An Evaluation of Risk Metrics

Vanguard Investment Counseling & Research

Author

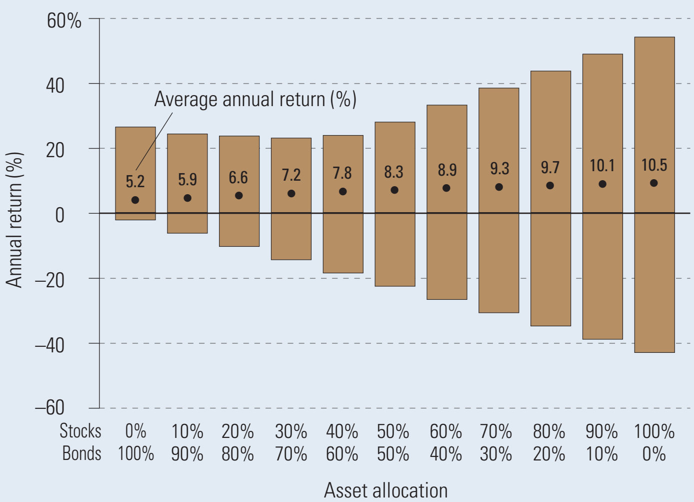
Frank J. Ambrosio, CFA

Executive summary

- Risk assessment is a critical element of effective investing.
- Risks can be measured in many different ways.
- Appropriate risk-measurement techniques depend on specific portfolio objectives.
- Even the best risk measures have limitations. Awareness of these limitations can help investors avoid common pitfalls.

Risk is the uncertainty of an outcome. Risk metrics quantify this uncertainty, providing an estimate of the range of possible outcomes. These measures are critical tools for portfolio construction and performance assessment because a principal assumption of investing is that to achieve a given level of return, investments with lower risk are preferred over those with higher risk. Investments that carry more risk typically offer a higher level of *expected* returns to compensate for the greater uncertainty. However, increased risk by its very nature means that these higher expected returns may not be realized in a given time period. Like any other tool, risk metrics have limitations.

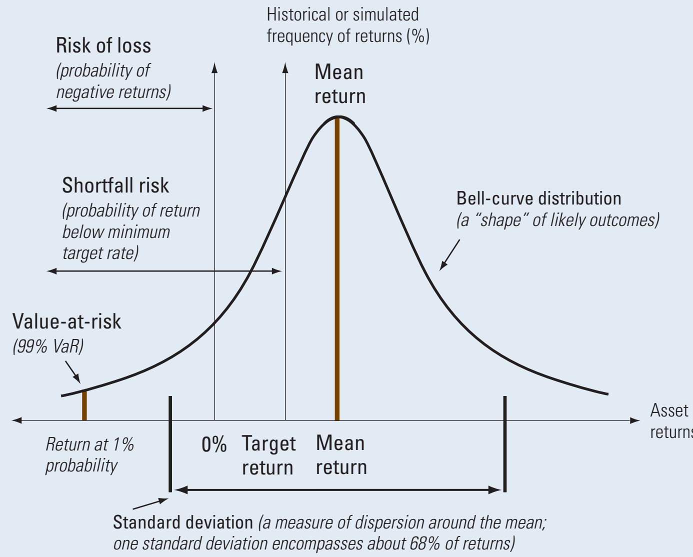
Figure 1. Range of calendar-year returns for various stock and bond allocations, 1926–2006



Benchmarks
Stocks = Standard & Poor's 500 Index, 1926–1970; Dow Jones Wilshire 5000 Composite Index, January 1, 1971–April 22, 2005; Morgan Stanley Capital International (MSCI) US Broad Market Index, April 23, 2005–December 31, 2006.
Bonds = Ibbotson Intermediate Government series, 1926–1972; Lehman Brothers Intermediate U.S. Treasury Index, 1973–2006.

Sources: Ibbotson Associates, Vanguard Investment Counseling & Research.

Figure 2. Various absolute risk measures for investment returns



Source: Vanguard Investment Counseling & Research.

Risk metrics are based on the historical performance of asset classes. Although historical performance is inescapably the foundation for our best estimates of future risk, the future will never repeat the past. The proverbial one-hundred-year storm can undermine expectations based on even the most sophisticated risk metrics.

Even if risk can't be predicted with certainty, risk metrics provide critical information to help answer the most important question for any investor: How should a portfolio be invested to gain the best chance of achieving its objectives? Risk metrics have also proven reliable for comparing the relative risks of different asset classes. For example, **Figure 1** displays the range of annual returns for various combinations of stocks and bonds over 81 years, providing a clear illustration that as the weighting of stocks in a portfolio increases, so does the volatility of the portfolio's returns.

Risk measures fall into two categories: absolute and relative. In this paper, we review widely used metrics, describing their uses and limitations. We also identify considerations for investors applying risk metrics to portfolio construction and oversight. Ideally, for major mandates such as pension or endowment funds, risk metrics will be incorporated into the policy statement. This practice not only provides guidance to portfolio managers but can also help fiduciaries in assessing portfolio performance over time.

Successful use of risk metrics depends on selecting measures that are consistent with a portfolio's objectives. It is also essential to understand the limitations of available data.

Absolute risk measures

Widely used measures of absolute risk include *standard deviation*, *value-at-risk (VaR)*, *risk of loss*, and *shortfall risk*. Here we define these measures and illustrate their uses and limitations.

Portfolio standard deviation

Standard deviation, a basic statistical tool, is widely used to measure the degree of fluctuation in a portfolio's return. The larger the standard deviation, the greater the magnitude of the fluctuations from the portfolio's average return. Consider a portfolio with an average return of 10% and a standard deviation of 15%. Its returns should fall between –5% and 25% in 68% of all observations.

Standard deviation can be a useful measure for portfolios such as pension and endowment funds that are concerned with the consequences of both positive *and* negative deviations from a specific return target or portfolio value. It is less appropriate for investors who worry more about negative returns than about upside deviations. In addition, standard deviation assumes that returns are normally distributed, which limits its use for investments with unusual return distributions.

Historical value-at-risk

Historical value-at-risk is based on a security's worst results over a given period. The measure can be derived from a fixed percentage of the worst observations—the worst 1% or 5%, for example—or a fixed number of those observations.

Historical VaR can provide a visceral sense of what risk means. The worst annual return for the stock market in the last 81 years was –43.1% (in 1931). In other words, the stock market's one-year VaR, based on roughly the worst 1% of observations, is –43.1%. This metric can be useful in explaining risk to people who are less familiar with investment theory.

Because historical VaR looks only at the worst observations without any regard for their frequency, many analysts prefer to use it in combination with other measures. Historical VaR is often employed in combination with risk of loss.

Risk of loss

Risk of loss is a useful counterpart to historical VaR, describing the frequency of bad results. It measures the percentage of outcomes below a certain total-return level, usually 0%. This risk metric is most often used to assess the likelihood that a portfolio will fall below a specific return or asset value target. Value-at-risk and risk of loss can be used by almost any portfolio as the ultimate test of risk tolerance. Would there be comfort if an extreme negative event occurred, or should risk levels be reduced?

Suppose an investor's objective is to maintain a portfolio's nominal value for a year. A risk-of-loss metric indicating, for example, that during the past 81 years, 30% of annual stock market returns were negative would be a useful piece of information.

Table 1. Example of risk metrics using broad asset-class data for 1926–2006

	Nominal total return			Real total return		
Asset class	Average annual return	Percentage of years with negative return	Highest one-year loss	Average annual return	Percentage of years with negative return	Highest one-year loss
Treasury bills	3.8%	0%	0.0%	0.8%	35%	–15.0%
Bonds (intermediate-term Treasury)	5.2%	9%	–2.3%	2.1%	38%	–14.5%
Stocks	10.5%	30%	–43.1%	7.2%	35%	–37.3%

Benchmarks

Treasury bills: Citigroup 3-Month Treasury Bill Index, 1926–2006.

Bonds: Ibbotson Intermediate Government series, 1926–1972; Lehman Intermediate U.S. Treasury Index, 1973–2006.

Stocks: S&P 500 Index, 1926–1970; Dow Jones Wilshire 5000 Index, January 1, 1971–April 22, 2005; MSCI US Broad Market Index, April 23, 2005–December 31, 2006.

Sources: Ibbotson Associates, Vanguard Investment Counseling & Research.

Risk of loss can also be measured in real terms: that is, the amount of the investment's return after inflation is removed. **Table 1** shows that the measurement of losses in real terms can be quite different from the measurement in nominal terms.

Historical VaR and risk of loss are most useful when they are derived from results over long time periods. If based on shorter, less-volatile periods, these measures may give investors a false sense of security. Generally, the longer the period, the more negative the worst potential loss could be at any point—something that is true looking either backward or forward in time.

Shortfall risk

Shortfall risk is the probability that an investment's value will be less than is needed to meet portfolio objectives. This probability can be measured using a variety of approaches, including historical time-path simulation and Monte Carlo simulation. The shortfall-risk metric is most often used to help create a comprehensive investment plan, taking into account current assets and estimated future liabilities in both the accumulation and spending phases. The metric might indicate, for example, that an investment portfolio stands a 25% chance of being depleted before the end of the liability funding period.

Shortfall risk could be used by an institution or individual spending, or expecting to spend, regularly from portfolio assets. Examples include endowments and foundations, pension plans, and people investing for retirement objectives.

Although the result is a seemingly simple probability percentage, the measure can be extremely complicated to calculate and understand. Minor changes to inputs and assumptions can cause wide differences in the results. Shortfall risk is only as good as the inputs, most of which are uncertain and subject to estimation errors.

Relative risk measures

Widely used measures of relative risk include *excess return*, *tracking error*, *Sharpe ratio*, *information ratio*, *beta*, and *Treynor ratio*.

Excess return

Excess return is a security's return above or below that of a benchmark or a theoretically risk-free asset such as Treasury bills. Excess return is simply the portfolio's or security's return minus that of the benchmark, making it easy to understand and calculate. If an actively managed mutual fund returns 11%, and its benchmark returns 10%, the fund's excess return is 1%.

Investment advisors use excess return to compare their portfolio returns with those of a benchmark. The metric's usefulness depends on two assumptions: that the portfolio's total risk (systematic and idiosyncratic) is similar to the benchmark's, and that the returns of the portfolio and benchmark always move in the same direction. Unless these conditions are met, excess return will provide limited insight. A portfolio might outperform a benchmark, for example, but take on more risk to do so. Excess return wouldn't account for this risk.

Tracking error

Tracking error is the standard deviation of excess return. Like portfolio standard deviation, tracking error assumes that returns are normally distributed. It combines both upside and downside risk. Consider an index fund that has no excess return relative to its benchmark when measured over a long period, but that produces an annualized tracking error of 10 basis points (0.1 percentage point). If the benchmark returns 10% per year, the fund's return should be between 9.9% and 10.1% (10 basis points on either side of the benchmark return of 10%) in 68% of the observed one-year time periods (one standard deviation).

Tracking error is commonly used to assess an index fund's success in matching its target index. Active managers who are closely tied to a benchmark (quantitative managers, for example) might describe their expected deviation from the benchmark in terms of tracking error. The metric could also be used to evaluate the success of a risk-controlled mandate, such as a balanced portfolio designed to generate returns within 50 basis points of a balanced benchmark's performance. Like portfolio standard deviation, tracking error is less useful to investors concerned mostly with downside risk.

Tracking error also is less directly relevant for actively managed portfolios that are not closely pegged to benchmarks. However, this metric is used in calculating the information ratio (discussed on the next page), which is often employed in comparisons of active managers.

Table 2. Sharpe ratios of various assets over portions of the period 1970–2006

	Sharpe ratios		
	1970–1981	1982–1999	2000–2006
Commodities	0.3984	0.1654	0.3356
Real estate	0.2577	0.4742	1.3920
International developed stock markets	0.0841	0.4809	0.0911
U.S. stock market	0.0301	0.7381	–0.0806
U.S. fixed income market	–0.1375	0.8193	0.9032
Long U.S. Treasury bonds	–0.2743	0.5647	0.5914

Benchmarks

Commodities: Goldman Sachs Commodity Index (GSCI) Total Return Index.

Real estate: National Association of Real Estate Investment Trusts (NAREIT) Equity REIT Index.

International stock markets: MSCI Europe, Australasia, Far East (EAFE) Index.

U.S. stock market: S&P 500 Index, 1970; Dow Jones Wilshire 5000 Index, January 1, 1971–April 22, 2005; MSCI US Broad Market Index, April 23, 2005–December 31, 2006.

U.S. fixed income market: Citigroup High Grade Bond Index, 1970–1975; Lehman Aggregate Bond Index, 1976–2006.

Long U.S. Treasury bonds: Ibbotson Long Government series, 1970–1972; Lehman Long U.S. Treasury Index, 1973–2006.

Source: Vanguard Investment Counseling & Research.

Sharpe ratio

The Sharpe ratio is a representation of the risk-adjusted return of a portfolio or security. The Sharpe ratio measures how much return is being obtained for each theoretical unit of risk. To calculate a Sharpe ratio, an asset's excess return versus a risk-free asset such as Treasury bills is divided by the standard deviation of the asset's returns.

Sharpe ratios can be negative if the asset underperforms the risk-free asset. For a longer-term measurement, the ratio generally falls in a range from 0 to +1; the higher the number, the better.

Sharpe ratios are used to compare investments from the same asset class or from asset classes with similar liquidity and valuation characteristics. Sharpe ratios are extremely time-period dependent and can vary greatly depending upon the length of the period used, as shown in Table 2.

In addition, an uncritical use of Sharpe ratios can lead to faulty conclusions. As Table 2 indicates, higher ratios often occur when an asset's performance is reaching its peak—as was true for commodities in 1981 and the stock market in 1999. If the Sharpe ratio were being used to select an asset for future outperformance, the metric could be counterproductive.

Information ratio

The information ratio is the risk-adjusted return of a portfolio or security versus a benchmark. To calculate the information ratio, an asset's excess return is divided by its tracking error relative to the benchmark. (The Sharpe ratio is actually an information ratio that uses the risk-free return as the benchmark.) This metric is typically used to measure a manager's skill versus peers.

An actively managed fund that has 100 basis points of excess return and 200 basis points of tracking error relative to its benchmark would have an information ratio of 0.5. A fund that had the same excess return and 400 basis points of tracking error would have an information ratio of 0.25. All else equal, higher information ratio values are preferred.

As with the Sharpe ratio, information ratios are extremely time-period dependent and can vary greatly depending upon the length of the period used. They tend to peak at the times of best performance, sending potentially misleading signals to investors about future returns.

Beta

Beta is the magnitude of an investment's price fluctuations relative to the ups and downs of the overall market. The market (or index) is assigned a beta of 1.00; if a portfolio has a beta of 1.20, its price would rise or fall by 12% when the market rises or falls by 10%. Beta is best used to measure the systematic or market risk of an investment and can be appropriate in evaluating an investment for possible inclusion in a diversified portfolio.

Hedge fund managers often cite beta as a measure of portfolio risk. For example, market-neutral managers try to offset the betas of their long positions with those of their short positions to reduce the portfolio's overall beta toward zero. In calculating betas, correct benchmark selection is critical. A portfolio beta derived from an index with different risk characteristics reveals little about the portfolio's potential volatility. Another caution: Betas are not stable over time. Frequent, even daily, rebalancing may be necessary to maintain a portfolio's beta.

Treynor ratio

The Treynor ratio is the risk-adjusted return of a portfolio or security versus the market. It is an asset's excess return versus a risk-free asset such as Treasury bills, divided by the asset's beta. The Treynor ratio is an appropriate measure of a portfolio's return per unit of risk. The measure has a drawback, however, in that it assumes a portfolio manager has diversified away all of the unsystematic (company-specific) risk, and that only systematic (market) risk is left. This limits the use of the Treynor ratio to comparisons involving extremely well-diversified portfolios.

An actively managed equity fund with a 1% excess return and a beta of 1.20 would have a Treynor ratio of 0.833. Higher Treynor ratios are preferred to lower ones.

The Treynor ratio is based on beta, so it shares beta's limitations. In addition, the Treynor ratio magnifies small differences in beta over time, particularly in the case of low-beta strategies such as those of market-neutral hedge funds. Dramatic changes in Treynor ratios may not necessarily reflect dramatic changes in risk.

Pitfalls of risk measures in a portfolio context

Risk metrics are critical tools for portfolio construction and oversight, but a simplistic use of these measures can lead to faulty conclusions. Potential pitfalls are related to the limitations of historical data, the integrity of the data, and challenges in adapting metrics to a specific portfolio. Maintaining awareness of what risk metrics can tell you—and what they can't—is the best way to sidestep these pitfalls.

Limitations of history

All risk metrics use historical data to make assumptions about future risk. But what if the future proves to be very different from the past? Until October 19, 1987, history suggested that the historical VaR for any one-day period in the U.S. stock market was –12.8%—the decline of the Dow Jones Industrial Average (DJIA) on October 28, 1929. On October 19, 1987, the DJIA fell by 22.6%, and the Wilshire 5000 Index lost 17.2% of its value.

Did stock market risk increase dramatically overnight? For an investor concerned with the worst possible one-day outcome, the answer might be yes. However, most portfolio managers take a longer-term view: Many mandates have a horizon measured in decades. For these portfolios, risk did not change quite so dramatically, especially considering that the total return of equities for all of 1987 was positive.

Data integrity

Data can present an incomplete picture of the opportunities available to real-world investors. In less-liquid markets, such as emerging markets or the small-cap stock arena, for example, higher transaction and management costs diminish the opportunities implied by cost-free benchmarks and risk metrics. This difference between expectations and reality can lead to flawed conclusions about an asset's optimal weight or a manager's success.

Quantitative risk metrics may also imply a precision and reliability that don't exist in the underlying data. Many investment-strategy databases have survivor bias and incubation bias that can overstate the performance of different strategies and managers. These risks are magnified when performance reporting is voluntary. In addition, the returns of some strategies—private equity, hedge funds—are based on relatively infrequent appraisal or mark-to-market methods, which can keep volatility artificially low. A study co-authored by Burton Malkiel shows that the full universe of hedge-fund returns, if made public and computed, would be 400–500 basis points lower than the reported returns (Malkiel, 2005).

Time-period dependency

As the October 1987 crash suggests, risk is time-period dependent. Risk metrics based on a longer time series are less influenced by short-term extremes, but the entire history of our financial markets is, in essence, just one time period. How do we address the risk of time-period dependency?

First, we could forecast risk through modeling based on a combination of historical and current information. This approach has its own shortcomings, however, depending upon the length of the forecast period and the number and accuracy of the variables involved. While the use of current data could possibly enhance accuracy, these shortcomings introduce an additional forecasting risk that must be taken into account with any solution.

Second, we could simulate the future using the available historical data by employing either a time-pathing or a Monte Carlo approach. The time-pathing approach keeps the historical data series in the order in which it occurred, but each simulation starts at a different point in time. As the series reaches the end date, the simulations loop around to the beginning of the data to fill in the end of the investment horizon. This approach is useful for selecting a subset from a long series of historical data.

The Monte Carlo approach uses many more simulations and fits the simulated data to the characteristics of the historical data, but does not use the actual historical sequences themselves. This approach can highlight the potential for extreme outcomes that did not actually occur in history. The Monte Carlo approach is useful for historical data that is distributed fairly normally, generally in a bell-shaped curve. This approach can employ additional techniques, such as vector auto-regression, to help ensure that the results include not only the normalized distribution of the returns, but also a sense of the environment in which these returns are achieved.

Ultimately, selecting the appropriate time period for a portfolio's risk metrics can become a question not only of time horizon but also of how the financial markets' performance is modeled over that period.

Benchmark selection and manager risk

Benchmark selection is an equally tricky adaptation issue. Relative risk measures such as excess return, tracking error, and beta are based in part on the performance of a benchmark. If the benchmark does not represent the portfolio effectively, these risk measures will fail to provide useful information.

Consider, for example, an equity mutual fund that is being compared with the Standard & Poor's 500 Index. The S&P 500 is dominated by large-capitalization stocks. If the equity fund happens to invest only in mid-cap stocks, its characteristics will differ significantly from those of the benchmark. If the fund further devotes itself to growth stocks, there will be an additional style difference between the benchmark and the fund. These differences would show up as a larger tracking error versus the benchmark and a more volatile excess return. Beta, which would have a value of 1 if the benchmark exactly matched the fund, might deviate significantly from 1 and would vary with time, possibly switching above and below 1.

Someone evaluating this mid-cap growth fund on the basis of the S&P 500 comparison might conclude that active management was responsible for the high degree of variation. Note, however, that the benchmark mismatch could produce similar results for a mid-cap growth *index* fund.

The question of benchmark selection is directly connected to the process of portfolio design. Ideally, portfolios are constructed in a "top-down" process—that is, the investor first clarifies the investment objective and establishes the level of risk that is deemed acceptable. The investor then determines the mix of asset classes most suitable to the objective, next chooses sub-asset classes, and finally selects specific investment vehicles. Among the many benefits of this process is that the investor gains a deep understanding of the portfolio and thus can readily evaluate which benchmarks are relevant, and which less so, for its performance.

Investors who lack such understanding of their portfolios are vulnerable to being misled by metrics that do not account for risks absent from market benchmarks, such as manager risk, factor risk, and investment-specific risk. In particular, choosing investment vehicles that are heavily dependent on a manager's performance may expose the flaws of portfolio-analysis tools. The manager selected for a strategy is a major influence on returns when the performance spread between managers is wide, as in the private equity, hedge fund, or real estate markets. In that situation, using index returns as a benchmark can lead to unrealistic expectations and to portfolio allocations that are unproductive over time.

By the same token, using quantitative methods on their own can lead investors into some serious traps. A degree of qualitative adjustment to the inputs is appropriate and prudent when comparing asset classes over several different lengths of time, when juxtaposing different valuation measurements, and, most importantly, when choosing vehicles to implement specific investment strategies.

Precise objective is key to success

The more precisely you define your objective, the better your ability to select metrics that can help you evaluate your portfolio's performance and risk exposures. Examples of objectives and possible primary risk measurements include:

- *Objective:* Reach a targeted real-return level to meet current annual distributions. *Possible key risk measure:* Find the investment or portfolio with the lowest percentage of real-return observations below the target—a risk-of-loss metric.
- *Objective:* Minimize present value of future investment contributions. *Possible key risk measure:* Measure standard deviation of returns relative to a target return.
- *Objective:* Accumulate wealth to meet a future level of real liabilities. *Possible key risk measure:* Measure the shortfall risk of each potential investment or portfolio and choose the one with the lowest probability of missing the target liability level.
- *Objective:* Match the return of a specified benchmark. *Possible key risk measure:* Measure the tracking error of the portfolio versus that benchmark.
- *Objective:* Maximize risk-adjusted performance. *Possible key risk measure:* Look for an investment or combination of investments with the highest Sharpe ratio.

Conclusion

Risk measurement is a vital part of portfolio construction and oversight. The investment policy statement should articulate a portfolio's risk-management goals and identify the metrics used to evaluate its success in reaching them. The appropriate metrics depend on the portfolio's objective. In addition to choosing the right quantitative tools, it is important to use them with a qualitative understanding of their limitations. Blind reliance on sometimes-flawed measures can lead to suboptimal investment decisions.

The difficulty in measuring risk only serves to underscore the value of a top-down approach to asset allocation. Because major asset classes like stocks, bonds, and cash have long return histories and well-defined relative risk relationships, portfolio construction should begin with a decision about how best to combine these classes based on the portfolio's objective and the level of risk deemed acceptable. Decisions about specific investments should be made only at the end of this process and again should incorporate a risk analysis. Portfolio benchmarks should be carefully chosen to match the investment strategy. Following the process of first determining the objectives, then determining the risk measures, will result in a deeper understanding of the risks being taken in the investment portfolio and a better ability to assess its subsequent performance.

Additional reading/references

Clarke, Roger G., 1998. Alternative Measures of Risk. In *Investment Management*, Peter L. Bernstein and Aswath Damodaran (eds.). New York: John Wiley & Sons. p. 81-98.

Hirt, Geoffrey A., and Stanley B. Block, 2006. *Fundamentals of Investment Management*. Boston, Mass.: McGraw-Hill/Irwin. 710 p.

Malkiel, Burton G., and Atanu Saha, 2005. Hedge Funds: Risk and Return. *Financial Analysts Journal* 61(6):80-8.

Molitor, Jeffrey S., 2006. *Evaluating managers: Are we sending the right messages?* Valley Forge, Pa.: Investment Counseling & Research, The Vanguard Group. 13 p.

Nawrocki, David N., 1999. A Brief History of Downside Risk Measures. *Journal of Investing* 8(3):9-25.

Sharpe, William F., 1994. The Sharpe Ratio. *Journal of Portfolio Management* 21(1):49-58.

Stocks, Mary Ellen, and Christopher to, 1997. Value at Risk for the Asset Manager. *Journal of Performance Measurement* 1(4):35-43.

P.O. Box 2600
Valley Forge, PA 19482-2600

Connect with Vanguard® > www.vanguard.com

Vanguard Investment Counseling & Research

Ellen Rinaldi, J.D., LL.M./Principal/Department Head

Joseph H. Davis, Ph.D./Principal
Francis M. Kinniry Jr., CFA/Principal

Frank J. Ambrosio, CFA
John Ameriks, Ph.D.
Donald G. Bennyhoff, CFA
Maria Bruno, CFP®
Scott J. Donaldson, CFA, CFP
Michael Hess
Julian Jackson
Colleen M. Jaconetti, CFP, CPA
Kushal Kshirsagar, Ph.D.
Karin Peterson LaBarge, Ph.D.
Christopher B. Philips, CFA
Glenn Sheay, CFA
Kimberly A. Stockton
Yesim Tokat, Ph.D.
David J. Walker, CFA

FLGERM 0307