RISK MANAGEMENT

Evaluation of Value-at-Risk Models Using Historical Data

Darryll Hendricks
Economic Policy Review
Federal Reserve Bank of New York
vol. 2, no. 1 (April 1996):39–67

Many institutions have adopted value at risk (VAR) as a way to measure portfolio risk. Hendricks discusses the VAR methodology and the various VAR approaches in use today. Using simulated portfolios examined over a long period of time, he looks at how the various VAR approaches perform in measuring risk and how they differ among themselves. Overall, at a 95 percent confidence level, all the VAR approaches examined measure risk accurately, but at a 99 percent confidence level, VAR measures tend to understate the actual risk.

The surge in financial market trading activity and the increasing complexity of financial instruments make it difficult for traders, managers, analysts, and investors to know the financial risks borne by an institution. Value at risk (VAR) addresses this uncertainty by providing a measure of how much portfolio value could decline over a specified period of time (at some level of confidence) as a result of movements in the financial markets. A daily VAR of $10,000,000 at a 95 percent confidence level means that 95 percent of the time, the portfolio is expected to lose no more than $10,000,000 in one day. Given the growth in the use of VAR models, the author examines how well VAR models perform in practice.

Darryll Hendricks is at the Federal Reserve Board of New York. The abstract was prepared by Bruce D. Phelps, CFA, Ark Asset Management.

The author investigates the three most common VAR models: (1) the equally weighted variance–covariance approach, (2) the exponentially weighted variance–covariance approach, and (3) the historical simulation approach. The first two approaches are similar in that they assume price changes are normally distributed and serially independent. But empirical evidence shows that distributions of financial price changes have fatter tails than those predicted by the normal distribution, which may cause VAR measures to underestimate true portfolio risk because the VAR measure is designed explicitly to capture performance at the tails of price change distributions (i.e., the "bad" outcomes). The first two approaches differ from each other in how past data are to be weighted in estimating the variances and covariances of future market movements. The equally weighted approach assumes that data are to be weighted equally over a historical period whereas the exponentially weighted approach gives more weight to recent observations relative to earlier observations.

The third VAR approach uses historical data to simulate the portfolio's performance. In other words, how would the portfolio perform if the market behaved exactly as it did in the past? The portfolio performance data are then used to calculate VAR measures at the desired confidence level. An attractive feature of this VAR method is that it does not make the assumptions of normality or serial independence. Instead, it assumes that the market will behave as it did in the past. If the historical period contains price-change distributions with fat tails, then their impact on the portfolio's performance will be reflected in the VAR measure.

Hendricks also investigates how well the risk measures produced by various VAR models compare with the actual performance of different portfolios. Do the approaches measure risk accurately? Do the approaches perform differently at a 95 percent confidence level than at a 99 percent confidence level? What are the trade-offs in VAR performance if long-term or short-term historical observation periods are used as inputs to the VAR model?

To answer these and other questions, the author generates 1,000 randomly selected foreign exchange portfolios (without option

positions) and calculates VAR estimates for each portfolio during the period from 1983 to 1994. For each portfolio and each day in the sample period, the author calculates 12 one-day VAR estimates that are variations of the three basic VAR approaches described previously. The author then assesses the performance of each VAR approach—each measured separately at the 95 percent and 99 percent confidence levels—using nine performance criteria.

Overall, no single VAR approach is clearly superior to the others. All the VAR approaches accurately measure the level of risk at the 95 percent confidence level and produce estimates roughly similar in average size. At the 99 percent level, VAR measures are somewhat less accurate and tend to understate risk. As expected, the historical simulation approach, which does not assume normality, produces larger risk measures at the 99 percent level than do the variance–covariance approaches. The exponentially weighted approach tends to track portfolio risk over time better than the other two approaches. The results also show that the largest daily portfolio losses can be several times larger than the VAR measure. This finding highlights the fact that VAR measures are not to be treated as "upper bounds" for expected portfolio losses. Other risk measurement methods (e.g., scenario analysis and stress testing) must be used to gauge maximum portfolio losses over a specified holding period.