Morningstar Risk Model
Methodology

Morningstar Research
May 2023

Contents

1 Introduction
2 Model Highlights
3 Universe Construction
3 Factor Selection
4 Factor Premium Estimation
5 Forecast Factor Comovement and Residual Volatility
7 Local-Currency Versions of Risk Models
7 Conclusion
8 References

Appendix A
Risk Model Definitions
Appendix B
Estimation Universe Construction Rules
Appendix C
Equity Factor Exposure Definition
Appendix D
Fixed-Income Factor Definition
Appendix E
Cross-Sectional Regression
Appendix F
Forecast Factor Comovement and Residual Volatility
Appendix G
Frequently Asked Questions

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 portfolios. 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 forecast distribution widens, it indicates more uncertainty about the future return potential of the portfolio. As tail probabilities increase, it 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.

The model seeks to identify a small number of independent, latent sources of return. Movements in these sources drive movement in a comparably small number of interpretable factors. An example of a factor is the exposure to a particular currency: For instance, how much does an increase in the euro/U.S. dollar exchange rate drive an increase in the value of a stock or bond? 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 differentiated, interpretable, responsive, and predictive model. We began with the following assumptions about asset returns, which shaped our methodological choices.

- A small number of independent sources of market movement 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 not controversial, although some or all of them are often ignored for convenience by risk-modeling practitioners.

Model Highlights

Several features make the Morningstar Risk Models unique:

1) We are holdings-based.

Our risk models are entirely holdings-based. When looking at portfolios, holdings-based models will provide more accurate outputs for risk prediction, factor attribution, risk decomposition, and sensitivity analysis. Holdings-based models do not assume that the past equals the future as they allow for the fact that securities may change, managed products may change, and portfolios may change over time. They also enable new securities or funds to be covered immediately.

2) We can forecast the full probability distribution of future returns with non-normal distributions.

Our risk models are agnostic to any particular risk metric a user wishes to employ. With the advanced features of our model, 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 models.

3) We use proprietary fundamentals-based factors that 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.

4) We make no assumption that comovement 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 comovements of returns, enabling the construction of portfolios that can control tail risk.

5) We can customize each methodological decision at scale.

Historically, risk model users are reliant on the decisions of the risk model providers. With Morningstar's Risk Model technology platform, we can construct and build entire histories of new models within a matter of hours. We have deployed a suite of risk models specific to asset class, region, and currency and can build customized models for clients. Descriptions for each risk model are found in Appendix A.

6) We offer integrated and robust risk analysis workflows.

While risk models themselves offer exposures, premia, and forecasts, these outputs are usually most valuable when placed within other workflows or modules. Morningstar offers users the ability to decompose risk or attribute returns to factors and holdings through time and across many instruments. Morningstar also offers a full complement of scenario analysis capabilities including historical scenarios, predefined macro-financial scenarios, or market-driven scenarios.

Universe Construction

We define an estimation universe of investable companies with reliable data on which to build the model. Securities outside the estimation universe—generally illiquid assets with small market capitalizations—are relegated to the extended universe. We use only securities in the estimation universe to derive model parameters. This ensures the model parameters are not influenced by illiquid assets with unreliable data.

Exhibit 1 Estimation and Coverage Universe for the Morningstar Global Equity Risk Model

Estimation Universe	Coverage Universe
Approximately 11,000 stocks	Approximately 44,000 stocks
(Curated broad group of large, liquid stocks)	(Small, illiquid stocks)

Source: Morningstar.

We aim for a broad selection of companies across countries that are liquid and large enough to be investable for local and international investors. Our liquidity and market capitalization thresholds are time-varying depending on the market condition, which ensures the estimation universe selects investable securities at any time point. Appendix B details the exact rules we use to filter our estimation universe for each model.

All market returns and factor inputs are converted into a denomination of the model currency when appropriate. For example, the outputs of the Morningstar United Kingdom Equity Risk Model measure risk with a British pound..

Factor Selection

There are many ways to estimate the comovement 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; 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 comovement.

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 comovement.

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

- An **asset return** is the return of an investable security over a 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 that represents how much a particular factor has influenced asset returns for a particular 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:

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

Each model has a specific list of factors tailored to the model's asset class, region, and currency. For equity securities, the factors fall naturally into five distinct groups: style, sector, region, currency, and an equity market factor. The equity market factor results from our estimation methodology and captures the common equity market movement globally or for a specific region. For fixed-income securities, we currently group factors into duration and credit. A detailed treatment for each factor can be found in Appendix C.

Factor Premium 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, to estimate the factor premium f_t.

$$r_t = X_t f_t + \varepsilon_t$$

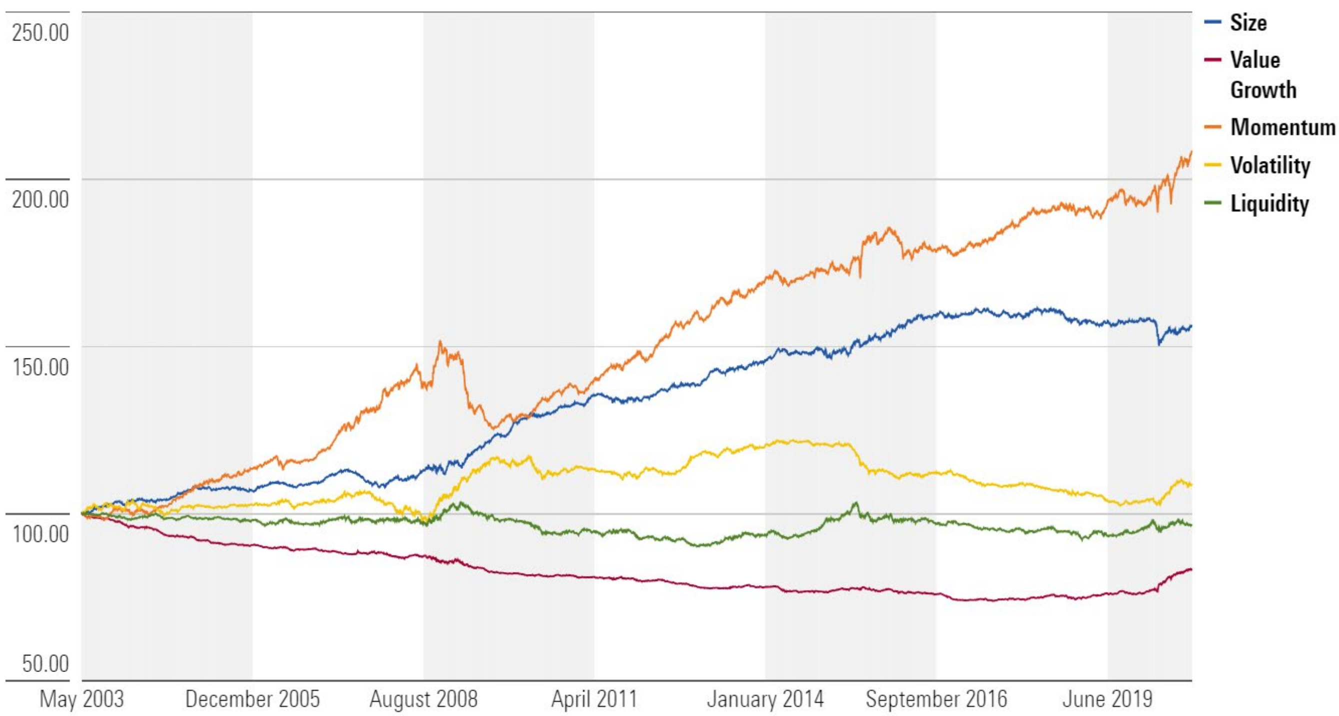
Where

$r_t = (n \times 1)\ vector\ of\ returns\ between\ time\ t\ and\ t+1$
$X_t = (n \times m)\ matrix\ of\ securities' exposures\ to factors\ at\ time\ t$
$f_t = (m \times 1)\ vector\ of\ factor\ premiums\ between\ time\ t\ and\ t+1$
$\varepsilon_t = (n \times 1)\ vector\ of\ error\ terms\ between\ time\ t\ and\ t+1$

To improve the statistical properties of the estimated factor premia, the exposure table contains an intercept term and a column of 1, and certain conditions are imposed on premium estimates using a constrained regression. A detailed description of the methodology can be found in Appendix E.

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

Exhibit 2 Historical Time Series of the Factor Premia for the Morningstar Global Equity Risk Model in USD



Source: Morningstar.

Exhibit 2 shows the cumulative return of the Morningstar Global Equity Risk Model style factor premia in U.S. dollars for the past five years.

Forecast Factor Comovement and Residual Volatility

Morningstar Risk Model supports a variety of methods to forecast the comovement of the factor premia. The simplest method is to estimate the sample covariance matrix of the factor premia. This rests on the assumption that the covariances between factors are stable over time. To accommodate time-varying covariance structure, another conventional method is to overweight more recent observations by exponentially weighting the observations, where a parameter "half-life" controls how quickly the impact of earlier observations is reduced. We can further enhance the flexibility of the method by using different half-life parameters in estimating the correlation matrix and the variance of premia and combining the two to create a covariance matrix. Longer horizons typically improve the estimation of the correlation structure, and shorter horizons tend to be better for variance forecast. Note that different optimized parameter settings can be applied to support specific applications.

In addition to factor premia, the cross-sectional regression produces residual terms for each stock in a particular period, which represents the return not explained by the fundamental factors. We model the volatility of this idiosyncratic portion of return using an exponentially weighted sample variance. The half-life parameter is typically set at short horizons, such as one month, to reflect the observation that the idiosyncratic risk of stocks is more responsive to short-term events.

In addition to this conventional covariance-matrix-based volatility estimation, we also support a method for directly estimating higher moments, which is described in another document.

Exponentially Weighted Moving Average Volatility Forecasts

In the exponentially weighted moving average, or EWMA, volatility forecast, the more recent returns are given higher weight in the estimation. A portfolio's variance at time t, V_t^P, is modeled as $V_t^P = (x_t^P)^T \cdot F_t \cdot x_t^P + (w_t^P)^T \cdot \Delta_t^{\mathrm{p}} \cdot w_t^P$, where the x_t^P are the portfolio's net exposures to the risk factors, w_t^P are the portfolio's holdings weights, F_t is a factor premia EWMA covariance matrix estimate, and Δ_t^P is the diagonal matrix with residual EWMA variance estimates along its diagonal. With a portfolio's net exposures and holdings weights assumed constant, the portfolio's EWMA volatility forecast over a forecast horizon $[t + 1, t + H]$ (that is, a horizon of size H) is then $\sqrt{H \times V_t^P}$.

The factor premia EWMA covariance estimate, F_t, is formed by combining the premia EWMA correlation matrix estimate, C_t, with the EWMA standard deviation estimates, so that $F_t = \Sigma_t \cdot C_t \cdot \Sigma_t$, where Σ_t is a diagonal matrix with premia standard deviations along its diagonal. C_t and Σ_t are estimated separately with a different half-life and a window of historical data. The correlation matrix has more parameters, and longer half-life and trailing window improves the accuracy of estimate, whereas the standard-deviation estimates benefit from a shorter half-life to allow for more timely response to market conditions.

The EWMA covariance estimate is further enhanced through two processes. First, we account for the impact of autocorrelation in factor premia by applying Newey-West adjustments to both the correlation matrix estimate C_t and the premia standard deviation estimate Σ_t. Second, the volatility forecast is scaled to remove the bias in forecast. The bias statistic applies the Mahalanobis distance to measure the gap between the forecast covariance matrix and the realized covariance in recent periods.

The residual variance of each security Δ_t^i, is first estimated as EWMA variance over a 300-day historical window. If a security's residual variance is missing (mostly due to insufficient returns), a cross-sectional regression is used to impute a variance estimate. This EWMA estimate is then adjusted to remove any bias in forecast measured against the realized volatility over a one-day period. Next, this adjusted residual volatility forecast is further modified to account for the impact of autocorrelation in residual returns. Specifically, a Newey-West adjustment and a bias correction measured against the 20-day realized volatility are applied. Note the portfolio residual volatility forecast is the weighted sum of the individual securities residual volatility forecast, and the residual returns across securities are assumed to have no correlation.

The half-lives used in the EWMA estimates were selected to maximize the likelihood of observed premia and residuals over a given forecast horizon via back-testing. Further details of the method are given in Appendix F, along with the optimal half-lives, for different forecast horizons.

Local-Currency Versions of Risk Models

Risk models must be interpretable by investors in their own local currencies. Morningstar approaches this problem straightforwardly by recalculating factor exposures and by re-estimating factor premia using local-currency returns in separate model runs. For example, momentum exposures may use EUR-based returns when a euro model version is chosen and may use U.S. dollar-based returns when a U.S. dollar model version is chosen. These exposures are built from the ground up using the local-currency returns. In the same vein, euro-based returns are used in the factor premia estimation when a euro model version is chosen, and U.S. dollar-based returns are used in the factor premia estimation when a U.S. dollar model version is chosen.

While this implies that the factor exposures and factor premia may differ between local-currency versions of the same model, it does represent the most direct approach to building a risk model artifact that a local investor would desire.

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 important, 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.

Some of our risk models include factors unique to Morningstar. These factors that spring from our analyst-driven research and our institutional portfolio holdings database are uncorrelated with more-traditional factors and are helpful tools for investors to use when tailoring their portfolio to suit their risk preferences. In addition, we also include some more-standard, academically backed factors in our risk models.

No risk model is perfect. Our aim has been to emphasize interpretability, responsiveness, and predictive accuracy, and in doing so, we believe we have developed a unique framework for building risk models. We recognize there are many decisions to make when constructing a risk model: from universe selection to individual factor calculations to forecasting methods. Our framework allows us to quickly spin up new models to better match the model with our users' definition of risk.

References

Chen, Y., Härdle, W., & Spokoiny, V. 2007. "Portfolio Value at Risk Based on Independent Component Analysis." *J. Computational and Applied Mathematics,* Vol. 205, No.1, P. 594.

Chen, Y., Härdle, W., & Spokoiny, V. 2010. "GHICA—Risk Analysis With GH Distributions and Independent Components." *J. Empirical Finance,* Vol. 17, No. 2, P. 255.

Chin, E., Weigend, A.S., & Zimmermann, H. 1999. "Computing Portfolio Risk Using Gaussian Mixtures and Independent Component Analysis. Computational Intelligence for Financial Engineering." Proc. IEEE/IAFE 1999 Conf.

Cover, T.M., & Thomas, J.A. 2012. *Elements of Information Theory.* (New York: Wiley)

Dor, A.B., Dynkin, L., Hayman, J., Houweling, P., Leeuwen, E.V., & Penninga, O. 2007. "DTS (Duration Times Spread)." *J. Portfolio Management,* Vol. 33, No. 2.

Ghalanos, A., Rossi, E., & Urga, G. 2015. "Independent Factor Autoregressive Conditional Density Model." *Econometric Reviews,* Vol. 34, No. 5, P. 594.

Heston, S.L., & Rowenhorst, K.G, 1994, "Does Industry Structure Explain the Benefits of International Diversification?" *J. Financial Economics*, Vol. 26, No.1, P. 3.

Hyvärinen, A., Karhunen, J., & Oja, E. 2004. *Independent Component Analysis, Vol. 46.* (New York: Wiley)

Hull, J.C., 2009. *Options, Futures, and Other Derivatives.* (New Jersey: Prentice Hall)

Jensen, M.B., & Lunde, A. 2001. "The NIG‐S&ARCH Model: A Fat-Tailed, Stochastic, and Autoregressive Conditional Heteroskedastic Volatility Model." *Econometrics J.,* Vol. 4, No. 2, P. 319.

Madan, D.B., & Yen, J.-Y.J. 2007. "Asset Allocation With Multivariate Non-Gaussian Returns." *Handbooks in Operations Research and Management Science,* Vol. 15, P. 949.

Morningstar. 2019. "Morningstar Quantitative Equity Ratings Methodology." https://www.morningstar.com/content/dam/marketing/shared/pdfs/Research/QER_Methodology-08-22-2019final.pdf

Morningstar. 2018. "Morningstar Style Box™ Methodology." https://www.morningstar.com/content/dam/marketing/shared/research/methodology/678263-StyleBoxMethodolgy.pdf

Nelson, C.R., & Siegel, A.F. 1987. "Parsimonious Modeling of Yield Curves." *J. Business*, Vol. 60, No. 4, P. 473.

Shephard, N.G. 1991. "From Characteristic Function to Distribution Function: A Simple Framework for the Theory." *Econometric Theory,* Vol. 7, No. 4, P. 519.

Sustainalytics ESG Risk Research. 2018. https://www.sustainalytics.com/esg-ratings/

Appendix A: Morningstar Risk Model Definitions

Morningstar Global Equity Risk Model

The Morningstar Global Equity Risk Model captures risk premia across the global equity universe.

Factors

The model is defined by 37 factors across style, sector, region, and currency.

- Equity Market Factor
- Style: Economic Moat, Financial Health, Liquidity, Momentum, Ownership Risk, Ownership Popularity, Size, Valuation, Valuation Uncertainty, Value-Growth, Volatility
- Sector: Basic Materials, Energy, Financial Services, Consumer Defensive, Consumer Cyclical, Technology, Industrials, Healthcare, Communication Services, Real Estate, Utilities
- Region: Developed Americas, Developed Europe, Developed Asia Pacific, Emerging Americas, Emerging Europe, Emerging Asia Pacific, Emerging Middle East
- Currency: Australian dollar, British pound, Canadian dollar, euro, Japanese yen, New Zealand dollar, Swiss franc

Data Availability

The model generates daily data from Jan. 1, 2003, to the present day.

Currency

The model is available in five currencies: Australian dollar, British pound, Canadian dollar, euro, Hong Kong dollar, Japanese yen, Singapore dollar, South African rand, Swiss franc, and U.S. dollar.

Morningstar Standard Factor Model

The Morningstar Standard Factor Model captures risk premia across the global equity universe using industry-standard style factors.

Factors

The model is defined by 33 factors across style, sector, region, and currency.

- Style: Yield, Size, Volatility (standard model), Quality, Liquidity, Value-Growth (standard model), Momentum
- Sector: Basic Materials, Energy, Financial Services, Consumer Defensive, Consumer Cyclical, Technology, Industrials, Healthcare, Communication Services, Real Estate, Utilities
- Region: Equity Market Factor, Developed Americas, Developed Europe, Developed Asia Pacific, Emerging Americas, Emerging Europe, Emerging Asia Pacific, Emerging Middle East
- Currency: Australian dollar, British pound, Canadian dollar, euro, Japanese yen, New Zealand dollar, Swiss franc

Data Availability

The model generates daily data from Jan. 1, 2003, to the present day.

Currency

The model is available in U.S. dollar.

Morningstar Global Multi-Asset Risk Model

The Morningstar Global Multi-Asset Risk Model captures risk premia across global equity and fixed income.

Factors

The model is defined by 62 factors across style, sector, region, currency, and rates.

- Style: Yield, Size, Volatility (standard model), Quality, Liquidity, Value-Growth (standard model), Momentum
- Sector: Basic Materials, Energy, Financial Services, Consumer Defensive, Consumer Cyclical, Technology, Industrials, Healthcare, Communication Services, Real Estate, Utilities
- Region: Equity Market Factor, Developed Americas, Developed Europe, Developed Asia Pacific, Emerging Americas, Emerging Europe, Emerging Asia Pacific, Emerging Middle East
- Currency: Australian dollar, British pound, Canadian dollar, euro, Japanese yen, New Zealand dollar, Swiss franc
- Yield Curve: Duration (USD, EUR, GBP, CHF, and CAD), Convexity (USD, EUR, GBP, CHF, and CAD)
- Carry
- Spread Factors (see Appendix D for definition)

Data Availability

The model generates daily data from Jan. 1, 2003, to the present day.

Currency

This model is available in the U.S. dollar currency.

Morningstar United Kingdom Equity Risk Model

The Morningstar United Kingdom Equity Risk Model captures equity factor risk premia in the United Kingdom.

Factors

The model is defined by 30 factors across style, sector, and currency.

- Equity market factor
- Style: Economic Moat, Financial Health, Liquidity, Momentum, Ownership Popularity, Ownership Risk, Size, Valuation, Valuation Uncertainty, Value-Growth, Volatility
- Sector: Basic Materials, Telecommunications, Consumer Cyclical, Consumer Defensive, Healthcare, Industrials, Real Estate, Technology, Energy, Financial Services, Utilities
- Currency: Australian dollar, Canadian dollar, euro, Japanese yen, New Zealand dollar, Swiss franc, U.S. dollar

Data Availability

The model generates daily data from Jan. 1, 2006, to the present day.

Currency

This model is available in British pound and U.S. dollar.

Morningstar Eurozone Equity Risk Model

The Morningstar Eurozone Equity Risk Model captures equity factor risk premia across the eurozone region.

Factors

The model is defined by 23 factors across style, sector, and currency.

- Equity market factor
- Style: Economic Moat, Financial Health, Momentum, Size, Value-Growth, Volatility
- Sector: Basic Materials, Telecommunications, Consumer Cyclical, Consumer Defensive, Healthcare, Industrials, Real Estate, Technology, Energy, Financial Services, Utilities
- Currency: Australian dollar, British pound, Japanese yen, Swiss franc, U.S. dollar

Data Availability

The model generates daily data from Jan. 1, 2006, to the present day.

Currency

This model is available in euro and U.S. dollar.

Morningstar North America Standard Factor Model

The Morningstar North America Standard Factor Model captures risk premia across the North America region.

Factors

The model is defined by 26 factors across style, sector, currency, and equity market.

- Equity Market Factor
- Style: Yield, Size, Volatility (standard model), Quality, Liquidity, Value-Growth (standard model), Momentum
- Sector: Basic Materials, Energy, Financial Services, Consumer Defensive, Consumer Cyclical, Technology, Industrials, Healthcare, Communication Services, Real Estate, Utilities

Data Availability

The model generates daily data from Jan. 1, 2006, to the present day.

Currency

The model is available in U.S. dollar and Canadian dollar.

Morningstar Developed Europe Equity Risk Model

The Morningstar Developed Europe Equity Risk Model captures risk premia across the developed Europe region.

Factors

The model is defined by 26 factors across style, sector, currency, and equity market.

- Equity Market Factor
- Style: Economic Moat, Financial Health, Liquidity, Momentum, Size, Value-Growth, Volatility, Ownership Popularity, Ownership Risk
- Sector: Basic Materials, Energy, Financial Services, Consumer Defensive, Consumer Cyclical, Technology, Industrials, Healthcare, Communication Services, Real Estate, Utilities
- Currency: British pound, U.S. dollar, Japanese yen, Australian dollar, Swiss franc

Data Availability

The model generates daily data from Jan. 1, 2008, to the present day.

Currency

The model is available in euro and U.S. dollar.

Morningstar Japan Equity Risk Model

The Morningstar Japan Equity Risk Model captures equity factor risk premia in Japan.

Factors

The model is defined by 24 factors across style, sector, currency, and equity market.

- Equity market factor
- Style: Economic Moat, Financial Health, Liquidity, Momentum, Size, Value-Growth, Volatility (standard model)
- Sector: Basic Materials, Telecommunications, Consumer Cyclical, Consumer Defensive, Healthcare, Industrials, Real Estate, Technology, Energy, Financial Services, Utilities
- Currency: Australian dollar, British pound, euro, Swiss franc, U.S. dollar

Data Availability

The model generates daily data from Jan. 1, 2010, to the present day.

Currency

This model is available in Japanese yen.

Morningstar Global Equity Risk Model—ESG

The Morningstar Global Equity Risk Model—ESG captures the risk premia across the global equity universe for 38 factors. The model includes an ESG (environmental, social, and governance) Rating factor that measures how well a company manages ESG-related issues according to Sustainalytics ESG rating criteria. A higher exposure indicates better management of ESG issues.

Factors

The model is defined by 38 factors across style, sector, region, and currency.

- Equity Market Factor
- Style: Economic Moat, Financial Health, Liquidity, Momentum, Ownership Risk, Ownership Popularity, Size, Valuation, Valuation Uncertainty, Value-Growth, Volatility, ESG Rating
- Sector: Basic Materials, Energy, Financial Services, Consumer Defensive, Consumer Cyclical, Technology, Industrials, Healthcare, Communication Services, Real Estate, Utilities
- Region: Developed Americas, Developed Europe, Developed Asia Pacific, Emerging Americas, Emerging Europe, Emerging Asia Pacific, Emerging Middle East
- Currency: Australian dollar, British pound, Canadian dollar, euro, Japanese yen, New Zealand dollar, Swiss franc

Data Availability

The model generates daily data from Aug. 17, 2009, to the present day.

Currency

The model is available in U.S. dollar.

Morningstar Emerging-Markets Equity Risk Model

The Morningstar Emerging-Markets Equity Risk Model captures risk premia across the emerging markets in Latin America, Asia-Pacific, Europe, the Middle East, and Africa.

Factors

The model is defined by 30 factors across style, sector, currency, and equity market.

- Equity Market Factor
- Style: Economic Moat, Financial Health, Liquidity, Momentum, Size, Value-Growth, Volatility, Ownership Popularity, Ownership Risk
- Sector: Basic Materials, Energy, Financial Services, Consumer Defensive, Consumer Cyclical, Technology, Industrials, Healthcare, Communication Services, Real Estate, Utilities
- Region: Emerging Americas, Emerging Europe, Emerging Asia Pacific, Emerging Middle East and Africa
- Currency: Australian dollar, British pound, euro, Japanese yen, Swiss franc

Data Availability

- The model generates daily data from Jan. 1, 2008, to the present day.

- **Currency**
- The model is available in U.S. dollar.

Appendix B: Estimation Universe Construction Rules

We outline the estimation universe logic for each model and provide an illustration for the logic below.

Morningstar Global Equity Risk Model

For all securities, we calculate the liquidity and percentile rank of market capitalization at world and country levels at daily frequency. Liquidity is the median dollar trading volume over the past 91 trailing days. We define the percentile rank of market capitalization as {total market capitalization of companies whose market capitalization is greater than or equal to the company in question}/{market capitalization of all companies}*100, where the stock with the largest market capitalization gets the smallest percentile rank. We include the stock of a company into the estimation universe if it satisfies the following four requirements:

- It is among the most-liquid 60% of global stocks
- It is among the most-liquid 60% of country stocks
- It has a percentile rank of global-market capitalization $\leq$ 98.5
- It has a percentile rank of country-market capitalization $\leq$ 97

For the U.S. stocks, if the number of stocks that satisfies the four criteria is less than 2,000, we add the most liquid U.S. stocks back into the universe to achieve this total. These U.S. stocks must have a U.S.-size rank $\leq$ 2,000 and have a percentile rank of U.S.-market capitalization $\leq$ 99. The size rank is a rank based on the sum of a stock's market-capitalization rank, where the stock with largest market capitalization gets a rank of 1, and its liquidity rank, where the stock with highest liquidity gets a rank of 1. The stock with the lowest sum is given a size rank of 1.

Further, we include shares having a foreign ownership limit, like China A stocks, and we do not use free-float-adjusted market cap.

Morningstar Standard Factor Model

Equity

The equity estimation universe follows the same logic as the Morningstar Global Equity Risk Model. The one key difference is that the Standard model standardizes equity exposures at the region level whereas the Global Equity model standardizes at the global level.

Morningstar Global Multi-Asset Risk Model

Equity

The equity estimation universe follows the same logic as the Morningstar Global Equity Risk Model.

Fixed Income

There is no estimation universe for the fixed-income portion of the risk model. Factor premiadefinition is in Appendix D.

Morningstar Global Equity Risk Model—ESG

The equity estimation universe construction applies the same logic as those of the Morningstar Global Equity Risk Model among the stocks with Sustainalytics ESG Rating. The coverage universe of the model is restricted to stocks with a Sustainalytics ESG Rating. Please refer to the ESG model white paper for more details.

Morningstar United Kingdom Equity Risk Model

Requirements:

- Securities listed on London Stock Exchange and Alternative Investment Market
- Industry Classification is not Asset Management
- No ADRs

Filters:

- Market capitalization > GBP 1 million
- Liquidity > GBP 10,000
- Size rank ≤ 300
- Sector-size rank ≤ 30
- Sector-market capitalization coverage < 95%

Morningstar Eurozone Equity Risk Model

Requirements:

- Securities listed on following stock exchanges: Athens Exchange S.A. Cash Market, Boerse Berlin, Boerse Berlin—Freiverkehr, Boerse Berlin—Regulierter Markt, Boerse Frankfurt—Freiverkehr, Boerse Frankfurt—Regulierter Markt, Boerse Muenchen, Boerse Muenchen—Freiverkehr, Boerse Muenchen—Regulierter Markt, Boerse Stuttgart, Boerse Stuttgart—Freiverkehr, Boerse Stuttgart—Regulierter Markt, Borsa Italiana S.P.A., Cyprus Stock Exchange, Deutsche Boerse Ag, Deutsche Boerse Mid-Point Cross, Irish Stock Exchange, Istanbul Stock Exchange, Kanmon Shohin Torihikijo, Luxembourg Stock Exchange, Malta Stock Exchange, Mercado Continuo Espanol, Nagoya Seni Torihikijo (Textile Exchange)—Chubu Commodity Exchange, Nasdaq Omx Nordic, Nyse Euronext—Euronext Amsterdam, Nyse Euronext—Euronext Brussels, Nyse Euronext—Euronext Lisbon, Nyse Euronext—Euronext Paris, Osaka Seni Torihikijo (Textile Exchange), Posit—Asia Pacific, Stoxx European Indexes, Stoxx Indexes, Stoxx Limited—Volatility Indexes, Swiss Tax Authorities, Wiener Boerse Ag
- Securities from the following countries: Austria, Belgium, Cyprus, Germany, Spain, Estonia, Finland, France, Greece, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Portugal
- Industry Classification is not Asset Management

- No ADRs

Filters:

- Market capitalization > EUR 1 million
- Liquidity > EUR 10,000
- Size rank ≤ 550
- Sector-size rank ≤ 30
- Country-size rank ≤ 75
- Sector-country-size rank ≤ 2

Morningstar Europe Equity Risk Model

Requirements:

- Securities listed on following stock exchanges: Vienna Stock Exchange, Euronext Brussels, Copenhagen Stock Exchange, Helsinki Stock Exchange, Euronext Paris, Deutsche Borse Xetra, Frankfurt Stock Exchange, Borse Stuttgart, Borse Munchen, Borse Berlin, Irish Stock Exchange, Borsa Italiana, Euronext Amsterdam, Oslo Stock Exchange, Euronext Lisbon, Madrid Stock Exchange, Stockholm Stock Exchange, First North, Nordic Growth Market, SIX Swiss Exchange, London Stock Exchange
- Securities from the following countries: Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom
- Industry Classification is not Asset Management
- No ADRs

Filters:

- Market capitalization > EUR 10 million
- Liquidity > EUR 10,000
- Size rank ≤ 1,300
- Sector-size rank ≤ 100
- Country-size rank ≤ 75
- Sector-country-size rank ≤ 5

Morningstar North America Standard Factor Model

Requirements:

- Securities listed on following stock exchanges: New York Stock Exchange, Nasdaq, American Stock Exchange, NYSE Arca Exchange (NYSE—ARCA, OTC Pink (OTCPK), OTC Bulletin Board (OTCBB), Better Alternative Trading System (BATS), Toronto Stock Exchange, TSX Venture Exchange, Canadian National Stock Exchange
- No ADRs

Filters:

- Market capitalization > USD 50 million

- Liquidity > USD 10,000
- Size rank ≤ 2,500
- Sector-size rank ≤ 300
- Country-U.S.-size rank ≤ 2,000
- Sector-Country-size rank ≤ 300
- Sector-market-cap % Canada (Exclude) ≤ 0.55
- Sector-market-cap % USA (Exclude) ≤ 0.70

Morningstar Japan Equity Risk Model

Requirements:

- Securities listed on Tokyo Stock Exchange, Jasdaq, Osaka Securities Exchange, Nagoya Stock Exchange, Fukuoka Stock Exchange, Sapporo Securities Exchange
- Industry Classification is not Asset Management
- No ADRs

Requirements:

- Most-liquid 60% of Japan stocks
- Percentile rank of Japan-market capitalization ≤ 97

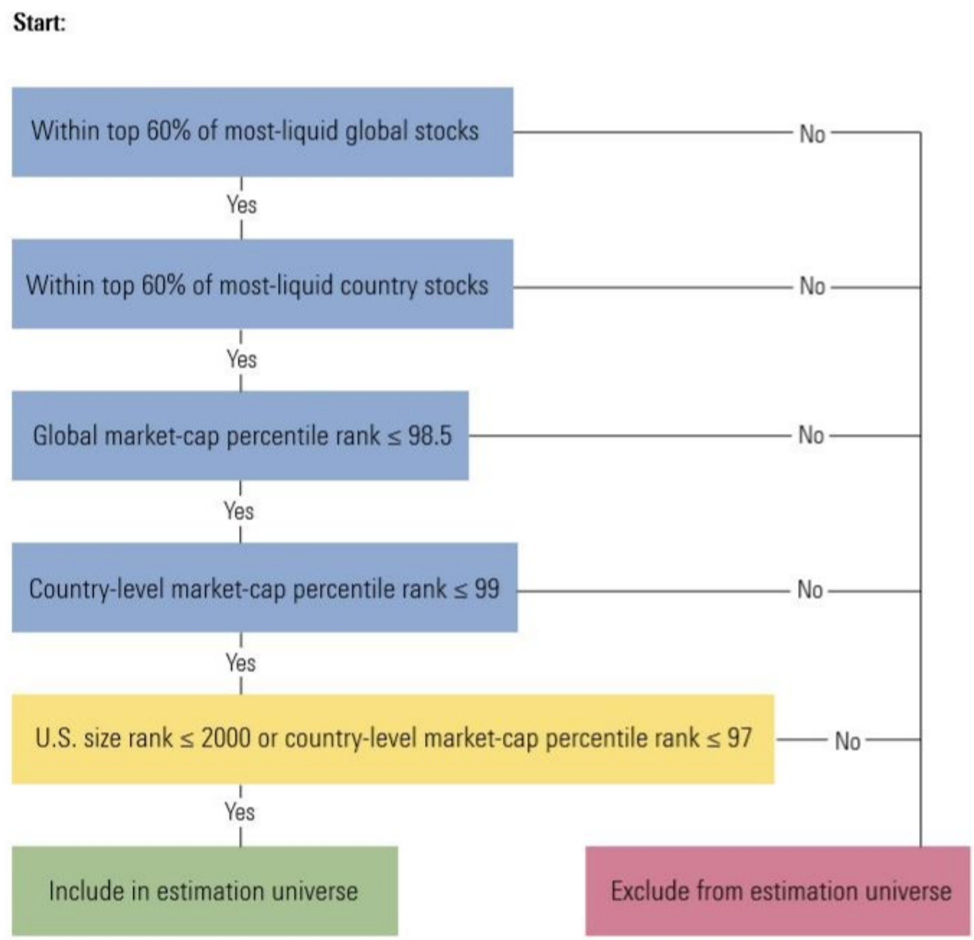
Morningstar Emerging Markets Equity Risk Model

The model uses the daily constituents of the Morningstar Emerging Markets Index as the estimation universe. The weights of the stocks in the estimation universe are directly obtained from the index. The construction rules of the index follow a logic similar to the one outlined for the Morningstar Global Equity Risk Model. Further details about the index, including detailed construction and maintenance rules, can be obtained from the Morningstar Indexes website.

An example of the estimation universe logic is depicted below:

Liquidity = the median dollar trading volume over past 91 trailing days
*Percentile rank of market capitalization = {total market capitalization of companies whose market capitalization is greater than or equal to the company in question} / {market capitalization of all companies}*100*
Size rank = re-ranked(market capitalization rank + liquidity rank)

Exhibit 3 Estimation Universe Construction Logic



Source: Morningstar.

Appendix C: Equity Factor Exposure Definitions

Interpretation

Style Factors

Our style factors are normalized by subtracting the cross-sectional mean and then dividing by the cross-sectional standard deviation, so a score of 0 can always be interpreted as the average score, and a nonzero 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 4 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 durability 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	Situation in which 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.
Value-Growth (standard model)	Uses the Morningstar Style Box raw style score for calculating value/growth characteristics. Higher scores imply firms that are more growth and less value oriented
Momentum	Total return momentum over the horizon from negative 12 months through negative two months. Higher scores imply greater return momentum.
Volatility	Total return volatility as measured by largest minus smallest one-month returns in a trailing 12-month horizon. Higher scores imply greater return volatility.
Volatility (standard model)	A combination of three volatility proxies (1) Idiosyncratic volatility (IVOL, 50%): the volatility of residual returns over the past six months; (2) Total volatility (TVOL, 25%): the volatility of daily total returns over the past six months; (3) MAX5 (25%): the average of the highest five daily returns over the past 1 month.
Quality	A quality firm is one with high profitability and low financial leverage. High scores imply high quality firms.
Yield	A measure of a firm's total yield (dividend plus buyback). High scores imply high yielding firms.
ESG Rating	ESG Rating factor describes how well a company manages ESG-related issues according to Sustainalytics ESG rating criteria. A higher exposure indicates better management of ESG issues.

Source: Morningstar.

Sector Factors

Our sector factors measure the economic exposure of a company to the 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.

Region Factors

Our region factors represent the economic exposure of a company to the 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.

Currency Factors

Our currency factors represent the economic exposure of a company to major currencies, excluding U.S. dollars. We perform a time-series regression analysis to find the exposures of an individual company's return denominated in U.S. dollar currency to the following list of currency returns: Australian dollar, British pound, Canadian dollar, euro, Japanese yen, New Zealand dollar, and Swiss franc. We calculate the return of these currencies against the U.S. dollar.

Style Factor Definitions

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 1,400 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, refer to the Morningstar Quantitative Equity Ratings methodology document cited in the References section.

The factor is unbounded, and higher scores indicate cheaper stocks. A score of 0 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, refer to the Morningstar Quantitative Equity Ratings methodology document cited in the References section.

The factor is unbounded, and higher scores indicate more-uncertain valuations.
A score of 0 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 durability of a firm's competitive advantages. We arrive at a moat score using an algorithm that extrapolates from the roughly 1,400 Morningstar 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, refer to the Morningstar Quantitative Equity Ratings methodology document cited in the References section.

The factor is unbounded, and higher scores indicate stronger and more-sustainable competitive advantages. A score of 0 indicates an average level of competitive advantages.

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 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 factor is unbounded, and higher scores indicate stronger financial health. A score of 0 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 factor relies on current portfolio holdings information and the raw 36-month Morningstar Risk rating. 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:

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

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

The ownership risk score for stock n is the weighted average of each manager m's 36-month Morningstar Risk score multiplied by the relative weight held in stock. After raw scores are calculated, ownership risk scores are cross-sectionally normalized.

The factor is unbounded, and higher scores indicate stronger ownership preference for risk. A score of 0 indicates an average level of ownership preference for risk.

Ownership Popularity

The ownership popularity factor represents the growth in the popularity of a particular stock from the perspective of fund manager ownership. It 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 three months.

The factor is calculated in the following manner:

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

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

Where:

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

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

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 held in stock. After raw scores are calculated, ownership popularity scores are cross-sectionally normalized.

The factor is unbounded, and higher scores indicate stronger ownership preference. A score of 0 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 factor is unbounded, and higher scores indicate smaller market capitalization. A score of 0 indicates an average level of market capitalization.

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.

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

The factor is unbounded, and higher scores indicate higher liquidity. A score of 0 indicates an average level of liquidity.

Value-Growth

Value-growth 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 factor is unbounded, and higher scores indicate higher growth expectations and less value exposure. A score of 0 is average.

Value-Growth (standard model)

Value-growth reflects the aggregate expectations of market participants for the future growth and required rate of return for a stock. For this version used in the Standard Factor model, we use the raw style score from the Morningstar Style Box as the input for calculating the value-growth exposure of stocks. The raw style score is calculated as the difference between a stock's growth score and value score:

$$Raw\ Style\ Score = Growth\ Score - Value\ Score.$$

The value score is the weighted average of a stock's prospective earnings, book value (BV), revenue (R), cash flow (CF), and dividend (D), all scaled by the current price of the stock:

$$Value\ Score = \left[w_E \times \frac{E}{P_t} + w_{BV} \times \frac{BV}{P_t} + w_R \times \frac{R}{P_t} + w_{CF} \times \frac{CV}{P_t} + w_D \times \frac{D}{P_t}\right].$$

The growth score of a stock is the weighted average of the growth rates in a company's earnings (E), book value (BV), revenue (R), cash flow (CF), and dividend (D):

$$Growth\ Score = [w_E \times E_{growth} + w_{BV} \times BV_{growth} + w_R \times R_{growth} + w_{CF} \times CF_{growth} + w_D \times D_{growth}].$$

The factor is unbounded, and higher scores indicate higher growth expectations and less value exposure. A score of 0 is average. For more details, refer to the Morningstar Style Box Methodology listed in the References section.

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, U.S. 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 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 0 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.

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

The factor is unbounded, and higher scores indicate higher volatility. A score of 0 indicates an average level of volatility.

Volatility (standard model)

The firm-specific volatility is a combination of three standardized volatility proxies:

$$Volatility\ Composite = 50\% * IVOL_z + 25\% * TVOL_z + 25\% * MAX5_z$$

(1) IVOL (six-month horizon, 50%):
Idiosyncratic volatility, or IVOL, is the capital asset pricing model's residual volatility over the past six months. We estimate a time-series regression of excess daily stock return against the value-weighted excess daily market return of the estimation universe. The IVOL is the standard deviation of the capital asset pricing model residuals. We standardize IVOL to obtain its z-score.

$$CAPM: r_{i,t} - r_{ft} = \alpha_{i,t} + \beta_t\ (r_{m,t} - r_{ft}) + \varepsilon_{i,t}$$
$$IVOL: \sigma_{i,t} = std(\varepsilon_{i,t})$$

(2) TVOL (six-month horizon, 25%):
Total volatility, or TVOL, is defined as the volatility of a stock's daily returns over the past six months. We standardize TVOL to obtain its z-score.

$$TVOL = \sqrt{\frac{\sum_{t=1}^{N}(r_t - \overline{r_t})^2}{N-1}}$$

(3) MAX5 (one-month horizon, 25%):
MAX5 is defined as the average of the highest five daily returns over the past one month. We standardize MAX5 to obtain its z-score.

The factor is unbounded, and higher scores indicate higher volatility. A score of 0 indicates an average level of volatility.

Quality

We define a quality score of a stock as the equally weighted z-score of a company's profitability (trailing 12-month return on equity) and the z-score of its financial leverage (trailing 12-month debt/capital). The z-score is with respect to all the stocks in the global universe.

$$Quality = \frac{1}{2}\left[ROE_z + \left(1 - \frac{Total\ Debt_t}{Total\ Capital_t}\right)_z\right].$$

where ROE is the trailing 12-month return on equity and the subscript z indicates a z-score.

The factor is unbounded, and higher scores indicate higher quality. A score of 0 indicates an average level of quality.

Yield

The yield factor is as a total yield, which is the sum of trailing 12-month buyback and dividend yield of a company. Higher values indicate larger, positive yield exposure:

$$Total\ Yield = Buyback\ Yield_{ttm} + Dividend\ Yield_{ttm}$$

The factor is unbounded, and higher scores indicate higher yield. A score of 0 indicates an average level of quality.

ESG Rating

The ESG Rating factor is constructed based on Sustainalytics ESG Rating, which measures how well companies manage various ESG-related issues. Sustainalytics' ESG Rating starts from August 2009. Before Sept. 20, 2019, the rating is based on their best-in-class rating framework, where each industry has a different rating structure. A higher ESG Rating indicates a company ranks better according to ESG criteria when compared with their industry peers. From Sept. 20, 2019, Sustainalytics provides a new ESG Risk Rating, where a higher ESG Risk Rating indicates a company has higher ESG risks when compared with all other stocks in the rating universe. In the ESG Risk framework, a lower rating indicates a company ranks better according to the ESG criteria. The rating is on an absolute scale, so a company's ESG Risk Rating is not compared with industry peers but the overall universe.

To provide a consistent interpretation of the ESG factor, the following transformations have been applied. First, the total ESG Rating before Sept. 20, 2019, is standardized within an industry every day so that the industry average rating is zero and the standard deviation of the rating is one. Second, the ESG Risk Rating from Sept. 20, 2019, is transformed by (100—ESG Risk Rating). Lastly, both the standardized total ESG Rating and the transformed ESG Risk Rating are standardized among the estimation universe of the risk model every day to ensure the average ESG Rating of the universe is zero and the standard deviation of the rating is one. With these transformations, a higher ESG Rating indicates better ESG performance among industry peers before Sept. 20, 2019, and among the rating universe after this date.

Sector Factor Definitions

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

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

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

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

$$r_t^k = weekly\ return\ on\ the\ kth\ sector\ benchmark\ (for\ example, Basic\ Materials)$$

$$constraints{:}\ 0 < \ \beta_k^i < 1;\ \sum\nolimits_k \beta_k^i = 1$$

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 U.S. dollars. Market capitalizations were also converted to dollars 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. If a stock has less than one year of history, we do not run the regression and instead default to the stock's Morningstar sector classification. We employ a Bayesian prior that presupposes that companies should be entirely exposed to the sector to which they are assigned.

Sectors

Below is the complete list of sectors available to be included in the multivariate regression. Note, depending on the factor list of each model, only a subset could be used.

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

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 Factor Definitions

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

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

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

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

$$r_t^k = weekly\ return\ on\ the\ kth\ region\ benchmark\ (for\ example, Developed\ North\ America)$$

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

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 U.S. dollars. Market capitalizations were also converted to dollars 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. If a stock has less than one year of history, we do not run the regression and instead default to the stock's Morningstar region classification based on country of domicile. 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

Below is the complete list of regions available to be included in the multivariate regression. Note, depending on the factor list of each model, only a subset could be used.

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

Exhibit 5 Map of Markets to Regions

Region	Market List		
Developed Asia Pacific	Australia	Japan	Singapore
	Hong Kong	New Zealand	
Developed Europe	Austria	Finland	Italy
	Belgium	France	Netherlands
	Switzerland	United Kingdom	Norway
	Germany	Greece	Portugal
	Denmark	Ireland	Sweden
	Spain	Israel	
Developed North America	United States	Canada	
Emerging Asia	China	South Korea	Thailand
	Indonesia	Malaysia	Taiwan
	India	Philippines	Vietnam
Emerging Europe	Bulgaria	Lithuania	Romania
	Czech Republic	Luxembourg	Russia
	Estonia	Latvia	Turkey
	Hungary	Malta	
	Iceland	Poland	
Emerging Latin America	Argentina	Colombia	Venezuela
	Brazil	Mexico	
	Chile	Peru	
Emerging Middle East & Africa	United Arab Emirates	Kuwait	Qatar
	Bangladesh	Morocco	Saudi Arabia
	Bahrain	Nigeria	South Africa
	Cyprus	Oman	
	Egypt	Pakistan	

Source: Morningstar as of May 2021.

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 Factor Definitions

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

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

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

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

$$r_t^k = weekly\ return\ on\ the\ kth\ exchange\ rate\ return\ (for\ example, \%\ change\ in \frac{EUR}{USD})$$

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 U.S. dollars.

Currencies

Below is the complete list of currencies available to be included in the multivariate regression. Note, depending on the factor list of each model, a subset of these currencies may be used. For example, the Morningstar U.K. Equity Risk Model includes the U.S. dollar but not the British pound.

- euro
- Japanese yen
- British pound
- Swiss franc
- Canadian dollar
- Australian dollar
- New Zealand dollar
- U.S. dollar

Interpretation

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

Appendix D: Fixed-Income Factor Definition

The factors driving fixed-income returns can best be understood by examining the basic bond valuation formula. A bond, in its simplest form, is similar to a loan contract where a company or government borrows money from investors. The investors are owed their initial investment (called the principal, denoted as M below), and they also (often) receive periodic payments called Coupons (C). The interest rate i below accounts for the time value of money, by discounting the future cash payments. We can express a bond's price in the following way:

$$Bond\ Price_{today} = \frac{C}{1+i_1} + \frac{C}{(1+i_2)^2} + \cdots + \frac{C}{(1+i_n)^n} + \frac{M}{(1+i_n)^n}$$

Bond valuation refers to calculating the present value of a bond's expected future cash flows. The present value of a bond is the sum of the present values of the cash flows, with discount rates given by the term structure of interest rates:

$$PV_{Bond} = \sum_{t=1}^{T} CF_t * e_t^{-r_t t}$$

The valuation formula states that the present value of a bond is the sum of cash flows at time t discounted by an appropriate discount rate. It makes it clear that changes in bond prices, and bond returns, are driven by changes in the expected cash flows and/or changes in the appropriate discount rates. For a risk-free fixed-coupon bond, the cash flows are the coupon and the principal payments, and the interest rates are the prevailing risk-free rates. The valuation formula becomes more complex when we leave the realm of risk-free fixed-coupon bonds because the cash flows may be state-dependent or history-dependent, and the discount rates include a spread to compensate investors for the additional risks.

In order to compute daily return, we use duration, convexity, and carry factors. When we combine these four prices together, we arrive at our daily price return, that is, R_t. We can express it as the following:

$$R_t = Carry_{return} + Duration_{return} + Convexity_{return}$$

Where

$Carry_{retun} = YTM * (t_2 - t_1)$ where $t_2 - t_1$ refers to the passage of time in years

$Duration_{return} = -D_{t-1} * Premia$ where $-D_{t-1}$ is the duration from the previous day

$Convexity_{return} = \frac{1}{2} * C_{t-1} * Premia^2$ where C_{t-1} is the convexity from the previous day

In the following sections, each component is discussed in detail.

Duration and Convexity Model

One of the main drivers of fixed-income securities are changes in the government yield curves. These changes impact the values of all fixed-income securities since they tell us about the prevailing rates for a specific tenor and currency pair. The Risk Model framework uses the two key factors—duration and convexity—to reflect the exposure of bond return sensitivity to benchmark interest-rate movements.

Factor premia, in this case, are the changes in the government yield curve. The risk model captures the interest-rate component of USD-, EUR-, GBP-, CHF-, and CAD-denominated bond returns.

In order to compute returns for fixed-income securities within the Risk Model framework, we need two components: exposure and premia. We get the exposure data from the previous day. Duration is defined as the sensitivity of a bond to change in interest rates. Additionally, convexity is defined as the change in duration with respect to change in interest rates. It is important to understand that both these concepts reflect a bond's price sensitivity to change in rates.

For premia, we take the average parallel shift between the five-year and 10-year Treasury curve. Mathematically, we can express it as follows:

$$Premia = 0.5 * ((5yr_{yield_t} - 5yr_{yield_{t-1}}) + (10yr_{yield_t} - 10yr_{yield_{t-1}}))$$

In the above equation, t refers to current day and $t - 1$ refers to the previous day.

Carry

Carry is the return earned due to the passage of time. In some sense, this is riskless, in that the carry earned over a period is known ex-ante. It is composed of coupon return and roll return.

The coupon accrues deterministically. The coupon return between t_1 and t_2 is:

$$r_{Coupon} = \frac{AI(t_2) - AI(t_1) + C(t_1, t_2)}{P(t_1) + AI(t_1)}$$

Where:

- $AI(t_i)$= The accrued interest at time i
- $C(t_1, t_2)$ = The coupon paid between t_1 and t_2
- $P(t_i)$= The clean price at time i

The roll return is the return due to the change in the clean price caused by the passage of time from t_1 to t_2, based on the yield curve at t_1. There are two drivers of roll return. First, we are discounting all future cash flows to t_2 instead of t_1. Second, these future cash flows occur at different points on the yield curve relative to t_2 than they do relative to t_1. Despite keeping the yield curve constant, the interest rate used to discount each cash flow is different. The roll return between t_1 and t_2 is:

$$r_{Roll} = \frac{P(t_2, y_1) - P(t_1, y_1)}{P(t_1, y_1) + AI(t_1)}$$

Where:

- $AI(t_i)$= The accrued interest at time i
- $P(t_{i,} y_j)$= The clean price at time i using the yield curve that prevailed at time j

Specifically:

- $P(t_{2,} y_1)$= The clean price at time t_2 using the yield curve that prevailed at time t_1
- $P(t_{1,} y_1)$= The clean price at time t_1 using the yield curve that prevailed at time t_1

Finally, carry return is the sum of coupon return and roll return:

$$r_{Carry} = r_{Coupon} + r_{Roll}$$

In our risk models, we capture this return by using yield to maturity. That is:

$$r_{Carry} \approx YTM_1 * (t_2 - t_1)$$

Where:

- YTM_i = The yield to maturity at time i
- $\left(t_j - t_i\right)$ = The time elapsed between time i and time j measured in years

This is algebraically equivalent to the formula above, up to a second-order approximation. In our risk models, we use yield to maturity (which is security-specific) as the exposure to the carry factor, and the elapsed time (which is common across securities) as the carry factor premium. In this way, the exposure of any fund or security to carry is its yield.

Appendix E: Cross-Sectional Regression

After deciding on the universe of securities to include in the model and gathering quality input data, the next important step in risk model construction is to run the cross-sectional regression. There are numerous techniques and specifications we can employ in the regression; the following method has been chosen to provide accurate, meaningful, and stable estimates of factor premia. Special care has been given to deal with the common multicollinearity issue associated with sector and region factors. As the sum of all sectors and regions are both the entire universe of securities, it is difficult to estimate the pure sector and region effects that are uncorrelated with each other. We apply a constrained regression to disentangle the sector and region effects from each other, as well as from the overall market movement.

The Constrained Regression

The return of a security r_i in the cross section can be explained as

$$r_i = \alpha + \sum_{m=1}^{M} X_{i,m} f_{\mathrm{m}}^{Style} + \sum_{s=1}^{S} X_{i,s} f_s^{Sector} + \sum_{r=1}^{R} X_{i,r} f_r^{Region} + \sum_{c=1}^{C} X_{i,c} f_c^{Currency} + \varepsilon_i \quad \text{(E1)}$$

where $X_{i,m}, X_{i,s}, X_{i,r}, X_{i,c}$ are security i's exposure to style factor m, sector s, region r, and currency c; $f_{\mathrm{m}}^{Style}, f_s^{Sector}, f_r^{Region}, f_c^{Currency}$ are factor premia for style m, sector s, region r, and currency c; M, S, R, C are the total number of style, sector, region, and currency factors in a particular model; α is the intercept; and ε_i is the residual term, representing a stock's specific return.

In the estimation, the market-cap-weighted average sector premia and region premia are both constrained to zero:

$$\sum_{s=1}^{S} u_s f_s^{Sector} = \sum_{r=1}^{R} \mathrm{v_r} f_r^{Region} = 0 \quad \text{(E2)}$$

where u_s and $\mathrm{v_r}$ are the market-cap weights of sector s and region r, respectively. This means certain sectors and regions earn positive returns and others earn negative, but the market-cap-weighted average sector and region returns are zero.

To understand the logic of these constraints, imagine an investor who has a portfolio that has the same sector and region composition as the entire market; the region and sector average return from this portfolio should not contribute extra return to the market because the sum of sectors and regions are both the market. But what captures the market return in this setting? It turns out that under certain conditions, the estimated α is a good proxy for the market.

The Equity Market Factor

The intercept is represented by a column of 1 in the exposure table, and it can be viewed as stocks' exposure to a factor. To what factor does every stock have the same level of exposure? It should be a factor that represents the equity market universe, and an exposure of 1 indicates membership in this universe. For this reason, the estimated α can approximate the overall equity market return; we name it the "equity market factor." The approximation becomes accurate with some additional conditions.

In addition to the constraints on sector and region premia, all style factor exposures are standardized cross-sectionally to have a market-cap-weighted mean of zero:

$\tilde{X}_m = \sum_{i=1}^{N} w_i X_{i,m} = 0$ (E3)

where
$\tilde{X}_m$ = market-cap-weighted average exposure of the estimation universe to factor m,
w_i = market-cap weight of security i,
$X_{i,m}$ = security i's exposure to style factor m.

This standardization ensures the overall market is style-neutral. Now, consider aggregating the market-cap-weighted estimation universe as

$r_E = \alpha + \sum_{m=1}^{M} \tilde{X}_m f_{\mathrm{m}}^{Style} + \sum_{s=1}^{S} u_s f_{\mathrm{s}}^{Sector} + \sum_{r=1}^{R} \mathrm{v}_{\mathrm{r}} f_{\mathrm{r}}^{Region} + \sum_{c=1}^{C} \tilde{X}_c f_{\mathrm{c}}^{Currency} +$
$\sum_{i}^{N} w_i \varepsilon_i$ (E4)

where r_E is the market-cap-weighted average return. Note, by equations (E2) and (E3), the second to the fourth items on the right-hand side become zero. The last term of weighted residuals equals zero because in a least-squares estimation the residual term is orthogonal to the independent variables including the intercept of 1s. Although we do not standardize the currency exposures, the impact of currency return is limited. Therefore, the estimated α can approximate closely the market-cap-weighted average return of the estimation universe.

Note that the regression has been weighted using the square root of the market-cap weight of each stock in the estimation universe. This is to reduce the uneven variability of the specific returns among stocks, which improves the statistical properties of premium estimates. In this case, the weighted sum of residuals in equation (E4) is only approximately zero.

To sum up, with the constrained regression, the sector and region premia measure the pure and uncorrelated sector and region returns relative to the overall market return, captured by α. $\alpha + f_s^{Sector}$ approximates the return of a geographically diversified portfolio of companies in sector s. "Geographically diversified" means the portfolio has the same market-cap-weighted region composition as the equity market universe and is free from any additional region effects. Similarly, $\alpha + f_r^{Region}$ gives the return of a portfolio of stocks that is sector diversified as the equity market universe.

Appendix F: Forecast Factor Comovement and Residual Volatility for Equity Models

Asset Returns Covariance Model

The risk model cross-sectional regression models stock returns at time t as

$$R_t = X_t \cdot f_t + S_t$$

where:

- R_t = N x 1 vector of asset returns at time t
- X_t = N x K matrix of asset-level factor exposures at time t
- f_t = K x 1 vector of factor returns (factor premia) at time t
- S_t = N x 1 vector of asset-level specific returns (residual returns) at time t

We model the N x N variance-covariance matrix of asset returns at time t, V_t, as:

$$V_t = X_t \cdot F_t \cdot X_t^T + \Delta_t$$

where:

- V_t = N x N variance-covariance matrix of asset returns
- F_t = K x K variance-covariance matrix of the factor returns (factor premia)
- Δ_t = N x N variance matrix of the specific returns S (diagonal matrix of specific variance)

That is, the covariances between factor premia are included in the model, but residual returns are assumed to be independent of each other and of the factor premia.

Portfolio Returns Variance Model

A portfolio is described at time t by an N x 1 vector w_t^P that gives the portfolio's holding-weights in N assets. The portfolio's K x 1 vector of factor exposures x_t^P is given by the product of the asset-level factor exposures X_t^T and the holdings weights w_t^P:

$$x_t^P = X_t^T \cdot w_t^P$$

The portfolio's return at time t, r_t^P, is modeled as

$$r_t^P = (x_t^P)^T \cdot f_t + (w_t^P)^T \cdot S_t$$

and the portfolio's variance at time t, V_t^P, is modeled as

$$V_t^P = (x_t^P)^T \cdot F_t \cdot x_t^P + (w_t^P)^T \cdot \Delta_t \cdot w_t^P$$

Exponentially Weighted Moving Average Estimates and Forecasts

For the EWMA model, we assume there is no autocorrelation in the portfolio returns so that the variance over the forecast horizon is the sum of variance estimates for each period of the forecast horizon. Additionally, the nature of the EWMA estimate is that forecasts beyond time t equal the estimate at time t.

So, to produce forecasts of portfolio volatility at time t, over the forecast horizon, [$t + 1$, $t +$ horizon], the task is to estimate F_t and Δ_t given historical data up to and including time t, and then, for a given portfolio: calculate V_t^P, multiply by the horizon length, and take the square root.

Our approach to modeling F_t, as an EWMA estimate, is to separately estimate the factor premia standard deviations and factor premia correlation matrix, over a historical window of fixed length, using a different half-life for each, and then combine them into a covariance matrix.

The factor-j premium standard deviation, $\sigma_{t,j}$, is estimated as follows.

$$\delta_1 = \left(\frac{1}{2}\right)^{\frac{1}{\tau_1}}$$

$$z_{1,t-i} = \left(\frac{1-\delta_1}{1-\delta_1^W}\right)\delta_1^i$$

$$m_{1,t} = \sum_{i=0}^{W-1} z_{1,t-i} \times f_{t-i}$$

$$\sigma_{t,j} = \sqrt{\sum_{i=0}^{W-1} z_{1,t-i} \times \left[f_{t-i} - m_{1,t}\right]_j^2}$$

where

- τ_1 is the half-life for standard deviation
- δ_1 is the decay rate for standard deviation
- W is historical data window size for covariance
- $z_{1,t-i}$ is the exponential weight for time $t - i$, for standard deviation
- $m_{1,t}$ is the K x 1 exponentially weighted mean premia vector estimate for time t for standard deviation
- $[v]_j$ denotes the j^{th} element of the vector v
- $\sigma_{t,j}$ is the exponentially weighted factor-j premium standard deviation estimate for time t

The factor premia correlation matrix, C_t, is estimated as follows.

$$\delta_2 = \left(\frac{1}{2}\right)^{\frac{1}{\tau_2}}$$

$$z_{2,t-i} = \left(\frac{1-\delta_2}{1-\delta_2^W}\right)\delta_2^i$$

$$m_{2,t} = \sum_{i=0}^{W-1} z_{2,t-i} \times f_{t-i}$$

$$\tilde{C}_t = \sum_{i=0}^{W-1} z_{2,t-i} \times \left(f_{t-i} - m_{2,t}\right) \cdot \left(f_{t-i} - m_{2,t}\right)^T$$

$$C_t = diag(\tilde{C}_t)^{-1/2} \cdot \tilde{C}_t \cdot diag(\tilde{C}_t)^{-1/2}$$

where

- τ_2 is the half-life for correlation
- δ_2 is the decay rate for correlation
- $z_{2,t-i}$ is the exponential weight for time $t - i$, for correlation
- $m_{2,t}$ is the K x 1 exponentially weighted mean premia vector estimate for time t for correlation
- $\tilde{C}_t$ is the K x K exponentially weighted factor premia covariance matrix estimate for time t, using the correlation half-life τ_2
- C_t is the K x K exponentially weighted factor premia correlation matrix estimate for time t
- $\mathrm{diag}(A)$ denotes the diagonal matrix of the matrix A.

Let Σ_t be the K x K diagonal matrix with vector $[\sigma_{t,1}, .., \sigma_{t,K}]$ along the diagonal. Then

$$F_t = \Sigma_t \cdot C_t \cdot \Sigma_t$$

so that the i-th, j-th entry of F_t is

$$[F_t]_{ij} = \sigma_{t,i} \times \rho_{t,ij} \times \sigma_{t,j}$$

The half-lives τ_1 and τ_2 are selected to maximize the back-test average of the average log-likelihood of the demeaned, observed premia over a given forecast horizon, H, assuming a multivariate Gaussian distribution with covariance F_t and mean of zero. That is

$$\begin{aligned}(\tau_1, \tau_2) &= \underset{(\tau_1,\tau_2)}{arg\,max} \frac{1}{|B|}\sum_{t\in B}\frac{1}{H}\mathcal{L}(f_{t+1}, .., f_{t+H}|F_t) \\ &= \underset{(\tau_1,\tau_2)}{arg\,max} \frac{1}{|B|}\sum_{t\in B}\left(-\frac{K}{2}log\,2\pi - \frac{1}{2}log(det(F_t)) - \frac{1}{2H}\sum_{i=1}^{H}(f_{t+i} - \bar{f}_{t,H})^T F_t^{-1}(f_{t+i} - \bar{f}_{t,H})\right)\end{aligned}$$

where

$$\bar{f}_{t,H} = \frac{1}{H}\sum_{i=1}^{H} f_{t+i}$$

and

- H is the number of periods in the forecast horizon
- $\bar{f}_{t,H}$ is the K x 1 vector of mean observed premia over the interval [$t + 1$, $t +$ horizon]
- B is the set of forecast start times included in the back-test
- $|B|$ denotes the cardinality of set B

The time periods used in the risk model loosely correspond to trading days, and (20, 60, 120, and 240) days correspond to (one, three, six, and 12) months.

Optimal Parameters

The half-lives were restricted to an integer number of periods. The historical data windows used were $W = 1{,}200$ and $W_r = 300$. The optimal half-lives, along with their mean log-likelihoods, are given in the following table. Note τ_3 is the half-life parameter of residual variance estimate, and refer to the section below on residual volatility forecast.

Exhibit 6 Optimal Half-Life Parameters

Horizon	τ_1	τ_2	τ_3	mean $\mathcal{L}$(premia)	mean $\mathcal{L}$(residual)
20	62	108	48	-9.213	-2.014
60	88	154	76	-12.365	-2.048
120	104	196	84	-14.757	-2.069
240	150	238	114	-16.881	-2.098

Source: Morningstar.

Enhanced Factor Covariance Forecast With Autocorrelation and Bias Correction

In the enhanced factor covariance model, variance-covariance matrix F_t is estimated in two steps. Step 1 is similar to the EWMA estimate model but includes the effects of autocorrelation. Step 2 uses a feedback mechanism to quantify local, broad equity market volatility deviations and then scales the Step 1 covariance matrix to compensate. Covariance forecasts over a horizon are obtained by scaling F_t by the forecast horizon.

In Step 1, EWMA variances and EWMA correlation matrixes are estimated with different half-lives and then combined. Autocorrelation is included in the variance and correlation estimates via the Newey-West estimator. The Newey-West estimator estimates autocorrelation-corrected covariance matrixes by combining lag-k autocovariance matrixes for each lag, up to a maximum number of lags. We estimate the autocovariance matrixes with exponentially decaying weights. Variances and correlation matrixes are readily extracted from the resulting covariance matrixes.

The feedback mechanism in Step 2 is based on the Mahalanobis distance between the sum-premia over the most recent x days and the multivariate Normal distribution defined by x times the forecast covariance from x-days prior. The multiperiod span allows the effects of autocorrelation to appear in the sum premia, although it does introduce a delay. If the model is correct, the Mahalanobis distance squared will follow a chi-squared distribution with K degrees of freedom and have mean K. Dividing by K gives a measure of broad-market bias, which is then exponentially smoothed to produce the broad-market bias correction for the Step 1 covariance matrix.

Step 1 Details

The covariance estimates of Step 1 are constructed from separately estimated EWMA variances and EWMA correlation matrixes. The variances are estimated using shorter half-lives and more lags, while the correlation matrixes are estimated using longer half-lives and fewer lags.

Let Σ_t be the diagonal matrix comprising the time-t factor-premia standard-deviation estimates along its diagonal, and R_t be the time-t factor-premia correlation matrix estimate. Then the Step 1 covariance matrix for time-t, denoted $\tilde{F}_t$, is given by

$$\tilde{F}_t = \Sigma_t \cdot R_t \cdot \Sigma_t$$

such that the i-th, j-th entry of $\tilde{F}_t$ is

$$[\tilde{F}_t]_{ij} = \sigma_{t,i} \times \rho_{t,ij} \times \sigma_{t,j}$$

where $\sigma_{t,i}$ is the time-t standard-deviation estimate for factor-i and $\rho_{t,ij}$ is the time-t correlation estimate between factor-i and factor-j.

To reduce notation, equations are given for the covariance matrix estimates, and then the factor variance and correlation matrix estimates are extracted from the covariance matrix estimates. To further reduce notational complexity, it is assumed that all premia are available. When missing, weights for missing premia are set to zero and the remaining weights, for a given lag, are rescaled to sum to one.

The Newey-West autocorrelation-corrected EWMA covariance matrix with half-life τ and maximum lags L, estimated at time t, is denoted $C_t(\tau, L)$ and is calculated as follows:

$$\delta = \delta(\tau) = \left(\frac{1}{2}\right)^{\frac{1}{\tau}}$$

$$w_{i,k,\delta} = \left(\frac{1-\delta}{1-\delta^{W-k}}\right) \times \delta^{W-1-i}$$

$$\mu_{t,k,\delta} = \sum_{i=0}^{W-1-k} w_{i,k,\delta} \times f_{t-i}$$

$$\eta_{k,\delta} = 1 - \frac{\sum_{i=0}^{W-1-k} w_{i,k,\delta}^2}{\left(\sum_{i=0}^{W-1-k} w_{i,k,\delta}\right)^2}$$

$$C_{t,k,\delta} = \sum_{i=0}^{W-1-k} \frac{w_{i,k,\delta}}{\eta_{k,\delta}} \times \left(f_{t-i} - \mu_{t,k,\delta}\right) \cdot \left(f_{t-i-k} - \mu_{t,k,\delta}\right)^T$$

$$C_t(\tau, L) = C_{0,t,\delta(\tau)} + \sum_{k=1}^{L} \frac{L+1-k}{L+1} \left(C_{k,t,\delta(\tau)} + C_{k,t,\delta(\tau)}^T \right)$$

where

- τ is the half-life
- δ is the decay rate
- W is the integer number of time periods included in the historical data window
- $w_{i,k,\delta}$ is the exponential weight for time $t-i$ through the historical window $[t-W+1, t]$, for lag-k autocovariance calculations, based on decay rate δ.
- f_{t-i} is the $(K \times 1)$ premia vector at time $t-i$
- $\mu_{t,k,\delta}$ is the $(K \times 1)$ exponentially weighted mean premia vector estimate for lag-k autocovariance calculations at time t, based on decay rate δ.
- $\eta_{k,\delta}$ is the weighted-sample bias normalization constant, for lag-k autocovariance calculations, based on decay rate δ.
- $C_{t,k,\delta}$ is the $(K \times K)$ exponentially weighted lag-k factor premia autocovariance matrix estimate for time t, based on decay rate δ.
- $C_t(\tau, L)$ is the $(K \times K)$ autocorrelation-corrected EWMA covariance matrix estimate for time t, based on half-life τ, accounting for autocorrelations up to lag L.

The factor variance and correlation matrix estimates are extracted from $C_t(\tau, L)$ as:

$$\Sigma_t = diag\left(C_t(\tau_1, L_1)\right)^{\frac{1}{2}}$$

$$R_t = diag\left(C_t(\tau_2, L_2)\right)^{-\frac{1}{2}} \cdot C_t(\tau_2, L_2) \cdot diag(C_t(\tau_2, L_2))^{-1/2}$$

where

- $\mathrm{diag}(A)$ denotes the diagonal matrix of the matrix A, equal to A on the diagonal and otherwise 0.
- τ_1 is the variance half-life
- L_1 is the maximum number of autocorrelation lags used to estimate the variance
- τ_2 is the correlation matrix half-life
- L_2 is the maximum number of autocorrelation lags used to estimate the correlation matrix

Step 2 Details

Step 2 entails estimating the local broad-market bias-statistic of the Step 1 covariance matrix estimate, then scaling the covariance matrix to remove the bias. The broad-market bias is assessed on recent sum premia so that the effects of autocorrelation are included. To assess forecasting, the Step 1 covariance matrix prior to the historic span is required, which creates a delay.

The Mahalanobis distance is used as a basis for the bias-statistic measure. The Mahalanobis distance can be seen as rotating the premia onto the eigenbasis of the covariance matrix, and then normalizing each rotated premia by the standard deviation of the covariance matrix in that direction—that is, the square root of the corresponding eigenvalue. The result is that, if the sum premia adhere to the multivariate normal distribution defined by the covariance matrix, the Mahalanobis distance will adhere to a chi-squared distribution with K degrees of freedom, which has mean K, where K is the number of risk factors.

Dividing by K gives a measure of the bias-statistic, which is then exponentially smoothed with a relatively short half-life. This is the broad-market covariance bias-feedback correction multiplier, m_t. The whole Step 1 covariance matrix, $\tilde{F}_t$, is then scaled by m_t to produce the final covariance estimate.

A caveat to this final point is that if the risk model includes multiple asset classes, then this multiplier is calculated and applied within just the equity asset class. In this case, the standard deviations of just the equity factors, which can be located along the diagonal of Σ_t, are scaled by $\sqrt{m_t}$.

Ignoring this case, or assuming all risk factors are for the equity market, the bias correction is calculated and applied as:

$$d_M(x, C) = \sqrt{x^T \cdot C^{-1} \cdot x}$$

$$\tilde{b}_t^2(h_b) = d_M^2\left(\sum_{i=0}^{h_b} f_{t-i}\,, h_b \times \tilde{F}_{t-h_b}\right) / \,K$$

$$m_t = m_t(h_b, \tau_b) = EWMA(\tilde{b}_t^2(h_b); \tau_b)$$

where

- $d_M(x, C)$ is the Mahalanobis distance of vector x from covariance matrix C
- $\tilde{b}_t^2(h_b)$ is the squared broad-market point bias-statistic for time t, calculated from the historic premia over the span $[t - h_b + 1, t]$ and the Step 1 covariance estimate at time $t - h_b$
- h_b is the bias feedback-correction horizon
- f_t is a $(K \times 1)$ vector of risk factor premia time t
- $\tilde{F}_t$ is the Step 1 covariance matrix at time-t
- τ_b is the bias feedback-correction half-life
- m_t is the broad-market covariance bias-feedback correction multiplier for time t

The final one-step-ahead covariance estimate at time t, F_t, is then

$$F_t = m_t \times \tilde{F}_t$$

Residual Volatility Forecast

The residual variance matrix, Δ_t, is diagonal because residuals for different stocks are assumed independent of each other. Additionally, the expected residual return is assumed to be zero for all stocks. Exposures and/or returns are not always available for all stocks, so the model must account for missing residuals. The residual variance for stock s, at time t, $\sigma_{t,s}^2$, is estimated as follows.

$$\delta_3 = \left(\frac{1}{2}\right)^{\frac{1}{\tau_3}}$$

$$z_{3,t-i} = \begin{cases} \delta_3^i, & if\ S_{t-i}\ is\ available \\ 0, & if\ S_{t-i}\ is\ missing \end{cases}$$

$$\sigma_{t,s}^2 = \begin{cases} \left(\sum_{i=0}^{W_r-1} z_{3,t-i} \times [S_{t-i}]_s^2\right) / \left(\sum_{i=0}^{W_r-1} z_{3,t-i}\right), & if \left(\sum_{i=0}^{W_r-1} z_{3,t-i}\right) \geq 0.5 \\ missing, & else \end{cases}$$

where

- τ_3 is the half-life for residual standard deviation
- δ_3 is the decay rate for residual standard deviation
- W_r is historical data window size for residual standard deviation
- $z_{3,t-i}$ is the unnormalized exponential weight for time $t-i$, for residual standard deviation

Then Δ_t is the N x N diagonal matrix with vector $[\sigma_{t,1}^2, .., \sigma_{t,N}^2]$ along the diagonal.

The half-life τ_3 is selected that maximizes the back-test average of the cross-sectional winsorized mean of available average-daily log-likelihoods of observed residual returns over a given forecast horizon, H, assuming a multivariate Gaussian distribution with covariance Δ_t and mean of zero. That is

$$\tau_3 = \underset{\tau_3}{arg\ max} \frac{1}{|B_r|} \sum_{t \in B_r} \frac{1}{|R_t|} \sum_{s \in R_t} g(L_t)$$

where

- H is the number of periods in the forecast horizon
- B_r is the set of forecast start times included in the residual back-test
- R_t is the set of calculable log-likelihoods for forecast time t
- $g(\ L_t)$ performs a 1% lower-side winsorization on the non-missing elements of vector L_t, and
- L_t is an N x 1 vector of the average daily log-likelihood, over forecast horizon H, of the residual for each stock, for forecast time t, with element s, for stock s, given by

$$[L_t]_s = \begin{cases} -\frac{1}{2} log\ 2\pi\sigma_{t,s} - \frac{1}{2\sigma_{t,s}^2} \frac{1}{|U_{t,s}|} \sum_{i \in U_{t,s}} [S_{t+i}]_s^2\ , & if\ \sigma_{t,s}^2\ is\ available, and\ |U_{t,s}| > 0 \\ missing, & else \end{cases}$$

where

- $U_{t,s}$ is the set of times, relative to forecast time t, for which the residual is available for stock s.

Residual Volatility Forecast With Autocorrelation and Bias Correction

The residual returns of the stocks in the coverage universe are assumed independent of each other. This results in the residual variance matrix, Δ_t, being diagonal. The residual variance estimates that form the diagonal of Δ_t, are based on EWMA population variance estimates for each stock. The Newey-West estimator, with EWMA autocorrelation estimates, is used to correct the variance estimates for autocorrelation. The autocorrelation correction is a multiplicative correction, calculated for each stock. The autocorrelation-corrected variance estimates are applicable over multiperiod forecast horizons. The exposures and/or returns are not always available for all securities, so cross-sectional regressions are used to estimate missing residual variances and autocorrelation corrections.

The nature of the residuals in the estimation universe and in the remainder of the coverage universe is different. The stocks outside the estimation universe tend to have smaller market capitalizations and have nonzero cross-sectional averages because they have not been used to estimate the premia. So, the missing-value regressions are fitted separately for the estimation universe and for the remainder of the coverage universe.

The EWMA half-lives define what is the trend and what is noise. Volatility estimates for individual securities become more stable with longer half-lives, which helps estimate the relative volatilities between securities. However, local market conditions can change rapidly and cause broad-market volatility deviations that evolve faster than the longer half-lives can respond to.

To compensate for the broad-market local volatility changes, broad-market bias patterns in the recent-historical, variance estimates are estimated and corrected for. Corrections are first estimated from and applied to the population variance estimates, using one-day-ahead bias statistics, and then from and to the autocorrelation-corrected variance estimates, using 20-day spans to estimate the bias patterns. The bias statistics provide a measure of the realized volatility divided by the predicted volatility, for some set of predicted volatilities, with variances combined over time, assets, or both. In each case, the corrections are estimated as linear models, with parameters fitted cross-sectionally each day and then exponentially smoothed with relatively short half-lives. Again, the bias-correction processes are performed effectively separately on the estimation universe and on the remainder of the coverage universe.

The resulting residual volatilities are given by

$$\sigma^2(t,s) = F_{20}(t,s) \times C_{NW}(t,s) \times F_1(t,s) \times \sigma^2_{pop}(t,s)$$

where

- $\sigma^2_{pop}(t,s)$ is the residual population variance estimate for security s at time t, with missing values replaced by a cross-sectional regression
- $F_1(t,s)$ is a one-step-ahead population-variance bias-statistic regression correction, for security s at time t
- $C_{NW}(t,s)$ is the Newey-West autocorrelation correction for security s at time t, with missing values replaced by a cross-sectional regression

- $F_{20}(t,s)$ is a 20-steps-ahead autocorrelation-corrected-variance bias-statistic regression correction, for security s at time t
- $\sigma^2(t,s)$ is the residual variance estimate for security s at time t

The initial EWMA population variances, $\sigma^2_{pop,0}(t,s)$, are estimated as:

$$\delta_3 = \left(\frac{1}{2}\right)^{\frac{1}{\tau_3}}$$

$$w_3(t-i,t,s) = \begin{cases} \frac{(1-\delta_3)}{1-\delta_3^{W_r}} \times \delta_3^i & , if\ S(t-i,s)\ is\ available \\ 0 & , if\ S(t-i,s)\ is\ missing \end{cases}$$

$$m(t,s) = \sum_{i=0}^{W_r-1} w_3(t-i,t,s)$$

$$\mu(t,s) = \frac{1}{m(t,s)} \times \sum_{i=0}^{W_r-1} w_3(t-i,t,s) \times S(t-i,s)$$

$$C(t,s) = 1 - \frac{\sum_{i=0}^{W_r-1} w_3^2(t-i,t,s)}{\left(\sum_{i=0}^{W_r-1} w_3(t-i,t,s)\right)^2}$$

$$\sigma^2_{pop,0}(t,s) = \begin{cases} \frac{1}{C(t,s)} \times \left(\left(\sum_{i=0}^{W_r-1} \frac{w_3(t-i,t,s)}{m(t,s)} \times S^2(t-i,s)\right) - \mu^2(t,s)\right) & , \quad if\ m(t,s) \geq \frac{1}{2} \\ missing & , \quad else \end{cases}$$

where

- τ_3 is the half-life for residual variance
- δ_3 is the decay rate for residual variance
- W_r is historical data window size for residual variance
- $w_3(t-i,t,s)$ is the normalized exponential weight for time $t-i$, for residual variance calculations at time t for security s
- $m(t,s)$ is the total normalized exponential weight with available data for residual variance calculations at time t for security s
- $S(t,s)$ is the residual, for time t, security s
- $\mu(t,s)$ is the weighted-sample mean, for time t, security s
- $C(t,s)$ is the weighted-sample bias normalization constant, for time t, security s, which accounts for the effective sample size
- $\sigma^2_{pop,0}(t,s)$ is the initial residual population variance estimate for security s at time t and may contain missing values

To estimates missing $\sigma^2_{pop,0}(t,s)$ from the estimation universe, a constrained least squares regression model is fitted at each time, regressing available $\log\left(\sigma_{pop,0}(t,s)\right)$ estimates from the estimation universe against the risk factor exposures, excluding the volatility composite factor, under zero-sum constraints on the sector parameters and on the region parameters. The fitted model is then used to fill in missing values over the estimation universe. The regression takes the form of

$$y_t = X_t b_t + e_t$$

under the constraint

$$C b_t = 0$$

where

- y_t is an $(N_t^A \times 1)$ vector comprising the available $\log\left(\sigma_{pop,0}(t,s)\right)$ estimates at time t
- N_t^A is the number of securities in the estimation universe at time t with available $\sigma_{pop,0}(t,s)$ estimates and available exposures for all risk factors, excluding the volatility composite factor
- X_t^A is an $(N_t^A \times K')$ matrix of the factor exposures, excluding the volatility composite factor, for the set of securities in the estimation universe at time t with available $\sigma_{pop,0}(t,s)$ estimates and available exposures for all risk factors, excluding the volatility composite factor
- K' is the number of risk factors in the model, excluding the volatility composite factor
- C is a $(2 \times K')$ constraint matrix, that is 1 for sector risk factors in the first row and for region risk factors in the second row, and otherwise 0.
- b_t is a $(K' \times 1)$ vector of regression parameters
- e_t is an $(N_t^A \times 1)$ vector of regression residuals

The constraints serve to push the mean fitted variance into the intercept and stops numerical instabilities due to ill-conditioning and collinearity. This reduces the misfit when exposures are missing, noting that the intercept is always one, so is never missing.

Then the fitted regression volatilities for securities in the estimation universe are

$$\sigma_{pop,reg}(t,s) = \underbrace{exp(X_{st}^F b_t)}_{inverse\ expected\ value} \times \underbrace{exp\left(\frac{\sigma_{e_t}^2}{2}\right)}_{lognormal\ scale\ bias\ correction}$$

where

- X_{st}^F is a $(1 \times K')$ vector of the risk factor exposures for the regression parameters, for time t, security s, where security s is in the estimation universe, with missing exposures set to zero.
- $\sigma_{e_t}^2$ is the variance of the regression residuals at time t
- $\sigma_{pop,reg}(t,s)$ is the regression-fitted residual population volatility estimate for security s at time t

The first term of $\sigma_{\text{reg},nt}$ is the inverse transformation of the location and the second term is the lognormal scale bias correction.

An identical process is then applied to produce $\sigma_{pop,reg}(t,s)$ for the remainder of the coverage universe. that is, for the set of securities that are outside the estimation universe, but in the coverage universe.

Then, the final population volatility estimates, using the regression to fill missing values, and denoted $\sigma_{pop}(t,s)$, are

$$\sigma_{pop}(t,s) = \begin{cases} \sigma_{pop,0}(t,s) & , when\ available \\ \sigma_{pop,reg}(t,s) & , else \end{cases}$$

A one-step-ahead bias-statistic regression is used to correct for local broad-market bias patterns in the volatility estimates. Let $b_1(t,s)$ be the one-step-ahead residual standardized by the current volatility estimate:

$$b_1(t,s) = \frac{S(t+1,s)}{\sigma_{pop}(t,s)}$$

If the variance estimate, $\sigma^2_{pop}(t,s)$, is an unbiased forecast, then $b_1^2(t,s)$ has an expected value of one. By fitting a model for $\mathrm{E}[b_1^2(t,s)]$ over the estimation universe at each time, broad bias patterns can be estimated and then corrected for. For example, if $\mathrm{E}[b_1^2(t,s)]$ equals 4, then to correct the bias, $\sigma_{pop}(t,s)$ is multiplied by 2. A regression is fitted each time, comprising a constant and slope, with respect to standardized log-transformed volatility estimates, with the standardization based on market cap, for the estimation universe, and then an additional constant and alternative slope for the remainder of the coverage universe. The regression takes the form:

$$y_t \ = X_t b_t + e_t$$

where

$$X_t = \begin{bmatrix} 1_E & z_t^e & 0_E & 0_E \\ 1_C & 0_C & 1_C & z_t^c \end{bmatrix}$$

$$b_t = \begin{bmatrix} p_{0,t} \\ p_{z^e,t} \\ p_{c,t} \\ p_{z^c,t} \end{bmatrix}$$

and

- y_t is an $(N_t \times 1)$ vector comprising $b_1^2(t,s)$ for securities in the coverage universe at time t that satisfy: $b_1^2(t,s) < 1{,}000$
- e_t is an ($N_t \times 1$) vector of regression residuals for time t
- N_t is the number of securities in the coverage universe at time t that satisfy: $b_1^2(t,s) < 1{,}000$
- N_t^e is the number of securities in the estimation universe at time t that satisfy: $b_1^2(t,s) < 1{,}000$
- $N_t^c = N_t - N_t^e$ is the number of securities in the coverage universe, excluding the estimation universe, at time t that satisfy: $b_1^2(t,s) < 1{,}000$

- 1_E and 0_E are $(N_t^e \times 1)$ column vectors of 1's and 0's, respectively.
- 1_C and 0_C are $(N_t^c \times 1)$ column vectors of 1's and 0's, respectively.
- z_t^e is an $(N_t^E \times 1)$ vector comprising standardized $\log\left(\sigma_{pop}(t,s)\right)$ estimates, given the market cap for security s at time t, truncated at +/- 3, for securities in the estimation universe. A further regression is used to obtain the mean and standard deviation used in the standardization.
- z_t^c is an $(N_t^E \times 1)$ vector comprising standardized $\log\left(\sigma_{pop}(t,s)\right)$ estimates, given the market-cap security s at time t, truncated at +/- 3, for securities in the remainder of the estimation universe. A further regression is used to obtain the mean and standard deviation used in the standardization.

- $p_{0,t}$ is the scalar constant-term coefficient for time t
- $p_{z^e,t}$ is the scalar coefficient for explanatory variable z_t^e for time t
- $p_{c,t}$ is the scalar additional constant-term coefficient, for the remainder of the estimation universe, for time t
- $p_{z^c,t}$ is the scalar coefficient for explanatory variable z_t^c for time t

The aim of including z_t^e and z_t^c in the regression is to account for bias patterns associated with the estimated variance: Low variances tend to be too low and high variances tend to be too high. This pattern is amplified by the estimated autocorrelation corrections. One explanation for this pattern is that the further a variance estimate deviates from its mean, the higher the chance that the value was partially due to the current sample and thus not be expected to repeat at such a large deviation into the future. Since variances tend to reduce as market-cap increases, z_t^e and z_t^c are based on market cap. z_t^e and z_t^c are reestimated each time to allow for time variation in the size of this effect.

z_t^e and z_t^c are constructed to provide standardized deviations from the expected log-volatilities, given each security's market cap. They are constructed and applied separately, z_t^e from and for the estimation universe and z_t^c from and for the remainder of the coverage universe. In each case, the construction process is the same, so only the process to construct z_t^e is described. Let E_t be the set of securities in the estimation universe at time t.

Element z_{st}^e of z_t^e, corresponding to stock s, is constructed as

$$x_{st} = log\left(\sigma_{pop}(t,s)\right)$$

$$\tilde{z}_{st}^e = \frac{x_{st} - E[x_{st}|mkt\ cap, x_{st} \in E_t]}{std[x_{st}|mkt\ cap, x_{st} \in E_t]}$$

$$z_{st}^e = max(-3, min(\tilde{z}_{st}^e, 3))$$

with $E[x_{st}|mkt\ cap, x_{st} \in E_t]$ and $std[x_{st}|mkt\ cap, x_{st} \in E_t]$ estimated from the simple linear regression

$$x_{st} = c_{0,t}^e + c_{m,t}^e \times m_{st}^e + \epsilon_{st}, \qquad for\ s \in E_t$$

where

- m^e_{st} is the log of the market-cap for stock s at time t, winsorized at 1% and 99%, over E_t.
- $c^e_{0,t}$ and $c^e_{m,t}$ are the regression coefficients at time t
- ϵ_{st} is the regression residual

Then

$$E[x_{st}|mkt\ cap, x_{st} \in E_t] = c_0 + c_m \times m^e_{st}$$

and

$$std[x_{st}|mkt\ cap, x_{st} \in E_t] = E[\epsilon^2_{st}]$$

The truncation of $\tilde{z}^e_{st}$ to +/- 3 aims to restrict the influence of more extreme $\tilde{z}^e_{st}$ values. The same process is applied to construct z^c_{st}, using the set of securities outside the estimation universe, but in the coverage universe.

To further restrict unduly influential points, values of y_t exceeding 1,000 are removed from the regression. Even so, linear regressions can potentially produce regions with a negative fit, while volatility estimates must be positive. To strictly enforce positive volatilities, the constraints: $|p_{z^e,t} \times z^e_{st}| < 0.9 \times p_{0,t}$ and $|p_{z^c,t} \times z^c_{st}| < 0.9 \times (p_{0,t} + p_{c,t})$ are enforced. Since $|z^e_{st}| < 3$ and $|z^c_{st}| < 3$, the constrained $p_{z^e,t}$ and $p_{z^c,t}$ become

$$\tilde{p}_{z^e,t} = max(-0.3 \times p_{0,t}, min(p_{z^e,t}, 0.3 \times p_{0,t}))$$

and

$$\tilde{p}_{z^c,t} = max(-0.3 \times (p_{0,t} + p_{c,t}), min(p_{z^c,t}, 0.3 \times (p_{0,t} + p_{c,t})))$$

The time series of parameters are then smoothed using an EWMA, with half-life τ_4:

$$\begin{aligned} p^{smoothed}_{0,t} &= EWMA(p_{0,t}; \tau_4) \\ p^{smoothed}_{z^e,t} &= EWMA(\tilde{p}_{z^e,t}; \tau_4) \\ p^{smoothed}_{c,t} &= EWMA(p_{c,t}; \tau_4) \\ p^{smoothed}_{z^c,t} &= EWMA(\tilde{p}_{z^c,t}; \tau_4) \end{aligned}$$

where τ_4 is chosen to be relatively short. Then the model for the recent variance bias-statistic is

$$\hat{b}^2_1(t,s) = \begin{cases} p^{smoothed}_{0,t-1} + p^{smoothed}_{z^e,t-1} \times z^e_{st}, & if\ s \in E_t \\ p^{smoothed}_{0,t-1} + p^{smoothed}_{c,t-1} + p^{smoothed}_{z^c,t-1} \times z^c_{st}, & if\ s \notin E_t \end{cases}$$

noting the smoothed parameters are shifted back a step to avoid data leakage, while the explanatory variable $z_{d,nt}$ is calculated from the most recent volatility estimate.

$\hat{b}^2_1(\mathrm{t}, s)$ is an estimate of the recent variance bias. That is, multiplying $\sigma^2_{pop}(t,s)$ by $\hat{b}^2_1(t,s)$ brings the expected variance bias back to one. So, $F_1(t,s)$, the one-step-ahead population-variance bias-statistic regression correction, is set to $\hat{b}^2_1(\mathrm{t},\mathrm{s})$:

$$F_1(t,s) = \hat{b}_1^2(t,s)$$

A similar process is used to include autocorrelation corrections into the volatility estimates: initial EWMA estimates, constrained cross-sectional regression to estimate missing values, then a bias-statistic-feedback regression to correct for local broad-market bias patterns in the now-autocorrelation-corrected volatility estimates.

The initial EWMA autocorrelation corrections, $C_{NW,0}(t,s)$, are estimated as:

$$\delta_5 = \left(\frac{1}{2}\right)^{\frac{1}{\tau_5}}$$

$$w_5(t-i,k,t,s) = \begin{cases} \frac{(1-\delta_5)}{1-\delta_5^{W_{NW}-k}} \times \delta_5^i & , if\ S(t-i,s)\ and\ S(t-i-k,s)\ are\ available \\ 0 & , else \end{cases}$$

$$m(k,t,s) = \sum_{i=0}^{W_{NW}-1-k} w_5(t-i,k,t,s)$$

$$\mu_{0k}(t,s) = \sum_{i=0}^{W_{NW}-1-k} \frac{w_5(t-i,k,t,s)}{m(k,t,s)} \times S(t-i,s)$$

$$\mu_{kk}(t,s) = \sum_{i=0}^{W_{NW}-1-k} \frac{w_5(t-i,k,t,s)}{m(k,t,s)} \times S(t-i-k,s)$$

$$m_2(k,t,s) = \sum_{i=0}^{W_{NW}-1-k} \frac{w_5(t-i,k,t,s)}{m(k,t,s)} \times S(t-i,s) \times S(t-i-k,s)$$

$$C(k,t,s) = 1 - \frac{\sum_{i=0}^{W_{NW}-1-k} w_5^2(t-i,k,t,s)}{\left(\sum_{i=0}^{W_{NW}-1-k} w_5(t-i,k,t,s)\right)^2}$$

$$R(k,t,s) = \begin{cases} \frac{1}{C(k,t,s)} \times (m_2(k,t,s) - \mu_{0k}(t,s) \times \mu_{kk}(t,s)) & , \quad if\ m(0,t,s) \geq \frac{1}{2} \\ missing & , \quad else \end{cases}$$

$$\rho(k,t,s) = \frac{R(k,t,s)}{R(0,t,s)}$$

$$C_{NW,0}(t,s) = 1 + 2\sum_{k=1}^{L} \frac{L+1-k}{L+1} \rho(k,t,s)$$

where

- τ_5 is the half-life for the autocorrelation correction
- δ_5 is the decay rate for the autocorrelation correction
- W_{NW} is historical data window size for the autocorrelation correction
- L is the number of autocorrelation lags included

- $w_5(t-i,k,t,s)$ is the normalized exponential weight for time $t-i$, for lag-k residual autocorrelation calculations at time t for security s, with values corresponding to missing data set to 0
- $m(k,t,s)$ is the total normalized exponential weight with available data for lag-k residual autocorrelation calculations at time t for security s
- $S(t,s)$ is the residual, for time t, security s
- $\mu_{0k}(t,s)$ is the lag-0 weighted-sample mean, over data available for lag-k residual autocorrelation calculations at time t for security s
- $\mu_{kk}(t,s)$ is the lag-k weighted-sample mean, over data available for lag-k residual autocorrelation calculations at time t for security s
- $m_2(k,t,s)$ is the lag-k weighted-sample second raw auto-moment, over data available for lag-k residual autocorrelation calculations at time t for security s
- $C(t,s,k)$ is the weighted-sample bias normalization constant, for time t, security s, lag k, which accounts for the effective sample size
- $R(k,t,s)$ is the lag-k residual autocovariance at time t for security s
- $\rho(k,t,s)$ is the lag-k residual autocorrelation at time t for security s
- $C_{NW,0}(t,s)$ is the initial Newey-West autocorrelation correction estimate for security s at time t, and may contain missing values

Missing $C_{NW,0}(t,s)$ are estimated using the same process as for missing $\sigma_{pop,0}(t,s)$. That is, a constrained least squares regression model is fitted at each time, regressing available $\log(C_{NW,0}(t,s))$ estimates from the estimation universe against the risk factor exposures, excluding the volatility composite factor, under zero-sum constraints on the sector parameters and on the region parameters. The fitted model is then used to fill in missing values over the estimation universe. The regression takes the form of

$$y_t = X_t b_t + e_t$$

under the constraint

$$C b_t = 0$$

where

- y_t is an ($N_t^A \times 1$) vector comprising the available $\log\left(C_{NW,0}(t,s)\right)$ estimates at time t
- N_t^A is the number of securities in the estimation universe at time t with available $C_{NW,0}(t,s)$ estimates and available exposures for all risk factors, excluding the volatility composite factor
- X_t^A is an ($N_t^A \times K'$) matrix of the factor exposures, excluding the volatility composite factor, for the set of securities in the estimation universe at time t with available $C_{NW,0}(t,s)$ estimates and available exposures for all risk factors, excluding the volatility composite factor

and otherwise, the variable descriptions remain the same as for the missing $\sigma_{pop,0}(t,s)$ regression, noting that N_t^A will differ when $\tau_3 \neq \tau_5$, since the set of available $C_{NW,0}(t,s)$ will be different. The fitted regression volatilities for securities in the estimation universe also take the same form:

$$C_{NW,reg}(t,s) = exp(X^F_{st}b_t) \times exp\left(\frac{\sigma^2_{e_t}}{2}\right)$$

Again, an identical process is then applied to produce $C_{NW,reg}(t,s)$ for the remainder of the coverage universe. Then, the final autocorrelation-correction estimates, with missing values filled by the regression fits, denoted $C_{NW}(t,s)$, giving

$$C_{NW}(t,s) = \begin{cases} C_{NW,0}(t,s) & , when\ available \\ C_{NW,reg}(t,s) & , else \end{cases}$$

The final step is another bias-statistic-feedback regression, this time applied to the one-step-ahead-bias- and autocorrelation-corrected variance estimates, denoted $\sigma^2_{NW}(t,s)$:

$$\sigma^2_{NW}(t,s) = C_{NW}(t,s) \times F_1(t,s) \times \sigma^2_{pop}(t,s)$$

Since the autocorrelation-correction incorporates L lags, where L is in the order of 5, a multiperiod residual return is needed to assess the bias. A 20-step span is used. Let $b_{20}(t,s)$ be the 20-step simple residual return standardized by its 20-step volatility estimate based on $\sigma_{NW}(t,s)$:

$$b_{20}(t,s) = \frac{\sum_{h=1}^{20} S(t+h,s)}{\sqrt{20} \times \sigma_{NW}(t,s)}$$

Daily regressions are applied to estimate the broad-market bias patterns in $b_{20}(t,s)$, taking the same form as those for $b_1(t,s)$, but with superscripts NW added to the explanatory variable and parameters:

$$y_t = X^{NW}_t b^{NW}_t + e_t$$

where

$$X^{NW}_t = \begin{bmatrix} 1_E & z^{e,NW}_t & 0_E & 0_E \\ 1_C & 0_C & 1_C & z^{c,NW}_t \end{bmatrix}$$

$$b^{NW}_t = \begin{bmatrix} p^{NW}_{0,t} \\ p^{NW}_{z^e,t} \\ p^{NW}_{c,t} \\ p^{NW}_{z^c,t} \end{bmatrix}$$

The regressand, y_t, now comprises $b_{20}(t,s)$, instead of $b_1(t,s)$, and $z^{e,NW}_t$ and $z^{c,NW}_t$ are now constructed from $\sigma_{NW}(t,s)$, instead of $\sigma_{pop}(t,s)$. Otherwise, the variables and associated constraints take the same meaning as for the $b_1(t,s)$ regression, *mutatis mutandis*.

The time series of parameters are smoothed using an EWMA, with half-life τ_6:

$$p^{NW,smoothed}_{0,t} = EWMA(p^{NW}_{0,t}; \tau_6)$$

$$p_{z,t}^{NW,smoothed} = EWMA\left(\tilde{p}_{z,t}^{NW};\tau_6\right)$$

where τ_6 is chosen to be relatively short, noting that there is an in-built 20-step delay. Then the model for the recent autocorrelation-corrected variance bias-statistic is

$$\hat{b}_{20}^2(t,s) = \begin{cases} p_{0,t-20}^{NW,smoothed} + p_{z^e,t-20}^{NW,smoothed} \times z_{st}^{e,NW} &, \quad if\ s \in E_t \\ p_{0,t-20}^{NW,smoothed} + p_{c,t-20}^{NW,smoothed} + p_{z^c,t-20}^{NW,smoothed} \times z_{st}^{c,NW} &, \quad if\ s \notin E_t \end{cases}$$

which constitutes the autocorrelation-corrected-variance bias-statistic regression correction:

$$F_{20}(t,s) = \hat{b}_{20}^2(t,s)$$

Appendix G: Frequently Asked Questions

What is the portfolio coverage threshold for calculating forecasts?
We calculate risk exposures for all equity portfolios, but we make forecasts only if we can generate risk forecasts for at least 80% of the portfolio. We can calculate roughly 10,000 equity funds. Note, this excludes money market funds and funds of funds but includes exchange-traded funds and any equity separately managed accounts with holdings information.

I see stock-level exposures are centered around mean 0 with a standard deviation of 1, but this does not appear to be the case for portfolio exposures. Why is that?
Portfolios are specific subsets of stocks. These subsets are often not equally weighted. Also, the subsets are usually tilted toward large-cap and more-liquid stocks. Furthermore, some stocks are never held by portfolios or indexes for which we have portfolio information. All these factors would contribute to the fact we would never expect portfolios to be centered around mean 0 with a standard deviation of 1.

What is the calculation date of the factor exposure data points?
The risk factors are recalculated daily. For portfolios, we use the most recent portfolio holdings information and assume the portfolio weightings do not change.

Why do region and sector exposures not sum to 1?
Region and sector exposures sum to 1 when we include the intercept term of the Bayesian regression. However, we currently do not display the intercept.

Why do some premia that I observe differ across the risk model options?
Premia depend on what sets of controls are used in the model and the universe over which the model is applied. For example, in some risk models, value shows up with a large premium, and in others, the size premium may be small. In the Morningstar Global Equity Risk Model, value generates a high mean return. This is not the case in the Morningstar U.K. Equity Risk Model.

Do you model equity and fixed-income securities independently?
Yes, we model equity and fixed-income securities independently, capturing the common risks in equities with the equity risk factors using yield-curve factors to capture the impact of interest-rate movements on bonds.

What types of fixed-income bonds do you cover?
Currently, we cover corporate, sovereign, and municipal bonds denominated in five major currencies (USD, EUR, GBP, CHF, and CAD). In future releases, we plan to improve the coverage by adding bonds denominated in more currencies, and to improve the explanatory power of our model by introducing new risk factors to capture the effects of credit, liquidity, prepayment, and interest-rate volatility risks.

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 Risk Model.

For More Information

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

22 West Washington Street
Chicago, IL 60602 USA