Look-Ahead Benchmark Bias in Portfolio Performance Evaluation

GILLES DANIEL, DIDIER SORNETTE, AND PETER WOEHRMANN

GILLES DANIEL
is a research assistant in the department of management, technology, and economics at UBS Investment Bank in Opfikon, Switzerland.
gillesdaniel@ethz.ch

DIDIER SORNETTE
holds the chair of Entrepreneurial Risks in the department of management, technology, and economics at ETH in Zürich, Switzerland.
dsornette@ethz.ch

PETER WOEHRMANN
is a visiting assistant professor in the department of management science and engineering at Stanford University in Palo Alto, CA.
peterw1@stanford.edu

Market professionals and financial economists strive to estimate the performance of mutual funds, hedge funds, and basically any financial investment, and to quantify the return–risk characteristics of investment strategies. Having selected the funds and/or strategies of interest, a time-honored approach consists of quantifying their past performance over some time period. A large literature has followed this route, motivated by the eternal question of whether some managers/strategies systematically outperform others, with its implications for market efficiency and investment opportunities.

Backtesting investment performance may appear straightforward and natural at first sight. However, a significant literature has unearthed, studied, and tried to correct for ex post conditioning biases, which include survival bias, look-ahead bias, and data-snooping, which continue to pollute even the most careful assessments. In this article, we present a dramatic illustration of a variant of look-ahead bias that we refer to as *look-ahead benchmark bias*, which surprised us by the large amplitude of the overestimation of expected returns of up to 8% per annum. This overestimation is comparable to the largest amplitudes of the survival biases and look-ahead biases found for mutual funds or hedge funds. We demonstrate the generic nature of look-ahead benchmark bias by studying the performance of portfolios investing solely in regular stocks using very simple strategies, such as buy-and-hold, Markowitz optimization, or random stock picking.

The look-ahead benchmark bias that we document is strongly related to look-ahead bias and to survival bias, but has no particular relation to data-snooping, which we will therefore not discuss further.[1]

The standard survivorship bias refers to the fact that many estimates of persistence in investment performance are based on datasets that only contain funds in existence at the end of the sample period.[2] The corresponding survivorship bias is caused by the fact that poor-performing funds are less likely to be observed in datasets that only contain the surviving funds, because the survival probabilities depend on past performance. Perhaps less appreciated is the fact that stocks themselves also have a large exit rate and hence also suffer from survival bias. For instance, Knaup [2005] examined the business survival characteristics of all establishments that started in the U.S. in the late 1990s when the boom of much of that decade was not yet showing signs of weakness and found that if 85% of firms survive more than one year, only 45% survive more than four years. Bartelsman, Searpetta, and Schivardi [2003] confirmed that a large number of firms in a group of 10 OECD countries enter and exit most markets every year. Data covering the first part of the 1990s show the firm turnover rate (entry plus exit rates) to be between 15% and 20% in the business sector of most countries (i.e., a fifth of firms are either recent entrants or will close down within the year). And this phenomenon of firm exits is not confined to small firms. Indeed, in the exhaustive Center for Research in Security Prices (CRSP) database of about 26,000 listed U.S. firms, covering the period from January 1927 to December 2006, we find that, on average, 25% of names disappeared after 3.3 years, 75% disappeared after 14 years, and 95% disappeared after 34 years.

The standard look-ahead bias refers to the use of information in a simulation that would not be available during the time period being simulated, usually resulting in an upward shift of the results. An example is the false assumption that earnings data become available immediately at the end of a financial period. Another example is observed in performance persistence studies, in which it is common to form portfolios of funds/stocks based upon a ranking performed at the end of a first period, together with the implicit or explicit condition that the funds/stocks are still in the selected ranks at the end of the second testing period. In other words, funds/stocks that are considered for evaluation are those that survive a minimum period of time after a ranking period (Brown et al. [1992]). This bias is not remedied even if a survivorship-free database is used, because it reflects additional constraints on ranking.

More generally, the fact that a dataset is survivorship free does not imply that standard methods of analysis do not suffer from ex post conditioning biases, which in one way or another may use (often implicit or hidden) present information that would not have been available in a real-time situation.

Previous works have investigated both survivorship and look-ahead biases. Brown et al. [1992] showed that survivorship in mutual funds can introduce a bias strong enough to account for the strength of the evidence favoring return predictability previously reported. Carpenter and Lynch [1999] found, among other results, that look-ahead-biased methodologies, which require funds to survive a minimum period of time after a ranking period, materially bias statistics. ter Horst, Nijman, and Verbeek [2001] introduced a weighting procedure based on probit regressions, which models how survival probabilities depend upon historical returns, fund age, and aggregate economy-wide shocks, and which provides look-ahead-bias-corrected estimates of mutual fund performance. Baquero, ter Horst, and Verbeek [2005] applied the methodology of ter Horst, Nijman, and Verbeek to hedge fund performance, which requires a well-specified model that explains survival of hedge funds and how it depends upon historical performance. ter Horst and Verbeek [2007] extended the look-ahead-bias correction method of Baquero, ter Horst, and Verbeek to hedge funds by correcting separately for additional self-selection biases that plague hedge fund databases; underperformers do not wish to make their performance known and funds that performed well have less incentive to report to data vendors to attract potential investors.

The major part of the literature is devoted to assess the look-ahead bias on actively managed investment funds. We study how the backtesting of investment strategies on biased stock price databases is effected. We add to the literature by focusing on the look-ahead bias that appears when the assets used to test portfolio performance are selected on the basis of their relationship with the benchmark to which the performance is compared. In the next section, we provide a specific, straightforward implementation using the S&P 500 Index as the benchmark over the period from January 2001 to December 2006. We then offer a more systematic illustration of the look-ahead benchmark bias over different periods from 1926 to 2007. The substantial difference in performance—of up to 8% between portfolios with and without look-ahead bias—provides an indication of the bias in the performance of the backtest of an active investment strategy, as it is commonly carried out.

Following that, we show how passive strategies perform better after cleaning the database with respect to this look-ahead bias. Under quite general assumptions, we then provide an analytical prediction of the look-ahead bias happening to the mean-variance investment rule that might be applied in a mutual fund. In particular, we discuss under what conditions naive diversification would be favorable. The same methodology can be applied to give decision support to the hedge fund manager regarding whether she should equally allocate money among alternative investment strategies. Finally, we extend the empirical evidence by using random strategies, proposed as a simple and efficient test of the value added by a given strategy, which take into account all possible biases, including those too difficult to address or that are even unknown to the analyst.

AN ILLUSTRATION USING THE S&P 500 AS THE BENCHMARK

Consider a manager who wants to backtest a given trading strategy, namely, on a pool of stocks, such as the constituents of the S&P 500 Index, over a given period, say, January 2001 to December 2006. To do so, the natural approach would be the following:

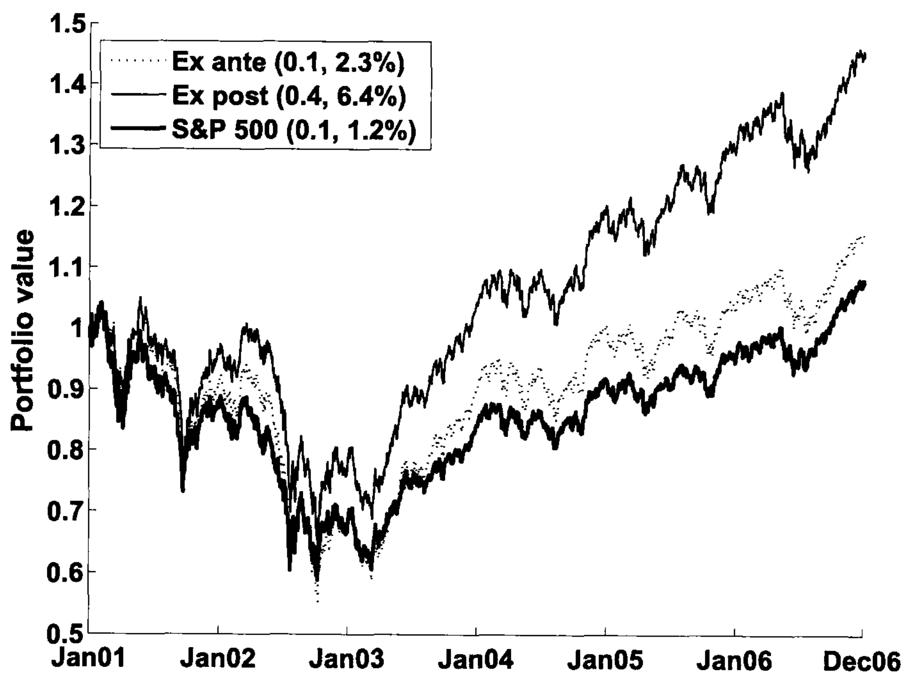
1. Obtain the list of constituents of the S&P 500 Index at the end of December 2006.
2. Retrieve the closing price time series from January 2001 to December 2006 for each stock.
3. Backtest the strategy on that dataset by comparing it with, for instance, the S&P 500 benchmark.

This approach introduces a formidable bias, however, and can easily lead to erroneous conclusions. Exhibit 1 dramatically illustrates the effect by comparing the performance of the two investment portfolios.

The ex post portfolio, which is subject to look-ahead bias, is $1 of an equally weighted portfolio invested in the 500 stocks constituting the S&P 500 Index at the end of the period (December 29, 2006). We hold it from January 1, 2001, until December 29, 2006. The ex ante portfolio consists of $1 of an equally weighted portfolio invested in the 500 stocks constituting the S&P 500 Index at the beginning of the period (January 1, 2001). We also hold it from January 1, 2001, until December 29, 2006. Both investments are buy-and-hold strategies and should have similar performance if the constituents of the S&P 500 Index remain unchanged over the period.[3] The list of constituents of the S&P 500 Index is updated, however, usually on a monthly basis, to account for changes in eligibility criteria due to market moves with one of the driving criteria being market capitalization.[4] Consequently,

EXHIBIT 1

Evolution of $1 Invested in Equally Weighted Ex Post and Ex Ante Buy-and-Hold Portfolios, January 2001–December 2006



Note: The two equally weighted buy-and-hold portfolios are composed of the 500 constituents of the S&P 500 Index as of January 1, 2001 (ex ante portfolio), and December 29, 2006 (ex post portfolio). For reference, we also plot the historical value of the actual S&P 500 Index, normalized to one on January 1, 2001. The performance of the three portfolios is reported in the upper left panel of the exhibit with their annualized Sharpe ratios (using a zero risk-free interest rate) and their continuously compounded average annual returns.

the list of constituents cannot contain, almost by definition, a stock that crashed in the recent past. In such a case, the stock would likely be passed in terms of capitalization by another stock in the same industry, leaving the index to be replaced by the other stock. The only difference in the two portfolios is that the ex post portfolio uses look-ahead information; that is, it knows on January 1, 2001, what the list of constituents of the S&P 500 Index will be at the end of the period (December 29, 2006). This apparently innocuous look-ahead bias leads to a huge difference in performance, as shown in Exhibit 1. The ex post and ex ante portfolios have an annual average compounded return of 6.4% and 2.3%, respectively, and a Sharpe ratio (not adjusted for the risk-free rate) of 0.4 and 0.1, respectively. The ex post portfolio has a significantly better return but, even more important, it exhibits a larger risk-adjusted return.

For reference, we also plot the historical value of the actual S&P 500 Index, normalized to one on January 1, 2001. The index's performance is slightly worse than that of the ex ante portfolio and could be due to the different weighting and/or to the effect reported by Cai and Houge [2007].[5]

Many managers would have been happier to report Sharpe ratios in the range obtained for the ex post portfolio, especially over this turbulent time period. Investment strategies exhibiting this kind of performance would fuel interpretations that this is evidence of a departure from the efficient market hypothesis and/or of the existence of arbitrage opportunities. Other pundits would observe that this look-ahead bias is so obvious that no one would fall into such a trivial trap. Such a reasonable assessment collides with one simple, but often overlooked, operational limitation of backtests, which is that the changes in the constitution of financial indices are not recorded in most standard professional databases, such as Bloomberg, Reuters, Datastream, or Yahoo! Finance.[6] As the standard goal for investment managers is to at least emulate or beat a reference index, backtests on comparative investments should use a set of assets defined at the beginning of the period. But because the list of index constituents is very challenging to retrieve, it is common practice to use the set of assets constituting the reference benchmark at the present time, rather than at the beginning of the period.[7] Thus, the kind of look-ahead bias that we report here will automatically pollute the conclusions, with sometimes dramatic consequences, as illustrated in Exhibit 1. We refer to this as look-ahead benchmark bias.

A part of the overperformance of the ex post portfolio versus the S&P 500 Index can be attributed to the fact that the former is equally weighted while the latter is value weighted. This does not explain, however, the look-ahead effect as shown by the large difference between the equally weighted ex post and ex ante portfolios. For instance, consider the Dow Jones Industrial Average (DJIA). While the annual mean return of the price-weighted DJIA index from January 2001 to September 2007 was slightly below that of the price-weighted ex post portfolio at 3.2% and 3.8%, respectively, the difference was much larger for the period from February 1973 to September 2007—5.7% for the price-weighted DJIA versus 7.8% for the price-weighted ex post portfolio.

THE EXTENT OF LOOK-AHEAD BENCHMARK BIAS

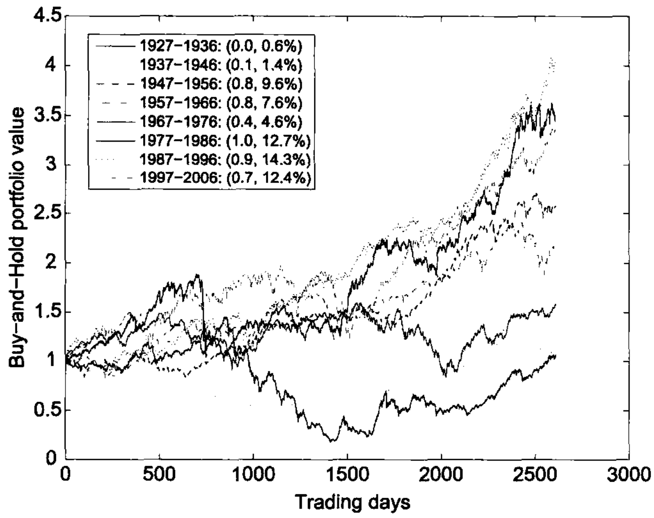
We use CRSP data from which we extract the daily close price, daily split factor, and number of outstanding shares on a monthly basis for all U.S. stocks from January 1927 to December 2006. This represents a total of 26,892 stocks.

We decompose the time interval from January 1927 to December 2006 into eight periods of 10 years each. For each period, we monitor the value of two portfolios. At the beginning of each 10-year period, the ex post (ex ante) portfolio invests $1 equally weighted in the 500 largest stock capitalizations as determined at the end (start) of the 10-year period.[8] The ex post portfolio has, by definition, look-ahead benchmark bias, while the ex ante portfolio is exempt from the bias and could have been implemented in real time. Exhibit 2 plots the evolution of the value of the two portfolios. The insets show that the Sharpe ratio and continuously compounded average annual returns are much larger for the ex post compared to the ex ante portfolio.

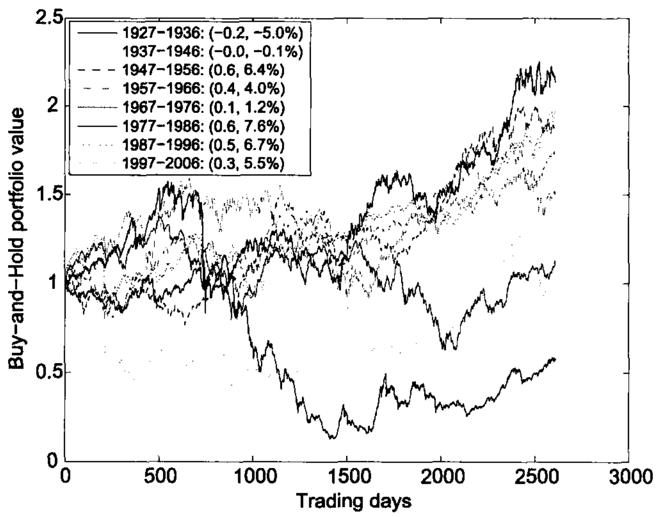
Exhibit 3 shows the evolution of the value of an investment that is long $1 in the ex post portfolio and short $1 in the ex ante portfolio. In other words, it shows the ratio of the value of the ex post portfolio divided by the value of the ex ante portfolio for each of the eight periods. This hedged long–short portfolio can only be implemented ex post when backtesting, and not in real time. Its performance is consistently good over the eight periods from 1926 to 2006,[9] with less risk and better return than the unbiased ex ante portfolio, thus demonstrating the signif-

EXHIBIT 2

Evolution of $1 Invested in Ex Post and Ex Ante Portfolios over Successive 10-Year Periods, 1927–2006



Panel A: Biased Ex Post Pickup



Panel B: Unbiased Ex Ante Pickup

Note: For each 10-year epoch, we plot the evolution of the ex post portfolio (Panel A) that invests $1 equally weighted in the 500 largest stock capitalizations as determined at the end of the 10-year period and the evolution of the ex ante portfolio (Panel B) that invests $1 equally weighted in the 500 largest stock capitalizations as determined at the start of the 10-year period. Note the different ranges of the vertical scales in the two panels. The inserts give the Sharpe ratio (with zero risk-free rate) and the continuously compounded average annual return for each 10-year period. The discrepancy between the two figures helps visualize the extent of the survival bias for the 500 largest capitalizations throughout time. The total return for each period is the second figure in each inset and allows us to sort the portfolios from the best performing ex post (14.3% for 1987–1996) to the worst performing ex post (0.6% for 1927–1936).

icance of look-ahead bias. The result is robust with respect to the number of stocks selected in the two portfolios.

Exhibit 4 tests for different means in the ex ante and ex post portfolios. In the first two decades, the means cannot be distinguished, but in recent decades the means differ significantly, confirming that the ex post portfolios have statistically significant higher returns than their ex ante counterparts.

ESTIMATION OF THE BIAS IN ESTIMATIONS OF LOOK-AHEAD AND SURVIVAL BIASES

The true value of the survivorship bias can be determined analytically for active investment strategies. Investigating Markowitz optimal portfolios, we find that the random nature of optimal portfolio weights increases the survivorship bias we have seen before for fixed weights. The level of the entries in the covariance matrix of asset returns has an impact on the amount of the bias. This is relevant because a database with survival bias has a covariance matrix with smaller covariance terms, which tends to enhance the difference between the true and the biased dataset. These calculations can also be used for the survival bias, as we will discuss at the end of the section.

Next, we will calculate the estimation error of the Sharpe ratio based on data with bias versus the Sharpe ratio based on data without bias. The derived expression depends on the true expected returns and covariance matrices of both datasets. Consider an economy characterized by a vector process of N asset excess returns $\{R_t, t = 1, \ldots, T\}$ that are normally distributed. Let μ be the vector of mean excess return, Σ the covariance matrix of the excess returns, and ω the vector of portfolio weights.

The Markowitz optimization program consists of finding ω, maximizing the following risk-adjusted excess return:[10]

$$U(\omega) = \omega'\mu - \frac{\gamma}{2}\omega'\Sigma\omega \tag{1}$$

where γ is the risk aversion coefficient. The optimal weights are

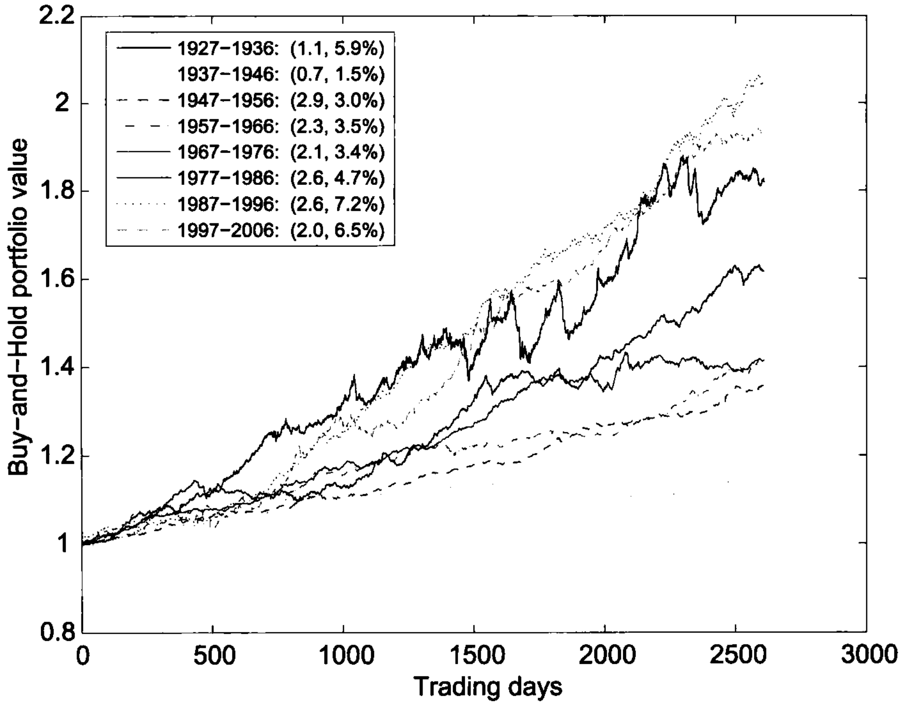
$$\omega^* = \frac{1}{\gamma}\Sigma^{-1}\mu \tag{2}$$

with

$$U(\omega^*) = \frac{1}{2\gamma}\omega'\Sigma^{-1}\omega = \frac{1}{2\gamma}S_*^2 \tag{3}$$

Exhibit 3

Evolution of Investment Composed of $1 Long Ex Post Portfolio and $1 Short Ex Ante Portfolio



Note: We plot the evolution of the value of an investment long $1 in a portfolio equally weighted on the 500 largest stock capitalizations at the end of each period and short $1 in a portfolio equally weighted on the 500 largest stock capitalizations at the beginning of each period, with compounded returns. The inset shows the Sharpe ratios (with zero risk-free interest) and compounded annual return for the eight periods.

Exhibit 4

Two-Sample *t*-Tests of Mean Returns for Ex Post and Ex Ante Portfolios

Subsample	Test statistic	Significance level
1927–1936	0.7201	0.4807
1937–1946	0.3708	0.7151
1947–1956	1.4125	0.1749
1957–1966	1.9151	0.0715
1967–1976	3.2103	0.0049
1977–1986	2.3461	0.0306
1987–1996	2.6059	0.0179
1997–2006	1.9879	0.0622

Note: Means have unknown, but common, standard deviations and the means are hypothesized to be identical. The test uses real one-year interest rates.

where S_* is the Sharpe ratio given by

$$S_*^2 = \omega' \Sigma^{-1} \omega \tag{4}$$

Sample estimations of the mean excess return, covariance matrix, and optimal Markowitz weights are

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t, \quad \hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})',$$
$$\hat{\omega} = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu} \tag{5}$$

Then, the sample excess returns $\hat{\mu}$ are distributed according to a multivariate normal distribution,

$$\hat{\mu} \sim N\left(\mu, \frac{\Sigma}{T}\right) \tag{6}$$

and the sample covariance matrix is distributed according to

$$T\hat{\Sigma} \sim W_N(\Sigma, T-1) \tag{7}$$

where W_N is the Wishart distribution.

Let index 1 refer to data with look-ahead bias and index 2 to data without the bias. The true bias is $U(\omega_1^*) - U(\omega_2^*)$, when one only has to assess the estimated bias $U(\hat{\omega}_1) - U(\hat{\omega}_2)$. To assess how much the bias can be under- or overestimated, the relevant measure is

$$\Delta = [U(\omega_1^*) - U(\omega_2^*)] - [U(\hat{\omega}_1) - U(\hat{\omega}_2)] \tag{8}$$

which can be expressed explicitly; similar calculations can be found in Kan and Zhou [2007].

$$\Delta = \frac{1}{\gamma}(1-k)[S_{*,1}^2 - S_{*,2}^2] = \frac{1}{\gamma}(1-k) \times (\omega_1' \Sigma_1^{-1} \omega_1 - \omega_2' \Sigma_2^{-1} \omega_2] \tag{9}$$

where

$$k = \frac{T}{T-N-2}\left(2 - \frac{T(T-2)}{(T-N-1)(T-N-4)}\right) < 1 \tag{10}$$

Suppose that the biased data is such that $1'\mu_1 > 1'\mu_2$ and $1_N' \Sigma_1 1_N < 1_N' \Sigma_2 1_N$, which is usually the case, as just shown (large returns and smaller risks for the look-ahead-biased data).

Then, $\Delta > 0$ (i.e., the true bias is larger than estimated from the data). If the sample size T is not too large compared with the number N of assets, then the effect of the bias on the Sharpe ratio is generally underestimated. This underestimation is also found for the equally weighted portfolio, but its magnitude is larger for the Markowitz rule.

Other effects occur. For instance, consider portfolios of less than 100 assets and having the same expected returns and variances, and with zero covariances. Then, the expected Sharpe ratio is slightly better for naïve diversification compared to the sample-based Markowitz optimal portfolio. Now, assume the expected returns and their covariance matrix were based on data with survival bias, where the mean and variances are higher than with bias-free data. Then, the expected measure of the Sharpe ratio is higher for the sample-based Markowitz-optimal portfolio compared to the naïve diversification—that is, the order flips. In the literature on survival bias, the methodology is often to compare the performance measures estimated by sample versions of expected returns and their covariance matrix based on both a clean and a biased database. As our calculations have shown (and which also straightforwardly apply to survival bias), the assessment of the impact of survivorship bias on the performance figures is itself biased. Under mild assumptions, when reading the literature one could conclude that the bias is worse than thought.[11]

CONSTRAINED RANDOM PORTFOLIOS AND PROPOSED TESTING METHODOLOGY

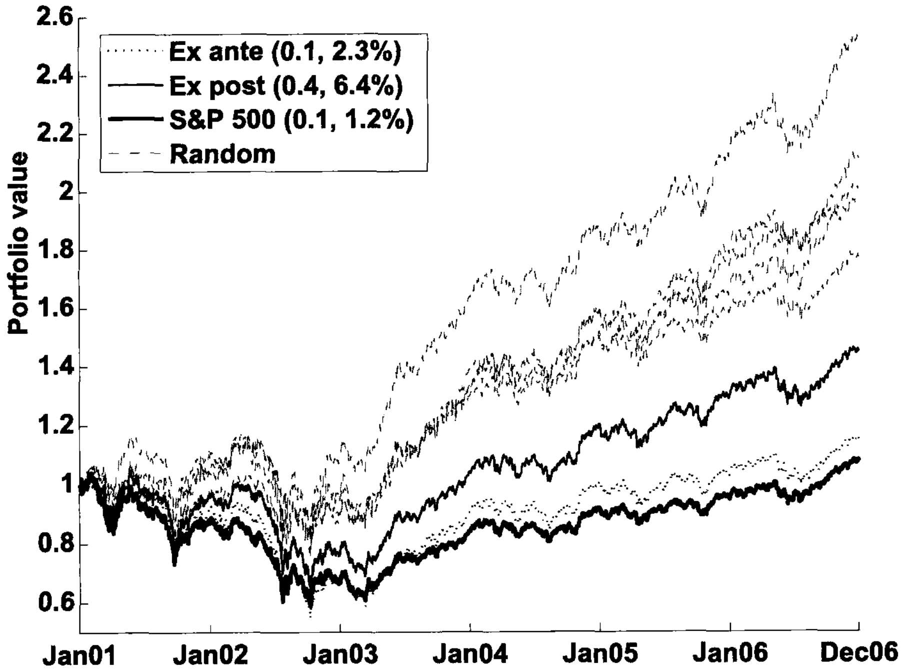
Reverting to the period from January 1, 2001, to December 29, 2006, shown in Exhibit 1, we can further test the amplitude and robustness of the impact of look-ahead benchmark bias. Investing in the 500 constituents of the S&P 500 Index at the end of the testing period (December 29, 2006) amounts to biasing the stock selection toward good performers. We illustrate that, as a consequence, non-informative random strategies exhibit very good to extremely good performance. We generate 10 portfolios, each implementing a random strategy. A random strategy opens only long positions (we buy first and sell later) on a subset of the 500 stocks, with an average leverage of 0.8 and an average duration per deal of nine days, both of which are common values. Given these constraints, the choices of stock and of the timing are random at each time step. We do not further specify the algorithm because all possible specific implementations give similar results.

Exhibit 5 plots the evolution of the value of \$1 invested on January 1, 2001, in each of the 10 random portfolios. These random portfolios provide an average compounded annual return of 9.1% ± 4% with an excellent Sharpe ratio (with zero risk-free interest) of 2 ± 0.8. The random portfolios strongly outperform the S&P 500 Index and diffuse around the look-ahead index with an upward asymmetry. Similar results are obtained with other parameters of the random strategies.

In practice, it may not be possible to completely exclude the presence of look-ahead bias. Therefore, we suggest using constrained random portfolios from the same database as the benchmark to test the value of proposed investment strategies and to assess the probability that the performance of a given strategy can be attributed to chance. Because the same look-ahead bias will impact both the random portfolios and the proposed strategies, it should be possible to detect the presence of

EXHIBIT 5

Evolution of $1 Invested in 10 Random Portfolios Comprising Stocks of a Biased Database, January 2001–December 2006



anomalously large gains that could result from a large amplitude in the look-ahead bias, and to quantify the real value, if any, of the proposed strategy over the random portfolios. This added value can be a useful metric of the performance of the proposed strategy. In order for this methodology to work, the constrained random portfolios should imitate as closely as possible the properties of the trading strategy about to be tested, such as its leverage, mean invested time, and turnover.[12]

CONCLUSIONS

We have reported a surprisingly large look-ahead benchmark bias that results from information on the future ranking of stocks in a benchmark index at the end of the testing period. We have argued that this look-ahead benchmark bias is present due to the need for strategies or investments to prove themselves against benchmark indices and because changes in composition of benchmark indices over the testing period are typically unavailable. It is difficult, if not impossible, to completely exclude any look-ahead bias in simulations of the performance of investment strategies. One way to address these biases is, first, to recognize their existence, and then to model how survival probabilities depend on historical returns, fund age, and aggregate economy-wide shocks.

But one can never be 100% certain that all biases have been removed. In certain scientific fields concerned with forecasting, such as in earthquake prediction, the community has recently evolved to recognize that only real-time procedures can avoid such biases and test the validity of models.[13] Actually, real-time testing is a standard of the financial industry, as cautious investors only invest in funds that have a proven track record established over several years. As shown in the academic literature however, success does not equate to skill and may not be predictive of future performance because luck and survival bias are both prevalent forces in the industry.[14] A large and growing literature on how to test for data snooping and fund performance is available.[15] The problem is more generally related to the larger issue of validating models.[16]

Practitioners should be careful to test for the presence of look-ahead bias in their dataset prior to backtesting their trading strategy. How should investors adjust the returns of a backtested portfolio and discount the results to account for the look-ahead problem? What practical steps could the money management industry take to minimize the problem in future? To address these questions, we suggest a simple and practical approach: compete with random strategies that retain the identical statistical signature of the original strategy, without the intelligence. Random portfolios are a universal method to assess any optimized portfolio.[17] In the case of the Markowitz optimization approach, which is widely used in the industry, these questions are answered by our theoretical estimation of the bias amplitude, which provides a rule-of-thumb to correct for the underestimation of the reported bias. In the absence of clean bias-free datasets, the bias should be fully disclosed. Independent advisors could use our methods to report approximate correction factors for the main investment approaches, as well as for passive investments. In addition, we foresee the development of diagnostic metrics of the presence of look-ahead benchmark bias, associated with the relative performance of different industry sectors.

ENDNOTES

We are grateful to Patrick Burns, Riley Crane, Yannick Malevergne, and Jason Zweig for useful feedback.

[1]Lo and McKinlay [1990], White [2000], and Sullivan, Timmermann, and White [1999].

[2]See, for example, Brown et al. [1992] and Grinblatt and Titman [1992].

[3]Because the S&P 500 Index is not equally weighted, we should expect a slight discrepancy between the evolution of the ex ante portfolio and the index.

[4]Only firms with market cap in excess of US$ 4 billion can be included in the S&P 500 Index. This criterion is monitored and may lead to an index exclusion if it is not subsequently met. No quantitative rule provides how the index members are selected based on the criterion. S&P states that "the index is a gauge of the U.S. equities market, including 500 leading companies in leading industries of the U.S. economy."

[5]The look-ahead benchmark bias documented here is related to the work of Cai and Houge [2007] who studied how additions and deletions affect benchmark performance. Studying changes to the small-cap Russell 2000 Index from 1979 to 2004, Cai and Houge [2007] found that a buy-and-hold portfolio significantly outperforms the annually rebalanced index by an average of 2.2% over one year and by 17.3% over five years. These excess returns result from strong positive momentum of index deletions and poor long-run returns of new issue additions.

[6]Because the benchmark is observed continuously, real-time assessment of performance does not suffer from this problem. We only refer to backtesting which uses a recorded time series of the benchmark and present knowledge of its constituents.

[7]Standard & Poor's provides the list of constituents of the S&P 500 Index only from January 2000. Reuters, Bloomberg, and Datastream provide only incomplete data. In fact, it appears that both the CRSP and Compustat databases are necessary to retrieve the list of constituents of the S&P 500 Index at any given point in time, and these databases are usually not accessible to practitioners.

[8]Between 1926 and 1954, the S&P Composite—not the 500—was the standard benchmark and had a smaller number of stocks. For our analysis, it does not matter whether an index is adopted as the benchmark by the majority of the market participants. In line with our methodology, we consider any index a benchmark.

[9]The 1937–1946 period exhibits the smallest gain of 1.5% with a significant reduction in risk having a Sharpe ratio of 0.7.

[10]In this contribution we interpret the Markowitz problem to find an efficient portfolio with respect to a specific level of risk aversion described by quadratic utility.

[11]See, for example, Brown et al. [1992], Brown, Goetzmann, and Ross [1995], Brown et al. [1997], Brown, Goetzmann, and Ibbotson [1999], Carhart et al. [2002], Carpenter and Lynch [1999], and Elton, Gruber, and Blake [1996].

[12]See, for example, Burns [2006].

[13]Jordan [2006], Schorlemmer et al. [2007].

[14]Barras et al. [2007].

[15]Lo and McKinlay [1990], Romano and Wolf [2005], and Wolf [2006].

[16]Sornette et al. [2007] and references therein.

[17]Burns [2006].

REFERENCES

Baquero, G., J. ter Horst, and M. Verbeek. "Survival, Look-Ahead Bias, and the Persistence in Hedge Fund Performance." *Journal of Financial and Quantitative Analysis*, Vol. 40, No. 3 (2005), pp. 493–517.

Barras, L., O. Scaillet, and R. Wermers. "False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas." Working Paper, March 2007.

Bartelsman, E., S. Scarpetta, and F. Schivardi. "Comparative Analysis of Firm Demographics and Survival: Micro-Level Evidence for the OECD Countries." Working Paper, OECD Economics Department Working Paper No. 348, 2003.

Brown, S.J., W.N. Goetzmann, and R.G. Ibbotson. "Offshore Hedge Funds: Survival and Performance, 1989–95." *Journal of Business*, 72 (1999), pp. 91–117.

Brown, S.J., W.N. Goetzmann, R.G. Ibbotson, and S.A. Ross. "Survivorship Bias in Performance Studies." *Review of Financial Studies*, Vol. 5, No. 4 (1992), pp. 553–580.

——. Rejoinder: "The J-Shape of Performance Persistence Given Survivorship Bias." *Review of Economic Statistics*, 79 (1997), pp. 167–170.

Brown, S.J., W.N. Goetzmann, and S.A. Ross. "Survival." *Journal of Finance*, 50 (1995), pp. 853–873.

Burns, P. "Random Portfolios for Evaluating Trading Strategies." Working Paper, 2006. Available at www.ssrn.com/abstract=881735.

Cai, J., and T. Houge. "Index Rebalancing and Long-Term Portfolio Performance." Working Paper, 2007. Available at www.ssrn.com/abstract=970839.

Carhart, M.M., J.N. Carpenter, A.W. Lynch, and D.K. Musto. "Mutual Fund Survivorship." *Review of Financial Studies*, 15 (2002), pp. 1439–1463.

Carpenter, J.N., and A.W. Lynch. "Survivorship Bias and Attrition Effects in Measures of Performance Persistence." *Journal of Financial Economics*, Vol. 54, No. 3 (1999), pp. 337–374.

Elton, E., M. Gruber, and C. Blake. "Survivorship Bias and Mutual Fund Performance." *Review of Financial Studies*, 9 (1996), pp. 1097–1120.

Grinblatt, M., and S. Titman. "The Persistence of Mutual Fund Performance." *Journal of Finance*, 47 (1992), pp. 1977–1984.

Jordan, T.H. "Earthquake Predictability: Brick by Brick." *Seismological Research Letters*, Vol. 77, No. 1 (2006), pp. 3–6.

Kan, R., and G. Zhou. "Optimal Portfolio Choice with Parameter Uncertainty." *Journal of Financial and Quantitative Analysis* 42 (2007), pp. 621-656.

Knaup, A.E. "Survival and Longevity in the Business Employment Dynamics Data." *Monthly Labor Review*, May 2005, pp. 50–56.

Lo, A.W., and A.C. McKinlay. "Data-Snooping Biases in Tests of Financial Asset Pricing Models." *Review of Financial Studies*, Vol. 3, No. 3 (1990), pp. 431–467.

Romano, J.P., and M. Wolf. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica*, Vol. 73, No. 4 (2005), pp. 1237–1282.

Schorlemmer, D., M. Gerstenberger, S. Wiemer, DD. Jackson, and D.A. Rhoades. "Earthquake Likelihood Model Testing." *Seismological Research Letters: Special Issue on Working Group on Regional Earthquake Likelihood Models (RELM)*, Vol. 78, No. 1 (2007), p. 17.

Sornette, D., A.B. Davis, K. Ide, K.R. Vixie, V. Pisarenko, and J.R. Kamm. "Algorithm for Model Validation: Theory and Applications." *Proceedings National Academy of Sciences USA*, Vol. 78, No. 16 (2007), pp. 6562–6567.

Sullivan, R., A. Timmermann, and H. White. "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap." *Journal of Finance*, 54 (1999), pp. 1647–1691.

ter Horst, J.R., T.E. Nijman, and M. Verbeek. "Eliminating Look-Ahead Bias in Evaluating Persistence in Mutual Fund Performance." *Journal of Empirical Finance*, 8 (2001), pp. 345–373.

ter Horst, J.R., and M. Verbeek. "Fund Liquidation, Self-Selection, and Look-Ahead Bias in the Hedge Fund Industry." *Review of Finance*, Vol. 11, No. 4 (2007), pp. 605–632.

White, H. "A Reality Check for Data Snooping." *Econometrica*, 68 (2000), pp. 1097–1126.

Wolf, M. "The Skill of Separating Skill from Luck." Working Paper, University of Zurich, 2006.

The Journal of Portfolio Management

ASSET–LIABILITY MANAGEMENT

Widening corporate bond spreads have caused a dislocation between corporate-bond-based pension discount rates and the rates of commonly used interest rate hedging tools. As a result, widening spreads have brought the issue of how to manage liability credit risk to the forefront for plan sponsors. Whereas managing liability interest rate risk via interest rate swaps and/or Treasuries is relatively straightforward, managing liability credit risk is more challenging for three reasons: 1) the credit component of liability returns are not investable, 2) no capital-efficient risk management tool exists to hedge liability credit risk, and 3) the connection between credit spreads and the returns of common risky assets (i.e., equities) is relatively reliable, especially during periods of economic stress when the values of risky assets typically fall as credit spreads widen. In order to construct efficient liability-driven solutions and avoid poor funding ratio outcomes, it is thus essential to view liability credit spread risk from a total portfolio perspective inclusive of risky assets. Meder recommends that, from a long-term policy perspective, plan sponsors should generally avoid credit risk in the liability hedge. From a tactical perspective, however, adding credit risk to the liability hedge when credit spreads are wide and expected to narrow can improve funding ratio outcomes, but the amount of credit risk taken must be appropriately scaled to the total portfolio.

The objective of this article is to shed light on the potential benefits of asset-liability management techniques, originally developed for institutional money management, in a private wealth management context. The authors show that much of the complexity of optimal asset allocation decisions for private investors can be captured through the addition of a single state variable—liability value—which accounts in a parsimonious way for investors' specific constraints and objectives. An asset-liability management approach to private wealth management has a direct impact on the selection of asset classes because it requires a consideration of the liability-hedging properties of various asset classes, that would, by definition, be absent from an asset-only perspective. An asset-liability perspective also leads to the use of the liability portfolio as a benchmark, or numeraire, acknowledging that, for private investors, terminal wealth per se is not as important as the investor's ability to achieve goals, such as preparing for retirement or buying property.

PERFORMANCE EVALUATION

Performance of investment managers is predominantly evaluated against targeted benchmarks, such as stock, bond, or commodity indices. But most professional databases do not retain time series for companies that drop from the database and do not necessarily track changes in the benchmarks over time. Consequently, standard tests of portfolio performance suffer from *look-ahead benchmark bias*, meaning that a given strategy is naively backtested against the assets constituting the benchmark of reference at the end of the testing period (i.e., now), rather than at the beginning of the period. In this article, the authors report that look-ahead benchmark bias can be surprisingly large in portfolios of common stocks—up to 8% per annum when the S&P 500 Index is the benchmark. Using CRSP data for the running top 500 U.S. capitalizations over the period 1927–2006, the authors demonstrate that look-ahead benchmark bias can account for a gross overestimation in performance metrics, such as

The Journal of Portfolio Management

the Sharpe ratio, as well as an underestimation of risk measured, for example, by peak-to-valley drawdowns. A general methodology to test investment strategy properties is advanced by the authors in the context of several random strategies having similar investment constraints.

FIXED INCOME

JEFFREY M. MERCER, MARK E. MOORE, AND DREW B. WINTERS

Studies show that riding the Treasury bill yield curve consistently provides higher returns than a matched-horizon buy-and-hold strategy and this article confirms earlier findings. Using Federal Reserve (FRED) interest rate data on 91- and 182-day T-bills and GovPX interdealer tick data over the period January 2001–September 2007, the authors find that no interdealer sales of 182-day T-bills occurred at the time needed to complete a ride, suggesting that no trader benefited through the interdealer market. They also show that selling the seasoned bills at the end of the ride in the new 91-day on-the-run secondary market or its when-issued market would have provided higher returns than the returns computed using the FRED data. But to generate $1 million of annual riding returns would require capturing 85% of the available market volume every week. The authors conclude that riding the T-bill yield curve continues to appear viable across time because of transaction volume limitations.

REAL ESTATE

JACK CLARK FRANCIS AND ROGER G. IBBOTSON

The authors study the annual returns of U.S. real estate over the 31-year period starting in 1978. With aggregate private real estate worth over $30 trillion and representing 60% of U.S. acreage, residential real estate—consisting primarily of single-owner-occupied houses—has by far the largest dollar value of any category of real estate. For the study period, business (commercial) real estate was the best-performing sector with a 9.99% average annual return. In comparison, farm real estate and residential real estate had average annual returns of 8.76% and 5.68%, respectively. All real estate sectors, as well as stock, bond, and commodity markets outperformed the average annual inflation rate of 4.01% over the period. Because physical real estate is not liquid and valuations are often based on appraisals rather than trades, its annual return can be hard to measure precisely and may be difficult to achieve. All physical real estate sectors are correlated with inflation, but equity REITs are more correlated with stock markets. Equity REITs substantially outperformed physical real estate over the sample period, and mortgage REITs and hybrid REITs suffered badly from the subprime mortgage crisis. All categories of real estate performed well during the two periods 1978–1979 and 2000–2005, but after enjoying decades of subsidized returns, residential real estate crashed during the subprime mortgage crisis, opening the door to the eventual fall of the commercial (business) real estate market.