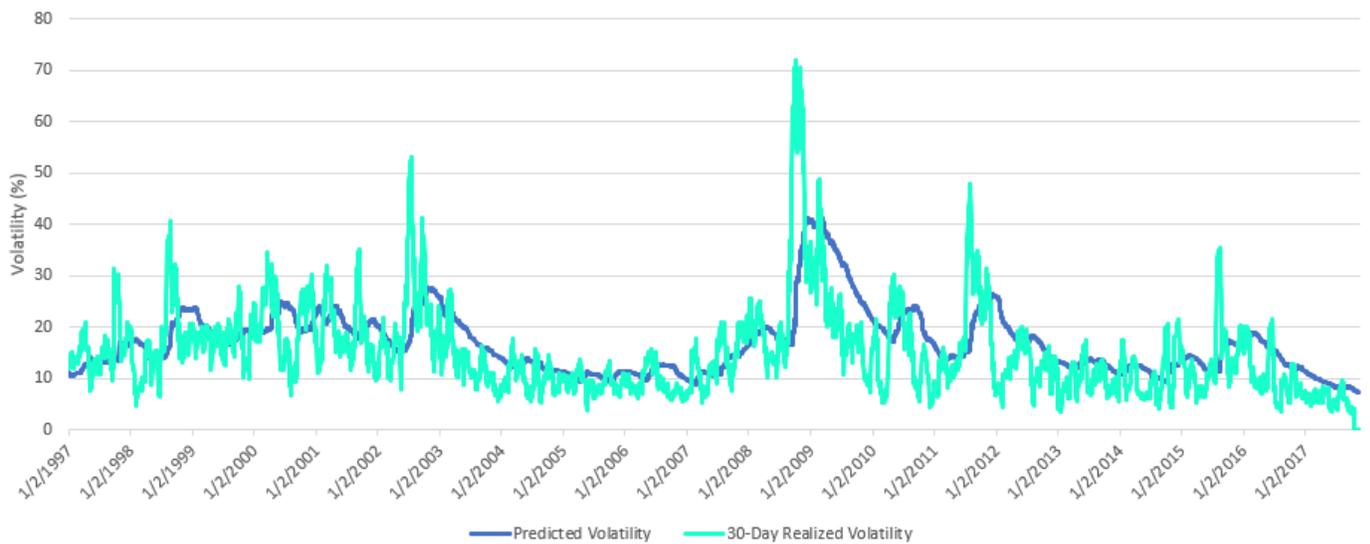# How to Evaluate a Risk Model

Quantitative Measure

Risk model providers often commonly report the average R^2 value of the asset returns model. Some models, such as statistical models, will consistently have greater R^2 values than others. However, strong explanatory power from a returns model does not necessarily translate into an accurate risk model. Low R^2 values, for example, may simply indicate little common behavior within the market, not that the model is failing in any way. The ultimate test of a risk model lies in the testing of its risk forecast against realized values.

Axioma Research Team

This article was originally published on Sept. 26, 2008, and has been revised to reflect an updated chart.

Visual Aids

One way in which to test the risk forecasts of a model is by visually comparing *ex-ante* model predictions of risk for portfolios, individual assets, or factors, with their realized (forward-looking) values. For example, Figure 1 shows the 30-day realized risk for the U.S. Russell 1000 benchmark versus risk forecasts from the Axioma U.S. Short-Term Model. The model risk appears to do a reasonable job of tracking the overall trend while avoiding the short-lived shocks.

Predicted vs Realized Volatilities: Russell 1000 Index and the Axioma Worldwide Equity Risk Model (WW4)

Bias Statistics

To complement visual aids, bias statistics provide hard numbers to assess how well a model is estimating risk. Testing the accuracy of each factor variance and correlation is cumbersome, if not outright infeasible. In contrast, the bias statistic measures the overall quality of the model's full covariance matrix in one go. Given a history of returns, $r_t, t = 1, \dots, T$, and a history of predicted risks, $\sigma_t, t = 1, \dots, T$, we compute the standardized return, $Z_t = r_t/\sigma_t$. Depending on the type of risk one is trying to test, this can be the standardized return of a single asset or a whole portfolio. It can be calculated in terms of total, common factor (or a subset thereof), or specific standardized return, in either active or absolute mode.

The *bias statistic* is defined as the standard deviation of the standardized returns:

$$\text{Bias Statistic} = \sqrt{\text{var}[Z_t]}\,.$$

If the risk forecasts are accurate, then the bias statistic should be close to one. If the bias statistic is significantly below one, then the model has over-predicted risk. Conversely, if the bias statistic is significantly greater than one, then the model has under-predicted risk. Assuming a large sample size and normally-distributed returns, the bias statistic has a 95% confidence interval of

$$\left[1 - \sqrt{2/T}, 1 + \sqrt{2/T}\right].$$

When the computed bias statistic lies within this interval, the null hypothesis that the model is unbiased cannot be rejected. Conversely, when the computed bias statistic lies outside this interval, the null

hypothesis that the model is unbiased is rejected (i.e. the model has a *bias* towards over-predicting or under-predicting risk over time).

Predicted Beta

It is often of importance to predict the beta of a portfolio accurately. The beta of a portfolio as determined by a risk model can be computed as follows:

$$\beta = \frac{w^T Q m}{m^T Q m},$$

where w is a vector of portfolio weights, m is a vector of market weights, and Q is an asset covariance matrix.[1] Similar visual aids and test statistics can be used to test beta forecasts against realized (forward-looking) beta values.

Risk Forecast Changes and Induced Turnover

Satisfactory bias statistics may indicate that risk forecasts have achieved some degree of accuracy, but do not ensure their stability over time. Risk predictions cannot be ``too volatile''; excessive day-on-day change in forecasted risk is likely to lead to continual portfolio rebalancing with all of the expense that this implies. A useful model, therefore, should strike a balance between accuracy or responsiveness of risk forecasts and smoothness of volatility over time.

Forecast variability can be defined as the average relative forecast change, FC: the absolute change in forecast risk from one period to the next, divided by the forecast for the earlier period, averaged over multiple periods. Obviously, excessive changes in risk forecast have an adverse effect on portfolio turnover. A very rough rule of thumb is that the amount of turnover in an active portfolio due to risk forecast changes is approximately $6 * FC$.

Ultimately, the question of what constitutes excessive forecast change and turnover is a subjective one. The ultimate authority is, of course, the model user.

Qualitative Measures

While a risk model's ability to predict risk is crucial, certain other qualitative features of a model are often of equal importance. Risk models are often used for purposes other than determining the predicted risk of some portfolio. Factor risk models are used in risk decomposition to determine the exact sources or risk. Similarly, factor risk models are used in performance attribution to determine the sources of portfolio return. Some risk models are more amenable to these tasks than others. For example, statistical models alone are of little value for these tasks since knowing that the risk of a portfolio is coming from "statistical factor #3" provides no information to the portfolio manager.

---

[1] The market, m, can be arbitrarily chosen by the user with the restriction that the market portfolio must not be the zero vector. For a factor risk model, the covariance matrix, Q, corresponds to $B\Sigma B^T + \Delta^2$.

How to Evaluate a Risk Model

Factor transparency, standard industry classifications, and model structure are all very important in determining the effectiveness of using a factor risk model for risk and performance attribution. Simple factors that are clearly defined allow the portfolio manager to clearly understand exposures to these sources of risk. While the factors themselves are important, exactly how they are used to construct a model has a significant effect on a portfolio manager's ability to interpret the results. In a multi-country fundamental model, the market return may be explained by a market factor, by industries, by countries, or by some combination of the three. The industry and country returns can each be in excess of the market return, or the factors can be constructed such that countries explain the residuals from industries or vice-versa. The different variants can produce significantly different risk decompositions and performance attribution results. The variant that provides the most intuitive results may depend on the specific investment process and manager preference.

Intangibles

In addition to the quantitative and qualitative measures of a risk model, it is also important to consider how a risk model is used within an application. A risk model may have certain properties that allow it to be used in customized ways, particularly when used in software produced by the risk model vendor.

Some software systems provide the ability to view a risk-decomposition according to a set of user-defined factors. This custom risk-decomposition methodology can be used, for example, to decompose the risk estimated by a statistical model according to a set of user-defined fundamental or identifiable factors. The ability to use a customized decomposition can thus be used to provide intuitiveness to a statistical model or to compare sources of risk as determined by multiple models.

There is much debate regarding the significance of countries versus industries in explaining returns and value of industry versus country diversification in an international portfolio. Depending on the investment process, it may be more appropriate to consider risk in different ways; industry risk can be relative to country risk, country relative to industry risk, or both relative to a market. Depending on the particular modeling choice, risk models provide risk estimates where the factors can be interpreted in one of these three ways. Axioma's regional models use country and industry factors that are both relative to the market. When analyzing a portfolio with Axioma Portfolio, these regional models can be used to view risk decompositions in any of the three ways, providing additional flexibility and intuitiveness to understanding the sources of risk.

Risk models often underestimate the risk of portfolios optimized using the same risk model. For example, if a portfolio manager maximizes expected return with a limit on the tracking error with respect to a benchmark of 2%, the realized tracking error is typically considerably larger than 2%. There are several reasons for this, and an optimizer using a risk model should have a methodology in place to adjust for this bias and correctly estimate risk for optimized portfolios. Axioma's Alpha Factor methodology does exactly this. Backtest results show optimized portfolios generated when using the Alpha Factor methodology produced significantly enhanced Sharpe and Information Ratios.

Furthermore, when using the Alpha Factor methodology, the realized risk of optimized portfolios was found to be more in-line with the predicted risk.

Asset owners may require industry or sector exposure reports according to specific standard industry classifications. Using a risk model that contains an industry classification different than that required by the asset owner can make risk management more difficult. If the classification is different, then the portfolio manager must then go through extra efforts to ensure that the risks according to the risk model also meet the sector exposure requirements of the asset owner.

Conclusion

Regardless of the method used to evaluate a risk model, one should use portfolios and risk measurements of personal interest. Risk model providers often provide results of risk model tests on standard benchmarks and equal-weighted portfolios whose assets are randomly selected. The results of these tests may not be representative of the results on actual portfolios, particularly ones generated by an optimizer. Additionally, when looking at either quantitative or qualitative risk measures, the user could consider either active or total risk as appropriate. The risk estimation quality, risk decompositions, and performance attribution results can be considerably different when considering active versus absolute portfolio holdings. In order to get a true understanding of the value and intuitiveness of the results produced by a risk model, it is crucial to look at actual portfolios generated with the user's particular investment process.