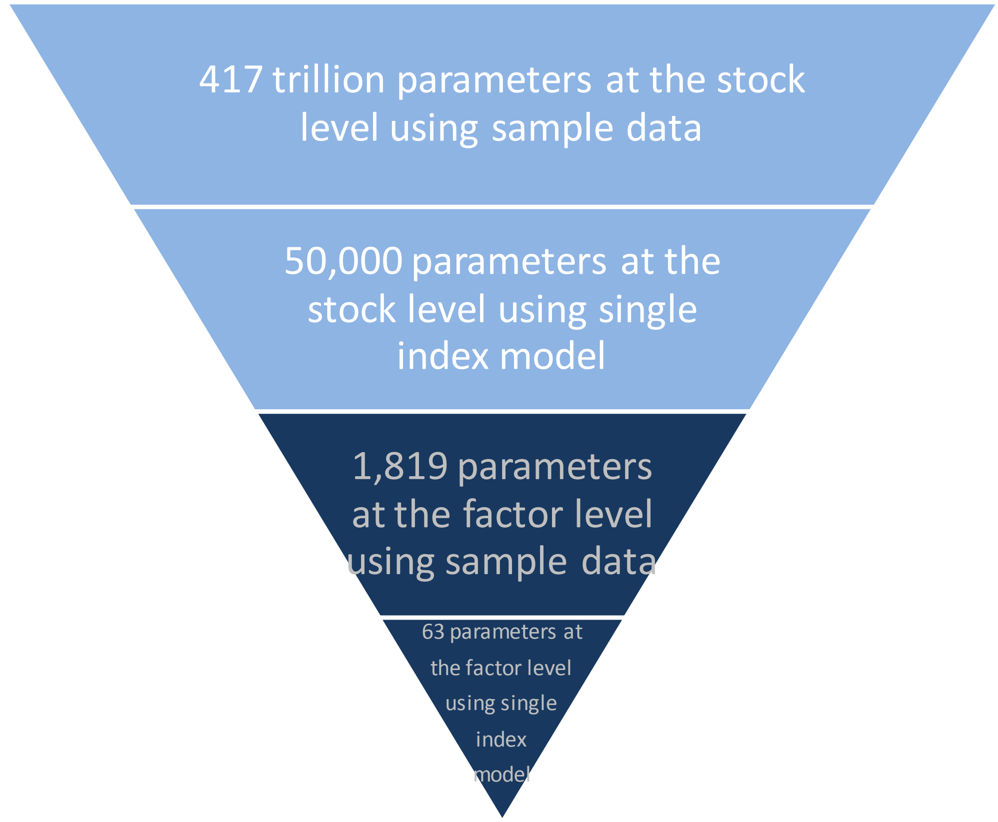

*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

## The benefits of the single index model

The single index model for factor covariance, coskewness, and cokurtosis matrices stands out in almost every case. There are five major benefits of using the single index model.

- It is intuitive, by shrinking factor risk and distribution toward a single "hyper" index.
- It is relatively easy to implement.
- It incorporates macro information in the factor weighting decision process.
- It substantially reduces the parameter dimensionality (see Figure 67).
- It largely removes the need to predict factor returns, which remains one of the most controversial topics among quantitative investment managers.

**Figure 67: Reduced dimensionality**

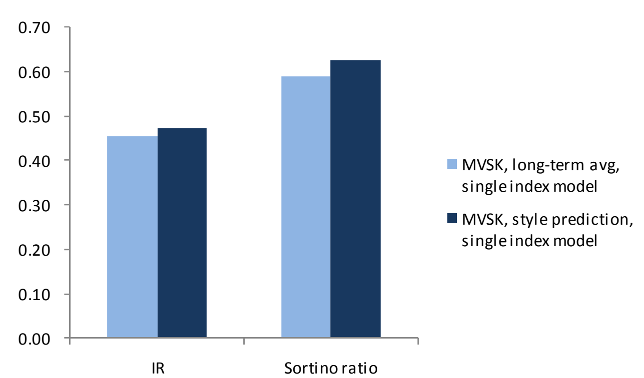



*Source: Deutsche Bank Quantitative Strategy*

### Style timing? Why even bother?

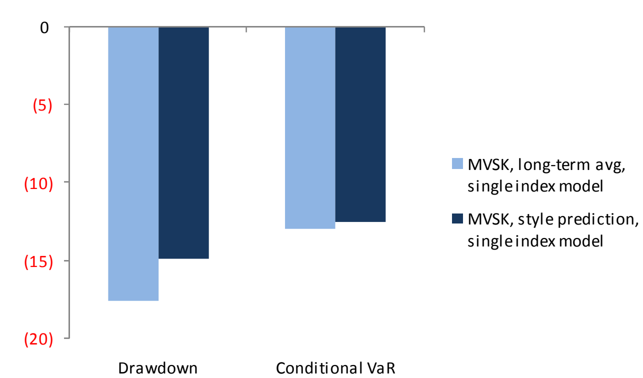
When using the single index model for factor higher moments (covariance, coskewness, and cokurtosis), factor timing becomes much less important. As shown in Figure 68, the MVSK model using naïve factor return prediction has almost the same IR as the model using our LINEAR style timing model (0.46 versus 0.47), while the Sortino ratio is only marginally lower (0.59 versus 0.62). On the tail risk side, the MSVK model with style timing only reduces downside risk from 17.6% to 14.9% and conditional VaR from 13.0% to 12.6% (see Figure 69). This again highlights the difficulty of estimating factor returns (i.e., factor timing). However, higher moments are generally considered to be easier to estimate and predict.

**Figure 68: IR and Sortino ratio**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 69: Drawdown and conditional VaR**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

# Portfolio performance

The ultimate test rests on portfolio performance. In this section, we build a long/short market neutral portfolio for each one of the 124 models discussed in the previous sections. For simplicity, we only show five key models in this section.

- The EQW model, which assumes no style rotation skill and no prediction of the factor covariance matrix, equally weighting all 12 factors.
- The LTA_COV model, which uses a naïve approach to estimate factor returns (historical averages) and a naïve sample factor covariance matrix.
- The LTA_SF model, which uses a naïve approach to estimate factor returns (historical averages) and the single index factor covariance matrix estimator.
- The MVSK_LTA_SF model, which uses a naïve approach to estimate factor returns (historical averages) and the single index model to estimate factor covariance/coskewness/cokurtosis matrices.
- The LN_SF model, which uses a full-fledged style prediction model (LINEAR model) to forecast factor returns and the single index factor covariance matrix estimator.
- The MVSK_LN_SF model, which uses a full-fledged style prediction model (LINEAR model) to forecast factor returns and the single index model to estimate factor covariance/coskewness/cokurtosis matrices.

## Portfolios setup

We structure the five portfolios with identical constraints:

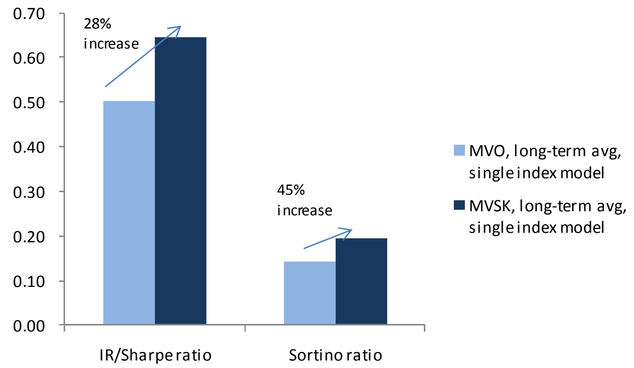
- Long/short market neutral strategy
- Maximize expected alpha
- 2x leverage, i.e., for $1 capital, the strategy invests in $1 long and $1 short
- Target annualized volatility of 4%
- Beta neutral
- Sector neutral
- Maximum holding of 2% (and maximum short of -2%)
- Turnover constraint: 20% one-way turnover per month

## Portfolio performance

The portfolio performance statistics mostly confirm our model backtesting (as measured by portfolio IR/Sharpe ratio, Sortino ratio, drawdown, conditional VaR, and overall distribution). In this section, we only provide a brief summary.

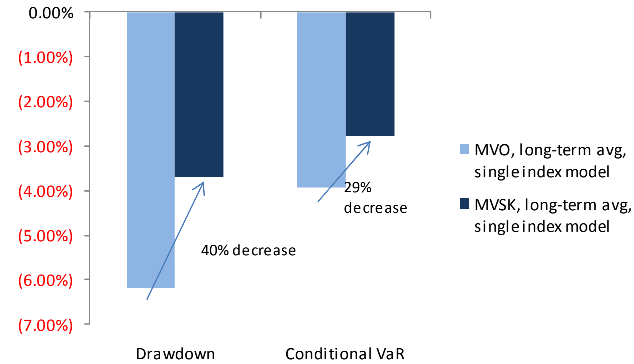
As shown in Figure 70 and Figure 71, for managers who do not follow factor timing (i.e., no explicit prediction of future factor returns), the MVSK type of optimal factor weight model outperforms the MVO model (i.e., traditional GK/QHS approach) by 28% in portfolio IR/Sharpe ratio (and 45% in Sortino ratio), while lowers drawdown by 40% (and conditional VaR by 29%).

**Figure 70: IR/Sharpe ratio and Sortino ratio**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

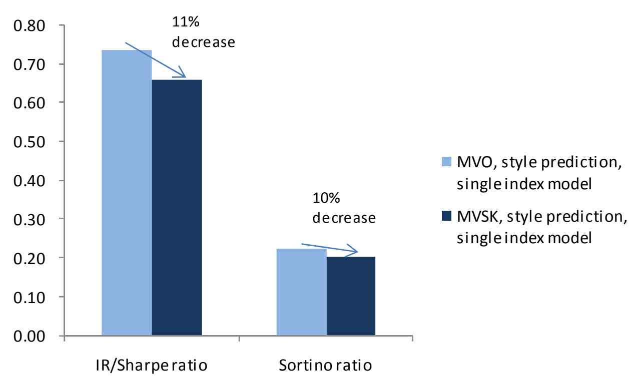
**Figure 71: Drawdown and conditional VaR**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

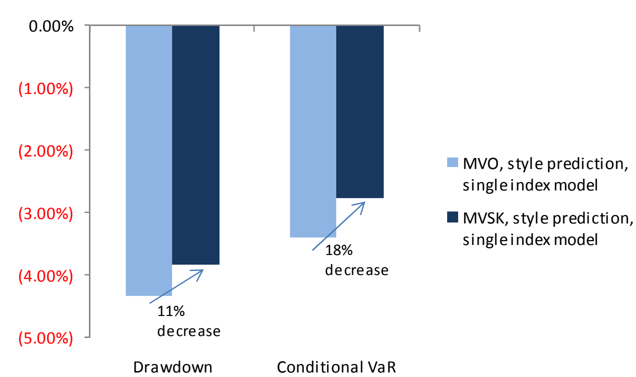
As shown in Figure 72 and Figure 73, for managers who follow factor timing (e.g., using our LINEAR style rotation model), the MVSK type of optimal factor weight model underperforms the MVO model (i.e., traditional GK/QHS approach) by 11% in portfolio IR/Sharpe ratio (and 10% in Sortino ratio). However, the MVSK model does achieve its main goal of reducing tail risk. In fact model drawdown was reduced by 11% (and conditional VaR was cut down by 18%).

**Figure 72: IR/Sharpe ratio and Sortino ratio**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

**Figure 73: Drawdown and conditional VaR**




*Source: Bloomberg Finance LP, Compustat, IBES, Russell, S&P, Thomson Reuters, Deutsche Bank Quantitative Strategy*

# Conclusion

First, let's review the different components of the optimal weighting decision in Figure 74. We would like to summarize our research below.

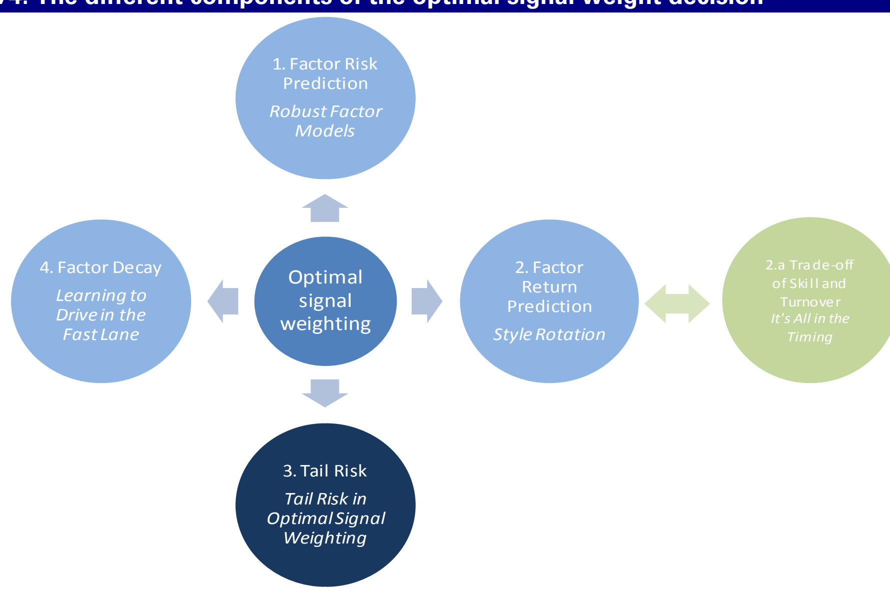
**Factor covariance estimation using the single index model**
A better factor risk model using the single index model essentially raises our model IR (and Sortino ratio) by 45% (and 53%), while reducing drawdown (and conditional VaR) by 17% (and 13%). The single index model is easy to implement at the factor covariance level.

**Factor return prediction using macroeconomic data**
A better factor timing model can effectively increase model IR (and Sortino ratio) by 36% (and 41%), while reducing drawdown (and conditional VaR) by 22% (and 7%). The argument in favor or against factor return prediction is more philosophical; therefore, we will leave the decision to the managers.

**Incorporating tail risk estimation using the single index model**
Estimating the tail distribution is arguably more involved. Optimization by the PGP algorithm is also not trivial. However, the benefits are significant.

- Factor timing or style rotation is no longer essential. MVSK models with and without factor return prediction make little difference in IR, Sortino ratio, drawdown, or conditional VaR. This essentially eliminates the controversy of estimating future factor returns.
- After incorporating tail risk in our signal weighting decision, our model's downside risk (as measured in drawdown or expected shortfall) declines significantly, while IR (and Sortino ratio) can be maintained at roughly the same level (or increased moderately).
- MVSK models tend to have much lower factor turnover (and portfolio turnover) than the comparable MVO models; therefore, the after cost performance is even more attractive.

**Figure 74: The different components of the optimal signal weight decision**




*Source: Deutsche Bank Quantitative Strategy*

# Bibliography

Alvarez, M., Luo, Y., Cahan, R., Jussa, J., and Chen, Z. [2011]. "Portfolios under Construction: Learning to Drive in the Fast Lane", Deutsche Bank Quantitative Strategy, April 26, 2011

Brandt, M., Santa-Clara, P., and Valkanov, R. [2009]. "Parametric portfolio policies: Exploring characteristics in the cross section of equity returns", *Review of Financial Studies* 22:3411-47.

Cahan, R., Luo, Y., Jussa, J., and Alvarez, M. [2010]. "Portfolios under Construction: It's all in the timing", Deutsche Bank Quantitative Strategy, August 19, 2010

Chunhachinda, P., Dandapani, K., Hamid, S., and Prakash, A.J. [1997]. "Portfolio selection and skewness: Evidence from international stock markets", *Journal of Banking and Finance* 21:143-167.

Davies, R.J., Kat, H.M., and Lu, S. [2004]. "Fund of hedge funds portfolio selection: A multiple-objective approach", Working Paper, ISMA Centre, University of Reading.

Dittmar, R. [2002]. "Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns", Journal of Finance 57:369-403.

Fabozzi, F.J., Kolm, P.N., Pachamanova, D.A., and Focardi, S.M. [2007]. *Robust Portfolio Optimization and Management*, John Wiley & Sons, Inc.

Goldberg, L., Hayes, M., Menchero, J., and Mitra, I. [2009]. "Extreme risk analysis", MSCI Barra Research Insight, April 2009.

Greene, W. [2007]. *Econometric Analysis*, Prentice Hall, 6th Edition.

Harvey, C., Liechty, M., Liechty, J., and Muller, P. [2010]. "Portfolio selection with higher moments", *Quantitative Finance* 10(5) 469-485.

Harvey, C. and Siddique, A. [2000]. "Conditional skewness in asset pricing tests", *Journal of Finance* 55:1263-95.

Hong, Y., Tu, J., and Zhou, G. [2007]. "Asymmetries in stock returns: statistical tests and economic evaluation", *Review of Financial Studies* 20:1547-81.

Horvath, P. and Scott, R. [1980]. "On the direction of preference for moments of higher order than the variance", *Journal of Finance* 35:915-19.

Jorion, P. [1985]. "International Portfolio Diversification with Estimation Risk", *Journal of Business* 58:259–78.

Keating, J. and Shadwick, W.F. [2002]. "A universal performance measure", *The Journal of Performance Measurement*, 6(3).

Kim, T., and White, H. [2004]. "On more robust estimation of skewness and kurtosis", *Finance Research Letters* 1:56-73.

Kimball, M. [1993]. "Standard risk aversion", *Econometrica* 61"589-611.

Ledoit, O., and Wolf, M. [2003] "Honey, I shrunk the sample covariance matrix", Working Paper, November 2003.

Ledoit, O., and Wolf, M. [2004a]. "Honey, I shrunk the sample covariance matrix: problems in mean-variance optimization", *Journal of Portfolio Management* 30:110–19.

Ledoit, O., and Wolf, M. [2004b]. "A well-conditioned estimator for large-dimensional covariance matrices", *Journal of Multivariate Analysis* 88:365–411.

Luo, Y., Cahan, R., Jussa, J., and Alvarez, M. [2010]. "GTAA/Signal Processing: Style Rotation", Deutsche Bank Quantitative Strategy, September 7, 2010.

Luo, Y., Cahan, R., Alvarez, M., Jussa, J., and Chen, Z. [2011]. "Portfolios under Construction: Robust factor models", Deutsche Bank Quantitative Strategy, January 24, 2011.

Malevergne, Y., and Sornette, D. [2005]. "High-order moments and cumulants of multivariate Weibull asset returns distributions: Analytical theory and empirical tests: II", *Finance Letters* 3(1), special issue on "Modeling of the equity market", edited by Fabozzi, F.J., Focardi, S.M., and Kolm, P.N., pp. 54-63.

Martellini, L., and Ziemann, V. [2010]. "Improved estimates of higher-order comoments and implications for portfolio selection", *Review of Financial Studies* 23, 4, 1467-1502.

Mitton, T. and Vorkink, K. [2007]. "Equilibrium underdiversification and the preference for skewness", *Review of Financial Studies* 20:1255.

Plantinga, A., van der Meer, R., and Sortino, F. [2001]. „The impact of downside risk on risk-adjusted performance of mutual funds in the Euronext markets", July 19, 2001, SSRN Working Paper.

Prakash, A.J., Chang, C.H., Pactwa, T.E. [2003]. "Selecting a portfolio with skewness: Recent evidence from U.S., Europe, and Latin American equity markets", *Journal of Banking and Finance* 27:1375-90.

Sharpe, W.F. [1963]. "A simplified model for portfolio analysis", *Management Science* 9(2): 277-93

Sortino, F. and Price, L, [1994]. "Performance measurement in a downside risk framework", Journal of Investing, Fall 1994, 59-65.

Sun, Q. and Yan, Y. [2003]. "Skewness persistence with optimal portfolio selection", *Journal of Banking and Finance* 27: 1111-21.

Taleb, N.N. [2007]. *The Black Swan: The Impact of the Highly Improbably*, Random House.

# Appendix 1

## Important Disclosures

Additional information available upon request

**For disclosures pertaining to recommendations or estimates made on a security mentioned in this report, please see the most recently published company report or visit our global disclosure look-up page on our website at http://gm.db.com/ger/disclosure/DisclosureDirectory.eqsr.**

## Analyst Certification

The views expressed in this report accurately reflect the personal views of the undersigned lead analyst(s). In addition, the undersigned lead analyst(s) has not and will not receive any compensation for providing a specific recommendation or view in this report. Yin Luo/Rochester Cahan/Miguel-A Alvarez/Javed Jussa/Zongye Chen

## Hypothetical Disclaimer

Backtested, hypothetical or simulated performance results discussed on page 10 herein and after have inherent limitations. Unlike an actual performance record based on trading actual client portfolios, simulated results are achieved by means of the retroactive application of a backtested model itself designed with the benefit of hindsight. Taking into account historical events the backtesting of performance also differs from actual account performance because an actual investment strategy may be adjusted any time, for any reason, including a response to material, economic or market factors. The backtested performance includes hypothetical results that do not reflect the reinvestment of dividends and other earnings or the deduction of advisory fees, brokerage or other commissions, and any other expenses that a client would have paid or actually paid. No representation is made that any trading strategy or account will or is likely to achieve profits or losses similar to those shown. Alternative modeling techniques or assumptions might produce significantly different results and prove to be more appropriate. Past hypothetical backtest results are neither an indicator nor guarantee of future returns. Actual results will vary, perhaps materially, from the analysis.

# Regulatory Disclosures

## 1. Important Additional Conflict Disclosures

Aside from within this report, important conflict disclosures can also be found at https://gm.db.com/equities under the "Disclosures Lookup" and "Legal" tabs. Investors are strongly encouraged to review this information before investing.

## 2. Short-Term Trade Ideas

Deutsche Bank equity research analysts sometimes have shorter-term trade ideas (known as SOLAR ideas) that are consistent or inconsistent with Deutsche Bank's existing longer term ratings. These trade ideas can be found at the SOLAR link at http://gm.db.com.

## 3. Country-Specific Disclosures

**Australia:** This research, and any access to it, is intended only for "wholesale clients" within the meaning of the Australian Corporations Act.
**EU countries:** Disclosures relating to our obligations under MiFiD can be found at http://globalmarkets.db.com/riskdisclosures.
**Japan:** Disclosures under the Financial Instruments and Exchange Law: Company name - Deutsche Securities Inc. Registration number - Registered as a financial instruments dealer by the Head of the Kanto Local Finance Bureau (Kinsho) No. 117. Member of associations: JSDA, The Financial Futures Association of Japan. Commissions and risks involved in stock transactions - for stock transactions, we charge stock commissions and consumption tax by multiplying the transaction amount by the commission rate agreed with each customer. Stock transactions can lead to losses as a result of share price fluctuations and other factors. Transactions in foreign stocks can lead to additional losses stemming from foreign exchange fluctuations. "Moody's", "Standard & Poor's", and "Fitch" mentioned in this report are not registered credit rating agencies in Japan unless "Japan" is specifically designated in the name of the entity.
**New Zealand:** This research is not intended for, and should not be given to, "members of the public" within the meaning of the New Zealand Securities Market Act 1988.
**Russia:** This information, interpretation and opinions submitted herein are not in the context of, and do not constitute, any appraisal or evaluation activity requiring a license in the Russian Federation.

## North American locations

**Deutsche Bank Securities Inc.**
60 Wall Street
New York, NY 10005
Tel: (212) 250 2500

**Deutsche Bank Securities Inc.**
225 Franklin Street
25th Floor
Boston, MA 02110
Tel: (617) 988 8600

**Deutsche Bank Securities Inc.**
222 South Riverside Plaza
30th Floor
Chicago, IL 60606
Tel: (312) 537-3758

**Deutsche Bank Securities Inc.**
3033 East First Avenue
Suite 303, Third Floor
Denver, CO 80206
Tel: (303) 394 6800

**Deutsche Bank Securities Inc.**
1735 Market Street
24th Floor
Philadelphia, PA 19103
Tel: (215) 854 1546

**Deutsche Bank Securities Inc.**
101 California Street
46th Floor
San Francisco, CA 94111
Tel: (415) 617 2800

**Deutsche Bank Securities Inc.**
700 Louisiana Street
Houston, TX 77002
Tel: (832) 239-4600

## International locations

**Deutsche Bank Securities Inc.**
60 Wall Street
New York, NY 10005
United States of America
Tel: (1) 212 250 2500

**Deutsche Bank AG London**
1 Great Winchester Street
London EC2N 2EQ
United Kingdom
Tel: (44) 20 7545 8000

**Deutsche Bank AG**
Große Gallusstraße 10-14
60272 Frankfurt am Main
Germany
Tel: (49) 69 910 00

**Deutsche Bank AG**
Deutsche Bank Place
Level 16
Corner of Hunter & Phillip Streets
Sydney, NSW 2000
Australia
Tel: (61) 2 8258 1234

**Deutsche Bank AG**
Level 55
Cheung Kong Center
2 Queen's Road Central
Hong Kong
Tel: (852) 2203 8888

**Deutsche Securities Inc.**
2-11-1 Nagatacho
Sanno Park Tower
Chiyoda-ku, Tokyo 100-6171
Japan
Tel: (81) 3 5156 6701

## Disclaimer