# APPENDIX 26B

# NUTS AND BOLTS OF PERFORMANCE EVALUATION AND BENCHMARKING IN THE PRIVATE REAL ESTATE ASSET CLASS

## APPENDIX OUTLINE



## LEARNING OBJECTIVES

After reading this appendix, you should understand:

- Some basic management principles as applied to real estate investment management in the private real estate asset class.
- Some important technical considerations in attempting to quantitatively evaluate and compare investment managers.
- Some basic considerations regarding how to treat investment risk in the investment management evaluation process.

This appendix will take the interested reader to some additional depth in the subject of evaluating and benchmarking investment management in the private real estate asset class. We begin in the first section with some broader and basic management principles. Subsequent sections will go into more depth relevant to quantitative evaluation and comparison techniques. Ultimately we will return to a broader consideration of the role and usage of formal investment management evaluation.

## 26B.1 Matching Evaluation, Responsibility, and Authority

In selecting (or constructing) a benchmark for use in manager performance evaluation, it is important to adhere to the basic management principle of equating responsibility with authority. In general, an investment manager should not be held responsible for, or evaluated on the basis of, factors that are beyond his decision authority or control, especially if the client has explicitly denied the manager discretion over such factors.

This issue commonly arises in the area of segment allocation. Asset market segments in private real estate typically correspond to space market segments characterized by property type and geographic location. Due to the need for local expertise, it is not uncommon for real estate investment managers to be specialized within one or a few market segments. For example, a manager may be hired by a client for the specific purpose of placing capital into the apartment property segment, perhaps within a proscribed geographic region. It would generally be unfair or misleading to benchmark such a manager's performance to an index

that included other types of property besides apartments, and/or other geographic regions beyond the manager's purview.

On the other hand, if a manager is given discretion to choose allocational weights across more than one property market segment, then his benchmark should include all of the segments in the manager's potential choice set. In this case, the question arises as to what segment weightings to use in the manager's benchmark. A numerical example of this issue can be seen in our previous story of Bob and Sue.[1] Recall that both Bob and Sue had discretion to allocate within the industrial and office segments. Bob chose a 90/10 allocation, while Sue chose a 10/90 allocation. What should be the segment weightings in an appropriate benchmark index for evaluating their performances?

In the private real estate asset class, there is probably no good general answer to this question. The appropriate benchmark will depend on the client's overall investment strategy and the client's particular objectives for the given manager. However, a couple of general principles can be enumerated:

- The segment weights in the manager's benchmark should be mutually agreed upon by the client and the manager in advance, that is, at the outset of the management contract period. (Midcourse changes should be made only with the consent of both parties.)
- It should be possible for the manager to at least approximately replicate his benchmark's segment allocation weights, if he wants to.[2]

Both of these principles generally require that the benchmark segment weights be constant over the period of the manager's contract. Unfortunately, the segment weights in a peer-universe-based index such as the NCREIF or IPD indices are not constant over time. However, segment allocations change only gradually in the typical private property return index. Furthermore, a manager's custom benchmark can be defined using property-level (i.e., within-segment) returns from the peer universe index, rebalanced to reflect constant segment weights agreed upon by the manager and client.

In the absence of a compelling reason otherwise, the simplest segment weights to use in the manager's custom benchmark would be equal weights across all the segments within the manager's purview. Alternatively, in the stock market, the capital asset pricing model (CAPM) is often invoked as a basis to suggest that the segment weights in a manager's benchmark index should reflect the **market portfolio**. In real estate, this would reflect the market-value-based segment weights of all commercial property in the country (or at least of all institutional quality commercial properties, for an institutional core manager, for example). However, as we saw in Chapter 22, the CAPM does not work very well as a descriptor of reality *within* the private real estate asset class. There may therefore be less theoretical rationale for the use of market-based weights in a manager's benchmark in private real estate investment, as compared to stock market investments.[3]

On the other hand, a client and/or manager may feel that market weights make at least as good a benchmark as any other. In particular, market weights approximately reflect the relative abilities of different segments to absorb new investment without price distortion. Market weights may also approximate the relative ease for the manager in finding assets for

[1] See sections 26.1.2 through 26.1.4.

[2] In other words, if the manager does not wish to place a bet against the segment allocation in his benchmark, he should not have to.

[3] In other words, there is perhaps less theoretical reason within the private real estate asset class as compared to the stock market, to view the market portfolio as reflecting an optimal segment weighting in a normative sense. Furthermore, from a practical perspective, it is difficult to estimate precisely market portfolio weights within the commercial property asset class. For an example of an attempt to quantify the geographic allocation across regions in the United States, see Mahoney, Malpezzi, and Shilling (2000).

sale in the different segments. In this sense, market weights may approximate something like passive index segment weightings.[4]

## 26B.2 The Problem of Statistical Significance

When people seek to evaluate investment manager performance quantitatively as compared to a benchmark, the underlying purpose of such evaluation is usually perceived to be the identification of managers with superior versus inferior investment abilities. Investors want to place their capital with superior managers and avoid using inferior managers. Rigorously speaking, **superior investment ability** is the ability to *consistently* beat an appropriate benchmark, as opposed to a random outcome in which a manager happened to beat the benchmark over a given period of historical time. Similarly, inferior ability is the opposite, the tendency to get consistently beaten by the appropriate benchmark, as distinguished from a random streak of bad luck.

This perception of the purpose of performance evaluation raises a fundamental conceptual problem in practice. This problem exists whether evaluation is being done in the context of the public securities market or the private real estate market. The problem is a lack of **statistical significance** in ex post performance differentials. What does this mean?

A basic characteristic of risky assets is that their realized returns vary randomly around their rational ex ante expectations. This **randomness in ex post returns** is simply a reflection of the risk in the assets from an investment perspective. Performance evaluation must be based on ex post returns, the only kind of returns that can be quantified objectively and that reflect the actual realized performance of the manager. The existence of randomness in these returns means that performance evaluation is an exercise in statistical inference. Based on a sample of time (the history over which the performance evaluation is being conducted), we are trying to infer what is the central tendency of the manager's differential performance relative to the benchmark over all (or any given interval of) time.[5] The greater the randomness in ex post returns, and the shorter the historical time sample for which we have differential performance data, the less accurately we can infer the central tendency from the available empirical evidence.

Consider a simple numerical example. Suppose a manager's quarterly holding period returns are compared against those of an appropriate benchmark over a three-year period. That is, we have 12 observations (or sample drawings) about the underlying "true" (or population) difference between manager i's quarterly return and the benchmark's quarterly return. (Label this true difference $E[r_i - r_B]$.) Suppose we have very favorable circumstances for inferring this difference. In particular, suppose the true difference is constant over time, and we observe the realized returns to both the manager and the benchmark without error. Suppose also that the true quarterly volatility is only 5 percent in both $r_i$ and $r_B$, and the correlation between the manager's returns and the benchmark returns is 90 percent. Then manager i's average quarterly return would have to exceed the average benchmark return by over 130 basis points (an annualized return difference of over 5 percent) before we could conclude with statistical significance that the central tendency of the manager's return exceeds that of the benchmark's return, in other words, that the manager can beat the benchmark as a result

[4]Recall, however, that specialized local expertise is necessary to enter into specific real estate market segments. One should be wary of defining a benchmark that encourages a manager to go into territories she is not familiar with. If market weightings are to be used, they should be applied only across the segments in which the manager has sufficient expertise and resources to operate effectively.

[5]Another way of putting this is as follows: Suppose a manager beat his benchmark by 50 basis points over the past three years. What does that imply about how we can rationally expect this same manager to perform relative to the same benchmark over the *next* three years?

of skill rather than a random outcome.[6] But it is very rare to find a manager's return beating the benchmark by this much.[7]

As a result of considerations such as these, at least one major national real estate investment industry association, that of Australia, formally recommended that a **health warning** should accompany the presentation of real estate investment performance results. The suggested wording was:[8]

*Past investment performance is not an adequate test of comparative performance, nor a reliable indicator of the expected absolute level of returns in the future.*

## 26B.3 Implications of the Lack of Statistical Significance: Will the Real Purpose of Performance Evaluation Please Stand Up

At first glance, it would seem that the difficulty of making rigorous inferences about managers' investment abilities on the basis of the evidence in their realized performances would render quantitative performance evaluation pointless. Could a simple statistical point really undercut an endeavor that so many people spend so much effort on in the real world? Probably not, for at least a couple of reasons, described here.

***Including Nonquantitative Considerations*** Investors rarely base their judgments about manager performance solely on the basis of quantitative analysis. Managers generally explain and describe their investment and management philosophy, strategy, tactics, and management procedures in some detail for their clients or potential clients. Investors make judgments in part based on these **manager stories**. If the story makes sense, the investor will factor this nonquantitative information in with the quantitative evidence. The two types

[6]Note that for this type of inference we use the arithmetic average return rather than the geometric average. The computations behind this conclusion are as follows. If $s_i$ is the manager's volatility, $s_B$ is the benchmark's volatility, and $C$ is the correlation between the two, then the following formula gives the standard deviation of the difference between the manager's return and the benchmark's return, $r_{i,t} - r_{B,t}$:

$$\sigma_{i-B} = (\sigma_i^2 + \sigma_B^2 - 2C\sigma_i\sigma_B)^{(1/2)} = [0.05^2 + 0.05^2 - 2(0.9)(0.05)(0.05)]^{(1/2)} = 2.24\%$$

Dividing this by the square root of $N - 1$, where $N$ is the number of observations, gives the standard error of the observed average difference:

$$Standard\ Error = \sigma_{i-B}/[(N-1)^{(1/2)}] = 2.24\%/3.32 = 0.675\%$$

Two standard errors are required for statistical significance (95% confidence), which would be $2 \times 0.675\%$, or 1.35% per quarter. This equates to about 5.5% annualized. Note also that, although the arithmetic average return is, in principle, more appropriate for this type of statistical comparison, it is still necessary to consider the effect of volatility on the arithmetic average return, if the two performance series being compared display different volatility (which is not the case in the present numerical example). One (somewhat crude or informal) way to adjust for the effect of volatility on arithmetic mean returns is to simply use the geometric mean returns of the two series being compared, even though the confidence limits are computed as above, based on arithmetic means.

[7]In the 47 portfolios they examined, Myer and Webb (op. cit.) found only 16 whose returns exceeded the NPI with statistical significance. (A more recent study, reported in Myer and Webb (1997) found similar results in this regard.) On the other hand, it should be noted that tests of performance rank persistence ("hot hands") are more robust than tests of absolute differential performance, and there is some statistically significant evidence of persistence in investment performance rank, at least in real estate opportunity funds. See Appendix 26A for a definition of opportunity funds, and see Hahn et al. (2005) and Canizo & LaFever (2005) for some studies suggesting such persistence in private real estate.

[8]The warning label presented here is taken from the First Edition (1998) of the *Australian Investment Performance Measurement and Presentation Standards*, published by the Property Council of Australia.

of information together may allow a substantially more accurate judgment to be made about a given manager's ability, at least relative to his competitors.

***Ex Ante Role of Performance Evaluation*** The critique of performance evaluation from a statistical significance perspective in section 26B.2 viewed the purpose of such evaluation from a rather narrow, ex post perspective. It suggested that, if performance evaluation is for the purpose of inferring, ex post, which managers are truly superior and which are truly inferior (based on skill), then it will often be very difficult to achieve this purpose. But this is certainly not the only, and perhaps not even the primary, purpose of investment manager performance evaluation (even though many practitioners may consciously perceive it as such). In fact, performance evaluation serves an ex ante purpose that may be more important than its ex post purpose. To see this, consider the following characterization of the relationship between investors and managers.

Suppose the typical investor hires the typical manager in a three-year contract at the end of which the manager's performance will be formally evaluated. An incentive fee or bonus based on this retrospective evaluation may (although need not necessarily) be included in the management contract. The performance evaluation that occurs at the end of the contract formally (and empirically) applies to the preceding three years, the years during which the manager worked for the client. In this formal ex post usage, the evaluation probably will lack statistical significance. But the mere prospect that the evaluation *will be performed* has an impact on the manager's incentives and behavior ex ante, that is, *prior to* the expiration of the contract. By working harder (or "smarter") for the client, the manager increases the likelihood (ex ante) of beating her benchmark.[9] This increases the likelihood that the manager will be rehired for a subsequent multiyear period (and it increases the manager's expected earnings from any incentive fee component of the contract based on the formal evaluation). Thus, on an ex ante basis, the use of formal performance evaluation helps to set up an **interest alignment** between the manager and the investor, thereby reducing agency costs in the investment industry.

Ex post, water is over the dam, and often little can be concluded, rigorously speaking. The real purpose (or more important role) of formal, quantitative investment performance evaluation probably lies in the ex ante role just described, particularly in combination with the (often nonquantitative) plausibility of the managers' stories. This is true both in real estate and securities investment management.

The importance of the **ex ante role of performance evaluation** does not mean that the statistical problems with ex post inference can be ignored, however. The role of randomness in ex post returns, and also of ex post return measurement difficulties especially in the private real estate asset class, should be carefully considered.[10] This is particularly important in constructing investment management contracts that have **incentive fee** provisions based on performance evaluation (e.g., the use of performance **bogeys**). Too much randomness and lack of precision in the relationship between the ex post performance measure and the actual investment skill and diligence applied ex ante by the manager will act to demoralize or discourage good managers. The results could be perverse from the perspective of aligning the interests of the investor and manager. For example, if too large a portion of the manager's overall fee is based on

[9]Incidentally, since this is as true for a peer universe benchmark as it is for a passive index, from this ex ante incentive perspective, peer-universe-based benchmarks are generally as functional as passive indices.

[10]As noted in Chapter 25, temporal lag bias in real estate returns may not be a prohibitive problem in performance evaluation, as such bias may largely cancel out in the performance differential between the manager and his benchmark. It is important, however, to ascertain whether the degree of lagging is similar in the manager's reported returns and those of the benchmark. If the lag is similar, then smoothing or lagging will probably not much affect the performance evaluation. However, random noise (purely random return measurement error as distinct from temporal lag bias) may still pose a serious measurement problem in the performance differential. Such measurement error is an additional source of randomness beyond the underlying volatility in the true returns across time. Noise will be a particular problem if either the manager's portfolio or the benchmark contains a relatively small number of individual properties or appraisers.

ex post performance relative to a bogey, then the best managers may shun investors who demand such contracts, or both sides may end up just "gaming" the incentive.[11]

The importance of minimizing randomness in the ex post return differential between the manager and an incentive contract bogey highlights the importance of selecting an appropriate benchmark for the manager's performance evaluation. Our previous point about matching responsibility with authority is obviously important in this regard.

## 26B.4 Adjusting for Risk

Risk and return go hand in hand in the capital markets. The market naturally provides investors with greater average returns, over the long run, for more risky investments. This means that, on average or from an ex ante perspective, managers can increase their returns (and increase their probability of beating any given fixed reference point) by taking on more risk. But obviously, this is not what investors have in mind when they aim at rewarding managers for beating reference points or providing managers with incentives to work more diligently. Superior (or even just diligent) investment management does not consist simply of using the capital market's return/risk trade-off to provide higher average returns for the investor. Therefore, it is essential to control for risk in investment performance evaluation.

### WHAT ABOUT THE SHARPE RATIO?

The Sharpe ratio, which we defined in Chapter 21, is one quantitative risk-adjusted return measure that is rather widely reported in private real estate investment. It is computed as the risk premium per unit of risk, where *risk* is defined as the volatility of the subject manager or portfolio. While this measure is of some general interest, to our knowledge it has not been widely used in formal performance evaluation in the private real estate asset class. This is probably for the best, as there are several theoretical and practical problems with the Sharpe ratio for use in the typical real estate management situation.

The most serious theoretical problem with the Sharpe ratio is that the volatility of the manager's portfolio is generally not the measure of risk in that portfolio that matters to the capital market. Therefore, it is not the measure of risk that will be priced. As a result, the Sharpe ratio is likely to be misleading as an adjustment for a portfolio's return to account for its risk. Related to this is the fact that any one real estate manager or portfolio will typically be responsible for only a small fraction of an investor client's total wealth portfolio. In this case, the denominator in the Sharpe ratio does not measure the contribution of the subject portfolio to the risk that matters to the investor, which is the volatility in the investor's total wealth portfolio (see Chapters 21 and 22).*

In addition to this theoretical problem, the Sharpe ratio faces another, more practical problem particularly within the private real estate equity quadrant. As we described in Chapter 25, private real estate returns must be based on appraisals or whole asset transaction prices that are "noisy." We noted there that noise affects the volatility of asset returns. The greater the noise in the valuations, the greater the volatility in the return time series derived from those valuations. But the greater the number of properties in the portfolio, the less the effect of noise. (Remember the Square Root of *n* Rule.) So the effect of random error on the volatility of a portfolio or index will vary, depending on how many individual properties are in it. Of course, different managers will typically hold different numbers of properties. And any given manager will typically hold a different number of properties (usually fewer) than what is in her benchmark. Thus, noise will differentially affect managers, and their benchmarks, and this will bias any performance comparison or evaluation based on the Sharpe ratio.

*A point raised in defense of the Sharpe ratio in this regard is that the risk that actually matters to the capital market may be *proportional to* an asset's volatility, if all the relevant assets (in the manager's choice set) have the same correlation with the market risk benchmark. In the case of private real estate assets, no one is really sure what this benchmark is, or what the "true" correlation between real estate and such a risk benchmark would be anyway, so the assumption of equal correlation is implicitly applied, allowing the use of the Sharpe ratio (as a kind of "better than nothing" fallback). The weakness in this argument is apparent.

[11]In the extreme, if too much random measurement error exists in quantifying the performance differential between the manager and the benchmark, then poor managers will have almost as good a chance of earning an incentive fee as superior managers.

Broadly speaking, this can be done in either of two ways, either by adjusting the performance measure, or by restraining the manager's leeway to modify the risk in her portfolio away from the risk in her benchmark. Let's briefly consider each of these approaches in the context of the private real estate asset class.

In the first approach, the manager's ex post return and that of her benchmark are quantitatively adjusted for risk. The performance evaluation is made using **risk-adjusted returns** rather than unadjusted raw return results. This approach is treated extensively in mainstream investments textbooks and has found some application in real world practice in the stock market. It includes measures such as the **Sharpe ratio**, the Jensen alpha, the Treynor ratio, and various others.[12] The advantage of this approach is that, if it can be done well enough, it reduces the need for the manager's benchmark to have the same risk as that of the manager's portfolio.

Unfortunately, there are some problems with quantitative risk adjustment for manager performance evaluation. In principle, the key to this technique is that the measure of risk in the manager's and benchmark portfolios (for example, the beta, if a CAPM definition of risk is being used) should reflect the amount of risk *as it matters to the capital market*. In theory, this is the only way the risk adjustment can reflect the **market price of risk**, that is, the way the market would adjust the expected return risk premium, to reflect the amount of risk in the portfolio. But practical problems in risk measurement, as well the fundamental theoretical state of the art in equilibrium asset price modeling (as described in Chapter 22), make this type of risk adjustment an inexact science, especially for private real estate. This problem, combined with the previously described difficulties with statistical precision in the estimation of performance differentials even without adjusting for risk, takes the wind out of the sails of formal, quantitative risk adjustment in investment performance evaluation. This is probably the main reason this approach is not very widely used in real world practice at least for private real estate.

There are even greater difficulties, both theoretical and practical, in applying formal quantitative risk adjustment to investment in the private real estate market. These include the data problems described in Chapter 25, as well as the theoretical problems described in Chapter 22. In particular, recall that a single-index CAPM-type model does not work well within the private real estate asset class. We do not know how to distinguish the beta of one type of property from that of another, and there seems to be little clear or stable differences in expected returns across different property segments (within the core institutional quality property market, apart from the difference between the "institutional" and "noninstitutional" market segments). More sophisticated multifactor risk models may improve this situation, but the practical development of such models is still in a rather embryonic stage. Furthermore, multifactor risk models do not provide a one-dimensional measure of risk that can be used to develop a single, unambiguous measure of a portfolio's risk-adjusted return.[13]

This does not mean that one can ignore the issue of portfolio risk in real estate performance evaluation. What it means is that one must fall back onto the second general approach for controlling for risk. For want of a better label, we will call this the **nonquantitative risk control**, to distinguish it from the previous technique. The basic idea in this approach is to constrain the manager to investing only in assets that are of essentially equal risk, which is also the risk of the manager's benchmark. Then we do not need to adjust for risk quantitatively. We can compare directly the unadjusted returns.

---

[12]For a good presentation of these measures and their strengths and weaknesses, see Brodie, Kane, and Marcus (1999). These types of measures are not widely employed in private real estate investment performance evaluation, although Sharpe ratios are often reported. (See the boxed feature on page 6.)

[13]A Jensen alpha measure can be calculated for a portfolio using a multifactor risk model, but such a measure only identifies the sign of the portfolio's risk-adjusted return (i.e., it tells whether the return is above or below the security market hyperplane). It does not generally allow a rank-ordering of the risk-adjusted return performance of two portfolios that have the same signed alpha but different amounts of risk.

In private real estate, the implementation of this approach typically requires restricting the manager regarding the use of leverage and regarding investments in **noninstitutional properties** or in terms of investment style (as previously described). For example, development projects or properties in need of rehabilitation are generally perceived to contain more risk than high-quality, fully operational properties. Properties that are smaller or of lower quality and less prime locations than traditional institutional core real estate assets may contain more risk.[14] Of course, the manager's benchmark must conform to the same constraints in this regard. Effectively, this confines the ability to rigorously control for risk to the institutional core of the private property market because benchmark indices of periodic returns do not exist as yet for noninstitutional commercial property in the United States.

## KEY TERMS

market portfolio
superior investment ability
statistical significance
randomness in ex post returns
health warning
manager stories
ex ante role of performance evaluation
interest alignment
incentive fee
bogeys
risk-adjusted returns
market price of risk
nonquantitative risk control
Sharpe ratio
noninstitutional properties

[14]See our discussion in Chapter 22 regarding the nature of risk in real estate, as it matters to the asset market.