Morningstar Global Equity Risk Model
Methodology

Morningstar Data+Analytics
June 6, 2025

Contents

Introduction

Risk is inherent to investing. Developing a prospective view of risk allows investors to analyze and control the risk of their investments and make investment decisions tailored to their individual risk preferences. Risk models are tools to provide such views on risk. A risk model attempts to forecast the distribution of future asset returns, which contains the information needed to assess the riskiness of a portfolio. As the forecasted distribution widens, it indicates more uncertainty about the future returns of the portfolio. With this forecast, investors are empowered to evaluate the riskiness of their portfolios.

Morningstar Global Equity Risk Model (the risk model) is a multifactor risk model built upon a large universe of global stocks to capture the key drivers of risks in equity markets around the globe. Combined with Morningstar's industry-leading portfolio holdings data, the risk model enables the holdings-based portfolio analytics that are highly sought after by asset-management professionals. Users of the risk analytics tools can look through the holdings of a portfolio to analyze a portfolio's factor exposures, forecast the risk of the portfolio, conduct return and risk attribution, and stress-test their portfolio through different types of models across a wide range of prebuilt and custom scenarios.

This methodology document contains the details of the risk model, starting from an introduction of the multifactor modeling framework. Next, we discuss how the universe of stocks covered by the model is selected and how we define the factors in the model, and then we give an overview of the statistical approach of risk estimation. Currency modeling is another key feature of the model and will be discussed in detail. The remainder of the document will introduce the analytical tools based on the risk model, including risk forecasting, factor attribution, and scenario analysis.

Before introducing the details of the model, we highlight some key features of the risk model and how they can facilitate users in the application of the model in their portfolio management workflow.

Model Highlights

Several features make the Morningstar Global Equity Risk Model attractive:

1) A wide coverage of global stock markets

The model covers stocks listed across 123 countries among seven regions. By market capitalization, over 99.8% of Morningstar Global Markets Index and 99.9% of Morningstar US Markets Index is covered by the model.

2) Blend of industry-standard factors with proprietary research

The risk model combines well-known industry-standard factors with Morningstar proprietary research, achieving greater interpretability of the model while providing unique insights. For example, the value-growth factor is calculated based on Morningstar's Style Box Methodology, and the sector and region exposure of a company is estimated with both traditional classification and a company's return dynamics.

3) Flexible configuration for currency modeling

The risk model is offered in multiple currency versions, allowing users in different countries to view the portfolio risk from their home or preferred currency. We model the currency position of a stock based on the currency denomination, which enables a clear interpretation of the currency exposure. At the fund level, we further take into account the currency hedging of the managed investments, a valuable yet rarer feature in risk models.

4) Holdings-based risk analysis

When analyzing portfolios, holdings-based models tend to provide more-accurate risk prediction, factor attribution, risk decomposition, and scenario analysis. Holdings-based models can quickly recognize changes in portfolio holdings and measure and update portfolio attributes accordingly. They also cover new securities or funds without a long history in risk analysis.

5) Full suite of analytical tools based on the risk model

Based on the risk model, Morningstar provides a range of advanced risk-management functions, such as factor attribution, risk forecast, and scenario analysis, which used to be only accessible to the most sophisticated investors through other platforms. Our tools streamline complex risk-management workflows and allow users to extract insights of their portfolios from multiple lenses.

Model Structure

Based on widely applied research in finance, we propose a fundamental factor model structure. The model assumes there exists a parsimonious set of factors that capture the risks of the equity markets and estimates the returns and risks of the factors. The risk model assumes the following structure to explain the stock returns R_t, as

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

where

R_t = N x 1 vector of security returns between time t and $t+1$

X_t = N x K matrix of security-level factor exposures at time t

f_t = K x 1 vector of factor returns (factor premia) between time t and $t+1$

E_t = N x 1 vector of security-level specific returns (residual returns) between time t and $t+1$

In essence, the model attempts to explain stock returns with the differences in a range of characteristics of the stocks, captured by the exposure matrix X_t. The relevant characteristics are determined by the factors we select based on economic theory and empirical research. At each estimation point in time, a regression is then used to estimate the factor premium, f_t. We can interpret the premium as the returns that can be generated by factor portfolios—portfolios that can be composed to have exposure only to one factor (refer to Factor Premium Estimation section for more information). A good risk model under this framework should be able to capture most of the comovement of stock returns with the selected factors, rendering little correlation among the security-level specific returns.

Once the factor premia are estimated for the entire sample period, we can use the time series of factor premia to forecast the future comovement among them, represented by F_t, the K x K covariance matrix. Assuming the security-level-specific return or residual returns are uncorrelated among themselves, we can also forecast the volatility of the residual returns, resulting in a diagonal N x N covariance matrix, Δ_t. With these components, we model the N x N variance-covariance matrix of stocks 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)

This is a simple yet remarkably powerful framework for risk modeling as we can use a small set of K factors to capture the variance-covariance of N securities, which is a key element for both risk forecasting and portfolio optimization in Markovitz's mean-variance framework. Direct calculation of the variance-covariance matrix is often challenging when N, the number of securities, is large relative to the sample period of data we have. This is often the case in financial markets where we have tens of thousands of securities, but returns of those securities are often constrained to hundreds of months or thousands of days.

This multifactor modeling framework carries directly to portfolio-level analytics. A portfolio can be described at time t by an n x 1 vector $w_{P,t}$ that contains the portfolio's holding weights in n assets. A portfolio's factor exposures $X_{P,t}$ is given by the product of the security-level factor exposures $X_{i,t}^T$ and the holding weights $w_{P,t}$:

$$X_{P,t} = X_{i,t}^T \cdot w_{P,t}$$

and the portfolio's variance at time t, $V_{P,t}$, can be modeled using the same framework as

$$V_{P,t} = X_{P,t}^T \cdot F_t \cdot X_{P,t} + w_{P,t}^T \cdot \Delta_t \cdot w_{P,t}$$

This means, under this linear multifactor model, we can forecast any security's or portfolio's risk using the same factor covariance matrix and residual variance.

Model Universe Selection

Model universe is the selection of stocks that can be analyzed by a risk model, and it is fundamentally determined by the regional target of the model. For the global equity model, we target a wide range of countries across the globe (see Appendix B for the full list of countries).

Once the country or region target is selected, we need to determine whether a security is suitable for statistical analysis. 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 or highly volatile stocks. The estimation universe and the extended universe together is the model coverage universe.

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

Estimation Universe	**Coverage Universe**
Approximately 13,500 stocks	Approximately 42,000 stocks
(Curated broad group of large, liquid stocks)	(Small, illiquid stocks in addition to those in the estimation universe)

Source: Morningstar Data+Analytics. Data as of March 1, 2025.

Specifically, we determine the estimation universe based on the trading volume, market capitalization, and the country classification of a company. Our liquidity and market-cap thresholds are time-varying depending on the market condition, which ensures the estimation universe selects investable securities specific at that point in time. See Appendix A for the detailed universe selection process.

Two important issues for a global equity model are the treatment of the multiple share classes of a company listed on multiple stock exchanges and the country classification of a company. We will discuss these two issues in details next.

Company-Level Modeling and Primary Share Class Selection

Our equity risk model employs a fundamental factor methodology where we identify the factors using company characteristics, such as company profitability, market capitalization, and financial positions. Since these attributes are largely the same across a company's multiple share classes, share-class-level modeling will introduce complexity in handling highly correlated factor attributes. After careful consideration of the trade-offs between model complexity and explanatory power, we adopt a company-level modeling approach for model efficiency and interpretability.

A company-level modeling approach requires the selection of a representative share class for market-based information, such as price and trading volume. The selection process is as follows:

- ► Share classes within a company are ranked based on two criteria: trading volume and market capitalization.
- ► These rankings are evaluated at each time point.
- ► The share class with the highest joint rank (that is, highest combined trading volume and market capitalization) is designated as the main share class.

This selection process is dynamic, allowing the main share class to change over time in response to shifts in trading volume and market capitalization.

Security Country Classification

The country classification of a stock is a critical data point utilized in universe selection and the construction of security exposures within the risk model. To assign a country to each security, the risk model employs the Morningstar Business Country Classification Methodology, which integrates multiple factors to account for securities that may trade on both domestic and international stock exchanges.

The following factors are used in determining the country of a security:

- ► Country of incorporation
- ► Country of primary headquarters
- ► Country of primary exchange listing
- ► Geographic source of revenues

These factors are listed in order of priority for business country classification determination. If the country listed for the first three factors are consistent for a company, this will be the resulting business country assignment. In cases where the first three factors yield inconsistent results, the last factor will be considered to determine the appropriate business country assignment. If the geographic revenue of the company is not determined or mentioned, then primary headquarters is considered as business country.

This structured approach ensures that each security is assigned a business country classification that reflects its operational and market characteristics as accurately as possible.

Factor Construction

The next step in building a factor risk model is to determine the factors included in the model. We set out with several criteria:

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

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 average equity market returns. In aggregate, it is the single most important factor in explaining stock returns. We list the remaining factors in each category below and provide more details for each factor in Appendix B.

Style Factors

The style factors attempt to capture the well-known investment styles that investors employ. For example, value and growth investments have long been recognized as the most important investment approaches, and momentum is ubiquitous in all markets. Similarly, investors tend to look for high-quality stocks or stocks with high yield. Academic literature on factor investment also highlights the importance of small size premia, low volatility premia, and illiquidity premia. We therefore select the seven style factors listed in Exhibit 2, with more details of each factor in Appendix B.

Based on the specific definition of each factor, the factor exposures are constructed each day. We then normalize each factor exposure by subtracting the cross-sectional market-capitalization-weighted mean and then dividing by the cross-sectional standard deviation. This means a value of zero can be interpreted as the average score, and a nonzero score of n can be interpreted as being n standard deviations away from the mean. For some factors, we define the sign of the exposure to be consistent with the common definition. For example, the size factor exposure of companies with smaller market capitalization tends to be larger, which helps capture the small size premium.

Exhibit 2 Style Factors

Name	Description
Size	Uses the Morningstar Style Box raw Size score for security-level size characteristics. Higher scores imply smaller market capitalization.
Value-Growth	Uses the Morningstar Style Box raw Style score for calculating value/growth characteristics. Higher scores imply firms that are more growth-oriented and less value-oriented.
Momentum	Total return over the past 12 months excluding the most recent one month. Higher scores imply greater return momentum.
Volatility	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 one month.
Quality	A combination of return/assets ratio and debt/total invested capital. Higher scores imply higher-quality firms.
Yield	Standardized company total yield (dividend plus buyback). Higher scores imply higher yield.
Liquidity	Share turnover of a company. Higher scores imply higher liquidity.

Source: Morningstar Data + Analytics. Data as of March 1, 2025.

Sector Factors

Stocks within the same sector tend to be influenced by similar forces, and it is important to include sector factors in the risk model. Our sector factors measure the economic exposure of a company to the Morningstar sectors. We perform a Bayesian inference time-series regression of a stock's returns against sector benchmarks to find the exposures of an individual company to the sector returns. The prior distribution of the regression coefficient is based on the discrete sector classification from Morningstar. We enforce constraints that our sector exposures must sum to 1 and must individually be between 0 and 1.

Region Factors

Similarly, stocks within a region tend to be affected by similar regional influences. Our region factors capture a company's exposures to the region indexes. We perform a Bayesian time-series regression analysis to find the exposures of an individual company to the returns of a portfolio of stocks in the region with a prior value based on the discrete region classification of the stock by Morningstar. Similar to sector exposures, we enforce constraints that all region exposure must sum up to 1 and must individually be between 0 and 1.

Currency Factors

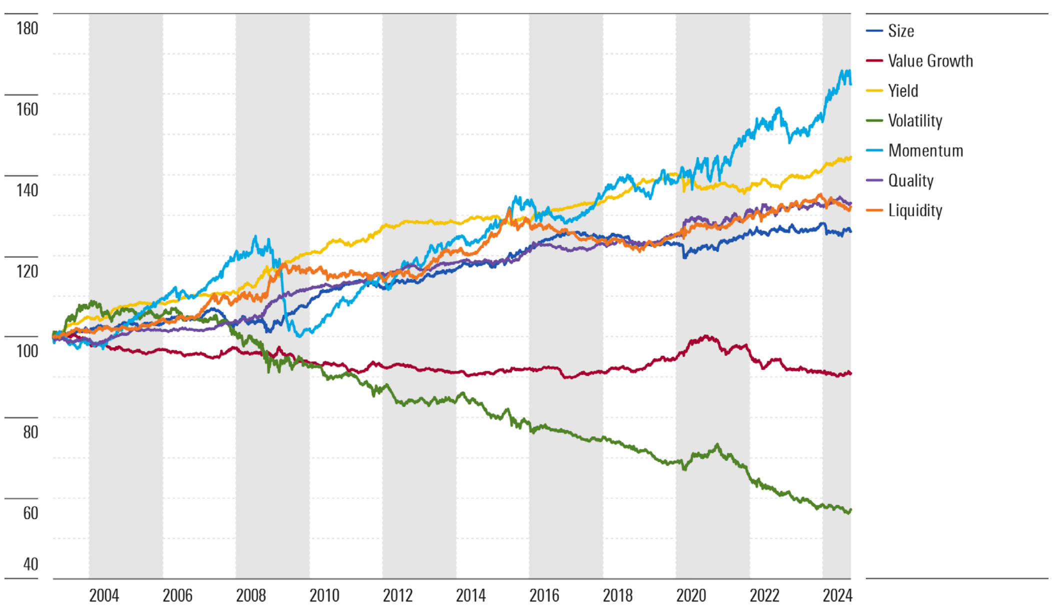
For global investments, the movements in currencies against an investor's local currency can represent significant risk. We attempt to capture such risks through currency risk factors. As the risk arises from movements in exchange rates, we use the changes in exchange rates directly as our factor premia. The factor exposure of a stock to the exchange rate is a dummy variable (either 0 or 1) denoting the currency of the security after currency conversion explained in Appendix C. Therefore, our factor designation captures the pure positional exposure of a security rather than economic exposure. The changes in the exchange rates become the factor premia, which are added to the factor variance-covariance matrix to account for interaction between exchange rates and other factors.

Factor Premium Estimation

Once model universe and factors are selected, we run constrained weighted cross-sectional regressions to estimate factor premia. The imposed conditions resolve the collinearity among sector and region factors, separate out pure factor premia, and provide an intercept that captures the average market return. Appendix D provides the technical treatment of the regression.

Exhibit 3 shows the cumulative return of the Morningstar Global Equity Risk Model style factors between 2003 and 2024.

Exhibit 3 Historical Time Series of the Factor Premia for the Morningstar Global Equity Risk Model



Source: Morningstar Data + Analytics. Data as of March 1, 2025.

Model Base Currency and Currency Factor Adjustment

Note the currency factors are not included in the cross-sectional regression discussed in the previous section and its appendix. Currency factor premia are the changes in the exchange rate between a currency and the model base currency. Therefore, the total return of a security can be represented as:

$$r_i = r_u + \varepsilon_i + r_c$$

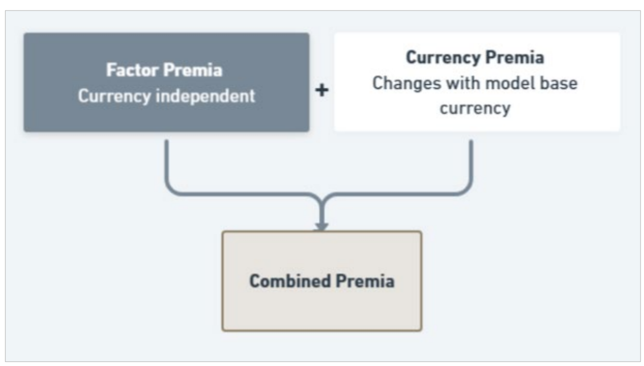
where

r_i: total security return
r_u: return from non-currency factors
r_c: currency returns
ε_i: residual return for security i

A base currency is the currency in which the risk of stocks and portfolio is measured, and investors should choose the base currency that is appropriate for their portfolio. The risk model currently supports five base currencies, including US dollar, Canadian dollar, euro, British pound, and Australian dollar.

As security returns are measured in a stock's local currency or a proxy major currency, the factor return and residual return of a stock estimated from the cross-sectional regression remain the same across models of different currencies. However, the currency factor premia will change depending on the base currency of the model. For example: In the USD base currency version, the premia reflect returns for each of the eight major currencies (for example, euro, Japanese yen, British pound) relative to the USD. Overall, the factors will include both the noncurrency factors and the currency factors that are model-base-currency dependent, as illustrated in the following chart:

Exhibit 4 Combining Factor Premia and Currency Returns



Source: Morningstar Data + Analytics. Data as of March 1, 2025.

Risk Forecast

With all the factor premia calculated, we can forecast the factor variance-covariance matrix, which is a critical input for portfolio risk forecast, attribution, and optimization. The risk model develops a proprietary methodology to account for several key features in volatility forecasting, such as time-varying volatility, auto-correlation in premia, and correction for forecast bias. This section summarizes the methodology, and a technical treatment can be found in Appendix E.

The starting point of a forecast can be the sample variance-covariance matrix of the factor premia. However, this rests on the assumption that the covariances between factors are constant over time. To accommodate time-varying covariance structure, we apply another popular method 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 further enhance the flexibility of the method by using different half-life parameters in estimating the correlation matrix and the variance of premia before combining the two to create a covariance matrix. Longer horizons are typically needed for a more precise estimation of the correlation structure due to the large number of parameters in a correlation matrix, and shorter horizons can provide variance forecast that reflects the current market condition. These estimates are referred to as exponentially weighted moving estimates.

The EWM 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 and the premia standard deviation estimate. 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.

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 the return using an exponentially weighted sample variance. The half-life parameter is typically set at short horizons to reflect the observation that the idiosyncratic risk of stocks is more responsive to short-term events.

Specifically, the residual variance of each security is first estimated as the EWM variance over a 300-day historical window. If a security's residual variance is missing (mostly owing to insufficient returns), a cross-sectional regression is used to impute a variance estimate. This EWM 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 EWM 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 E, along with the optimal half-lives, for different forecast horizons.

Portfolio Analytics With Risk Model

The ultimate application of the risk models is in portfolio analysis and construction. With security-level modeling and portfolio holdings information, we can apply the same factor framework in examining portfolio characteristics such as factor exposures and risks.

Using Morningstar's holdings data, the factor exposures of portfolios are primarily calculated as the weighted sum of security-level exposures. Residual return and residual variance are aggregated similarly using holdings weight. This aggregation process assumes residual returns and variances are independent of each other between stocks, which is a common simplifying assumption in portfolio risk models.

Many portfolios implement currency hedging strategies to mitigate currency risk, utilizing derivatives or other financial instruments to minimize the impact of exchange-rate fluctuations. For portfolios where hedging information is available, we adjust currency exposures to reflect the actual realized currency impact. Currently, our model applies corrections only for fully hedged portfolios. This approach ensures that the risk assessment more accurately represents the portfolio's true currency exposure.

For the majority of the equity portfolio universe, we are able to cover a significant part of the portfolio. In cases where a portfolio holds a large portion of securities that are not covered in the risk model, such as derivatives, or when holdings information is not available, the holdings-based exposure will be inaccurate. Consequently, any analytical results derived from these exposures should be interpreted carefully.

Portfolio Analytics Tools With Risk Model

Based on the risk model, we have developed a wide range of analytical tools to help investors analyze portfolios and make informed decisions:

► Morningstar Portfolio Risk Score. This is a single number that represents the expected risk of a portfolio and can be used by investors to assess whether the riskiness of the portfolio matches their risk profile. The methodology generates a score by analyzing thousands of portfolios and ranking them within the selected universe.

► Portfolio Factor Profile. This tool enables investors to rank and compare the factor exposure of a portfolio relative to a large universe of funds.

► Ex-Ante Risk Forecast and Decomposition. This tool breaks down the volatility forecast into its constituent factors and individual holdings using factor exposures and residual volatility. The analysis provides both total and marginal risk contributions at the portfolio level. Additionally, when compared against a benchmark, it can calculate active risk, offering insights into the sources of relative risk in the portfolio.

► Ex-Post Performance Attribution and Risk Decomposition. This tool enables users to calculate the contribution to portfolio return in terms of factor returns and residuals. It can also generate the contribution of active returns when the portfolio is compared against a benchmark.

► Macro-Financial Scenario Analysis. This tool calculates the impact of user-specified macroeconomic and financial system shocks on forecast factor exposures and volatilities. Users can specify multiple shocks at different points in time and calculate the subsequent expected risk and return of the portfolio.

► Market-Driven Scenario Analysis. This tool allows users to select a market index to determine the impact of user-specified market shocks on factor exposures, portfolio returns, value at risk, and conditional value at risk.

► Morningstar Historical Scenario Analysis. This tool allows users to simulate historical events and examine the impact of those events on their current portfolio.

Please refer to the methodology documents of each tool for more details.

Conclusion

The ability to model the risk of a portfolio is paramount to investment decision-making. Our fundamental factor-based methodology provides a way to capture the main risk factors of the global equity markets, allowing investors to understand the risk drivers of their portfolio and predict future volatility.

The holdings-based portfolio analytics distinguish our risk model from the more commonly used returns-based methodology and allow more-flexible and accurate risk analysis. The currency modeling framework captures the risk arising from major exchange-rate movements and enables investors to adjust the base currency of the model according to their preference.

Through our proprietary approach in forecasting the factor variance-covariance matrix and stock residual variance, the risk model enables sound risk forecasts for equity portfolios, which can be leveraged for portfolio optimization and index construction. The range of analytical tools based on the risk model provided by Morningstar rivals the offerings from industry leaders in the risk-management space and simplifies complex risk-management tasks for investors.

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 risk model system that can enable better risk management and investment decision-making. We recognize there are many decisions to make when constructing a risk model—from universe selection to individual factor calculations to forecasting methods—and we strive to continually improve the model with sound research, high-quality data, advanced technology frameworks, and robust processes.

References

Amihud, Y. 2002. "Illiquidity and Stock Returns: Cross-section and Time-series Effects." *Journal of Financial Markets,* Vol. 5, No. 1, P. 31.

Ang, A., Hodrick, R.J., Xing, Y., & Zhang, X. 2006. "The Cross-Section of Volatility and Expected Returns." *The Journal of Finance*, Vol. 61, P. 259.

Bali, T.G., Cakici, N., & Whitelaw, R.F. 2011. "Maxing Out: Stocks as Lotteries and the Cross-section of Expected Returns." *Journal of Financial Economics,* Vol. 99, No. 2, P. 427.

Banz, R.W. 1981. "The Relationship Between Return and Market Value of Common Stocks." *Journal of Financial Economics*, Vol. 9, No. 1, P. 3.

Fama, E.F., & French, K.R. 1992. "The Cross-Section of Expected Stock Returns." *The Journal of Finance*, Vol. 47, No. 2, P. 427.

Grinold, R.C., & Kahn, R.N. 2000. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk,* 2nd ed. New York: McGraw-Hill.

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

Jegadeesh, N., & Titman, S. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *The Journal of Finance*, Vol. 48, P. 65.

Markovitz, H. 1952. "Portfolio Selection." *The Journal of Finance*, Vol. 7, No. 1, P. 77.

Morningstar. 2024. Morningstar Quantitative Equity Ratings Methodology.

Morningstar. 2024. Morningstar Global Equity Style Box Methodology.

Novy-Marx, R. 2013. "The Other Side of Value: The Gross Profitability Premium." *Journal of Financial Economics*, Vol. 108, No. 1, P. 1.

Rosenberg, B. 1974. "Extra-market Components of Covariance in the Security Returns." *Journal of Financial and Quantitative Analysis*, Vol. 9, P. 263.

Appendix A: Universe Selection Process

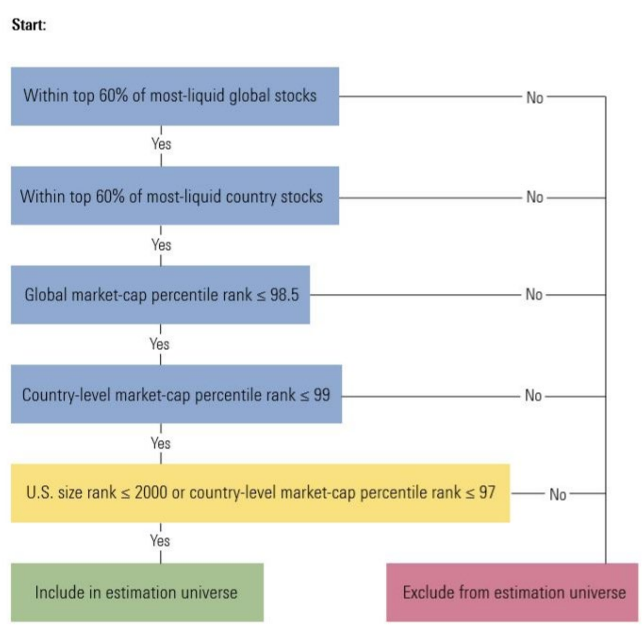
To construct the estimation universe, we rank securities by liquidity and market capitalization at global and country levels daily using a percentile ranking method. These metrics are converted to the model's base currency before ranking to ensure compatibility. Liquidity measures is the median dollar trading volume over the trailing 90 days. As a convention, securities with higher trading volume or market cap receive a higher rank, between 1 and 100, with 1 being the highest rank.

For a security to be included in the estimation universe, it must meet the following four criteria:
- 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 ≤ 98.5.
- It has a percentile rank of country-market capitalization ≤ 99.

In addition to the above filters, we also use size rank, which is a rank based on the sum of a stock's market-capitalization rank, where the stock with the largest market capitalization gets a rank of 1; and its liquidity rank, where the stock with the highest liquidity gets a rank of 1. The stock with the lowest sum is given a size rank of 1. If the number of stocks from the US that satisfy the above four criteria is less than 2,000, we add the most-liquid US stocks back into the universe to achieve this total. These US stocks must have a US-size rank ≤ 2,000 and have a percentile rank of US-market capitalization ≤ 99.

An example of the estimation universe construction logic is depicted below:

Exhibit A1 Estimation Universe Construction Logic



Source: Morningstar Data+Analytics. Data as of March 1, 2025.

Appendix B: Equity Factor Definition

Style Factor Definitions

Size

The size factor reflects the relative market cap of a stock in the estimation universe. We use the raw size score from the Morningstar Style Box as the input for calculating the size exposure of stocks. Raw size score is calculated by scaling a natural logarithm value of market cap. The number is scaled so that each mid-cap stock has a raw Y size score between 100 and 200. Hence, given a stock with a market cap of "cap":

$$Raw\ Y\ =\ 100\ \times\left(1+\frac{ln(cap)-ln(cap_1)}{ln(cap_2)-ln(cap_1)}\right)$$

where:

cap_2 = the market cap that corresponds to the breakpoint between large- and mid-cap stocks for the stock's respective style zone

cap_1 = the market cap that corresponds to the breakpoint between mid- and small-cap stocks for the stock's respective style zone

Raw Y is unbounded for large- and small-cap stocks. Mid-cap stocks are assigned a score between 100 and 200.

To be consistent with the well-documented small size premium (Banz, 1981; Fama and French, 1993), we apply a negative transformation to the raw size score such that small-cap companies have high size scores:

$$Size\ Score\ = -\text{Raw Size Score}$$

Scores are finally normalized within the estimation universe, yielding size exposure.

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 reflects the aggregate expectations of market participants for the future growth and required rate of return for a stock. 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:

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

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

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

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, local-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 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 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 is defined as the volatility of a stock's daily returns over the past six months. We standardize TVOL to obtain its z-score.

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

(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 asset) and the z-score of its financial leverage (trailing 12-month debt/invested capital). The z-score is with respect to all the stocks in the global universe.

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

where ROA is the trailing 12-month return on asset 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 defined as the 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.

Sector Factor Definitions

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

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

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

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

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

Benchmark Construction

Sector benchmark returns are calculated using a market-cap-weighting scheme using stocks from our estimation universe. Stocks are assigned to sectors on the basis of Global Sector ID. All returns are computed in local currency. Market capitalizations are 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 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 stock returns to a set of region benchmarks.

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

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

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

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

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 local currency. Market capitalizations are 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.

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

Exhibit B1 Region Market List

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	

* Russian stocks have been removed from the universe from May 2022.
Source: Morningstar Data + Analytics. Data as of March 1, 2025.

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

Currency Factor Definitions

Currency exposures are binary variable, indicating trading currency of the main share class for a company. If the main share class is traded in one of the eight currencies—US dollar, Canadian dollar, Singaporean dollar, euro, British pound, Japanese yen, Australian dollar, and South African rand—a value of 1 is assigned to the trading currency and 0 to all other. If the share class is not traded in any of these eight currencies, then currency of converted return is used.

Appendix C: Currency Denomination of Stocks

There are over 200 currencies based on the country classification of a stock in the global equity universe. To have a parsimonious set of currencies that captures the major risks in the currency market, we decided to either measure stock returns in their local currency if it belongs to one of the major world currencies or convert the returns to one closely linked major currency.

Major Currencies

We model eight major currencies in the global universe, including US dollar, Canadian dollar, euro, British pound, Japanese yen, Australian dollar, Singaporean dollar, and South African rand. The selection of currencies for our model is based on two primary criteria: their significance in the global market and their role as drivers of regional currency markets. This approach optimizes the balance between comprehensive coverage of major currency influences and model parsimony. As of 2024, stocks that are modeled in the eight currencies account for over 72% of the total market capitalization and over 75% of total global trading volume.

Currency Conversion of Other Currencies

For securities trading in currencies outside the major currency list, their returns are converted to one of the major currencies. The conversion methodology is determined by the economic and market characteristics of the currency as follows.

For securities traded in currencies that are market-determined but heavily influenced by a major currency, return is converted to the economically and spatially nearest major currency:

► Returns of securities traded in New Zealand dollar are converted to Australian dollars.
► Returns of securities traded in Swiss franc, Swedish krona, Norwegian krone, Bulgarian lev, Czech koruna, Danish krone, Hungarian forint, Polish zloty, Romanian leu, and Serbian dinar are converted to euro.

All other nonmajor currencies are converted to US dollars, and approximately 25% of the global equity market by market capitalization falls under this category. These currencies include:

► Those influenced by significant nonmarket factors, such as USD-pegged currencies, such as Chinese yuan, Hong Kong dollar, and Indian rupee.
► Currencies with large volatility, such as Turkish lira and Argentine peso.
► All other remaining currencies.

Appendix D: Constrained Cross-Sectional Regression

This appendix provides the details of the constrained cross-sectional regression and the interpretation of the market factor.

The Constrained Cross-Sectional Regression

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

$$r_i = \alpha + \sum_{m=1}^{M} X_{i,m}\, f_{\mathrm{m}}^{Style} + \sum_{s=1}^{S} X_{i,s}\, f_s^{Sector} + \sum_{r=1}^{R} X_{i,r} f_r^{Region} + \varepsilon_i \quad \text{(D1)}$$

where $X_{i,m}, X_{i,s}, X_{i,r}$ are security i's exposure to style factor m, sector s, and region r; $f_{\mathrm{m}}^{Style}, f_s^{Sector}, f_r^{Region}$ are factor premia for style m, sector s, and region r; M, S, R are the total number of style, sector, and region 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:

$$\sum_{s=1}^{S} u_s\, f_s^{Sector} = \sum_{r=1}^{R} \mathrm{v_r}\, f_r^{Region} = 0 \quad \text{(D2)}$$

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 total return from region or sector of this portfolio should equal to the total return of the market, because adding up all sectors or all regions gives the total equity market. Therefore, region and sector factors together should not contribute extra return on top of the market. Under these conditions, the estimated intercept α gives the market-cap-weighted average return of the equity 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:

$$\tilde{X}_m = \sum_{i=1}^{N} w_i X_{i,m} = 0 \quad \text{(D3)}$$

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

$$r_E = \alpha + \sum_{m=1}^{M} \tilde{X}_m f_{\mathrm{m}}^{Style} + \sum_{s=1}^{S} u_s f_{\mathrm{s}}^{Sector} + \sum_{r=1}^{R} \mathrm{v_r} f_{\mathrm{r}}^{Region} \quad \text{(D4)}$$

where r_E is the market-cap-weighted average return. Note, by equations (D2) and (D3), 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. 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 (D4) 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. $\alpha + f_{\mathrm{s}}^{Sector}$ 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.

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.

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

Security and Portfolio Return Variance 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 between time t and $t+1$
- 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) between time t and $t+1$
- S_t = N x 1 vector of asset-level specific returns (residual returns) between time t and $t+1$

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.

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

$$V_{t,P} = X_{t,P}^T \cdot F_t \cdot X_{t,P} + w_{t,P}^T \cdot \Delta_t \cdot w_{t,P}$$

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. In the remaining section, we introduce the methodology for forecasting the two components separately.

Factor Variance-Covariance Matrix Forecast Estimate

The variance-covariance matrix F_t is forecast in two steps. In Step 1, an EWM estimate is developed that includes the adjustment for autocorrelation. Step 2 uses a feedback mechanism to quantify local equity market volatility deviations and then scales the Step 1 covariance matrix to compensate for such deviation. Covariance forecasts over a horizon are obtained by scaling F_t by the forecast horizon.

Step 1: EWM Estimate With Autocorrelation Adjustment

Our approach to modeling F_t, as an EWM 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. 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 E1 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 Data + +Analytics. Data as of March 1, 2025.

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 EWM 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 EWM 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:

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

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

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: Forecast Bias Correction

In Step 1, EWM variances and EWM 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 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 Variance 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 nonmissing elements of vector L_t, andL_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

Next, the Newey-West estimator, with EWM 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 EWM 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 and separately on the estimation universe and on the remainder of the coverage universe.

The resulting residual volatilities are given by

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

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

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

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

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

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

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

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

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

$$Cb_t = 0$$

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

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

where

- X^F_{st} 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
- $\sigma^2_{e_t}$ 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 $\sigma_{\mathrm{reg},nt}$ 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 $b^2_1(t,s)$ has an expected value of one. By fitting a model for $\mathrm{E}[b^2_1(t,s)]$ over the estimation universe at each time, broad bias patterns can be estimated and then corrected for. For example, if $\mathrm{E}[b^2_1(t,s)]$ 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

$$X_t = \begin{bmatrix} 1_E & z^e_t & 0_E & 0_E \\ 1_C & 0_C & 1_C & z^c_t \end{bmatrix}$$

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

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

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

where

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

Then

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

and

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

The truncation of $\tilde{z}_{st}^{e}$ to +/- 3 aims to restrict the influence of more extreme $\tilde{z}_{st}^{e}$ values. The same process is applied to construct z_{st}^{c}, 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: $|p_{z^e,t} \times z_{st}^{e}| < 0.9 \times p_{0,t}$ and $|p_{z^c,t} \times z_{st}^{c}| < 0.9 \times (p_{0,t} + p_{c,t})$ are enforced. Since $|z_{st}^{e}| < 3$ and $|z_{st}^{c}| < 3$, 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 EWM, with half-life τ_4:

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

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

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

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.

$\hat{b}_1^2(\mathrm{t,s})$ is an estimate of the recent variance bias. That is, multiplying $\sigma_{pop}^2(t,s)$ by $\hat{b}_1^2(t,s)$ 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 $\hat{b}_1^2(\mathrm{t,s})$:

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

A similar process is used to include autocorrelation corrections into the volatility estimates: initial EWM 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 EWM autocorrelation corrections, $C_{NW,0}(t,s)$, are estimated as:

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

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

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

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

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

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

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

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

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

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

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 $\log\left(C_{NW,0}(t,s)\right)$ 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

$$Cb_t = 0$$

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:

$$C_{NW,reg}(t,s) = exp(X_{st}^F b_t) \times exp\left(\frac{\sigma_{e_t}^2}{2}\right)$$

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 $\sigma_{NW}^2(t,s)$:

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

Since the autocorrelation-correction incorporates L lags, where L is in the order of five, 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:

$$y_t = X_t^{NW} b_t^{NW} + e_t$$

where

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

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

The regress and y_t now comprises $b_{20}(t,s)$, instead of $b_1(t,s)$, and $z_t^{e,NW}$ and $z_t^{c,NW}$ 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 EWM, with half-life τ_6:

$$p_{0,t}^{NW,smoothed} = EWMA\left(p_{0,t}^{NW};\tau_6\right)$$
$$p_{z,t}^{NW,smoothed} = EWMA\left(\tilde{p}_{z,t}^{NW};\tau_6\right)$$

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

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

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

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

Appendix F: Morningstar Global Equity Proprietary Factor Risk Model

We have developed a new model, the Global Equity Proprietary Factor Risk Model, which incorporates additional factors to provide users with an alternative perspective on portfolio risk. This model enhances our suite of risk assessment tools by offering a unique approach to understanding and quantifying investment risk.

These new factors are developed by leveraging forward-looking analytics generated by Morningstar's research group and our proprietary database of mutual fund holdings. These factors exhibit low correlation with traditional risk factors and have demonstrated success in predicting future return distributions, making them a valuable complementary addition to our risk factor model.

The primary distinction in this model lies in the set of style factors utilized. All other aspects, including the estimation universe, model structure, regression setup, and benchmark process, remain consistent with our standard methodology. All the factors go through same standardization process as in the Global Equity Model.

The Equity Proprietary Factor Risk Model uses the following factors. Some of them are shared with the Global Equity Model:

- 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

The additional factors unique to this Model are the Economic Moat, Financial Health, Ownership Risk, Ownership Popularity, Valuation Uncertainty, and Valuation factors. It should be noted that the calculation methodologies for Size, Value-Growth, and Volatility factors have been slightly modified in this enhanced model. We explain below in detail the calculation of these factors.

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 the manager holds 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 the manager holds 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.

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.

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.

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.

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.

Exhibit F1 Proprietary Style Factors

Name	Description
Valuation	The ratio of Morningstar's quantitative fair value estimate for a company to its current market price. Higher scores indicate cheaper stocks.
Valuation Uncertainty	The level of uncertainty embedded in the quantitative fair value estimate for a company. Higher scores imply greater uncertainty.
Economic Moat	A quantitative measure of the strength and sustainability of a firm's competitive advantages. Higher scores imply stronger competitive advantages.
Financial Health	A quantitative measure of the strength of a firm's financial position. Higher scores imply stronger financial health.
Ownership Risk	A measure of the risk exhibited by the fund managers who own a company. Higher scores imply more risk exhibited by owners of the stock.
Ownership Popularity	A measure of recent accumulation of shares by fund managers. Higher scores indicate greater recent accumulation by fund managers.
Liquidity	Share turnover of a company. Higher scores imply more liquidity.
Size	Market capitalization of a company. Higher scores imply smaller companies.
Value-Growth	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-oriented 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.

Source: Morningstar Data + Analytics. Data as of March 1, 2025.

Appendix H: Contributors and Version History

Version 3.1, June 6, 2025

Patrick Wang, Ph.D.
Associate Director of Quantitative Research
patrick.wang@morningtar.com

Updates in This Version

- Corrected typos in Appendix B on the region and sector exposure calculation.

Version 3.0, March 15, 2025

Arpit Agrawal
Senior Quantitative Analyst
arpit.agrawal@morningstar.com

Patrick Wang, Ph.D.
Associate Director of Quantitative Research
patrick.wang@morningtar.com

Updates in This Version

- Material Change in Risk Model Methodology - to version 3.0. We have published separate papers covering the Global Equity models (this document) vs. the Global Multi-Asset Model.
- Updated title from "Morningstar Risk Model Methodology" to "Morningstar Global Equity Risk Model Methodology."
- Added sections: Model Structure, Model Base Currency and Currency Factor Adjustment, Portfolio Analytics With Risk Model.
- Updated sections: Introduction, Model Highlights, Model Universe Selection.
- Updated currency factor modeling methodology.
- Removed fixed-income factor definition.
- Removed universe selection rules of the regional risk models.
- Removed Frequently Asked Questions.

Version 2.5, July 18, 2024

Patrick Wang, Ph.D.
Associate Director of Quantitative Research
patrick.wang@morningtar.com

Updates in This Version

- Updated the Introduction and Model Highlights to better reflect the current state of the models.
- Corrected typos in the documents at various places.

Version 2.4, March 26, 2024

Patrick Wang, Ph.D.
Associate Director of Quantitative Research
patrick.wang@morningtar.com

Updates in This Version

- Added Appendix I: Contributors and Version History.

Version 2.3, Dec. 14, 2023

Morningstar Research

Updates in This Version

- Updated to include information on HaRBSA methodology.
- All versions of this document before March 26, 2024, are represented by Version 2.3 and prior.

About Morningstar® Data+Analytics™

Morningstar's Data and Analytics group creates one of the world's broadest and highest quality financial data assets in the world. The team is committed to the meticulous collection, organization, and dissemination of investment information that powers financial decisions globally.

Our data and analytics capabilities span millions of securities and entities across a diverse range of asset classes, including funds, equities, fixed income, and a variety of other financial products. Our unwavering commitment to high-quality data and analytics is the foundation for empowering investor success.

MORNINGSTAR®

22 West Washington Street
Chicago, IL 60602 USA