MORNINGSTAR®

Morningstar Global Risk Model
Methodology

Morningstar Research
27 May 2016

Content

1 Introduction
2 Model highlights
2 Universe Construction
3 Factor Selection
7 Factor Premia Estimation
8 Forecasting Factor Co-movement
10 Forecasting Idiosyncratic Movement
10 Aggregating Forecasts to Portfolios
14 Conclusion
15 References

Appendix A
16 Estimation Universe Construction Rules

Appendix B
18 Factor Exposure Definitions
Style Factors
Sector Factors
Region Factors
Currency Factors

Appendix C
26 Independent Component Analysis

Appendix D
28 Generalized Autoregressive Conditional Heteroskedastic Normal Inverse Gaussian

Introduction

Risk is inherent to investing. Developing a prospective view of risk allows investors to make investment decisions tailored to their individual risk preferences and ultimately increase the utility derived from their investment portfolio. A risk model forecasts the distribution of future asset returns. This distribution contains all the information needed to assess the riskiness of a portfolio. As the forecasted distribution widens, that indicates more uncertainty about the future return potential of the portfolio. As tail probabilities increase, that indicates the portfolio has higher risk of experiencing an extreme loss. With this forecast, investors are empowered to evaluate the riskiness of assets or portfolios of assets.

In essence, the model seeks to identify a small number of independent, latent sources of return. Movements in these sources drive the movement in a comparably small number of interpretable factors. An example of a factor is the exposure to particular industry currencies – for instance how much does an increase in the Euro/USD exchange rate drive an increase in the value of a stock? Movements in the factors drive asset returns.

Several methodological choices must be made when building a risk model. Our choices were made with the goal of creating a unique, interpretable, responsive, and predictive model. We began with the following assumptions about asset returns which shaped our methodological choices.

- There are a small number of independent sources of market movement which drive the majority of variation in asset returns
- Asset returns are not normally distributed
- The distribution of asset returns changes through time

These three concepts are well-recognized and non-controversial, although some or all of them are often ignored for convenience by risk modelling practitioners.

Authors

Patrick Caldon, PhD
Senior Quantitative Analyst
+61 2 9276 4473
patrick.caldon@morningstar.com

Lee Davidson, CFA
Head of Quantitative Research
+1 312 244-7541
lee.davidson@morningstar.com

Xin Ling
Senior Quantitative Developer
+1 312 696-6191
xin.ling@morningstar.com

Warren Miller, CFA
Head of Asset Management Software
+1 312 384-4902
warren.miller@morningstar.com

Harry Tan, CFA
Senior Quantitative Developer
+1 312 384-3832
harry.tan@morningstar.com

Andy Tang, CFA
Quantitative Analyst
+1 312 384-4839
andy.tang@morningstar.com

Model highlights

Several features make the Morningstar Global Risk Model unique:

1. **We use proprietary fundamental based factors which we believe are superior drivers of returns.**
   Morningstar's research group provides forward-looking ratings on assets which have been successful in predicting the future distribution of returns. Factors based on these ratings also tend to be uncorrelated with traditional risk factors, making them a complementary addition to our risk factor model. Likewise, we have distilled Morningstar's proprietary database of mutual fund holdings into factors which are also uncorrelated predictors of the future distribution of returns.

2. **We forecast the full probability distribution of future returns with non-normal distributions.**
   Our risk model is agnostic to any particular risk metric a user wishes to use. Volatility, conditional value at risk, downside deviation, interquartile range, skewness, kurtosis and many other measures can be calculated directly from the probability distribution that is output from our model.

3. **We accommodate a range of time horizons.**
   There is no need to guess whether a "short term" model or "long term" model matches your investment horizon. .

4. **We make no assumption that co-movement of returns is exclusively linear.**
   The common practice of building and analyzing only a covariance matrix misses the fact that stocks can experience tail events at the same time. Our model directly captures higher co-moments of returns, enabling the construction of portfolios which can control tail risk.

Universe Construction

We define an estimation universe of investible companies with reliable data on which to build the model. Stocks outside the estimation universe – generally illiquid stocks with small market capitalizations – are relegated to the extended universe. We only use stocks in the estimation universe to derive model parameters. This ensures the model parameters are not influenced by illiquid stocks with unreliable data. Further we assume that the factors driving return in small illiquid stocks, like currency and region, are the same factors which drive return for large stocks.

Exhibit 1 Estimation and Coverage Universe

Estimation Universe	**Coverage Universe**
Approximately 7,000 stocks	Approximately 44,000 stocks
(Curated broad group of large liquid stocks)	(Small illiquid stocks)

Source: Morningstar.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

We aim for a broad selection of companies across regions and sectors which are liquid enough to be investable for most investors and are large enough to represent a large portion of the investable universe. The filtering process below results in an estimation universe that is roughly 85% the size of the global equity market capitalization. Appendix A details the exact rules we use to filter our estimation universe.

All market returns and factor inputs are converted into a US dollar denomination when appropriate; in particular all market returns are calculated in US dollars except as otherwise noted. The outputs of the model measure risk with a US dollar numeraire.

Factor Selection

There are many ways to estimate the co-movement of asset returns. A naïve approach might be to calculate a sample covariance matrix using historical returns. Unfortunately, this solution suffers from the curse of dimensionality, i.e. the number of parameters in the covariance matrix is huge relative to the number of historical return observations. As a result, the covariance matrix will be dominated by noise and will poorly forecast future co-movement.

To remedy this problem we use a well understood approach to reduce the number of dimensions – factor modeling. By finding common factors that drive asset returns, we no longer need to model each asset individually. We can instead model a much smaller number of factors. This reduces the dimension of our problem to reasonable levels and allows us to generate estimates of future co-movement.

There are several key notions needed to understand the way this model works:

- An **asset return** is the return of an investible security over a time period
- A **factor** is an observable data point that appears to influence asset returns, like Liquidity or Sector.
- A **factor exposure** is a number that measures how much an asset's return is influenced by a factor. Exposures can be positive, negative or zero. Exposures change through time
- A **factor premium** is a number which represents how much a particular factor has influenced asset returns for a particular time period
- We will later introduce **sources**. These are unobservable phenomena discovered through statistical inference that drive some collection of factor premia.

We set out with several criteria when selecting factors for our model.

1. Our factors should have an economic basis and empirical relevance as predictors of the future distribution of asset returns
2. Our factors should be interpretable and lend insight to a risk attribution analysis
3. Our factor set should be parsimonious
4. Our factor exposures should be practical to calculate

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Ultimately, we arrived at 36 factors which fall naturally into four distinct groups: style, sector, region and currency. A short exposition on these factors is below. A more detailed treatment can be found in Appendix B.

Style Factors

Our 11 style factors are normalized by subtracting the cross-sectional mean and then dividing by the cross-sectional standard deviation, so a score of zero can always be interpreted as the average score, and a non-zero score of n can be interpreted as being n standard deviations from the mean. In addition, we modify the sign of our exposures so the premia associated with them are generally positive.

Exhibit 2 11 Style Factors

Name	Description
Valuation	The ratio of Morningstar's quantitative fair value estimate for a company to its current market price. Higher scores indicate cheaper stocks.
Valuation Uncertainty	The level of uncertainty embedded in the quantitative fair value estimate for a company. Higher scores imply greater uncertainty.
Economic Moat	A quantitative measure of the strength and sustainability of a firm's competitive advantages. Higher scores imply stronger competitive advantages.
Financial Health	A quantitative measure of the strength of a firm's financial position. Higher scores imply stronger financial health.
Ownership Risk	A measure of the risk exhibited by the fund managers who own a company. Higher scores imply more risk exhibited by owners of the stock.
Ownership Popularity	A measure of recent accumulation of shares by fund managers. Higher scores indicate greater recent accumulation by fund managers.
Liquidity	Share turnover of a company. Higher scores imply more liquidity.
Size	Market capitalization of a company. Higher scores imply smaller companies.
Value-Growth	Value-Growth, where a value stock has a low price relative to its book value, earnings and yield. Higher scores imply firms that are more growth and less value oriented.
Momentum	Total return momentum over the horizon from -12 months through -2 months. Higher scores imply greater return momentum.
Volatility	Total return volatility as measured by largest minus smallest 1month returns in a trailing 12 month horizon. Higher scores imply greater return volatility.

Source: Morningstar.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Sector Factors

Our 11 sector factors measure the economic exposure of a company to the 11 Morningstar sectors. We perform a Bayesian time-series regression analysis to find the exposures of an individual company to the sector return with a prior based on the discrete sector classification of Morningstar's data analysts. We enforce constraints that our sector exposures, including the intercept term, must sum to 1 and must individually be between 0 and 1.

- Basic Materials
- Energy
- Financial Services
- Consumer Defensive
- Consumer Cyclical
- Technology
- Industrials
- Healthcare
- Communication Services
- Real Estate
- Utilities

Region Factors

Our 7 region factors represent the economic exposure of a company to the 7 Morningstar regions. We perform a Bayesian time-series regression analysis to find the exposures of an individual company to the return of the portfolio of stocks in the region with a prior based on the discrete Region classification of Morningstar's data analysts.

- Developed Americas
- Developed Europe
- Developed Asia Pacific
- Emerging Americas
- Emerging Europe
- Emerging Asia Pacific
- Emerging Middle East

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Currency Factors

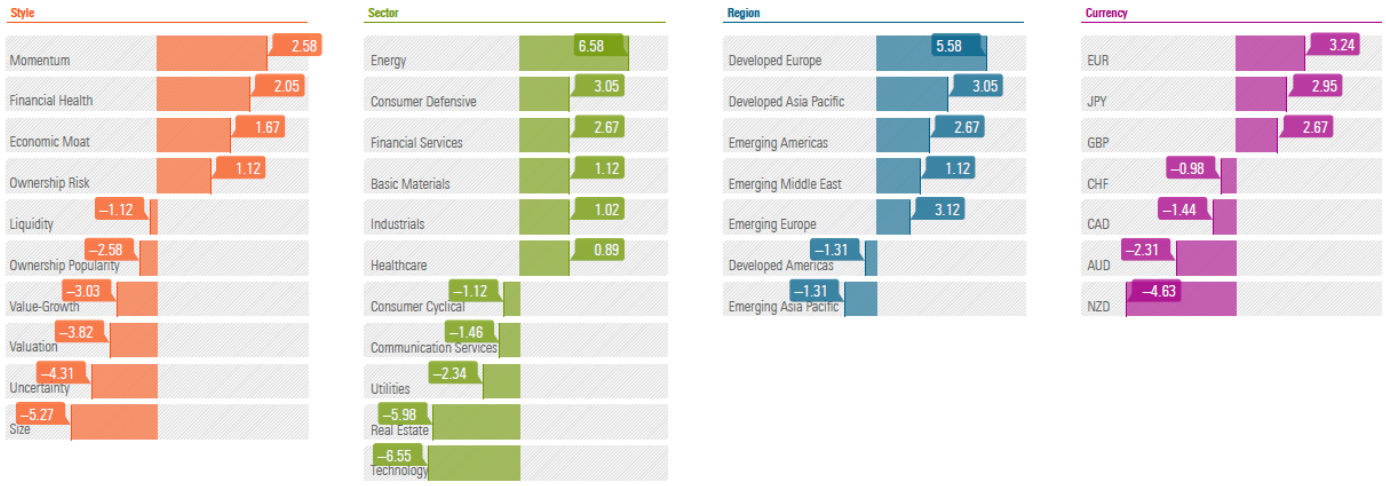
Our 7 currency factors represent the economic exposure of a company to 7 major currencies, excluding US dollars. We perform a time-series regression analysis to find the exposures of an individual company's return denominated in US dollar currency to the following list of currency returns. We calculate the return of these currencies against the US dollar.

- Euro
- Japanese Yen
- British Pound
- Swiss Franc
- Canadian Dollar
- Australian Dollar
- New Zealand Dollar

Factor Exposure Visualization

Morningstar's Risk Signature visualization shows current factor exposures for an asset or portfolio, and can include a benchmark for relative comparisons. It's a fast way to immediately understand the characteristics of a particular portfolio.

Exhibit 3 Risk Signature



Source: Morningstar.

Exhibit 3: In Morningstar's Risk Signature visualization, factor exposures for an asset or portfolio are show in bars which are grouped by factor type and sorted by exposure size. The relevant benchmark's exposures are represented by lines overlaid on each bar.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Factor Premia Estimation

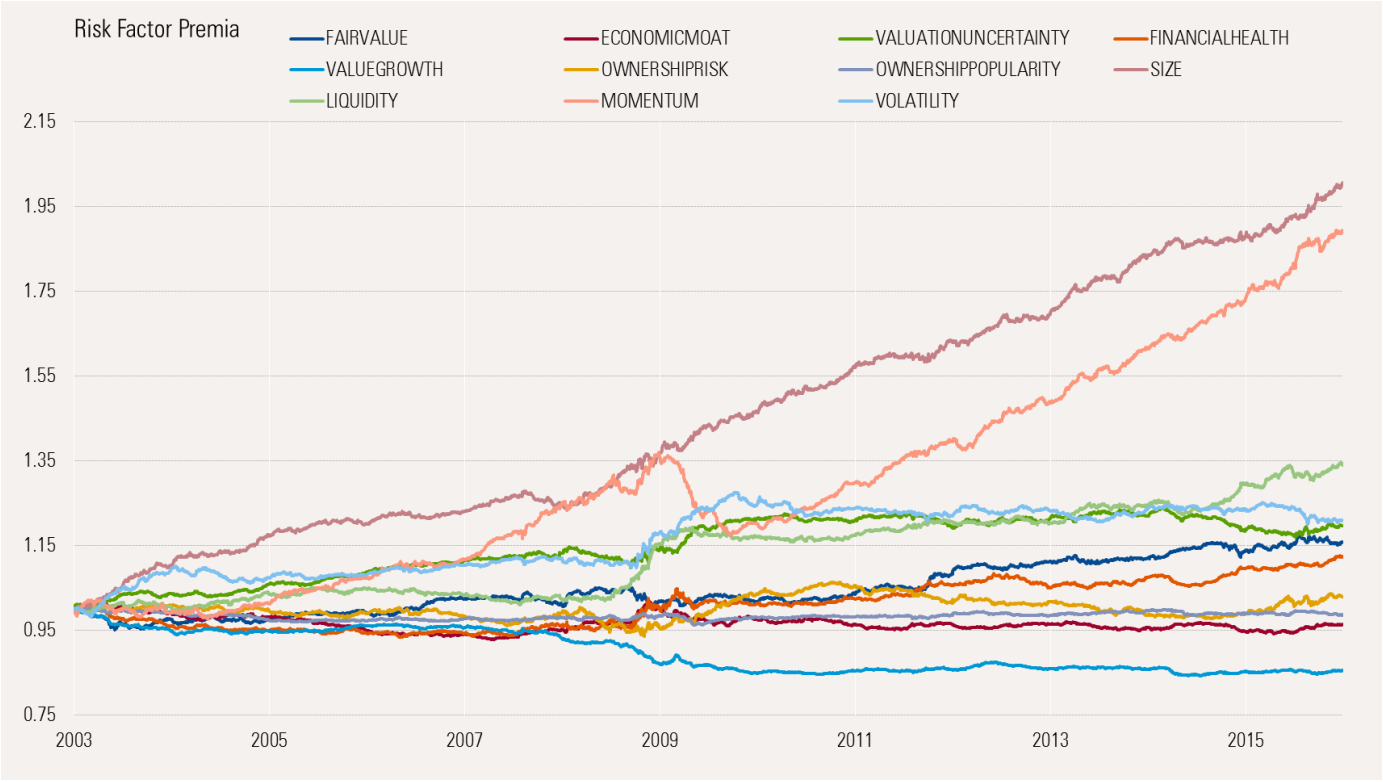
Given a collection of factor exposures, X_t, for a set of n stocks at time, t, we perform a cross-sectional regression of those exposures on total returns from t to $t+1$, r_{t+1}, to estimate the factor premia, f_t.

$$r_{t+1} = X_t f_t + \varepsilon_t$$

Where:
$r_t = (n \times 1)\ vector\ of\ stock\ returns\ at\ time\ t$
$X_t = (n \times m)\ matrix\ of\ stock\ exposures\ to\ factors\ at\ time\ t; m = 36$
$f_t = (m \times 1)\ vector\ of\ factor\ premia\ at\ time\ t$
$\varepsilon_t = (n \times 1)\ vector\ of\ error\ terms\ at\ time\ t$

By repeating this cross-sectional regression, we construct an historical time series of the factor premia. We use this time series of factor premia to analyze how each factor behaves in the context of the other factors by examining factor co-movement in the history.

Exhibit 4 Historical Time Series of the Factor Premia



Source: Morningstar.

Exhibit 4: Morningstar's Factor Premia visualization shows the cumulative return of our fundamental style factor premia through time.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Forecasting Factor Co-movement

Our co-movement forecasts are derived in a three-step process. We first perform a natural logarithm transformation of our factor premia. We then remove the statistical dependency of the factors using a technique known as Independent Component Analysis (ICA). Finally we forecast the statistically independent sources using a time-series forecasting technique called a Generalized Autoregressive Conditional Heteroskedastic Normal Inverse Gaussian GARCH-NIG distribution. This approach enables us to forecast an entire non-normal distribution of returns for individual assets and portfolios.

Independent Component Analysis

ICA separates a collection of signal sources into their statistically independent driving factors. Our factors share some mutual information which causes them to exhibit co-movement, and we need to model this co-movement – or lack of statistical independence - if we are to estimate the future distributions of portfolios.

ICA allows us to linearly transform our fundamental factors into latent "sources" which share minimal mutual information. If sources share mutual information then analysis of some of the sources can deliver useful information about other sources. This means in turn we cannot analyze the sources separately but must analyze them jointly. This joint analysis is technically difficult. The practical significance of the ICA transformation is that we no longer have to construct a joint distribution of the independent sources. Instead we use univariate forecasting techniques, which we aggregate into forecasts for individual assets and portfolios.

Specifically, ICA is a matrix factorization algorithm. It is used to decompose the matrix of demeaned historical financial log-premia, $\mathrm{F} - \bar{\mathrm{F}}$, into the product of a mixing matrix, A, and a matrix of a time series of our statistically independent sources, S.

$$F - \bar{F} = AS$$

Where:
$F = (m \times h)\ matrix\ of\ historical\ log\ factor\ premia$
$\bar{F} = (m \times 1)\ vector\ of\ unconditional\ log\ factor\ premia\ \ means$
$A = (m \times n)\ matrix\ of\ mixing\ coefficients$
$S = (n \times h)\ matrix\ of historical\ statistically\ independent\ sub - factors$

ICA is particularly well-suited for modeling financial returns or premia. ICA contrasts with principal component analysis (PCA) by finding the statistically independent sources as opposed to the sources which are merely uncorrelated. Given a sufficiently large quantity of normally distributed data, both PCA and ICA should converge to identical decompositions up to negation and reordering of the rows of A. But in the presence of non-normal data, ICA results in a more useful decomposition than PCA. It is generally non-controversial that returns and risk premia are non-normally distributed, and we find this to be true with our own premia data.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

There are several ways to characterize the estimation process for the mixing matrix, $\mathbf{A}$, and the sources, $\mathbf{S}$. A straightforward way is to imagine that we have some method to measure the non-normality of a sample of a weighted sum of the factors – we can do this by measuring the sample kurtosis, under the assumption that a very high sample kurtosis means the weighted sum is non-normal. Suppose we can somehow search the entire space of weights for the factors and find the weights which produce maximal non-normality - call this a source. We then apply a mathematical operation to remove the source we just discovered from the original collection of factors, and iterate the procedure on the factors with this source removed. This operation produces the inverse matrix of A. A refinement and formalization of these ideas is called "projection pursuit". ICA has some technical differences to projection pursuit but produces identical answers to under some formal conditions, most importantly that the factors are produced by a weighted sum of independent non-normal sources. A more detailed and technical description of the ICA methodology that we use in our risk model is found in Appendix X.

The Generalized Autoregressive Conditional Heteroskedastic Normal Inverse Gaussian (GARCH-NIG) Model

We now need a model to forecast the distribution of our statistically independent sources. ICA relies on the non-normality of the sources so a non-normal distribution seemed likely to be an appropriate model. We examined several widely used non-normal univariate distributions to model the sources. The distribution which struck the right balance of flexibility to model higher moments, mathematical and computational tractability, and good model fit statistics was the Normal Inverse Gaussian (NIG) Distribution.

The NIG distribution is a special case of the generalized hyperbolic distribution which has four parameters, allowing it to flexibly model skewed and fat-tailed data. The density function is given below, and straightforward expressions for the distribution's mean, variance, skewness and kurtosis in terms of these variables are in the appendix.

$$g\left(x;\bar{\alpha},\bar{\beta},\mu,\delta\right)=\frac{\bar{\alpha}}{\pi\delta}exp\left(\bar{\gamma}+\bar{\beta}\frac{(x-\mu)}{\delta}\right)\frac{\boldsymbol{K}_1\left(\bar{\alpha}\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right)}{\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}}$$

Here $\mathbf{K}_1$ is the modified Bessel function of the third order and index 1. The four parameters $\bar{\alpha}=\delta\alpha,\bar{\beta}=\delta\beta,\mu,\delta$ parameterize the distribution, where $x,\mu\in R,\delta\geq 0,0\leq|\beta|\leq\alpha$ and $\bar{\gamma}=\sqrt{\bar{\alpha}^2-\bar{\beta}^2}$. The expressions are written in terms of either α,β or $\bar{\alpha},\bar{\beta}$. The mean, variance and higher moments of the distribution have straightforward expressions in these variables given in the appendix.

Still, fitting an NIG distribution as above to our data did not result in the desired dynamic properties of our forecast. In particular the sources—like many asset returns—exhibit volatility clustering, so a large movement in a source at time t tends to be followed by another large movement at time $t+1$. The effect is that we know the distribution parameters are changing over time, i.e. if the distribution has a large dispersion at time t, then it will tend to have a large dispersion at time $t+1$. We chose a

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

methodology that generates a conditional forecast, i.e. a forecast that adapts to changing conditions to capture phenomena such as volatility clustering. This is accomplished by fitting an NIG distribution with zero mean and a variance of 1 to our historical normalized independent sources using maximum likelihood estimation. We then multiply our NIG variable by a parameter, σ_t, which evolves as a GARCH(1,1) process. This allows the variance of the distribution to change through time, while preserving the tail and skewness characteristics.

Forecasting Idiosyncratic Movement

In addition to the factor premia, our cross-sectional regressions produce residual terms for each stock in a particular time period which represent the portion of that stock's return not explained by the fundamental factors. In order to properly forecast the risk of the stock, we must also model this idiosyncratic portion of returns.

We choose to model these log-transformed idiosyncratic returns exactly as we model our statistically independent sources, using a GARCH-NIG model as described above. This allows us to have a time-varying forecast which accommodates the fat tails and skewness which may be present in the idiosyncratic returns.

Aggregating Forecasts to Portfolios

A portfolio's return is the weighted average of the returns of its holdings. Using this fact we map the return on a portfolio to a linear combination of independent components. We need not define a separate set of equations for individual stocks because stocks can be thought of as a portfolio that is 100% invested in a single holding.

A portfolio's return $\mathrm{p_t}$ is a weighted average of the returns of its holdings. $\mathrm{p_t = w_t^T r_t}$

An asset's return is an average of the returns of the factors weighted by the asset's exposure to the various factors plus the idiosyncratic return.

$$r_t = X_t f_t + \epsilon_t$$

A factor's return is an exponential transformation of the average of the statistically independent source returns weighted by its exposure to each source as determined by the mixing matrix.

$$f_t = exp(AS_t + \bar{f}) - 1 - \bar{f}$$

where $\mathbf{1}$ is a vector of ones. Therefore the portfolio's return is an average of the independent source returns weighted by the portfolio's exposure to each source plus an average of the idiosyncratic returns of its holdings weighted by the portfolio's asset values.

$$\begin{aligned} p_t &= w_t^T \left(X_t \left(exp\left(AS_t + \bar{f} \right) - 1 \right) + \epsilon_t \right) \\ &= w_t^T X_t exp\left(AS_t + \bar{f} \right) - w_t^T X_t 1 + w_t^T \epsilon_t \end{aligned}$$

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Analytically Aggregating Moments

It is reasonable to assume investors derive utility from more than just the mean and variance of portfolio returns. If we assume investors also care about skewness and kurtosis, then we must find the first four raw moments of the portfolios forecast distribution. Mean, variance, skewness and kurtosis can all be expressed as a function of a distribution's raw moments as follows.

$$Mean(X) = \mu_1$$

$$Variance(X) = \mu_2 - \mu_1^2$$

$$Skewness(X) = \frac{\mu_3 - 3\mu_1\mu_2 + 3\mu_1^3 - \mu_1^2}{(\mu_2 - \mu_1^2)^{1.5}}$$

$$Kurtosis(X) = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 4\mu_1^3\mu_1 + \mu_1^4}{(\mu_2 - \mu_1^2)^2}$$

Where:

$\mu_n = the\ nth\ raw\ moment\ of\ x = E[X^n]$

We can find the raw moments of the portfolio from our equation above. Each of the terms below can be expressed as a function of the moments of the NIG distribution for S_t and ϵ_t.

$$\mu_1 = E[p_t] = E\left[w_t^T X_t exp(AS_t + \bar{f}) - w_t^T X_t 1 + w_t^T \epsilon_t\right]$$

$$\mu_2 = E[p_t^2] = E\left[\left(w_t^T X_t exp(AS_t + \bar{f}) - w_t^T X_t 1 + w_t^T \epsilon_t\right)^2\right]$$

$$\mu_3 = E[p_t^3] = E\left[\left(w_t^T X_t exp(AS_t + \bar{f}) - w_t^T X_t 1 + w_t^T \epsilon_t\right)^3\right]$$

$$\mu_4 = E[p_t^4] = E\left[\left(w_t^T X_t exp(AS_t + \bar{f}) - w_t^T X_t 1 + w_t^T \epsilon_t\right)^4\right]$$

We can find the nth raw moment or exponential transformed moments of any NIG distributed random variable from its moment generating function. We replace the z with the moment order to get the exponential transformed moments.

$$MGF_x(z) = e^{\mu z + \delta\left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + z)^2}\right)}$$

$$x \sim NIG(\mu, \alpha, \beta, \delta)$$

We produce multi-horizon forecasts over time periods longer than days by composing the returns over several time periods.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Approximating the Conditional Forecast Distribution

It can be useful to produce visualizations of the forecast distribution. To produce the forecast distribution chart shown below, we need to draw the density function of a portfolio's forecast returns at a given time horizon.

We've shown that a portfolio can be modeled as a linear combination of the exponentially transformed statistically independent sources of returns. Sometimes it is simpler to find the distribution with parameters estimated using the method of moments to match the moments that are analytically derived in the sections above.

First note that the characteristic function of a sum of independent random variables is equal to the product of the characteristic functions of each variable. The characteristic function of an NIG distribution follows.

$$\varphi(z) = e^{-i\mu z+\delta(\gamma-\sqrt{\alpha^2-(\beta+iz)^2})}$$

The characteristic function of the sum of variables is as follows.

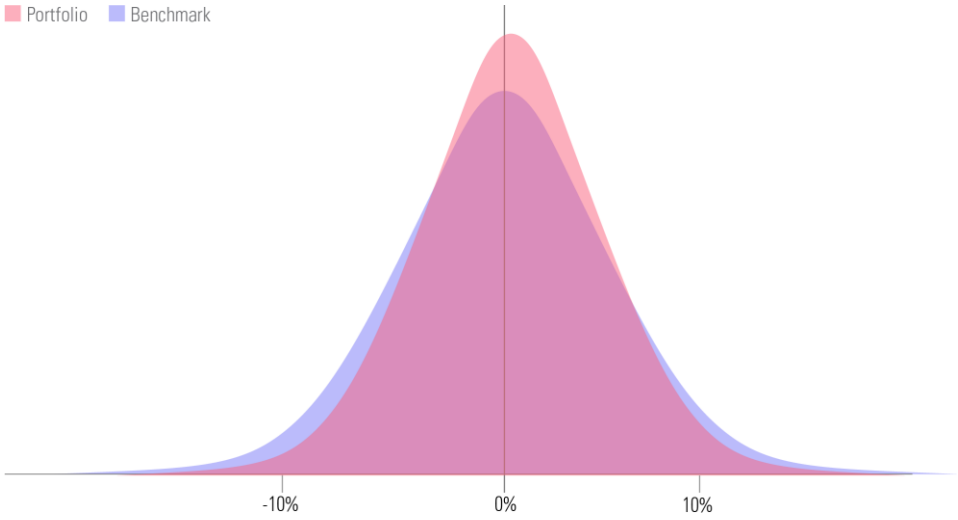
$$\varphi_p(v) = \prod_{j=1}^{L+N} \varphi_{z_{j,t}}(v)$$

Finally, we can take the inverse Fourier transform of this characteristic function to arrive at the density function for our portfolio.

$$f_p(r) = \frac{1}{2\pi}\int e^{-ivr}\varphi_p(v)dv$$

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Exhibit 5 Morningstar's Risk Forecast



Source: Morningstar.

Exhibit 5: Morningstar's Risk Forecast visualization displays the full probability distribution of a stock or fund relative to its most relevant benchmark along with summary statistics for a given time horizon. Data is hypothetical for illustration purposes.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Conclusion

The ability to model the risk of a portfolio is paramount to making investment decisions that maximize utility. Our fundamental factor-based methodology provides a way to forecast risk, but more importantly, it provides an intuitive interpretation of the mechanics behind the forecast. Monitoring factor exposures and making economically sound decisions about which exposures are prudent and which are worth avoiding is much easier when factor exposures are interpretable.

Our risk model includes factors unique to Morningstar. These factors which spring from our analyst-driven research and our institutional portfolio holdings database are uncorrelated with more traditional factors and helpful tools for an investor to use when tailoring their portfolio to suit their risk preferences.

Finally, we chose a non-traditional approach to modeling the future returns of our fundamental factors through the use of Independent Component Analysis and a Generalized Autoregressive Conditional Heteroskedastic Normal Inverse Gaussian process. This two-step procedure allows us to model the entire non-normal joint distribution of future returns. From this joint distribution, we can calculate virtually any risk statistic.

No risk model is perfect. Our aim has been to emphasize interpretability, responsiveness, and predictive accuracy, and in doing so, we believe we've developed a unique model. We are grateful to the risk modelers that paved the way for the methodological choices we've made along the way.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

References

Chen, Ying, Wolfgang Härdle, and Vladimir Spokoiny. "Portfolio value at risk based on independent component analysis." *Journal of Computational and Applied Mathematics* 205.1 (2007): 594-607.

Chen, Ying, Wolfgang Härdle, and Vladimir Spokoiny. "GHICA—Risk analysis with GH distributions and independent components." *Journal of Empirical Finance* 17.2 (2010): 255-269.

Chin, Elion, Andreas S. Weigend, and Heinz Zimmermann. "Computing portfolio risk using gaussian mixtures and independent component analysis."Computational Intelligence for Financial Engineering, 1999.(CIFEr) Proceedings of the IEEE/IAFE 1999 Conference on. IEEE, 1999.

Cover, Thomas M., and Joy A. Thomas. Elements of information theory. John Wiley & Sons, 2012.

Ghalanos, Alexios, Eduardo Rossi, and Giovanni Urga. "Independent factor autoregressive conditional density model." *Econometric Reviews* 34.5 (2015): 594-616.

Hyvärinen, Aapo, Juha Karhunen, and Erkki Oja. Independent component analysis. Vol. 46. John Wiley & Sons, 2004.

Jensen, Morten B., and Asger Lunde. "The NIG - S&ARCH model: A fat - tailed, stochastic, and autoregressive conditional heteroskedastic volatility model."*The Econometrics Journal* 4.2 (2001): 319-342.

Madan, Dilip B., and Ju-Yi J. Yen. "Asset allocation with multivariate non-gaussian returns." Handbooks in Operations Research and Management Science 15 (2007): 949-969.

Morningstar Quantitative Equity Ratings Methodology. 2012. <http://corporate.morningstar.com/US/documents/MethodologyDocuments/FactSheets/QuantitativeEquityCreditRatingsMeth.pdf>

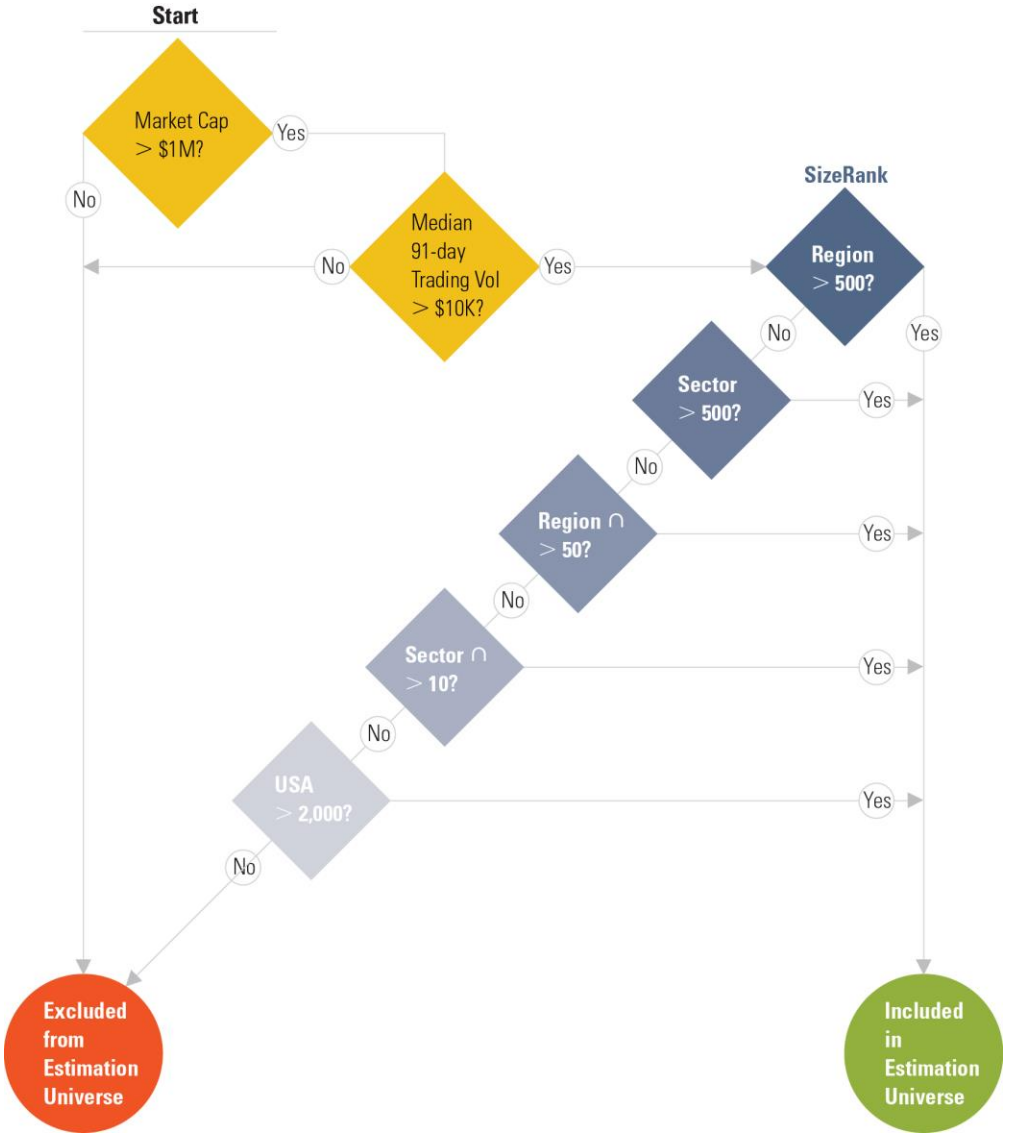
Shephard, Neil G. "From characteristic function to distribution function: a simple framework for the theory." *Econometric Theory* 7.04 (1991): 519-529.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Appendix A: Estimation Universe Construction Rules

$$MvRank = Rank\ of\ market\ capitalization\ (USD)$$
$$TvRank = Rank\ of\ trading\ volume\ (USD)$$
$$SizeRank = MvRank + TvRank$$

Exhibit 6 Estimation Universe Construction Logic



Source: Morningstar.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Appendix B: Factor Exposure Definitions

Style Factors

Valuation

The Valuation factor is the normalized ratio of Morningstar's Quantitative Fair Value Estimate to the current market price of a security. It represents how cheap or expensive a stock is relative to its fair value. We arrive at a Quantitative Fair Value Estimate using an algorithm that extrapolates from the roughly 1400 valuations our equity analyst staff assigns to stocks to a coverage universe of more than 45,000 stocks. For a detailed explanation of this methodology, please refer to the Morningstar Quantitative Equity Ratings methodology document cited in the References section.

The Valuation factor is unbounded, and higher scores indicate cheaper stocks. A score of zero indicates an average valuation.

Valuation Uncertainty

The Valuation Uncertainty factor is the normalized value of Morningstar's Quantitative Valuation Uncertainty Score. It represents the standard error of Morningstar's Quantitative Valuation, in other words, how unsure we are of a particular valuation. For a detailed explanation of this methodology, please refer to the Morningstar Quantitative Equity Ratings methodology document cited in the References section.

The Valuation Uncertainty factor is unbounded, and higher scores indicated more uncertain valuations. A score of zero indicates an average level of uncertainty.

Economic Moat

The Economic Moat factor is the normalized value of Morningstar's Quantitative Moat Score. It represents the strength and sustainability of a firm's competitive advantages. We arrive at a Quantitative Moat Score using an algorithm that extrapolates from the roughly 1400 Economic Moat ratings our equity analyst staff assigns to stocks to a coverage universe of more than 45,000 stocks. For a detailed explanation of this methodology, please refer to the Morningstar Quantitative Equity Ratings methodology document cited in the References section.

The Economic Moat factor is unbounded, and higher scores indicate stronger and more sustainable competitive advantages. A score of zero indicates an average level of competitive advantages.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Financial Health

The Financial Health factor is the normalized value of Morningstar's Quantitative Financial Health Score. It represents the strength of a firm's financial position. The Quantitative Financial Health Score is driven by market inputs, making it responsive to new information. It is calculated as follows.

$$QFH = 1 - \frac{(EQVOLP + EVMVP + EQVOLP \times EVMVP)}{3}$$

Where:
$EQVOLP = percentile\ rank\ trailing\ 300\ day\ equity\ return\ volatility$

$$EVMVP = percentile\ rank\ of\ \frac{Enterprise\ Value}{Market\ Capitalization}$$

The Financial Health factor is unbounded, and higher scores indicate stronger financial health. A score of zero indicates an average level of financial health.

Ownership Risk

The Ownership Risk factor represents, for a particular stock, the ownership preferences of fund managers with different levels of risk exposure. The Ownership Risk factor relies on current portfolio holdings information and the raw Morningstar 36-month Risk score. High Ownership Risk scores signify that those stocks are currently owned and preferred by fund managers with high levels of Morningstar Risk. If high-risk managers are purchasing these stocks, then those stocks are likely to be high-risk. A stock's characteristic is therefore defined by who owns it.

The Ownership Risk score is calculated in the following manner:

$$Ownership\ Risk_n = \sum_{m=1}^{M} v_{m,n} MRISK36_m$$

Where:

$$v_{m,n} = \frac{w_{m,n}}{\sum_{m=1}^{M} w_{m,n}}$$

$$MRISK36 = Morningstar\ Risk\ Score\ 36 - month$$

In words, the Ownership Risk score for stock n is the weighted average of each manager m's Morningstar Risk 36-month score multiplied by the relative weight he/she holds in stock. After raw scores are calculated, Ownership Risk scores are cross-sectionally normalized.

The Ownership Risk score is unbounded, and higher scores indicated stronger ownership preference for risk. A score of zero indicates an average level of ownership preference for risk.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Ownership Popularity

The Ownership Popularity factor represents the growth in the popularity of a particular stock from the perspective of fund manager ownership. The Ownership Popularity factor relies on current and past portfolio holdings information. High Ownership Popularity scores signify that more funds have gone long the stock relative to those that have gone short the stock in the past 3 months.

The Ownership Popularity score is calculated in the following manner:

$$Ownership\ Popularity_{\ n} = \frac{1}{T}\sum_{t=1}^{T}\frac{O_{n,t} - O_{n,t-1}}{O_{n,t-1}}$$

$$O_{n,t} = \sum_{m=1}^{M} v_{m,n,t} NetLong_{m,t}$$

Where:

$$v_{m,n,t} = \frac{w_{m,n,t}}{\sum_{m=1}^{M} w_{m,nt}}$$

$$NetLong_{m,t} = \begin{cases} -1 & if\ w_{m,n,t} < 0 \\ 0 & if\ w_{m,n,t} = 0 \\ 1 & if\ w_{m,n,t} > 0 \end{cases}$$

In words, the Ownership Popularity score for stock n is the average growth in ownership over the past three months. Ownership is the weighted average of each manager m's Net long score multiplied by the relative weight he/she holds in stock. After raw scores are calculated, Ownership Popularity scores are cross-sectionally normalized.

The Ownership Popularity score is unbounded, and higher scores indicated stronger ownership preference. A score of zero indicates an average level of ownership preference.

Size

The Size factor is the normalized value of the logarithm of a firm's market capitalization.

$$size_{i,t} = -ln\left(MV_{i,t}\right)$$

The Size factor is unbounded, and higher scores indicate smaller market capitalization. A score of zero indicates an average level of market capitalization.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Liquidity

The Liquidity factor is the normalized value of the stock's raw share turnover. The raw share turnover score is calculated as the logarithm of the average trading volume divided by shares outstanding over the past 30 days. It is essentially a churn-rate for a stock and represents how frequently a stock's shares get traded.

$$shareturnover_{i,t} = ln\left(\frac{1}{T}\sum_{t=1}^{T}\frac{V_{i,t}}{SO_{i,t}}\right), where\ T = 30$$

The Share Turnover factor is unbounded, and higher scores indicate higher liquidity. A score of zero indicates an average level of liquidity.

Value-Growth

It is a reflection of the aggregate expectations of market participants for the future growth and required rate of return for a stock. We infer these expectations from the relation between current market prices and future growth and cost of capital expectations under the assumption of rational market participants and a simple model of stock value. The Value-Growth factor is unbounded, and higher scores indicated higher growth expectations and less value exposure. A score of zero is average.

Momentum

The Momentum factor is the normalized value of the stock price's raw momentum score. The raw Momentum score is calculated as the cumulative return of a stock from 365 calendar days ago to 30 days ago. This is the classical 12-1 momentum formulation except using daily return data as opposed to monthly. To compute, US dollar currency returns are used.

$$momentum_{i,t} = \sum_{t-365}^{t-30}\left(ln\left(1 + r_{i,t}\right) - ln\left(1 + rf_{t}\right)\right)$$

The Momentum factor is unbounded, and higher scores indicate higher returns over the past year as well as a propensity for higher returns in the future. A score of zero indicates an average level of momentum.

Volatility

The volatility factor is the normalized range of annual cumulative returns over the past year. Each day, we compute the trailing 12-month cumulative return. Then, we look over the past year and identify the maximum and minimum 12-month cumulative returns. We compute the range by taking the maximum minus the minimum 12-month cumulative returns.

$$range_i = \left(ln\left(1 + r_{i,t}\right) - ln(1 + rf_t)\right)^{max} - \left(ln\left(1 + r_{i,t}\right) - ln(1 + rf_t)\right)^{min}$$

The Volatility factor is unbounded, and higher scores indicate higher volatility. A score of zero indicates an average level of volatility.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Sector Factors

Sector exposures are calculated based on a time-series regression of excess stock returns to a set of sector benchmarks.

$$r_t^i - r_t^f = \alpha^i + \beta_1^i \left(r_t^1 - r_t^f\right) + \cdots + \beta_k^i \left(r_t^k - r_t^f\right) + \varepsilon_t^i$$

$r_t^i = weekly\ return\ on\ the\ ith\ stock$

$r_t^f = weekly\ return\ on\ 3 - mo\ US\ TBill$

$r_t^k = weekly\ return\ on\ the\ kth\ sector\ benchmark\ (e.g. Basic\ Materials)$

$constraints: 0 < \beta_k^i < 1; \sum_k \beta_k^i = 1$

Benchmark construction

Sector benchmark returns are calculated using a market-cap weighting scheme using stocks from our estimation universe. Stocks are assigned to sectors on the basis of Global Sector ID. All returns are computed in USD. Market capitalizations were also converted to USD prior to benchmark constitution.

Regression setup

Regressions are five years in length and are run on a rolling, weekly frequency. In the case where a stock does not have five years of history, we run the time-series regression back to the inception date. We employ a Bayesian prior that presupposes that companies should be entirely exposed to the Sector to which they are assigned.

Sectors

- Basic Materials
- Energy
- Financial Services
- Consumer Defensive
- Consumer Cyclical
- Technology
- Industrials
- Healthcare
- Communication Services
- Real Estate
- Utilities

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Interpretation

Sector exposures are bounded between 0 and 1. They must jointly (including the intercept) sum to 1. Higher scores indicate higher levels of sensitivity to individual sectors.

Region Factors

Regional exposures are calculated based on a time-series regression of excess stock returns to a set of region benchmarks.

$$r_t^i - r_t^f = \alpha^i + \beta_1^i \left(r_t^1 - r_t^f\right) + \cdots + \beta_k^i \left(r_t^k - r_t^f\right) + \varepsilon_t^i$$

$r_t^i = weekly\ return\ on\ the\ ith\ stock$

$r_t^f = weekly\ return\ on\ 3 - mo\ US\ TBill$

$r_t^k = weekly\ return\ on\ the\ kth\ region\ benchmark\ (e.g.\ Developed\ North\ America)$

$constraints: 0 < \beta_k^i < 1; \sum_k \beta_k^i = 1$

Benchmark construction

Region benchmark returns are calculated using a market-cap weighting scheme using stocks from our estimation universe. Stocks are assigned to regions on the basis of company-level Country ID. All returns are computed in USD. Market capitalizations were also converted to USD prior to benchmark constitution.

Regression setup

Regressions are five years in length and are run on a rolling, weekly frequency. In the case where a stock does not have five years of history, we run the time-series regression back to the inception date. We employ a Bayesian prior that presupposes that companies should be entirely exposed to the Region in which their company-level Country ID belongs.

Regions

- Developed North America
- Developed Europe
- Developed Asia Pacific
- Emerging Latin America
- Emerging Europe
- Emerging Asia Pacific
- Emerging Middle East & Africa

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Exhibit 7 Map of countries to regions

Region	Country List		
Developed Asia Pacific	Australia	Japan	Singapore
	Hong Kong	New Zealand	
Developed Europe	Belgium	Finland	Italy
	Switzerland	France	Netherlands
	Germany	United Kingdom	Norway
	Denmark	Greece	Portugal
	Spain	Ireland	Sweden
Developed North America	United States	Canada	
Emerging Europe	Czech Republic	Poland	Turkey
	Hungary	Russian Federation	
Emerging Latin America	Brazil	Colombia	Peru
	Chile	Mexico	
Emerging Middle East & Africa	Egypt	Morocco	South Africa
	Israel		

Source: Morningstar.

Interpretation

Region exposures are bounded between 0 and 1. They must jointly (including the intercept) sum to 1. Higher scores indicate higher levels of sensitivity to individual regions.

Currency Factors

Currency exposures are calculated based on a time-series quantile regression of excess stock returns to a set of exchange rates.

$$r_t^i - r_t^f = \alpha^i + \beta_1^i(r_t^1) + \cdots + \beta_k^i(r_t^k) + \varepsilon_t^i$$

$$r_t^i = weekly\ return\ on\ the\ ith\ stock$$

$$r_t^f = weekly\ return\ on\ 3-mo\ US\ TBill$$

$$r_t^k = weekly\ return\ on\ the\ kth\ exchange\ rate\ return\ (e.g.\%\ change\ in \frac{EUR}{USD})$$

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Regression Setup

Regressions are five years in length and are run on a rolling, weekly frequency. In the case where a stock does not have five years of history, we run the time-series regression back to the inception date. Stock returns are calculated in US dollar currency returns.

Currencies

- Euro
- Japanese Yen
- British Pound
- Swiss Franc
- Canadian Dollar
- Australian Dollar
- New Zealand Dollar

Interpretation

Currency exposures are unbounded, but generally fall between -1 and 1. Higher scores indicate higher levels of sensitivity to individual exchange-rate fluctuations.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Appendix C: Independent Component Analysis

Independent Component Analysis (ICA) is a technique that was developed to solve the problem of blind source separation—an example of which is known as the cocktail party problem. In the cocktail party scenario an observer has set up several recording devices throughout a cocktail party to listen to the attendees. The voices are all jumbled together in the hubbub. The observer would like to extract the individual voices from the recordings without any additional knowledge..

ICA solves the cocktail party problem by linearly transforming the time series of recordings into a new set of time series which are each statistically independent from each other. In other words, these new time series are simply weighted sums of the original time series and the weights are chosen in such a way that none of the new time series have any mutual information—that is that there is no knowledge to be gained about any time series by looking at any other time series. In the cocktail party problem, these independent time series, or "sources", should represent individual voices. This mutual information criterion maps naturally to speakers—supposing one person in one part of the room is saying something at some particular instant should tell us nothing about what the other speakers are saying at that same instant.

In finance we care about returns. We cannot directly observe the independent latent sources of those returns, but we can observe asset returns—these act like our recording devices. Performing ICA on asset returns delivers us a set of statistically independent drivers of those returns that are linear combinations of the original returns. This makes it easy to transform returns into sources or vice versa. So the assumption is that there are a collection of independent sources which drive returns. There are two potential intuitive justifications for this assumption. First, some drivers clearly are independent—for instance natural disasters and technology innovation are truly independent. Second, the assumption that returns are a linear combination of univariate time series is a convenient and makes modeling possible, we can see this in the large collection of workable risk models which use techniques like factor analysis and PCA. Independence is the natural criteria which allows for rich univariate models.

The objective of ICA is to take an original $\mathrm{m} \times \mathrm{h}$ matrix $\mathbf{X}$ and find the weights A, that can be multiplied by the statistically independent sources S, to recover our original data, X, so $\mathrm{X} = \mathrm{AS}$. The matrix A is sometimes referred to as the mixing matrix.

There are several algorithms for performing ICA. These algorithms have inter-relationships and are equivalent with the addition of some technical criteria. There is a maximum likelihood formalism of ICA called Maximum Likelihood ICA in which a likelihood function for the mixing matrix is written down and standard ML techniques are employed. There is a technique called InfoMax which measures the joint entropy of the signals, and uses the fact that a set of signals which has maximum joint entropy is mutually independent. There is a technique called FastICA, which is independently derived but is similar to using stochastic gradient descent to maximize the likelihood, with a time-varying learning rate resulting in a fast solution. We use FastICA.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Suppose we know the pdf of the univariate sources by assumption. Call these pdfs $\mathrm{f_i}$. Suppose that $\mathrm{W} = (\mathrm{w_1}, \ldots \mathrm{w_m})^{\mathrm{T}}$ is the matrix A^{-1}. Then the log likelihood will take the form:

$$L = \sum_{t=1}^{h} \sum_{i=1}^{m} log\, f_i\left(w_i^{\mathrm{T}} x_t\right) + h\, log\, |\, det\, W|$$

The last term here is from the Jacobian to transform the random variables.

This formalism assumes we know the pdf of the univariate sources – of course this is not known in general. The sum of a collection of independent variables (given some nice properties like finite variance) will tend to be normally distributed thanks to the central limit theorem. Suppose we have a fixed unknown S and A, and we want to estimate one of the $\mathrm{w_i}$. Because of the central limit theorem, $\mathrm{w_i^T AS}$ will necessarily be more normal than all of the s unless it happens to be equal to one of them. So the collection of weights which maximizes the non-normality of $\mathrm{w_i^T AS}$ will recover the original source. It happens that for many purposes several high-kurtosis pdfs work approximately equivalently well to find non-normal sources.

The FastICA algorithm we use employs an approximation of the pdf shown below.

$$f(x) = \frac{1}{\alpha} log\, cosh(\alpha x)$$

This approximation tends to be a more robust measure of non-normality than other common measures, e.g. kurtosis.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

Appendix D: Generalized Autoregressive Conditional Heteroskedastic Normal Inverse Gaussian

We use Normal Inverse Gaussian (NIG) distribution to model the independent univariate time series. By assumption both the idiosyncratic risk and independent components are mutually independent. The NIG distribution is a special case of generalized hyperbolic (GH) distributions, and which have five parameters to capture the heaviness and skewness of a data set. It is possible to show that setting λ= -1/2 in the GH distribution, we obtain an NIG distribution. We estimate the NIG parameters with maximum likelihood, relying on the independence property to assume that the sum of the log-likelihoods is the log-likelihood of the joint distribution.

The density expression for the NIG distribution is:

$$f(x;\alpha,\beta,\mu,\delta) = \frac{\alpha\delta}{\pi} e^{\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)} \frac{\boldsymbol{K}_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\sqrt{\delta^2+(x-\mu)^2}}$$

The distribution has the following parameters:
μ: location parameter, for real μ
δ: scale parameter, for non-negative δ
α: tail heaviness parameter, and
β: asymmetry parameter for $0 \leq |\beta| \leq \alpha$.

We also use an auxiliary parameter $\gamma = \sqrt{\alpha^2 - \beta^2}$.

Here $\mathbf{K}_1$ is the modified Bessel function of the third order and index 1.

If x is NIG distributed, the shape parameters for a random variable s of the distribution are

$$E[s] = \mu + \frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}} = \mu + \frac{\delta\beta}{\gamma}$$

$$Var[s] = \frac{\delta\alpha^2}{\left(\sqrt{\alpha^2-\beta^2}\right)^3} = \frac{\delta\alpha^2}{\gamma^3}$$

$$Skew[s] = \frac{3\beta}{\alpha\sqrt{\delta\sqrt{\alpha^2-\beta^2}}} = \frac{3\beta}{\alpha\sqrt{\delta\gamma}}$$

$$Kur[s] = \frac{3(1+4\beta^2/\alpha^2)}{\delta\sqrt{\alpha^2-\beta^2}} + 3 = \frac{3(1+4\beta^2/\alpha^2)}{\delta\gamma} + 3$$

NIG models are by themselves insufficient to characterize the univariate time series. The volatility of real return series is not constant over time, whereas the second moment for f above, which does not include any autocorrelation, is constant. We want to capture behavior typical of financial markets, such as volatility clustering and the auto-correlated volatilities. Furthermore we want a model with few

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

parameters which is likely to be numerically robust. This model is in essence a simplified version of the model in Jensen and Lunde (2001) as we outline below. The simplification makes the model robust for estimation purposes.

We use a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) process to model the dynamic nature of our NIG distributed returns. This work is related to existing GARCH-NIG and ARCH-NIG models. Work which fits GARCH(1,1) models, e.g. Jensen and Lunde (2001) typically finds the volatility update equation which has the following form

$$\sigma_{t+1}^2 = a + b\sigma_t^2 + c\epsilon_t^2 .$$

Start by re-parameterizing the NIG distribution:

$$\bar{\alpha} = \delta\alpha$$
$$\bar{\beta} = \delta\beta$$
$$\bar{\gamma} = \sqrt{\bar{\alpha}^2 - \bar{\beta}^2} = \delta\gamma$$

This gives an alternative density function g and we say a random variable is "NIG"($\bar{\alpha}$,$\bar{\beta}$,μ,δ) distributed if it has this following density.

$$g(x;\bar{\alpha},\bar{\beta},\mu,\delta) = \frac{\bar{\alpha}}{\pi\delta} exp\left(\bar{\gamma} + \bar{\beta}\frac{(x-\mu)}{\delta}\right)\frac{\boldsymbol{K}_1\left(\bar{\alpha}\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right)}{\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}}$$

Given this rewrite, if s is a random variable we can write the conditional moments (omitting subscripts) as follows. Note that the skewness and kurtosis can be expressed only in terms of$\bar{\alpha}, \bar{\beta}$ and $\bar{\gamma}$.

$$E[s] = \mu + \frac{\delta\beta}{\gamma} = \mu + \frac{\delta\bar{\beta}}{\bar{\gamma}}$$

$$Var[s] = \frac{\delta\alpha^2}{\gamma^3} = \frac{\delta^4\alpha^2}{\bar{\gamma}^3} = \frac{\delta^2\bar{\alpha}^2}{\bar{\gamma}^3}$$

$$Skew[s] = \frac{3\beta}{\alpha\sqrt{\delta\gamma}} = \frac{3\bar{\beta}}{\bar{\alpha}\sqrt{\bar{\gamma}}}$$

$$Kur[s] = \frac{3(1+4\beta^2/\alpha^2)}{\delta\gamma} + 3 = \frac{3\left(1+4\bar{\beta}^2/\bar{\alpha}^2\right)}{\bar{\gamma}} + 3$$

We use a GARCH approximation, and allow the volatility σto progress as follows.

$$\sigma_t^2 = a + bb\sigma_{t-1}^2 + (1-\lambda)\epsilon^2{}_{t-1}c\epsilon_{t-1}^2$$

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

We then set:

$$s_t = \mu_t + \frac{\sqrt{\bar{\gamma}}\bar{\beta}}{\bar{\alpha}}\sigma_t + \sigma_t\eta_t$$

$$\eta_t \sim NIG(\bar{\alpha}, \bar{\beta}, -\frac{\sqrt{\bar{\gamma}}\bar{\beta}}{\bar{\alpha}}, \frac{\bar{\gamma}^{3/2}}{\bar{\alpha}})$$

$$\epsilon_t = \sigma_t\eta_t$$

Here s_t is a factor or residual innovation.

Why this parameterization? First notice that now everywhere that x occurs, it occurs in the context of $(x - \mu)/\delta$. This means that if $X \sim NIG(\bar{\alpha}, \bar{\beta}, \mu, \delta)$, then $(X - \mu)/\delta \sim NIG(\bar{\alpha}, \bar{\beta}, 0, 1)$. It then becomes possible to push all the time-varying portion of the distribution into δ, and leave $\bar{\alpha}, \bar{\beta}$ constant. The idea is to give η_t mean of 0 and variance of 1, which comes about as follows:

$$E[\eta_t] = \mu_t + \frac{\bar{\beta}}{\bar{\alpha}}\frac{\delta_t}{\sqrt{1 - (\bar{\beta}/\bar{\alpha})^2}}$$

$$= -\frac{\sqrt{\bar{\gamma}}\bar{\beta}}{\bar{\alpha}} + \frac{\bar{\beta}}{\bar{\alpha}}\frac{\bar{\gamma}^{3/2}/\bar{\alpha}}{\sqrt{1 - (\bar{\beta}/\bar{\alpha})^2}}$$

$$= 0$$

For variance we have:

$$VAR[\eta_t] = \frac{\delta\alpha^2}{\gamma^3} = \frac{\delta^4\alpha^2}{\bar{\gamma}^3} = \frac{\delta^2\bar{\alpha}^2}{\bar{\gamma}^3}$$

$$= \left(\frac{\bar{\gamma}^{\frac{3}{2}}}{\bar{\alpha}}\right)^2 \frac{\bar{\alpha}^2}{\bar{\gamma}^3}$$

$$= 1$$

This means in turn that:

$$s_t \sim NIG(\bar{\alpha}, \bar{\beta}, \mu_t, \sigma_t\frac{\bar{\gamma}^{\frac{3}{2}}}{\bar{\alpha}})$$

$$VAR[s_t|\sigma_{t-1}] = \sigma_t^2 = a + b\sigma_{t-1}^2 + c\epsilon_{t-1}^2$$

$$E[s_t|\sigma_{t-1}] = \mu_t + \sigma_t\frac{\sqrt{\bar{\gamma}}\bar{\beta}}{\bar{\alpha}} = 0$$

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

For most of our time series we subtract the mean from the dataset before fitting the distribution and so essentially assume that the unconditional expected return is zero, which gives that:

$$\mu_t = -\sigma_t \frac{\sqrt{\bar{\gamma}}\bar{\beta}}{\bar{\alpha}}$$

Alongside the pdf from the last section, a log-likelihood for particular sequences of volatilities and NIG parameters is straightforward, and the parameters of the distribution can be estimated with the following steps:

1. Take as inputs $s_0, \ldots, s_T$, and an initial guess for $\sigma_0, \bar{\gamma}, \bar{\alpha}, \bar{\beta}$ (we set μ_0 from d below for time 0).
2. Repeat for all $i > 0$:
   a. $$\epsilon_{i-1} = s_{i-1} - \sigma_{i-1}\frac{\sqrt{\bar{\gamma}}\bar{\beta}}{\bar{\alpha}} - \mu_{i-1} = s_{i-1}$$
   b. $$\sigma_i^2 = a + b\sigma_{i-1}^2 + c\epsilon_{i-1}^2$$
   c. $$\mu_i = -\sigma_i \frac{\sqrt{\bar{\gamma}}\bar{\beta}}{\bar{\alpha}}$$
3. Calculate the log-likelihood as:

$$\mathcal{L}(\bar{\alpha}, \bar{\beta}; \sigma_0, \boldsymbol{s}) = \sum_{i=0}^{T} log\, g(s_i; \bar{\alpha}, \bar{\beta}, \mu_i, \sigma_i \frac{\bar{\gamma}^{3/2}}{\bar{\alpha}})$$

Note that if $E[s_t]$ is assumed to be zero, then the future volatility estimate depends only on the historical returns and volatility, and the μ_i are only required to calculate the likelihood. We can then maximize this log-likelihood with a suitable optimizer, we use BFGS.

©2016 Morningstar, Inc. All rights reserved. The information in this document is the property of Morningstar, Inc. Reproduction or transcription by any means, in whole or part, without the prior written consent of Morningstar, Inc., is prohibited.

About Morningstar® Quantitative Research™

Morningstar Quantitative Research is dedicated to developing innovative statistical models and data points, including the Quantitative Equity Ratings and the Morningstar Global Risk Model.

For More Information

+1 312 244-7541
lee.davidson@morningstar.com



22 West Washington Street
Chicago, IL 60602 USA

©2016 Morningstar. All Rights Reserved. Morningstar's Credit Ratings & Research is produced and offered by Morningstar, Inc., which is not registered with the U.S. Securities and Exchange Commission as a Nationally Recognized Statistical Rating Organization ("NRSRO"). Unless otherwise provided in a separate agreement, you may use this report only in the country in which its original distributor is based. The information, data, analyses and opinions presented herein do not constitute investment advice; are provided solely for informational purposes and therefore are not an offer to buy or sell a security; and are not warranted to be correct, complete or accurate. The opinions expressed are as of the date written and are subject to change without notice. Except as otherwise required by law, Morningstar shall not be responsible for any trading decisions, damages or other losses resulting from, or related to, the information, data, analyses or opinions or their use. The information contained herein is the proprietary property of Morningstar and may not be reproduced, in whole or in part, or used in any manner, without the prior written consent of Morningstar. To order reprints, call +1 312-696-6100. To license the research, call +1 312-696-6869.