CAMBRIDGE ASSOCIATES LLC

BENCHMARKING:
An Introduction

Cambridge Associates LLC is a Massachusetts limited liability company headquartered in Boston, MA with branch offices in Arlington, VA; Dallas,TX; and Menlo Park, CA. Cambridge Associates Limited is a Massachusetts limited liability company headquartered in Boston, MA and registered in England and Wales (No. FC022523, Branch No. BR005540). Cambridge Associates Limited also is registered to conduct business in Sydney, Australia (ARBD 109 366 654). Cambridge Associates Asia Pte Ltd is a Singapore corporation (Registration No. 200101063G).

C O N T E N T S

Introduction

Indifference to benchmarking leads to sloppy and misleading performance measurement; that is, *misinformation* about whether one's investment decisions have succeeded or failed, which must inevitably degrade the quality of subsequent decision-making. At the end of the day, benchmarking is all about answering the question, "how are we doing?" in ways that are both accurate and relevant to the objectives of the portfolio being measured. A clear and thorough statement of portfolio objectives is therefore a prerequisite of accurate benchmarking—"what are we trying to achieve?" must precede "how are we doing?" Obvious as this seems, investors often measure their performance with yardsticks that are both inaccurate and irrelevant to their stated objectives, leading them to shift course at inopportune times. This is analogous to a physician prescribing the wrong course of treatment as a result of a misdiagnosis, thus risking further damage to the patient's health.

Moreover, the *purpose* for which the question is asked also has a bearing on how it should be answered. For example, when the object of the exercise is to establish yardsticks for computing performance-based compensation, then the benchmarks should be as transparent, objective, and precise as possible. However, for less exacting purposes, like providing feedback on the relative success of this or that investment approach, more ambiguous answers may prove perfectly adequate. In this context, it is important to acknowledge that almost no benchmarks, except perhaps custom-designed "normal" portfolio benchmarks of marketable securities (which are extremely difficult to build and maintain) are free of distortions, defects, and deficiencies. Slicing benchmark data into ever-finer slivers is therefore wasted effort that may lead to a delusory precision worse than the informed ambiguity of more qualitative judgements.

Benchmarking and performance measurement are often taken to be synonymous. They are not. Benchmarking is all about measuring performance *relative to* some relevant yardstick, but one can also just measure performance. For example, private equity investors typically translate internal rates of return (IRR) into time-weighted returns in order to include the results in their total portfolio performance measurement; however, time-weighted returns should not be used for benchmarking private equity portfolios.

Similarly, it may be important for university A to know how its endowment fund has performed compared to those of institutions B, C, D, E, F, and G, with which it competes for students, faculty, and research grants. A large endowment confers a competitive advantage, which may widen or narrow on the basis of relative endowment fund performance over time—hence the importance of knowing how one is doing compared to the competition. Valuable and instructive as such comparative analysis may be, however, it should not be set up as a benchmark, since it lacks virtually all the characteristics a benchmark should possess.

Defining Objectives

Primary Objectives

Most funds have a clear hierarchy of objectives. For a defined benefit pension fund, for example, the primary objective may be to maintain sufficient assets to meet or exceed the fund's liabilities (i.e., funded status), while for an endowment fund it should generally be to maximize sustainable distributions to the operating budget. The purpose of asset allocation is to construct a portfolio that maximizes the probability of realizing this primary objective, over relevant time horizons, within tolerable shortfall risk parameters. Funds are liable to veer off course when secondary objectives become primary—either *de jure* or *de facto* (e.g., minimizing the company's pension contributions or outperforming peer institutions).

Subordinate Objectives

Having constructed a "policy portfolio" designed to realize what is essentially a *financial* objective, the fund's managers can then concentrate on *implementing* these allocations as effectively as possible. At funds large enough to employ professional investment staff, the performance of the actual fund relative to that of the policy portfolio is often the yardstick against which the staff is measured (and their incentive compensation assessed), but this measure is equally relevant to smaller funds managed by investment committees. All subordinate objectives—for asset classes, sub-asset classes, or investment managers—are similarly defined in terms of performance relative to some component or sub-component of the policy portfolio, with benchmark indices used as proxies for the underlying investment.

Measuring Results: Time Horizons, Risk and Attribution Analysis

Defining the objective is a necessary but insufficient precondition to determining "how are we doing?" Before that question can be answered, investors must also define the time horizon over which results should be measured and the amount of risk that can be incurred in pursuit of the objective. In fact, the three most common mistakes investors make in measuring performance are the following: prescribing the wrong benchmark; underestimating the relevant time horizon; and mis-specifying the level of risk.

Time Horizons

Generally speaking, the achievement of primary objectives should be measured over far longer time horizons than are appropriate for subordinate objectives. However, different kinds of funds may have very different time horizons. In the two cases above, for example, the pension fund might want to measure success in meeting its primary objective over a much shorter time horizon than that of the endowment fund, because of a pressing need to maintain its funded status even over relatively short periods, while the endowment fund should be thinking in terms of the next ten, 20, and 50 years in

assessing whether its current mix of spending and asset allocation policy will enable it to maintain real spending distributions indefinitely.

Ideally, the time horizon for measuring the performance of the actual portfolio relative to policy should also be relatively long (i.e., ten years or more), so that capable fund managers have scope to add value by investing in assets whose superior returns may not be realized for many years (i.e., out-of-favor asset classes).[1] In practice, however, few fund managers and the committees to which they report can tolerate such underperformance longer than five years, at most. On the one hand, this is unfortunate, since longer would be better; on the other hand, those with fiduciary responsibility for institutional assets should never presume that just because a particular investment approach worked in the past it must therefore work in the future, and so they should not place too many large tactical bets whose success can be measured only over ten or more years. And the deficiency of measuring performance over too short a time horizon can be mitigated by attribution analysis, as discussed below.

In allocating assets to so-called alternative investments, investors should determine whether these allocations are permanent or opportunistic. If permanent, then the asset class should be included in the policy portfolio, weighted and benchmarked appropriately. If opportunistic, then performance should be measured as for any other tactical allocation: that is, define which permanent asset class the money has been diverted from, and the time horizon over which the efficacy of this re-allocation decision will be measured. When the time horizon has been underestimated, disappointing performance over a relatively short period often leads committees to pull the plug on programs likely to produce good results only over the long term. For example, it would not surprise us if many investors whose first foray into venture capital was in 1999-2000 decide after seven or eight years that they were mistaken in thinking this should be a permanent addition to their portfolio, because performance has been so disappointing. If they had studied the asset class more carefully before investing, however, they would perhaps have recognized that its returns are highly variable and abnormally distributed—meaning that one should expect long periods (longer than five years) of lackluster performance, long periods of modest performance, and occasional starbursts of spectacular performance on which the long-term record largely depends.

Risk

The appropriate time horizon for measuring the performance of individual investment managers also depends on the variability of their returns—but the variability not of their absolute returns but of their returns relative to those of the benchmark index against which they are measured. Thus, for example, an equity index manager that underperforms (or outperforms) the benchmark index by 80 basis points (bps)

[1] We have long argued that one way equity investors can outperform "the market," even if it is informationally efficient, is to adopt a different time horizon than that of most investors, and use this to their advantage (as Warren Buffett has done). The view that "value" investing wins over the long term has many adherents; however, few investors have the discipline to abide patiently in the wilderness, watching others whiz by them, during those periods when value is out of favor, and so fail to realize the benefits of value investing—if price (i.e., paying too much) is the enemy of growth stock investors, *time* is the enemy of value investors.

in one year should probably be fired, while an active manager expected to have far greater tracking error[2] should not. Before hiring any manager, an investor should have a clear idea of the conditions under which the manager is likely to perform better or worse than an appropriate benchmark index, and by how much for how long. If the benchmark has been carefully selected as a suitable proxy for the underlying investment, then the deviation of the manager's returns relative to that of the benchmark represents a kind of risk—risk of not achieving the returns of that asset class or sector—and so it is incumbent on the investor to understand the likely extent and duration of this "active risk" being incurred. Failure to specify this active risk accurately (or at all) often leads to managers being fired for failing to live up to unrealistic expectations. Managers themselves often exacerbate this problem by promising the moon in the hope of winning the business—and then worrying later about how to deliver. A close study of a prospective manager's historical tracking error may not prove sufficient—especially if the track record is relatively short—but is at least a sensible place to begin.

Attribution Analysis

While it is true that—as the old saying goes—you can't eat risk-adjusted performance, it is also important not to confuse brains with a bull market; that is, to understand the extent to which performance is attributable simply to taking on more or less risk. Not that more risk is necessarily bad and less risk good—in some asset classes careful, opportunistic management of risk exposure is a crucial component of portfolio management—but investors should realize that, for example, a bond manager that outperforms by maintaining high allocations to high-yielding low-quality credits will get slammed when investors become risk averse, or that an equity manager's results may not be attributable to superior sector or stock selection, but simply to taking on more market risk.

In general, attribution analysis is the last step in the performance measurement feedback loop and should help investors minimize the inevitable temptation to drive with their eyes clamped on the rear-view mirror. For example, for investors seeking to understand why so many previously reliable investment managers, applying apparently sound investment principles, were underperforming the S&P 500 so consistently during the period 1996-99, it was important to recognize the increasing concentration of the index in fewer and fewer names, many of which were technology stocks commanding ever-higher price-earnings multiples. Understanding that perhaps the problem was not so much with the managers as with the shifting structure and valuation of the index itself, an investor might hesitate to fire longstanding managers, and to accept a greater degree of tracking error for a longer period of time than had been previously anticipated.

In short, we should not expect too much even from careful benchmarking. Markets veer off in unexpected directions, managers go off fishing in new ponds, and the benchmarks themselves change character. However, careful attribution analysis gives investors insights into the how and why of what

[2] Tracking error is an unfortunate term—given the normal connotations of "error"—used to denote the variability of a manager's value-added relative to an appropriate benchmark.

happened, which should enable them to better assess whether they should remain on their current course, or whether they have made some mistakes that should be corrected.

Benchmarking Public Markets

The Composition of Benchmark Indices

Ideally, a benchmark index should fully and accurately represent the asset class in which the investor intends to invest and should also be itself investable. Thus, the S&P 500, for example, meets the second of these criteria (since numerous S&P 500 index funds are available), but fails the first, because it does not fully and accurately represent the asset class "U.S. equities" since it consists of only 500 stocks, selected by a committee to reflect the composition of the U.S. economy, and restricted to the mid- to large-cap sector of the market.[3] The Wilshire 5000 best meets the criterion of full and accurate representation of U.S. equities, but is somewhat less investable than the S&P 500 since the smallest 50% of names in this comprehensive index of U.S. public market equities are largely uninvestable by institutions.[4] For global ex U.S. equity markets, the recent debate over whether the weight of individual securities in market indices should reflect a company's total capitalization or only that portion of capitalization accessible to investors (i.e., the free float) speaks directly to the issue of investability—and on that basis, the free-float argument appears to have carried the day.

There are important distinctions among the various public market equity and fixed income indices, and investors cannot avoid making active decisions when they select among them. These selections should be *informed* and *deliberate* rather than careless and accidental—because the selection of benchmarks has consequences that should be clearly understood in advance. In particular, investors should be aware of the very substantial differences among competing U.S. equity style and capitalization indices, and of those occasional examples of indices that have become accepted as market standards, despite their unrepresentative characteristics.

Benchmarking Managers

Entire conferences are devoted each year to this topic, and cutting edge opinion seems to have come full circle in the past decade from the advocacy of "normal" portfolios as a standard of precise measurement, to the denigration of "style boxes" as inhibiting managers' flexibility in their stewardship of clients' assets.

[3] From 1990-99, for example, 195 issues were added to the S&P 500 replacing those dropped by the committee or de-listed as a result of mergers and acquisitions. Between the announcement of their addition to the index and the date of their actual inclusion, these new S&P 500 stocks outperformed the index itself by 7.5 percentage points, only about half of which advantage dissipated over the longer term.

[4] "Somewhat less investable" rather than "considerably less investable" since the bottom half of the securities in the Wilshire 5000 account for only 1.98% of the capitalization of the total index and so the fact that they are largely uninvestable is of little consequence.

Ideally, a manager's performance should be measured against that of the pond from which he fishes. The pond should not be so loosely defined as to include large areas into which the manager never paddles, nor so narrowly defined as to exclude significant pools into which he often drops his hook. In other words, the ideal benchmark consists of the manager's opportunity set. This applies whether the manager focuses consistently on one slice of a market (e.g., U.S. small-cap value equities), or roams around selecting stocks from a much larger universe (e.g., global equity markets).

In practice, it is not always easy to define that opportunity set fully and accurately—and prospective clients cannot always rely on managers to cooperate enthusiastically in the effort to do so. However, investors who refuse to consider managers that are more difficult to benchmark fall into the trap of becoming benchmarkers rather than investors. So, for example, U.S. equity managers who occasionally buy some non-U.S. securities, or ADRs, or even bonds should not be overlooked just because this amounts to fishing outside the prescribed pond. Such excursions may disqualify these managers from consideration for "core" U.S. equity mandates, but if they are viewed as "satellite" managers the relevant question is: have forays into other assets classes enabled them to outperform the U.S. equity market over time without additional, uncompensated risk?

The bottom line is this: where a manager has demonstrated strategic investment skill, and there are good reasons to believe this will persist, the client should have sufficient confidence to tolerate wide dispersion of returns relative to an imprecise benchmark. *What this requires, of course, is that investors spend far more time and resources gaining a thorough knowledge and understanding of their managers before they hire them.*

Before hiring an active manager to implement an allocation to an asset class or market sector, investors should have a clear understanding of four key issues:

- Does the opportunity set—the pond from which this manager fishes—correspond to the area in which we want to invest?
- Is there an investable market index that corresponds reasonably closely to that pond? If so, this index should become the benchmark. If not, what is the closest available index, or can the pond be defined by a blend of several indices?
- By how much, over various time periods, is the manager's performance likely to vary from that of the chosen benchmark?
- Since we could implement our allocation to this area by investing in the selected index (or blend of indices), what compelling evidence is there that the manager will outperform over time, net of fees, by a sufficient margin to compensate for the "active" risk we are incurring (i.e., the risk that active management will *detract* rather than add value)?

This last point is critical. Despite recent arguments that benchmarking ties managers into straitjackets that inhibit their ability to add value, nothing whatsoever prevents investors from hiring

managers whose returns vary widely from those of whatever index seems the most appropriate performance yardstick. But a manager that outperforms by 1,600 bps in a given year is probably susceptible to underperforming by 1,600 bps the next year, or year after that. Fine—if the client understands and accepts this and has good reason to believe that over, say, a ten-year period, the end result will be positive value added. However, it is folly to believe that *any* manager that outperforms an appropriate benchmark by a substantial margin over several years is immune from underperforming by a similar margin in a subsequent period. Clients desperately want to believe they can have an asymmetrical distribution of variability—deviation only on the upside—and managers are happy to foster this delusion, but it just ain't so.

Benchmarking Marketable Alternatives Programs

As we stress in our annual report, hedge funds are not an asset class *per se*, but rather a collection of investment products, pursuing many different investment strategies, sharing only similar legal and fee structures. Their returns are often attributable more to manager skill than to overall market direction and therefore do not lend themselves to benchmarking. This is especially the case for managers employing leverage and for those engaged in significant short selling and "arbitrage" strategies. Investors nevertheless require some way to measure the performance both of their hedge fund programs and of individual managers.

The purpose for which performance is being measured has a substantial bearing on which approach should be adopted. For example, when the object of the exercise is to establish yardsticks for computing performance-based compensation, then the benchmark should be as transparent, objective, and precise as possible—difficult to achieve in an area like hedge funds where the various manager "indices" are so badly corrupted by selection and survivor bias. For less exacting purposes, like providing feedback on the relative success of this or that investment strategy, more ambiguous answers may prove perfectly adequate. In the end, however, there is no way around the fact that all benchmarks for marketable alternatives have serious shortcomings—it comes down to figuring the least bad choice for a particular purpose.

As always, "how are we doing?" must be preceded by "what are we trying to do within what time period?" If, for example, the objective of the program is to realize a specific return, with volatility no greater than that of bonds, and little or no correlation of returns to those of the equity and bond markets, then the program's success in attaining all three of these objectives must be measured. Since the strategies employed to achieve such objectives are typically various forms of arbitrage, in which LIBOR or T-bills represent the manager's cost of capital, the "absolute return" is often defined in terms either of a premium over (e.g., +5%) or a multiple of (e.g., two or three times) the LIBOR or T-bill rate.

Of these two, we would advocate LIBOR or T-bill plus a premium, rather than a multiple of LIBOR or T-bill rates. This is because the alpha of a hedge fund strategy is not directly related to the

level of short-term interest rates, and consequently does not necessarily increase with rising interest rates. Consider the case of a diversified U.S. arbitrage manager recently earning a steady 10% annualized return. If T-bills currently yield 5.0%, this return might be construed as 2 times T-bills, or as T-bills + 5%. What happens, however, if T-bill rates double? Should the manager now be expected to earn 20% (i.e., 2 x 10%) or 15% (i.e., 10% + 5%)? The latter seems more logical, while the former could create a perverse incentive for a manager to incur greater risk to meet a relatively arbitrary target rate of return.

Investors should only measure the success of their marketable alternative programs against such benchmarks over relatively long periods of five years or more. This is because many arbitrage strategies will thrive in some periods and wilt in others, depending on economic and market conditions. For example, it was relatively easy for competent merger arbitrage managers to realize returns in the mid-teens in the period 1996-2000 because merger activity was strong, competition modest, and spreads relatively attractive. Today, however, huge amounts of money have poured into this space, and spreads have narrowed. In fact, a key question today is whether there are always some strategies that skillfull diversified arbitrage managers can pursue to earn that absolute return investors expect, or whether the vast amounts of capital pouring into every corner of this universe has itself arbitraged away the bases of return, rendering the conventional benchmarks unrealistic not just in the short run, but for the foreseeable future.

Where the program consists not of arbitrage strategies, but of long/short equity, the benchmark should again be tailored to match the program's purpose and objectives. These typically range from modest to aggressive, depending on whether the objective is to earn a return greater than that of the equity markets, or to mitigate some equity market risk. If, for example, a long/short hedge fund program has net average exposure to U.S. equities of 60%, then a sensible benchmark might be 60% Wilshire 5000 (or other U.S. equity index, if more appropriate) and 40% T-bills + premium. With such a benchmark, however, it is important to know the likely dispersion of net equity exposure around that 60% average so that one can gauge how much the actual returns are likely to vary from those of the benchmark over various periods. The greater the variability—especially over extended periods—the less accurate and useful the benchmark.

Where there is a clear mandate to construct a diversified program of multiple strategies, its results may be measured against those of a "policy portfolio" representing each strategy included in the mandate, weighted accordingly (e.g., 30% diversified arbitrage, 20% merger arbitrage, 10% distressed securities, and 40% long/short hedged equities). Since there is no "market" proxy for each of these strategies, the median manager return for each category should be used. The deficiency of this approach is that many marketable alternative strategies are worth pursuing only when the risk/reward ratio is favorable, and not otherwise, but their inclusion in the policy portfolio suggests (even if it does not enforce) a sense of permanence.

Another possibility is to regard funds-of-funds as an investable universe against which to measure an internally managed hedge fund program. The question here is whether to compile a reasonably broad

universe of funds-of-funds, some of whom may have investment profiles and objectives quite different from that of the program, or to identify just three or four whose characteristics are more closely aligned. With a broad universe, the opportunity set becomes somewhat distorted, but the results of a single fund-of-funds has less impact on the whole; on the other hand, a narrow universe may provide a more focused opportunity set, but outlier results will tend to skew the total. A reasonable compromise would be to select eight to ten funds-of-funds, all of which might qualify as viable investment alternatives to an internal program, and to measure results against those of this opportunity set.

Benchmarking Marketable Alternatives Managers

Since there are no "market" indices for hedge funds, the relative performance of individual managers must be measured against the median return for a relevant peer group. Even in narrowly defined categories (e.g., merger arbitrage), however, the number of holdings and the use of leverage and hedging techniques can vary dramatically, with the result that one ends up comparing apples and oranges. In addition, it can be difficult to assemble suitable peer groups, especially for managers with broad, loosely defined mandates, and we would stress again that all hedge fund manager "indices" are severely distorted by selection and survivor bias.

Other than an acute sense of competition vis-à-vis other funds, hedge fund managers themselves generally think only in terms of an absolute target rate of return on invested capital and are likely to express some impatience at discussions of benchmarking. However, we would regard the typical target returns (8% to 10% for diversified arbitrage managers and 10% to 15% for long/short equity managers) as laudable aspirations rather than as realistic expectations and would encourage investors to develop more modest yardsticks to measure "how are we doing?"

Benchmarking Private Equity

An ideal benchmark provides:

- A viable, investable, alternative to active management that is available at the time investment decisions are made—not just after the fact;
- An accurate, complete, and verifiable representation of the available investment opportunity set;
- Risk and return characteristics that can be accurately measured; and
- Constituents that can be parsed to allow for attribution analysis.

With the possible exception of the last point, none of the benchmarks for private equity investments fully satisfy these criteria. For example, while investors can buy all the securities included in the S&P 500 or FTSE 100 indices, they can't invest in all the managers included in the Cambridge

Associates or *Venture Economics* private equity benchmarks. In addition, no index encompasses all available funds, and the risk and return characteristics of these investments cannot be measured in the same terms as are applied to portfolios of marketable securities.

Although less than ideal, available benchmarks for U.S. venture capital and non-venture private equity nevertheless have sufficient breadth and depth to provide investors with reliable answers to "how are we doing?"—although not to two decimal places. This is becoming increasingly true also for European private investments, but benchmarks for Asian and emerging markets investments are less robust.

Who's Asking and Why?

How one measures something may be significantly affected by who is doing the measuring and for what purpose. For trustees sitting on a pension or endowment fund investment committee, for example, the primary version of "how are we doing?" may be "have we been rewarded for allocating capital to private equity?" For the investment staff, however, the key questions are "how has our program performed?" and "have we added value?" since these may have a direct bearing on their compensation, or at least their performance review. When the purpose of the measurement is to determine performance-based compensation, one obviously wants benchmark data to be as precise, reliable, transparent, verifiable, and consistent as possible. However, when the purpose is to provide feedback on the value added of, say, sector allocation decisions, precision and transparency are less critical.

Have We Been Rewarded for Allocating Capital to Private Equity Investments?

Assuming the investment objective is to earn higher returns that those of public equities, the benchmark should be public equities plus a percentage (e.g., 130% for venture, 120% for non-venture private equity). The more common practice of assigning public equities **plus** a premium (e.g., + 300 basis points for venture) ignores the fact that private market assets are high beta investments, with greater variability of returns than that of public markets. Consequently, one should expect them to **underperform** public markets during particularly weak environments (as venture capital did during the bear market decline of 2000-02), just as one should expect them to outperform when conditions improve.

Note, however, that the public equity baseline need not consist of market indices. After all, if an investor could identify active managers of, say, U.S. public equities capable of returning 130% of the index annually, compounded over many years, this should influence the decision whether or not to allocate capital to U.S. private equity investments. In other words, the yardstick for measuring whether allocations to private equity investments have been worthwhile should be some premium over the investor's target rate of return for public equities, including any assumption of alpha from active management.

Investors should avoid making this comparison over relatively short periods. Since the J-curve effect in private equity investing makes the results of the most recent three years' commitments meaningless, these should be excluded from any calculation. In addition, there may be long periods—as long as a decade—during which U.S. equities as a whole perform reasonably well while small-cap technology stocks, with which U.S. venture investments are most highly correlated, perform relatively poorly. As a result, venture investors may well find themselves underperforming public markets for extended periods, as they did, for example, in the mid-1980s. Consequently, investors who expect their U.S. venture portfolio to outperform their U.S. public equity portfolio in every five-year period have not read their history with sufficient care, are going to be disappointed, and will make poor decisions as a result. In short, one should not draw conclusions about the wisdom of allocating capital to private equity investments on the basis of less than ten years' results, since such investments require commitment, patience, and discipline to stay the course during long barren periods.

How Has Our Program Performed?

This may be computed by comparing the IRR of the total private equity program from its inception (but minus the last three years of commitments), to that of a composite composed of all funds included in the program's mandate (e.g., global or just U.S.?), weighted by the market capitalization of those funds.

This composite is as close a proxy as possible for "the market," and does in fact accurately represent the investment opportunity set. Ideally, investors might also consider excluding funds closed to new investors in which they *couldn't* invest, since these also fall outside the available opportunity set, but this would require customized benchmarking and might be neither practical nor worth the effort.

Attribution Analysis

Having measured performance relative to the available opportunity set, the private equity investor now needs to determine the sources of value added. Without such attribution analysis, there is no way to know which of the many decisions made in the course of constructing the program have added value, and how much.

Attribution analysis is like peeling an onion—one starts with the largest layer and works down to finer and finer levels of detail. In private equity, however, the quality of the data is such that one soon reaches a point of diminishing returns—one *could* keep peeling the onion further, but is unlikely to learn much of value from doing so. Moreover, it's not always possible to measure the *extent* of value added from each decision point; sometimes one can only infer by a process of elimination that such-and-such an activity did in fact add value. As noted above, this is problematic if one needs precise measurements for setting performance bonuses, for example, but may be all one needs for other, less rigorous purposes. The layers of the attribution onion worth peeling are:

- Have our allocation decisions added value?
- How have our portfolios performed?
- Have we added value through sector allocation?
- Have we added value through manager selection?

Have Our Allocation Decisions Added Value?

This may be measured by calculating the IRR of the private equity program, since inception (but minus the most recent three years of commitments), as above, but then comparing the results with those of the benchmark, market-cap weighted as before, but now on the basis of the *investor's* cash flow schedule (both investments and distributions), as opposed to that of the market.

Since this calculation eliminates differences in return attributable to differences in the timing of cash flows (over which the investor may have limited control), it will indicate, in the broadest possible terms, whether or not *allocation* decisions have added value.[5] However, it provides no indication of *which* allocation decisions have added value and how much. So, for example, the investor still has no idea whether manager selection has contributed more or less than such decisions as overweighting private equity and underweighting venture, nor whether overweighting the United States and underweighting Europe has added value, nor whether the concentration in early stage at the expense of late-stage venture has added value, nor whether the focus on health care rather than information technology has added value, and so on.

The next step, therefore, is to look at the performance of each of the program's sub-asset classes (i.e., U.S. venture, U.S. buyout, European buyout, etc.) relative to that of the available opportunity set in each sub-asset class, again adjusted to reflect the program's cash flows. This will indicate whether decisions to over- and under-weight commitments to each sub-asset class have added or detracted value. Although eminently logical, this exercise is not as straightforward as it appears because it ignores the messy realities of private equity investing. Assume, for example, that the CIO of a private equity

[5] Obviously, if one *does* have control over the timing of cash flows and wants to measure value added as a result of market timing decisions, then all the factors outlined above must be neutralized in order to isolate and measure the impact of these timing decisions alone.

investment program at a large foundation decides in December that next year's commitments should favor European buyouts funds relative to U.S. venture capital funds. As the year unfolds, however, she finds opportunities to invest with several U.S. venture managers she's been courting for years, while all the European buyout funds coming to market seem thoroughly second-rate. What should she do? This reflects the facts of life in private equity investing and subsequent benchmarking—one can't always invest where one wants, when one wants, but must still make investment decisions.

How Have Our Portfolios Performed?

This is a straightforward exercise in which the performance of each of the program's sub-asset class returns is compared to those of the relevant benchmark, once again applying the investor's cash flows for each sub-asset class to the benchmark data. This will indicate how the private equity program's U.S. venture investments, for example, have performed relative to the market, but it will not indicate the relative contribution of sector allocation (e.g., early- versus late-stage investments; health care versus information technology) and of manager selection.

For example, let's assume that our private equity program CIO decided in 1996 to overweight the health care sector in her U.S. venture allocations and to underweight information technology funds. Subsequent performance attribution analysis shows that her U.S. venture portfolio has outperformed the pooled mean. Was this because of that sector overweighting or of something else? We have to go to the next layer of the onion to learn that the answer is "something else", because that's where we see that 1996 information technology funds did much better than 1996 health care funds.

Have We Added Value through Sector Allocation?

To measure value added from sector allocation, one must compare the results of the pooled means of the different sectors over the measurement period. This is where attribution becomes increasingly fuzzy, however, since most funds are not sector specific, but are diversified across several sectors, and the data for all but one or two sectors (e.g., U.S. information technology) are too thin to support extensive analysis. In addition, we run again into the problem outlined above: in any given year, investors are presented with a limited range of options as to where they can deploy their capital—managers they like may be in sectors they don't and vice-versa, and trade-offs must be made.

Have We Added Value through Manager Selection?

Each manager's performance should be measured against that of the median manager (quartile rankings are also available) within the relevant sub-asset class, by vintage year. Although sector allocation decisions and manager selection decisions cannot be precisely disaggregated, the effects of each can be observed. For example, if the managers an investor has selected for U.S. venture investments in health care have performed well relative to other managers in that space, but health care funds in general have underperformed technology funds in general, then the investor can infer that over-allocating

to health care was not a good decision, but can't precisely compute the relative contribution of these manager selection and sector allocation decisions to the returns of the total venture portfolio.

Total Value vs IRR

In addition to IRR, manager performance should be measured on the basis of total value created over time. Total value is the ratio of distributions + residual net asset value to paid-in capital. Thus, for example, a fund with a value of 2.5 times paid-in capital in year seven may be regarded as more successful than one with a value of only 1.75 times paid-in capital in year seven. This reflects investors' recognition that a high IRR with a low-value creation is worse than a lower IRR with a higher-value creation over a longer period of time.

Difficulties and Distortions in Private Equity Benchmarking

- J-curve effects.
- Lack of standard accounting practices for write-downs.
- Reporting lag: lack of consistency in how to adjust for the one-quarter lag in data.
- Lack of transparency: investors cannot examine the constituents of the benchmark.
- Quality and consistency of data included in benchmarks.
- Shifting universes: the addition of new funds means that benchmark returns are constantly revised.
- Division of administrative responsibilities: should the investor or the custodian calculate the program's performance? If both, do they agree on methodology?
- Translation of IRR returns into time-weighted returns for inclusion in total portfolio reports.

Benchmarking Real Estate

Real estate is an asset class that includes both public and private market investments. Each has different benchmarks and portfolios invested in both should be benchmarked accordingly. Ideally, investors' time horizon should include a complete market cycle—which means a decade or longer.

As always, investors should articulate their investment objectives before attempting to measure "success." In the case of real estate, the investment rationale may include inflation protection or portfolio diversification. Although these are both difficult to measure quantitatively except over very long time periods, they should not be overlooked. In particular, inflation protection is like insurance in that its effectiveness can only be determined in times of trouble and developed markets have not suffered any serious attacks of unanticipated inflation in recent years. During periods of calm, those using real estate

primarily as an inflation hedge can gauge the cost of the insurance policy by calculating the difference between their real estate investment returns and the returns of whatever asset class(es) this money would otherwise have been invested in (e.g., MSCI World Index).

Public Real Estate

This is the easy part. Readily available market indices appropriately represent target trading universes and the Cambridge Associates REIT manager median is representative of the U.S. manager population. The familiar principles and best practices of benchmarking any marketable securities portfolio should be applied. One important caveat is that while data extend back to 1978, public real estate securities only became widely traded and institutionally owned in the mid-1990s. Prior to that the market was small (less than $10 billion) and largely in retail hands, and so any data from the pre-institutional period should be used with care.

In the United States, the Wilshire Associates Real Estate Securities Index (WARESI) most accurately represents the universe of securities from which managers select. In addition to REITs, it includes important public real estate companies that are legally organized under traditional corporate guidelines rather than as REITs. It excludes very small, thinly traded companies as well as those REITs that are only marginally connected to the real estate business, like communications tower companies and health care REITs. In these respects, the WARESI is an important improvement over NAREIT, which simply includes all listed companies, regardless of size or purpose, which happen to be organized in the legal form of a REIT.

Institutional interest in global property investing, combined with the growing adoption of REIT-like legislation and structures outside the U.S., has led to a proliferation of global property indices. The three most widely used indices are the Global Property Research (GPR) General Index, the FTSE/EPRA/ NAREIT Global Index, and the S&P/Citigroup World Property Index. Each of these index families includes coverage by region, country, and property type, though the depth and range varies. The GPR Global is the broadest index in terms of the number of constituents and total market cap, followed by S&P/Citigroup Index and FTSE/ EPRA. The FTSE/EPRA and S&P/Citigroup indices have similar regional weightings, with roughly 45%-48% North America, 20%-21% Europe, and 31%-35% Asia, while weightings in the GPR General is more evenly distributed, with 35% North America, 33% Europe, and 32% Asia.

Importantly, REIT-like legislation varies from country to country, and investors need to be aware that these indices include both REIT-like structures as well as more traditional listed real-estate firms, including some open- and closed-ended real estate funds. All three providers also offer REIT-only indices that exclude companies which are not structured as REITs. This limits the investable universe as many large markets (notably Germany, and until very recently the U.K.) have yet to formally adopt U.S.-type REIT structures. According to a 2005 ABN-AMRO survey, 75% of global real estate investors and fund

managers use some derivation of the FTSE/EPRA index, while 21% use GPR, and the remaining 4% of respondents reported using some other index.

Public market metrics should be used to analyze performance of public securities managers and strategy decisions within the public sector. These should be combined with private market metrics to evaluate the real estate portfolio as a whole, in proportion to the actual or policy allocations, depending on the question one is trying to answer.

Private Real Estate

All the principles and caveats applied to benchmarking private equity investments should be applied to benchmarking private real estate. Many private real estate funds are arranged in the same structure as private equity funds, and involve distinct J-curves. As with private equity partnerships, these real estate funds are best evaluated over the life of the fund through such measures as the IRR and total value over time, rather than through computation of a time-weighted average annual compound return.

As with private equity, it is particularly difficult to measure real estate returns in such a way as to determine whether the investor was sufficiently compensated for the risks incurred. There is a wide array of risk available in real estate investing, ranging from almost bond-like security to vehicles that rival the most aggressive venture funds, and investors should seek to understand how the risk of their own real estate portfolio compares to that of their benchmark index. Unfortunately, there is as yet no comprehensive market index that includes the mid- to high-risk, private-equity-style real estate funds.[6]

U.S. Real Estate

In the United States, the widely-used NCREIF Index (named for its sponsor, the National Council of Real Estate Investment Fiduciaries) is the most common benchmark for measuring the annual time-weighted returns of unleveraged, core (i.e., diversified, low risk), private real estate portfolios. For such portfolios, NCREIF's one significant shortcoming for benchmarking purposes is that it does not represent an investable universe; instead, it captures roughly $220 billion of privately held U.S. commercial property. However, the NCREIF Index allows for attribution analysis to a reasonable level of detail, according to two important real estate parameters, location and property type, with data back to 1978.

With some exertion and ingenuity, investors can overcome the fact that NCREIF results are unleveraged and time-weighted. For example, portfolio results can be de-leveraged; that is, reviewed without any of the debt-related cash flows for purposes of comparison with NCREIF. However, this may

[6] As noted below, Cambridge Associates has constructed such an index, but is still working to populate it with a sufficient number of funds to improve its representation of the available opportunity set.

do injustice to managers who incorporate leverage in their decision-making process and might have had a different debt/equity ratio under different market conditions. In short, all such adjustments have their shortcomings and are not recommended for regular portfolio review.

NCREIF's most significant limitation is its core orientation. For investors in non-core strategies, such as "value-added" and "opportunity" funds, NCREIF is not a reliable benchmark. If a strategy has higher risk than NCREIF, it ought to outperform NCREIF by a margin sufficient to compensate investors for incurring this added risk. The difficult question is "by how much?" A further complication is that most investors do not attempt the onerous and possibly misleading task of peeling out debt information, but simply conflate property selection risk and financial risk (i.e., leverage), although these are quite different.[7]

In practice, major real estate investors use a variety of methods, none of which is ideal, to benchmark non-core real estate investments. These fall into two categories: NCREIF-based and absolute return. A third possibility is to use the Cambridge Associates Real Estate Index, which is structured like the Cambridge Associates Private Equity Index, and should be similarly applied (see previous section Benchmarking Private Equity Investments). However, although this index has the virtue of specifically addressing non-core strategies, it does not yet represent the universe of institutional real estate managers as thoroughly as the private equity index does for its asset class, and so it may be less reliable for certain vintage years and strategies.

The NCREIF-based approach is probably most useful for determining whether a real estate program has succeeded in outperforming the market and may also be used to gauge whether a manager has met expectations in those cases where it is appropriate to tie these to NCREIF results. Value-added managers are typically expected to generate returns anywhere from 200 to 400 bps over NCREIF and opportunity fund managers, 400 to 600 bps over—but these are more-or-less arbitrary spreads designed, albeit crudely, to reflect the additional risk incurred, including leverage. This method is particularly useful for investors (like pension funds) that have predominantly core investments and report in annual terms. And if a program is also relatively large and mature, the distortion created by time-weighting the returns of the satellite, non-core funds will be mitigated.

While the NCREIF-based method has the benefit of relating results to underlying real estate fundamentals, it can be misleading to the extent that non-core managers may not be taking "directional" risk, but may be attempting to add value through such activities as property repositioning. In addition, non-core funds rely heavily on capital gains, which comprise only about 15% of the total return on

[7] An array of academic papers addresses the topic of private real estate risk premia. However, these tend to have unsatisfying theoretical or impractical results, and are not useful for the regular business of evaluating a portfolio and providing input for real-time decision making or compensation.

NCREIF and are based on appraisals, which are most likely to misrepresent actual prices during those periods when non-core buyers may be finding the best buying opportunities. Another issue is asset match: NCREIF includes only U.S. real estate in the five major property types: hotel, office, industrial, retail, and residential; whereas non-core managers also invest in niche properties, such as self-storage and senior housing, and may make investments outside the United States.

An alternative approach is to adopt an absolute return hurdle for the portfolio (e.g., 5% real return), the structure and level of which should be dictated by the purpose of the allocation to this asset class and the strategies employed. While absolute return neatly circumvents NCREIF's shortcomings, its own deficiency is that it establishes a benchmark return completely divorced from the variability of returns inherent in the real estate cycle, and is therefore most useful only over the long term. However, it is worth noting that most non-core private fund managers specifically target a specific absolute return for their funds and an important measure of how well a fund is performing is to compare actual results against this advertised target. Of course, when a fund has a preferred return or a promoted return, the fee structure itself implies an absolute return hurdle. Managers also tend to use the absolute return level to communicate their perception of the risk inherent in their strategy, but investors are cautioned to make their own risk assessment.

Non-U.S. Property

Investors attempting to benchmark non-U.S. investments run into many of the same issues as in the United States, and may have the additional problem of investing in countries for which there are no private real estate indices. In Europe, the IPD subscription service captures returns for seven countries, but with the exception of the United Kingdom, for which there is 20 years of data, the coverage for most countries covers less than ten years. Unlike NCREIF, IPD includes non-core investments, but returns for core and non-core investments are not broken out and IPD does not provide detail on the relative proportions of core and non-core data. This serves to make benchmarking even more difficult than if the index coverage core properties only, because one cannot assess the level of risk inherent in the index. Like NCREIF, the IPD index provides annual returns, relies on appraisals to determine property values, and is unleveraged.

In Asia, the IPD service captures returns in Japan and Australia and will likely include additional Asian countries in the future. Jones Lang LaSalle's REIS product is also a subscription service and covers the major cities in nine countries. However, these data include only core properties and therefore have the same virtues and deficiencies as the NCREIF data. Approximately ten years of data are available for most markets. As with IPD and NCREIF, REIS index returns are annual, based on property appraisals, and unleveraged.

PERFORMANCE MEASUREMENT HIERARCHY

	Primary Objective[1]	Total Portfolio	Asset Class	Manager
Time Horizon	Long-term (e.g., rolling ten-year periods)[2]	Intermediate-term (e.g., rolling five-year periods)	Shorter-term (e.g., rolling three-year periods)	Shorter-term (e.g., rolling three-year periods)
Risk Definition	Probability of Shortfall[3]	Variability of actual returns vs. those of the policy portfolio	Variability of sub-portfolio returns vs. those of asset class benchmark	Variability of returns relative to those of manager-specific benchmark
Measurement	Real value of fund, net of distributions and inflows[4]	Actual portfolio returns vs. those of the policy portfolio[5]	Actual sub-portfolio returns vs. those of asset-class benchmark	Manager returns vs. those of manager benchmark

[1] For most endowment funds this should be to maximize sustainable spending while maintaining the real value of the endowment (net of new inflows).

[2] It is appropriate to measure success in achieving the primary objective over multiple time horizons, of which ten years would be the shortest. As with financial equilibrium for the institution as a whole, the key here is to recognize the *direction* the fund is headed, rather than the specific place it is now. In other words, the key issue is whether the fund is moving towards or away from the maintenance of its real value, net of all distributions and new inflows.

[3] See our paper *Portfolio Risk Measurement.*

[4] Of course, one of the trustees most important fiduciary responsibilities is to determine what level of spending is compatible with maintenance of the real value of the endowment over the long term. Most generic studies of the question have tended to conclude that spending in excess of 4.5% to 5.0%, net of all endowment management costs, is likely to result in depletion of the real value of an endowment fund over time.

[5] In addition to simply measuring the results of the actual portfolio relative to that of the policy portfolio, investors should isolate and measure separately the effect of asset allocation decisions (i.e., over- or under-weighting asset classes relative to target allocations), of asset class structure decisions (e.g., overweighting smaller-cap stocks in the U.S. equity portfolio), and of active manager value added.