
Morningstar Fund Flow Model

Methodology

Morningstar Quantitative Research

15 January 2020

Version 1.1

Contents

- 1 Overview
- 1 Introduction
- 2 Universe Construction
- 3 Factor Exposures
- 4 Model Methodology
- 6 Forecast Factor Premiums
- 8 Morningstar SWOT Score
- 9 Conclusion
- 9 References

Appendix A: Fund Flow Factor Definitions

Appendix B: Model Definitions

Appendix C: Input Data FAQ

Appendix D: Model Testing

Appendix E: Fund Level Flow Forecast

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

Harshal Dalvi
Senior Data Scientist
+91 22 61217100
harshal.dalvi@morningstar.com

Som Dubey
Data Scientist
+91 22 61217100
som.dubey@morningstar.com

Madison Sargis
Associate Director of Quantitative Research
+1 312 244-7352
madison.sargis@morningstar.com

Timothy Strauts
Director of Quantitative Research
+1 312 384-3994
timothy.strauts@morningstar.com

Overview

The Morningstar Quantitative Research team has been researching drivers of fund flows since the 2015 launch of the "What Factors Drive Investment Flows" paper. Subsequently, we conducted a body of work researching fund flows for new funds, how fund flows influence fund closures, and the effects of manager changes on fund flows. In total, the work analyzes fund flows during all phases of a fund's lifecycle: launch, operational, closure, and transitional periods. Morningstar is interested in the full lifecycle of fund flows because they represent the aggregate decisions investors make to their portfolios.

Until now, we have been refreshing our flows models on an ad-hoc basis. Now we have implemented the fund flows model into Morningstar products so these insights can be gleaned in a live setting. The first model implemented is the Morningstar fund flow model for the United States:

- ▶ Provide fund growth rate factor exposures across the equity, fixed-income, and allocation asset classes
- ▶ Construct the monthly growth rate return associated with each factor
- ▶ Forecast fund flow growth rates up to one year in the future

Introduction

Morningstar generated a fund flows model to decompose fund flow growth rates into its underlying interpretable characteristics. From a technical perspective, we are modeling growth rates rather than dollar flows because growth rates behave like a stationary variable, similar to returns, rather than cash flows, which behave like nonstationary variables, similar to prices. As a result, our model attempts to decompose the reasons why funds grow and why funds shrink and then use such findings to forecast future growth rates.

Morningstar calculates the one-month growth rate by taking the one-month cash flow divided by the prior month-end's net assets. Asset-management firms are only required to disclose net assets and not investment flows. So, Morningstar calculates estimated flows by looking at the change month-to-month in net assets that cannot be explained by the fund's return. The calculation includes an adjustment for reinvested dividends, which can have a large impact on funds where dividend payouts are large and frequent. For a full explanation, please consult the Morningstar Cash Flow Methodology document listed in the References section.

For the remainder of the methodology, we discuss our coverage universe, factor selection, constructing factor premiums, and the forecast calculation.

Universe Construction

The Morningstar Fund Flow Model uses Morningstar’s global, survivorship-bias-free mutual fund database. In the U.S., the fund flow data set begins in 2003, but because of data availability, the ex-U.S. sample does not begin until 2008. Funds are grouped into one of three broad asset classes—equity, fixed-income, and balanced—because we observed that investor preferences can vary quite strongly depending on the asset classes. We excluded commodity, alternative, and leveraged funds from the model.

We do not restrict the model to a single share class of individual funds because we are interested in modeling the variation in fund flows between share classes, especially as it relates to cost. We recognize that an alternative method of analysis would select representative share classes before proceeding, such as choosing “retail” or “institutional” shares. However, it was not feasible to select share classes in this manner because of regional differences in share-class topology. Furthermore, we believe the primary differences between share-class selection approaches resolve to cost. Given that we included fees in all our regressions, we believe our analysis controls for any fund flow disparities related to fee differences (stemming from share-class differences or otherwise).

We further restrict our analysis to funds that possessed monthly returns and funds that are categorized into a Morningstar broad asset class. We further filter out any funds that have less than \$100 million in assets under management or are less than 36 months old. We remove funds that are not in rated Morningstar Categories, for example, trading–inverse equity.

For model estimation purposes, we add additional filters to only include funds that have more than \$250 million in AUM and are older than 60 months. The model parameters are then extended to the larger universe described above. A full set of criteria for each model is found in Appendix B.

The model's historical data does not suffer from survivorship bias. Morningstar’s global fund databases retain a full history of dead funds, and these funds are included in the historical data set, where appropriate. Moreover, our evaluation technique dynamically incorporates monthly changes in fund universe composition, providing a more holistic and realistic picture of historical performance. Each monthly snapshot captures any funds that were subsequently merged or liquidated away.

Factor Exposures

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

- ▶ A **growth rate** is the change in a fund's net assets over a time period not attributable to return of the portfolio.
- ▶ A **factor** is an observable data point that appears to influence growth rates, like net expense ratio or fund size.
- ▶ A **factor exposure** is a number that measures how much a share class' growth rate is influenced by a factor. Exposures can be positive, negative, or zero. Exposures change through time.
- ▶ A **factor premium** is a number that represents how much a particular factor has influenced share class growth rates for a particular time period.

At first launch, the model uses 19 factors that fall naturally into four distinct groups: performance, price, process, and parent. Exhibit 1 lists the entire set of factors below. A more detailed treatment can be found in Appendix A.

Exhibit 1 Set of Fund Flow Factor Exposures

Performance	Price	Process	Parent
12-Month Growth Rate	Net Expense Ratio	Index Fund	Asset-Weighted Manager Tenure
12-Month Excess Return		Fund of Funds	Average Net Expense Ratio Rank
Cumulative Outflows		Fund Age	Firm AUM
1 Star		Fund Age: 12-Month Growth Rate	Success Five-Year
2 Stars		Fund Size	
3 Stars		Number of Holdings	
4 Stars		Portfolio Concentration	
5 Stars		Socially Responsible Fund	
1-Month Category Return			

Source: Morningstar, Inc.

Model Methodology

To evaluate the fund-specific drivers of flows, we employ a series of monthly cross-sectional regressions. Each month, we regress the forward one-month fund flows on a set of contemporaneous explanatory variables. We believe that this model, as so constructed, offers a glimpse at the underlying decision-making process that investors go through when choosing to allocate their money. We purposefully re-estimate the models by asset class so that we are capturing the within-asset-class variation in fund flows rather than the between-asset-class variation. After the investor has made a decision to allocate to equity funds, we want to know how they go about choosing which equity funds to invest with.

We apply the following framework to regional and market-based subsets of the data. This framework is applied separately by asset class.

Factor Premiums Estimation

Given a collection of factor exposures, X_t , for a set of share classes at time, t , we perform a cross-sectional regression of those exposures on growth rates from t to $t + 1$, gr_{t+1} , to estimate the factor premiums, f_t .

$$gr_t = X_t f_t + \epsilon_t$$

Where:

$gr_t = (n \times 1)$ vector of share class growth rates at time t

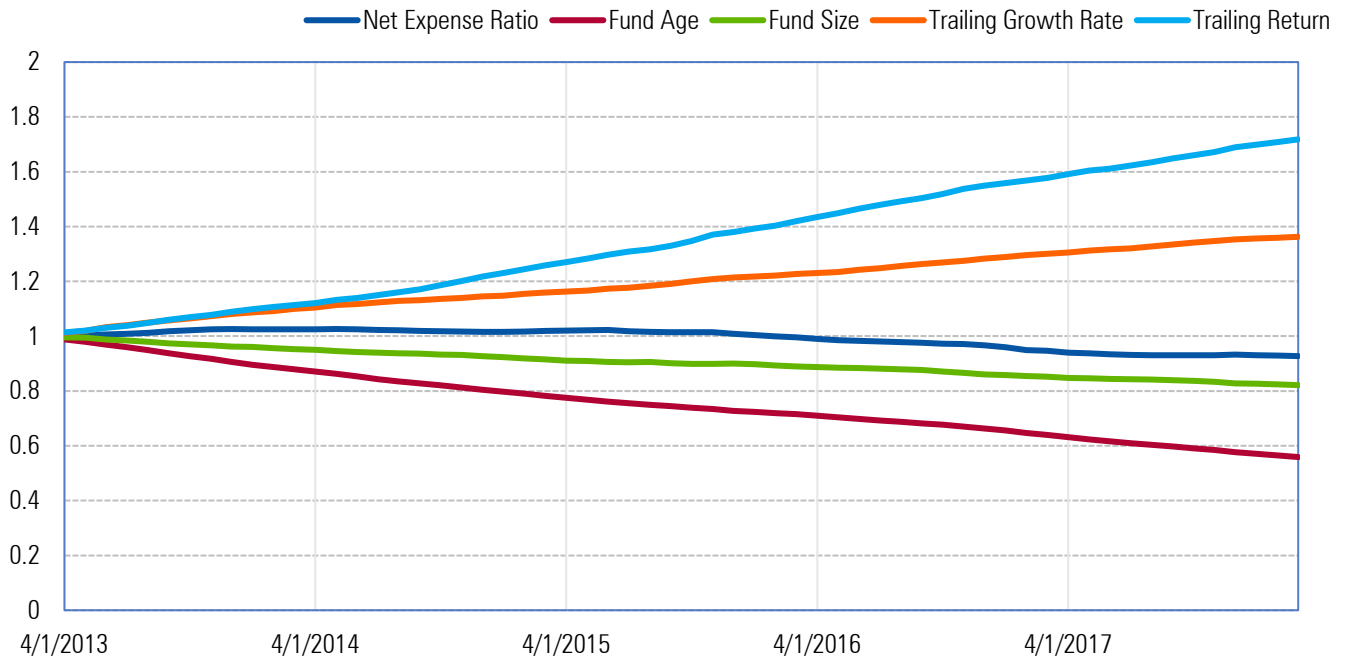
$X_t = (n \times m)$ matrix of share class exposures to factors at time t

$f_t = (m \times 1)$ vector of factor premiums at time t

$\epsilon_t = (n \times 1)$ vector of error terms at time t

By repeating this cross-sectional regression, we construct a historical time series of fund flow factor premiums. We use this time series of factor premiums to analyze how each factor behaves in the context of other factors by examining factor co-movement in the history. Below, we show the trailing five-year cumulative effect of a sample of factors.

Exhibit 2 Historical Cumulative Effect of Equity Factor Premiums for U.S. Equity



Source: Morningstar, Inc.

The model shows trailing returns and trailing growth rates contribute to positive growth rates. Additionally, large funds, older funds, and high-fee funds are growing at slower rates than their smaller, younger, cheaper peers. These findings have been previously well documented in previous fund flow papers.

Forecast Factor Premiums

Our growth rate forecast is a three-step process. We first reduce the dimensionality of factor premiums into principal components. Second, we forecast the components using a vector autoregressive model and forecast residuals using an exponential smoothing model. Third, we map a linear combination of the forecasted premiums to a security's factor exposures via the principal components.

Principal Component Analysis

Principal Component Analysis is performed on the regional factor premiums data to reduce the dimensionality of the analysis and to discern the driving macroeconomic factors of the model. The number of PCA components is determined by the variance explained threshold per each regional model.

The PCA decomposition takes the functional form

$$\underbrace{\mathbf{F}_m}_{t \times f} = \underbrace{\mathbf{PCS}}_{t \times nc} \times \underbrace{\mathbf{M}}_{nc \times f}$$

where f represents the factor premiums, nc represents the number of PCA components, \mathbf{PCS} is the underlying PCA components, and \mathbf{M} is the mixing matrix tying the two models together.

Vector Autoregressive Processes

An autoregressive process for a univariate time series y_t takes the following form

$$y_t = b + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + u_t,$$

where y_t , $\alpha_1 y_{t-1} \dots \alpha_p y_{t-p}$, and u_t are random variables. When multiple time series are considered, the notation in the equation above is extended to obtain a vector autoregressive, or VAR, process:

$$\mathbf{y} = \boldsymbol{\beta} + \sum_{i=1}^p A_i \mathbf{y}_{t-i} + \mathbf{u}$$

where $\boldsymbol{\beta}$ is a fixed K -dimensional vector of intercept terms, p is the number of lags, \mathbf{u} is a K -dimensional white noise or innovation process with a covariance matrix $\Sigma_{\mathbf{u}} = \mathbf{E}(\mathbf{u}_t \mathbf{u}_t^T) = \mathbf{Cov}(\mathbf{u}_t)$, and A_i are $K \times K$ coefficient matrices:

$$A_i = \begin{bmatrix} \alpha_{11} & \dots & \alpha_{1K} \\ \vdots & \ddots & \vdots \\ \alpha_{K1} & \dots & \alpha_{KK} \end{bmatrix}.$$

A VAR model is created with the PCA components calculated in the step outlined above to determine the white noise covariance matrix $\Sigma_{\mathbf{u}}$. The forecast for k -step ahead of y_{t+k} is given by

$$E(y_{t+k}) = \boldsymbol{\beta} + \sum_{i=1}^{k-1} A_i E(y_{t+k-i}) + \sum_{i=k}^p A_i y_{t-i} + \mathbf{v}_t$$

Forecasting Idiosyncratic Movements

In addition to the factor premia, our cross-sectional regressions produce residual terms for each share class in a particular time period. The residual is the difference between the predicted value and the observed value:

$$e(t) = y(t) - \hat{y}(t)$$

where:

$$e(t) = \textit{Residual}$$

$$y(t) = \textit{Observed Value}$$

$$\hat{y}(t) = \textit{Predicted Value}$$

The captured residuals are then used to predict the residual at time $T+1$. We choose to model the residuals through an exponential smoothing method, where the forecast values are based on the historical weighted averages. The weights decline exponentially as we move further into the past. In our current model, we have chosen $\beta = 0.9$.

$$y_{t+1} = \beta y_t + \beta(1 - \beta)y_{t-1} + \beta(1 - \beta)^2 y_{t-2} + \dots$$

Where:

$$y_{t+1} = \textit{forecasted residual}$$

$$y_t = \textit{residual}$$

$$0 \leq \beta \leq 1 \text{ is the Smoothing parameter}$$

Aggregating Forecasts to Funds

A fund's growth rate is the asset-weighted average of the growth rate of its underlying share classes. Using this fact, we map the growth rate of a fund to the linear combination of the principal components.

A fund's growth rate \mathbf{p}_t is a weighted average of the growth rates of its share classes $\mathbf{p}_t = \mathbf{w}_t^T \mathbf{g}r_t$.

A share class' growth rate is an average of the returns of the factors weighted by the security's exposure to the various factors plus the idiosyncratic growth rate.

$$gr_t = X_t f_t + \epsilon_t$$

A factor's growth rate is the average of the principal components' returns weighted by its exposure to each component as determined by the mixing matrix.

$$\underbrace{\mathbf{F}_m}_{t \times f} = \underbrace{\mathbf{PCS}}_{t \times nc} \times \underbrace{\mathbf{M}}_{nc \times f}$$

Therefore, the fund's growth rate is the average of the component growth rates weighted by the portfolio's exposure to each source.

$$\mathbf{p}_t = \mathbf{w}_t^T (X_t \mathbf{f}_t + \epsilon_t)$$
$$\mathbf{p}_t = \mathbf{w}_t^T (X_t (PCS \times M) + \epsilon_t)$$

Morningstar SWOT Score

Morningstar SWOT Score summarizes a fund's peer-relative growth potential relative to the importance of the fund at the firm. To proxy importance at a firm, we use the assets under management of the fund.

$$raw\ SWOT_t = gr_{t+12} AUM_t$$

Finally, we percentile the output using Morningstar Categories.

Conclusion

The suite of fund flow models is Morningstar's first set of statistical models to understand the drivers behind investors' decision-making process for when and why they invest. In the coming months, we will be releasing other lifecycle models and improving the existing models.

We expect that, over time, we will enhance the fund flows models to improve their performance. We will note methodological changes in this document as they are made. ■■

References

Chen, N., Roll, R., & Ross, S.A. 1986. "Economic Forces and the Stock Market." *Journal of Business*, Vol. 59, P. 383-403.

Fama, E.F., & French, K.R. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, Vol. 33, P. 3-56.

Fama, E.F. & MacBeth, J.D. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy*, Vol. 81, P. 607–636.

Morningstar Category Classifications. 2016.

http://morningstardirect.morningstar.com/clientcomm/Morningstar_Categories_US_April_2016.pdf

Morningstar Cash Flow Methodology. Oct. 31, 2011.

[https://www-](https://www-prd.morningstar.com/content/dam/marketing/shared/research/methodology/765555_Estimated_Net_Cash_Flow_Methodology.pdf)

[prd.morningstar.com/content/dam/marketing/shared/research/methodology/765555_Estimated_Net_Cash_Flow_Methodology.pdf](https://www-prd.morningstar.com/content/dam/marketing/shared/research/methodology/765555_Estimated_Net_Cash_Flow_Methodology.pdf)

Morningstar Quantitative Rating for Funds Input Data Methodology. June 26, 2017.

<https://corporate1.morningstar.com/ResearchLibrary/article/813699/morningstar-quantitative-rating-for-funds-input-data-methodology/>

Morningstar Rating for Funds. Sept. 30, 2016.

<https://www-prd.morningstar.com/content/dam/marketing/shared/research/methodology/813568-QuantRatingForFundsMethodolgy.pdf>

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. 2002. "A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods." *International Journal of Forecasting*, Vol. 18(3), P. 439-454.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. 1990. "STL: A Seasonal-Trend Decomposition." *Journal of Official Statistics*, Vol. 6(1), P. 3-73.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. 2001. "Prediction Intervals for Exponential Smoothing State Space Models," Monash Econometrics and Business Statistics Working Papers 11/01, Monash University, Department of Econometrics and Business Statistics.

Meeker, W. Q., Hahn, G. J., & Escobar, L. A. 2017. *Statistical Intervals: A Guide for Practitioners and Researchers*, 2nd ed. (John Wiley & Sons).

Appendix A: Fund Flow Factor Definitions

Asset-Weighted Manager Tenure

The firm-level tenure number is an asset-weighted average of the longest manager tenure of each fund assigned to the firm. The tenure number at the fund level is the number of months the current manager has been on the fund. For funds with more than one manager, the tenure of the manager who has been with the fund the longest is used in the calculation.

A higher score is indicative of a firm maintaining greater manager continuity across their fund lineup. On the margins, we find that investors prefer firms with longer-tenured managers. We see this as a proxy for firm quality.

Average Net Expense Ratio Rank

The firm-level fee number is an equal-weighted average of the net expense ratio equivalent data point ranked by the Morningstar Category of each fund assigned to the firm. Net expense ratios are defined per the definition in the preceding sections and ranked by category. Each fund's fee rank is then averaged to arrive at a firmwide estimate of the typical, relative cost of its fund lineup.

A higher score is indicative of a firm charging less than its competitors. On the margins, we find that investors prefer lower-fee firms, which we see as a proxy for investors' preference for firm quality and stewardship.

Category Average Return

This is the trailing one-month return of the fund's category. A higher score is indicative of a higher-performing category. Typically, fund flows are positively correlated with short-term performance, especially in the fixed-income asset classes.

Cumulative Trailing-12-Month Growth Rate

For each share class, we sum the trailing flows over the past 12 months and then divide by the starting assets. We convert assets and flows into the mode's denominated currency.

$$Growth Rate_{12 Mo} = \frac{\sum_{t=1}^{12} flows_t}{Net Assets_{t=0}}$$

A higher score indicates a fund with a higher 12-month growth rate. Typically, we find that funds that have recently gathered assets continue to do so in the near future.

Cumulative Trailing Outflows

For each share class, we sum the flows over the trailing consecutive months when the fund lost money. A lower score indicates a fund has lost more money than its peers. Funds that lose money typically continue to do so.

Firm AUM

The firm-level AUM number is a simple summation of each fund's AUM assigned to the firm. Firm-level AUM is expressed in U.S. dollars. A higher score indicates a larger firm. With respect to investment flows, firm size matters. Larger firms attract a larger share of the assets.

Fund Age

Fund age is measured as the number of months from inception to time t . Fund age was also similarly right-skewed and therefore it was necessary to log-transform it. A higher score indicates an older fund, whereas a negative score indicates a newer fund.

Fund age has one of the largest estimated effects on flows. There is a negative correlation between fund age and fund growth rates. The older the fund is, the slower we expect the fund to grow, all else equal. We find the reason being that fund launches are strategically marketed and often intentionally timed to meet optimal investor demand.

Fund of Funds

This is a categorical, dummy variable that indicates whether a fund is structured as a fund of funds—a fund that specializes in buying shares in other mutual funds rather than individual securities. Quite often, this type of fund is not discernible from its name alone but rather through prospectus wording (that is, the fund's charter). Because of coverage, we only consider funds of funds for the Allocation asset class.

Fund Size (AUM)

Fund size is measured as the fund's total market value of investments in USD. Not surprisingly, this data point had a heavy right-skewed distribution—there were much larger AUM funds than would be expected under a normal distribution. In order to better prepare these data for an OLS regression, we performed log-transformations on AUM. A higher score indicates a larger fund size.

A structural pattern exists between flows and AUM. There is a negative fund flow premium attached to fund size. We expect that larger funds in terms of AUM should on average receive higher flows in U.S.-dollar terms. However, given that we are modeling organic growth rates, we should expect the opposite—namely, larger funds will grow at a slower pace than small funds, on average. Therefore, the negative correlation that we observe between fund AUM and flows makes sense.

Index Fund

This is a categorical dummy variable that indicates whether a fund tracks an index. While an index typically has a much larger portfolio than a mutual fund, the fund's management may study the index's movements to develop a representative sampling and match sectors proportionately.

Global Net Expense Ratio Equivalent

The Global Net Expense Ratio Equivalent data point definition should be consistent with Morningstar's Expense Ratio definition: "the annual fee that all funds or ETFs charge their shareholders. It expresses

the percentage of assets deducted each fiscal year for fund expenses, including 12b-1 fees, management fees, administrative fees, operating costs, and all other asset-based costs incurred by the fund."

Moreover, this data point definition should hold across boundaries. The following is the logic for calculating a globally consistent expense ratio. Exhibit 3 outlines the calculation for an individual fund's Net Expense Ratio Equivalent (NER_{eq}).

Exhibit 3 Global Net Expense Ratio Equivalent

Single Fund:

$$Net\ Expense\ Ratio\ Equivalent\ (NER_{eq}) = \begin{cases} \text{Calculation} & \text{Domicile} \\ \text{Indirect Cost Ratio (ICR) or Management Expense Ratio (MER)} & \text{Australia} \\ \text{MER} & \text{Canada or New Zealand} \\ \text{Ongoing Charge + Performance Fee} & \text{Europe or United Kingdom} \\ \text{Japan After Tax Total Expense Ratio} & \text{Japan} \\ \text{Net Expense Ratio (NER)} & \text{United States} \\ \text{NER} & \text{Else} \end{cases}$$

Fund of Funds:

$$Net\ Expense\ Ratio\ Equivalent\ (NER_{eq}) = FoFexp_i + \sum_{i=1}^N w_i NER_{eq} \text{ when Fund of Funds} = \text{TRUE}$$

Source: Morningstar, Inc.

A high score indicates a more expensive fund. Investors, on average, prefer lower-fee funds, so there is a negative correlation between fees and fund flows.

Morningstar Rating for Funds

The Morningstar Rating™ is a quantitative assessment of a fund's past performance—both return and risk—as measured from 1 to 5 stars. It uses focused comparison groups to better measure fund manager skill on an after-fee basis. The peer group for each fund's rating is its Morningstar Category. Ratings are based on funds' risk-adjusted returns. Funds must first have a minimum three-year track record. Performance is then assessed after fees and on a risk-adjusted basis. The Morningstar Rating rewards long-term consistent performance, low volatility of returns, and low fees—aspirational areas for the typical investor. More information can be found in the link listed in the References section.

For the purposes of the model, we define dummy variables for the level of the star rating that take the value 1 when a fund is rated a specific level for month t and 0 otherwise. Star ratings can take on values of 1 star, 2 stars, 3 stars, 4 stars, and 5 stars. Note, new funds and share classes without star ratings are assigned 0 for each of the five binary factors. In order to have a baseline for comparison, we considered 3 stars, as it is the middle rating. To do this, we have excluded the "3-star" factor from U.S. flow models. In the rest of the regions, we have considered no rating as a baseline for comparison.

Star ratings are typically published three business days after month-end. It could be argued, therefore, that any results obtained by using star ratings as of time t to predict flows at time t+1 will not represent an appropriate insight. We recognize this and sought to insulate our study from this criticism in addition

to any potential look-ahead bias. Therefore, we use star ratings as of time t-1 to test the efficacy of the rating system in terms of predicting flows. Therefore, the rating information would have been available for nearly a month, allowing plenty of time for investors to act on this information.

Number of Holdings

This is the net number of holdings in the portfolio. It is also an alternative measure of portfolio concentration. A higher score indicates a larger number of holdings.

Portfolio Concentration

This is the aggregate assets, expressed as a percentage, of the fund's top 10 portfolio holdings. This figure is meant to be a measure of portfolio concentration, making it potentially indicative of manager conviction or fund risk. Specifically, the higher the percentage, the more concentrated the fund is in a few companies or issues, the more the fund is susceptible to the market fluctuations in these few holdings, and the more likely the manager has a strong belief in the prospects of these holdings. Cash and cash equivalents are generally not included in this calculation. (An exception is made for money market portfolios.) A higher score indicates a more concentrated portfolio.

Socially Responsible Fund

This is a categorical, dummy variable that indicates whether a fund has identified itself as socially conscious. This data point indicates if the fund selectively invests based on certain noneconomic principles. Such funds may make investments based on such issues as environmental responsibility, human rights, or religious views. A socially conscious fund may take a proactive stance by selectively investing in, for example, environmentally friendly companies or firms with good employee relations. This group also includes funds that avoid investing in companies involved in promoting alcohol, tobacco, or gambling, or in the defense industry.

Success Ratio

Success ratio measures the percentage of a branding company's open-end mutual funds with a Morningstar Category rank of less than 50 over the five-year period through the previous month's end. A higher score is indicative of a firm charging less than their competitors. On the margins, we find that investors prefer funds whose lineups consistently beat their peers. We use success ratio as a proxy for investor preference for firm quality.

Trailing-12-Month Excess Return

For each share class, we calculate the trailing-12-month cumulative U.S.-dollar return. For each category, we calculate the simple average of the trailing-12-month cumulative U.S.-dollar return for all share classes in the category. We then subtract the category average from each share class' return. A higher score indicates a fund with a higher excess return. Fund flows follow performance, so we expect to see a positive relationship between excess return and growth rates.

Appendix B: Model Definitions

Morningstar Global Flow Models

The Morningstar global flow models are specified by the below parameters.

Asset Classes

- ▶ Equity
- ▶ Fixed Income
- ▶ Allocation
- ▶ Alternatives, Commodity, Convertibles, Property, Tax-Preferred

Factors

- ▶ Structural: Index Fund, Fund of Funds (Allocation only), Socially Responsible Fund (Equity only), Net Expense Ratio, Firm Size, Fund Age, Fund Size
- ▶ Performance: Morningstar Rating, Cumulative 12-Month Flows, Cumulative Outflows, 12-Month Trailing Excess Return, Category Average Return, Cumulative 12-Month Flows:FundAge

Data Availability

The flow factors are available for funds meeting the below criteria:

- ▶ Fund Size > USD 100 million
- ▶ Fund Age > 36 months
- ▶ Category must be rated
- ▶ Historical data starts in January 2003

Estimation Universe

The flow factor model is estimated on a set of funds meeting the below criteria:

- ▶ Fund Size > USD 250 million
- ▶ Fund Age > 60 months
- ▶ Absolute Value (One-Month Growth Rate) < .15

Currency: U.S. dollar

Morningstar U.S. Flow Models

The Morningstar U.S. flow models are specified by the below parameters.

Asset Classes

- ▶ Equity
- ▶ Fixed Income
- ▶ Allocation

Factors

- ▶ Structural: Index Fund, Fund of Funds (Allocation only), Socially Responsible Fund (Equity only), Net Expense Ratio, Fund Age, Fund Size, Portfolio Concentration
- ▶ Parent: Firm Size, Asset-Weighted Manager Tenure, Average Net Expense Ratio Rank
- ▶ Performance: Morningstar Rating (excluded 3-star rating), Cumulative 12-Month Flows, Cumulative Outflows, 12-Month Trailing Excess Return, Category Average Return, Cumulative 12-Month Flows:FundAge

Data Availability

The flow factors are available for funds meeting the below criteria:

- ▶ Fund Size > USD 100 million
- ▶ Fund Age > 36 months
- ▶ Category must be rated
- ▶ Domicile = United States
- ▶ Historical data starts in January 2003

Estimation Universe

The flow factor model is estimated on a set of funds meeting the below criteria:

- ▶ Fund Size > USD 250 million
- ▶ Fund Age > 60 months
- ▶ Absolute Value (One-Month Growth Rate) < .15

Currency: U.S. dollar

Morningstar Canada Flow Models

The Morningstar Canada flow models are specified by the below parameters.

Asset Classes

- ▶ Equity
- ▶ Fixed Income
- ▶ Allocation

Factors

- ▶ Structural: Index Fund, Fund of Funds (Allocation only), Socially Responsible Fund (Equity only), Net Expense Ratio, Firm Size, Fund Age, Fund Size, Portfolio Concentration
- ▶ Performance: Star Rating, Cumulative 12 Month Flows, Cumulative Outflows, 12 Month Trailing Excess Return, Category Average Return, Cumulative 12 Month Flows: FundAge

Data Availability

The flow factors are available for funds meeting the below criteria:

- ▶ Fund Size > USD 100 million
- ▶ Fund Age > 36 months
- ▶ Category must be rated
- ▶ Domicile = Canada
- ▶ Historical data starts in January 2008

Estimation Universe

The flow factor model is estimated on a set of funds meeting the below criteria:

- ▶ Fund Size > USD 250 million
- ▶ Fund Age > 60 months
- ▶ Absolute Value (One-Month Growth Rate) < .15

Currency: U.S. dollar

Morningstar Australia Flow Models

Asset Classes

- ▶ Equity
- ▶ Fixed Income
- ▶ Allocation

Factors

- ▶ Structural: Index Fund, Fund of Funds, Socially Responsible Fund (Equity only), Net Expense Ratio, Firm Size, Fund Age, Fund Size
- ▶ Performance: Morningstar Rating, Cumulative 12-Month Flows, Cumulative Outflows, 12-Month Trailing Excess Return, Category Average Return, Cumulative 12-Month Flows: FundAge

Data Availability

The flow factors are available for funds meeting the below criteria:

- ▶ Fund Size > USD 100 million
- ▶ Fund Age > 36 months
- ▶ Category must be rated
- ▶ Region = Oceania
- ▶ Historical data starts in January 2008

Estimation Universe

The flow factor model is estimated on a set of funds meeting the below criteria:

- ▶ Fund Size > USD 250 million
- ▶ Fund Age > 60 months
- ▶ Absolute Value (One-Month Growth Rate) < .15

Currency: U.S. dollar

Morningstar Europe Flow Models

Asset Classes

- ▶ Equity
- ▶ Fixed Income
- ▶ Allocation

Factors

- ▶ Structural: Index Fund, Fund of Funds (Allocation only), Socially Responsible Fund (Equity only), Net Expense Ratio, Firm Size, Fund Age, Fund Size
- ▶ Performance: Morningstar Rating, Cumulative 12-Month Flows, Cumulative Outflows, 12-Month Trailing Excess Return, Category Average Return, Cumulative 12-Month Flows: FundAge

Data Availability

The flow factors are available for funds meeting the below criteria:

- ▶ Fund Size > USD 100 million
- ▶ Fund Age > 36 months
- ▶ Category must be rated
- ▶ Region = Europe
- ▶ Historical data starts in January 2008

Estimation Universe

The flow factor model is estimated on a set of funds meeting the below criteria:

- ▶ Fund Size > USD 250 million
- ▶ Fund Age > 60 months
- ▶ Absolute Value (One-Month Growth Rate) < .15

Currency: U.S. dollar

Appendix C: Input Data FAQ

How do we handle missing data?

In the case of missing data, we have a three-step process. First, we check if the fee data is missing. If so, we carry forward the fund's last calculated global net expense ratio equivalent for the trailing 12 months. Second, for share classes still missing fee data and all other data points, we cross-sectionally impute the median value of the Morningstar Category to which the fund is assigned. Third, in the rare case an entire category is missing data, we impute using the region-asset class group. We use believe Morningstar Categories represents the appropriate peer group to determine similar characteristics. Furthermore, the model aims to forecast growth rates within categories. Therefore, imputed values will be treated as the "average" and hence unlikely to sway the forecast output.

How do we handle outlier data?

All continuous explanatory variables are winsorized between the 98%-99% level and standardized to standard deviation units (mean 0, standard deviation 1) cross-sectionally by date and asset class. The dependent variable is winsorized at the 98% level but is not standardized. Winsorizing the outlier growth rate data reduces the leverage of extreme observations and increases the variance captured within the monthly cross-sectional regressions.

How do we normalize the input data?

After all data is calculated and collected, we cross-sectionally normalize the data by asset class to be mean zero and standard deviation 1. This puts everything into the same units (in terms of standard deviation), which makes the data a bit easier to interpret.

What are the investment types covered by the model?

The model covers exchange-traded funds, insurance funds, open-end funds, unit trusts, UK LP subaccounts, and VA subaccounts.

What does "average" stand for?

Average stands for an equally weighted average of all share classes given a branding ID.

Appendix D: Model Testing

Feature Selection

Below we show the summary mean and t-statistics for the cross-sectional regressions for each of the four global models: equity, fixed income, allocation, and other.

Exhibit 4 Regression Summary Output

	Equity	Fixed Income	Allocation	Other
Intercept	-0.54% (23.66)	-0.46% (13.05)	-0.40% (16.23)	-0.38% (6.40)
Index Fund	0.18% 10.15	0.05% 1.33	0.17% 4.21	-0.73% (4.42)
Socially Responsible Fund	0.07% 6.10	— —	— —	— —
Fund of Funds	— —	— —	0.24% 15.05	— —
Fund Age	-0.06% (11.27)	-0.08% (11.04)	-0.10% (16.04)	-0.02% (0.91)
Fund Size	-0.01% (2.08)	0.04% 5.44	0.00% 0.52	0.07% 2.51
Net Expense Ratio	-0.10% (10.14)	-0.09% (7.56)	-0.05% (7.03)	-0.01% (0.45)
Firm Size	0.07% 14.43	0.04% 3.26	0.04% 9.11	0.01% 0.73
1 Star	-0.26% (9.40)	-0.28% (8.89)	-0.24% (7.85)	-0.10% (0.46)
2 Stars	-0.20% (9.08)	-0.12% (4.18)	-0.13% (5.52)	-0.24% (2.91)
3 Stars	-0.05% (2.25)	0.11% 4.07	0.03% 1.59	0.00% (0.00)
4 Stars	0.26% 10.54	0.32% 10.91	0.27% 11.66	0.15% 1.67
5 Stars	0.72% 20.85	0.66% 20.06	0.68% 20.34	0.75% 6.72

Source: Morningstar Direct. Data as of Jun. 30, 2019.

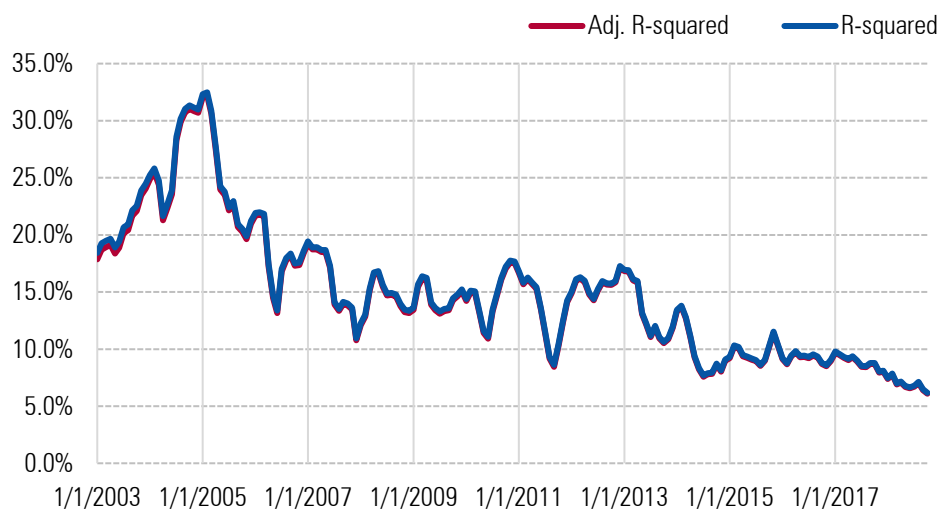
Exhibit 4 Regression Summary Output (Continued)

	Equity	Fixed Income	Allocation	Other
Cumulative 12-Month Flows	0.42%	0.39%	0.58%	0.41%
	36.51	31.05	38.41	12.60
Trailing-12-Month Excess Return	0.33%	0.26%	0.28%	0.49%
	39.24	29.31	31.05	18.96
Cumulative Outflows	-0.57%	-0.65%	-0.48%	-0.63%
	(52.25)	(64.81)	(55.59)	(39.52)
Category Average Return	0.18%	0.19%	0.04%	0.05%
	15.35	10.36	4.06	1.50
Cumulative 12-Month Flows: Fund Age	0.03%	0.05%	0.04%	0.09%
	4.55	6.17	3.14	2.89

Source: Morningstar Direct. Data as of Jun. 30, 2019.

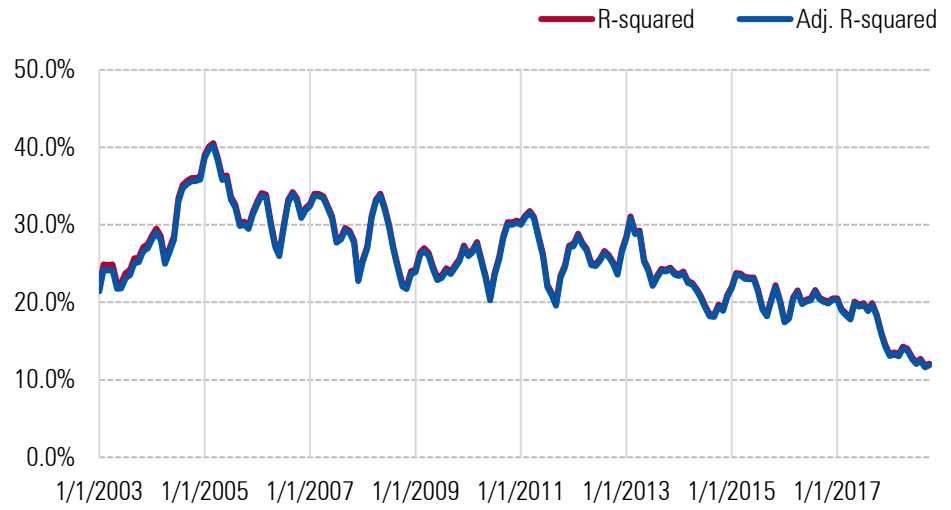
Model Fit

The simplest measure of model fit is the model's R-squared and adjusted R-squared. In Exhibits 5 to 8, we show both measures.

Exhibit 5 Global-Equity Model Fit

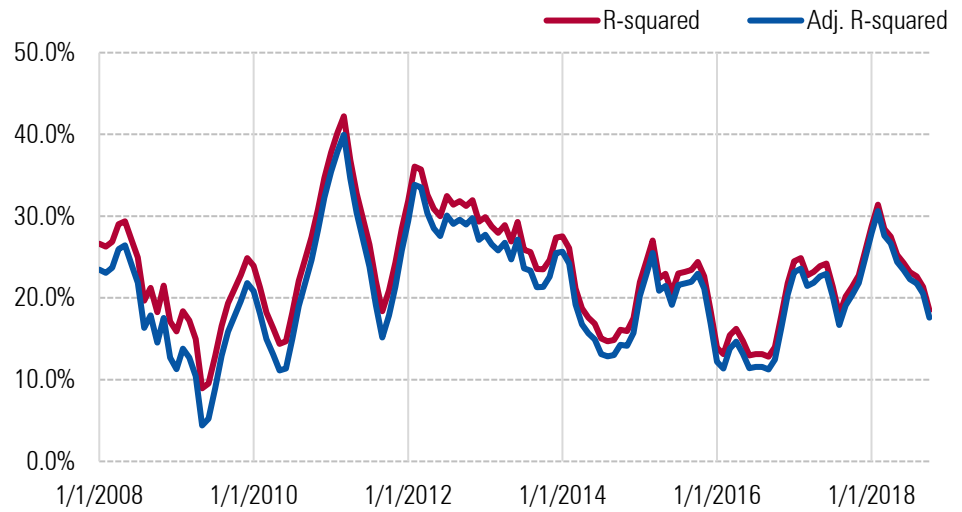
Source: Morningstar, Inc.

Exhibit 6 U.S. Equity Model Fit



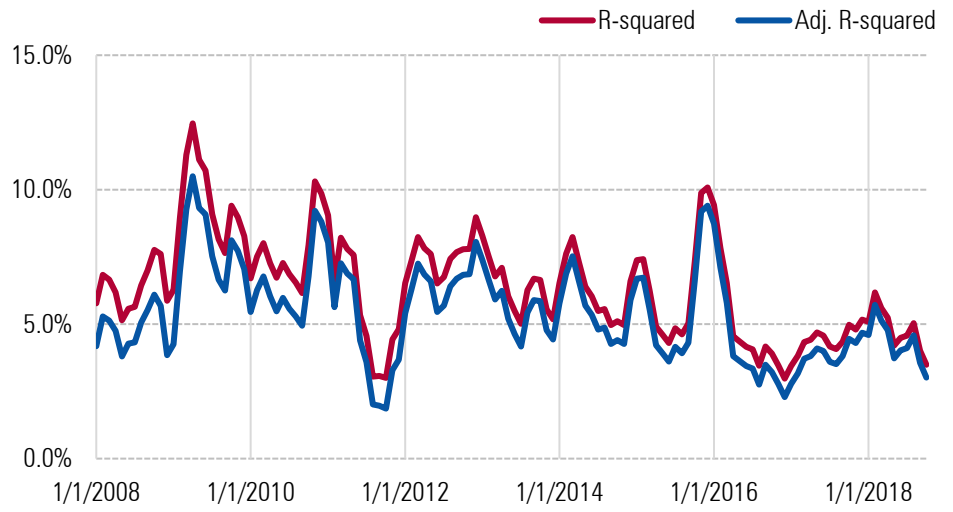
Source: Morningstar, Inc.

Exhibit 7 Canada Equity Model Fit



Source: Morningstar, Inc.

Exhibit 8 Europe Equity Model Fit



Source: Morningstar, Inc.

Appendix E: Fund Level Flow Model

The Morningstar Fund Flow Model framework produces two different lenses for analyzing fund flows. First, we produced the factor-level model to decompose growth rates into attributes. Second, we produce a growth-rate forecast based on trailing growth-rate trends. While the first factor model is produced at the share-class level to accommodate differences in fees and performance, the second model uses fund-level growth rates.

We apply the following framework to U.S. Equity, Fixed Income, and Allocation asset classes. The share classes' net assets and cash flow data is aggregated up for each fund to produce monthly historical organic growth rates. We filter out funds with less than 100 million in assets and feeder funds.

To predict the organic growth rate of mutual funds, we employ multistep-ahead forecasts using univariate Triple-Exponential Smoothing to forecast up to one year ahead. This framework is applied separately for each fund. Below we provide descriptions for the exponential smoothing models.

Exponential Smoothing:

Exponential Smoothing is a time-series forecasting method for univariate data. It is a procedure for continually revising a forecast in the light of more recent experience. Exponential Smoothing assigns exponentially decreasing weights as the observations become older. Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older.

There are three main types of Exponential Smoothing time-series forecasting methods: A simple method that assumes no systematic structure; an extension that explicitly handles trends; and the most advanced approach that adds support for seasonality.

Single-Exponential Smoothing:

Single-Exponential Smoothing is a time-series forecasting method for univariate data without a trend or seasonality used for short-range forecasting. This method usually forecasts just one month into the future.

$$l_t = \alpha Y_t + (1 - \alpha)l_{t-1}$$

$$F_{t+h} = l_t$$

Where:

l_t : Series level at time t

α : Smoothing factor

F_{t+h} : Forecast for h periods-ahead

α (alpha) parameter controls the rate at which the influence of the observations at prior time steps decay exponentially. α is often set to a value between 0 and 1. Large values mean that the model pays

attention to the most recent past observations, whereas smaller values mean more of the history is considered when making a prediction.

Double-Exponential Smoothing:

Double-Exponential Smoothing is an extension to Single-Exponential Smoothing that adds support for trends in the univariate time series. Exponential Smoothing with a trend works much like simple smoothing except that two components must be updated each period—level and trend. The level is a smoothed estimate of the value of the data at the end of each period. The trend is a smoothed estimate of average growth at the end of each period. The specific formula for simple Exponential Smoothing is:

$$\begin{aligned}l_t &= \alpha Y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad 0 < \alpha < 1 \\b_t &= \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad 0 < \beta < 1 \\F_{t+h} &= l_t + hb_t\end{aligned}$$

Where:

b_t denotes the slope at time t

β : trend-smoothing factor

F_{t+h} : Forecast for h periods-ahead

In addition to the alpha parameter for controlling smoothing factor for the level, an additional smoothing factor is added to control the decay of the influence of the change in trend called beta (β). The method supports trends that change in different ways: an additive and a multiplicative, which can be either a linear or exponential trend, respectively. Here, we restrict the trend to additive trend, assuming the data to follow a linear trend. Double-Exponential Smoothing with an additive trend is classically referred to as Holt's linear trend model.

Triple-Exponential Smoothing:

This method is used when the data shows trend and seasonality. To handle seasonality, a third parameter is added. The resulting set of equations is called the Holt-Winters (HW) method after the names of the inventors. There are two main HW models, depending on the type of seasonality.

$$\begin{aligned}l_t &= \alpha(Y_t - s_{t-m}) + (1 - \alpha)l_{t-1} \\b_t &= \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \\s_t &= \gamma(Y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \\F_{t+h} &= l_t + hb_t + s_{t+h-m}\end{aligned}$$

Where:

s_t : seasonal component of the series at time t

m : number of seasons in a year

γ : seasonality-smoothing factor

F_{t+h} : Forecast for h periods-ahead

Model Description:

We use the Triple-Exponential Smoothing with additive trend and additive seasonality to forecast organic growth rate.

To measure the model accuracy, we use the root mean square error, or RMSE. Since the seasonality is unknown for each fund, grid-search on the seasonality parameter is applied and the model with the least error (RMSE) is chosen for forecasting. The seasonality values considered for grid-search are three, six, and 12-month seasonality. The seasonality value with least error on the validation set is chosen, is given as an input to the model, and the growth rate is forecasted.

We also define forecast Uncertainty in terms of High, Medium, and Low based on back-test parameter RMSE and the percentage of similar directionality for each fund. We calculate the percentage of similar directionality based on 12 months of back-testing per the direction of actual and forecasted growth rate in terms of positive and negative growth.

If $RMSE \leq 2\%$ and $directionality \geq 50\%$, then that fund has Low Uncertainty. If $2\% < RMSE \leq 4\%$ and $directionality \geq 50\%$, then it has Medium Uncertainty, and the rest of the funds have High Uncertainty.

We then derive three additional outputs based on the model output.

Cumulative Growth Rate Forecast

We take the product of the monthly forecasted organic growth rates.

Cumulative Flow Forecast

We take the cumulative growth-rate forecast and multiply by the starting period's fund size.

Cumulative Flow Forecast Percentile

Within each Morningstar category, we percentile the funds' Cumulative Flow Forecast for a given forecast time horizon.

Prediction Intervals:

Prediction intervals are used to describe the uncertainty of the forecast outcomes. A prediction interval for a single future observation is an interval that will, with a specified degree of confidence, contain a future randomly selected observation from a distribution (Meeker et al., 2017).

Based on the analytical expressions for the variances of the forecast errors provided by Hyndman et al., (2002), prediction intervals to Exponential Smoothing models are calculated. The expressions to Prediction intervals for multistep-ahead Exponential Smoothing forecasts are as follows:

For 95% Confidence Interval, Prediction Intervals is $[\mu_h - 1.96\sqrt{v_h}, \mu_h + 1.96\sqrt{v_h}]$

where

μ_h : h period-ahead forecast mean

v_h : h period-ahead forecast variance

The expressions for μ_h and v_h are as follows:

$$\text{Variance: } v_1 = \sigma^2 \text{ and } v_h = \sigma^2(1 + \sum_{j=1}^{h-1} c_j^2)$$

Values of c_j for Triple-Exponential Smoothing model is:

$$\alpha(1 + j\beta) + \gamma d_{j,m} \quad d_{j,m} = 1 \text{ if } j = m(\text{mod}m) \text{ and } 0 \text{ otherwise}$$

The simplified expressions for variance for Triple-Exponential model, AA according to Hyndman et al. (2002) is:

$$v_h = \sigma^2 \left[1 + \alpha^2(h-1) \left\{ 1 + \beta h + \frac{1}{6} \beta^2 h(2h-1) \right\} + \gamma k \left\{ \gamma + \alpha [2 + \beta m(k+1)] \right\} \right]$$

$$\text{where } k = \lfloor (h-1)/m \rfloor$$

About Morningstar® Quantitative Research

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

For More Information

+1 312 244-7541

lee.davidson@morningstar.com



22 West Washington Street
Chicago, IL 60602 USA

©2020 Morningstar. All Rights Reserved. Unless otherwise provided in a separate agreement, you may use this report only in the country in which its original distributor is based. The information, data, analyses, and opinions presented herein do not constitute investment advice; are provided solely for informational purposes and therefore are not an offer to buy or sell a security; and are not warranted to be correct, complete, or accurate. The opinions expressed are as of the date written and are subject to change without notice. Except as otherwise required by law, Morningstar shall not be responsible for any trading decisions, damages, or other losses resulting from, or related to, the information, data, analyses, or opinions or their use. The information contained herein is the proprietary property of Morningstar and may not be reproduced, in whole or in part, or used in any manner, without the prior written consent of Morningstar. To license the research, call +1 312 696-6869.