

INTERNATIONAL REAL ESTATE REVIEW

2007 Vol. 10 No. 1: pp. 93 - 118

Indices for Investment Benchmarking and Return Performance Analysis in Private Real Estate

David Geltner

George Macomber Professor of Real Estate Finance, Department of Urban Studies & Planning, Massachusetts Institute of Technology, Cambridge, MA 02139; E-mail: dgeltner@mit.edu

David C. Ling

William D. Hussey Professor of Real Estate, Department of Finance & Real Estate, Warrington College of Business Administration, University of Florida, Gainesville, FL 32611-7160; E-mail: ling@ufl.edu

In U.S. commercial real estate markets, a single index product produced by the National Council of Real Estate Investment Fiduciaries (NCREIF) has been used for both asset class research and for agent evaluation benchmarking. While the NCREIF Index is an invaluable tool for the U.S. real estate investment industry, in some respects its construction is not optimized for *either* research or the benchmarking function, though the gap between the reality and the ideal is arguably greatest on the research side. This paper first discusses the purpose of the ideal research index, the optimal depth and breadth of market coverage, and the type of information included in the research index database. This discussion is followed by an analysis of the ideal agent evaluation benchmark index, including the definition, purpose, and uses of evaluation benchmarking, the characteristics of the ideal benchmark index, market coverage and information considerations, and a comparison of benchmark indices in the private real estate market to comparable indices used to benchmark public securities.

Keywords

benchmarking; return performance; agent evaluation; private real estate investment

Introduction and Definition of Index Types

Geltner and Ling (2000, 2001) argue that it is critically important to distinguish between two types of commercial real estate index products: (1) agent evaluation “benchmark” indices and (2) asset class “research” indices. Benchmark indices are designed and constructed for use in evaluating the return performance of an agent/manager relative to the performance of his or her peers. In contrast, the primary purpose of an asset class “research” index is to support quantitative analysis useful for property and portfolio level valuation and investment decision making.

There are at least two reasons why it is important to distinguish between benchmark indices and research indices. First, the ideal evaluation benchmark index differs in important technical respects from the ideal research index, which is not surprising given that their uses and functions are quite different. For example, a benchmark index should ideally include performance data on the manager’s peer universe; that is, all competing investment managers of a given investment type or style.

In contrast, the ideal research index should be constructed using stratified samples optimized for making statistical inferences about research oriented questions, for example, questions about real estate market movements, volatility, or correlations. Because trade-offs exist in the construction of any index product, Geltner and Ling (2000, 2001) argue that it is not possible to optimize a single index product simultaneously for both the benchmarking function and the broader and more diverse function of asset class research.

A second reason to distinguish between benchmark indices and return indices is that it may be useful to consider separate production of the two products. That is, it may make sense for different entities or organizations to specialize in the production and dissemination of one index or the other. Moreover, research indices have many of the characteristics of a “public good” in that the long-run collective benefit they provide to the real estate industry is greater than the short-run or private benefit they provide to any one firm or investor. As a result, a high quality research index may require subsidized or collective production. In sharp contrast, the growth in demand for formal agent evaluation benchmarking and performance attribution over the past two decades has made the production of such benchmarks, along with performance attribution and diagnostic services, viable as a commercial product.

In the U.S., a single index product produced by the National Council of Real Estate Investment Fiduciaries (NCREIF) has been used for both the benchmarking and asset class research functions. While the NCREIF Index

is an invaluable tool for the investment industry, in some respects it is not optimized for *either* the research or the benchmarking function, though we shall argue that it comes closer to the ideal as a benchmark index than as a research index. Interestingly, the NCREIF Index was originally established as a public good by collective industry action with the primary motivation being to foster asset class research, and NCREIF's mission statement reflects this goal. Nevertheless, the use of the NCREIF Index by the investment industry in the U.S. has become heavily focused on agent evaluation benchmarking. However, the NCREIF Index is also not ideal as an evaluation benchmark. It covers a relatively narrow segment of private property markets (the tax-exempt fiduciary branch) and is missing some key components of the evaluation function, such as the data necessary for property operational level performance attribution (at least until recently¹).

Six years ago we proposed that the NCREIF Index evolve into two families of index products, separate, but coordinated (Geltner and Ling, 2000). The first index product would focus primarily on, and be optimized for, agent evaluation and performance attribution. The second index product would be optimized for broader asset class research purposes. Despite its recent strong return performance, private commercial real estate still receives less respect and credibility among many investors than stock and bond investments. We believe a high quality research index will attract more academic and practitioner research. In turn, this will, over the longer run, build up the fundamental knowledge base of the asset class. The existence of such a knowledge base would add to the credibility of the real estate asset class and lead, we predict, to increased capital flows to the asset class.

NCREIF has made significant progress during recent years in including property operational and valuation data and expanding its database beyond fully leased, institutional quality, properties. Nevertheless, progress toward the production of publicly available return indices optimized for asset class research has been painfully slow, despite numerous demonstrations of the feasibility of creating a "CRSP for real estate."² Some progress has been recently made with the launch in 2006 of the MIT Center for Real Estate's Transactions-Based Index of Institutional Commercial Property Investment Performance (TBI), developed in cooperation with NCREIF and based on NCREIF data. The purpose of this index is to measure market movements and returns on investment based on transaction prices of properties sold from

¹ Subsequent to the dissemination of the Geltner-Ling report of 2000, NCREIF launched an effort to begin collecting operational data from its members. This data is now being compiled and made available to NCREIF members and subscribers.

² The Center of Research in Security Prices (CRSP) at the University of Chicago provides a database of sufficient quality to support a high level of academic and industry research into the equity asset class.

the NCREIF Index database. This transactions-based index is being provided free of charge for research purposes by the MIT Center for Real Estate as a service to the industry and academic research communities.³ The development and publication by MIT of this research oriented return index should provide an important new development in the commercial real estate industry.

This paper is divided into two main sections. The first broadens and deepens the analyses, first put forth by Geltner (2000) and Geltner and Ling (2000, 2001), of the ideal benchmarking and asset class research indices for private commercial real estate. In particular, the purpose of the ideal research index, the optimal depth and breadth of market coverage, and the type of information included in the research index database are discussed in detail. This analysis is especially timely given that, with the publication of MIT's transaction-based index, researchers now have access to a return index that is tailored to shed light on some research questions that are difficult to resolve using the appraisal-based NCREIF Index by itself. This discussion is followed by an analysis of the ideal agent evaluation benchmark index, including the definition, purpose, and uses of evaluation benchmarking, the characteristics of the ideal benchmark index, market coverage and information considerations, and a comparison of benchmark indices in the private real estate market to comparable indices used to benchmark public securities.

Definitions and Purpose of the Ideal Real Estate Research Index

The primary purpose of a private real estate research index is to support quantitative analyses that improves our understanding of the risk and return characteristics of direct investment in commercial real estate. As first suggested by Geltner and Ling (2000, 2001), high quality research indices are needed to support diverse and fundamental asset class research studies.⁴

³ Detailed information on MIT's transaction-based index can be found at <http://web.mit.edu/cre/research/credl/tbi.html>.

⁴ For the functions served by a research index it is probably more realistic and useful to speak in terms of a commercial real estate "research database," from which one or more types of specific indices can be constructed, rather than to speak in terms of a single "research index." Nevertheless, we are focusing in the present context on the index-oriented applications of such a database, so we shall use the term "research index" widely, keeping in mind that this may encompass a "family" of indices and index-related information products derived from a supporting database.

Characteristics of a flagship research index

A “flagship” research index must command and retain the respect of the mainstream academic and industry research communities. To succeed in this regard, Geltner and Ling (2000, 2001) argue that a research index must be:

- Based on state-of-the-art data and index construction methodologies;
- Compiled in an unbiased, neutral manner not susceptible to manipulation;
- Subject to public and academic scrutiny and criticism; and
- Accessible widely to industry and academic researchers at a reasonable cost.

Furthermore, the ideal asset class research index should be useful in answering the following questions:

- (1) What is the long-run (multi-decade) nature of the investment performance of the commercial real estate asset class compared to other major investment asset classes such as stocks and bonds?
- (2) What are the long-run total return and capital return time-series mean, volatility, and correlations both within the real estate asset class and between real estate and other asset classes?
- (3) What is the nature of the lead-lag relationships between real estate returns and other variables?
- (4) How can real estate returns be forecasted, and with what degree of accuracy over what time horizon?
- (5) When, historically, did market value turning points occur, and what were the duration and amplitudes of the historical “cycles” in various segments of the market?
- (6) What is the difference in the investment performance of various types of property, and different types of locations?

In short, the ultimate objectives of the asset class research index are to improve fundamental understanding of the risk and return characteristics of private market commercial real estate investments and to improve practical decision-making.

Market Coverage of an Ideal Research Index

The breadth of market coverage refers to the number of different property asset market segments represented by the index. For example, should the research index include only the classic four “food groups” (office, industrial, retail, apartment), or should it extend to other types of properties (hotels,

land, timber, golf courses, etc)? Should it break down into sub-types (CBD vs suburban office, garden vs high-rise apartments, etc)? Should the index include only the largest metropolitan areas, or should it extend to all 200+ metropolitan areas plus rural zones? Should the research index only include “institutional size” properties (e.g., \$10 million +) or smaller properties as well?

A useful way to approach the question of research index market breadth is to ask what types of research and investment decision questions could be addressed by an additional increment of index coverage. For example, improved investment performance information about market sectors that are beyond the current institutional holdings could improve the ability of institutions to invest in such new market segments. A relevant question thus becomes: to what extent would the inclusion of an additional market sector push out the efficient frontier; that is, how much additional diversification or improved risk/return opportunities would be provided by the additional coverage?

In the absence of better information, the intuitive answer to the question of ideal market coverage is to first “grab the low-hanging fruit.” That is, include all market segments that can be relatively easily and inexpensively covered. The relevant “ease” and “expense” includes both technical considerations and institutional or organizational considerations. In the U.S., there is essentially no commercial real estate performance index coverage beyond the existing holdings of the tax-exempt fiduciary branch of the real estate investment industry.⁵ However, other countries have been able to include numerous major branches of the real estate industry in their indexes. For example, in Britain, the Netherlands, and Australia the holdings of public real estate companies, insurance companies, and some wealthy private investors are included in the major index products.

It is important to note that advances in statistical methodologies and computer technology have made it easier to include appraised values and transaction prices in index construction. In effect, properties can be reappraised much less frequently than the reporting frequency of the index, and/or indices can be based directly on transaction price evidence without the need for individual property-level appraisals.⁶ Thus, it is possible to expand the market coverage of a research index by (i) including data-contributors that do not regularly or frequently appraise their properties (for example, REITs and insurance companies) and/or by (ii) basing the capital

⁵ Electronic databases recently compiled by firms such as Real Capital Analytics and Co-Star may provide some possibility to move beyond this state of affairs.

⁶ See, for example, Gatzlaff and Geltner (1998), Geltner and Goetzman (2000), Fisher et al. (2006), and Marcato (2005).

gain component of the research index at least partly on transaction price data that can be acquired through publicly-available sources.

Level and Type of Information Included in the Ideal Research Index

The types of information encompassed in an ideal commercial real estate index include:

- (1) Asset market prices: Both the levels and changes in property prices (values);
- (2) Asset market activity: The source and volume of transactions and capital flows;
- (3) Investment cash flows: Net operating income and capital expenditures;
- (4) Rental rates; and
- (5) Development industry information: Construction costs, activity levels, and land values.

Obviously, not all of this information can be presented in a single “index.” An entire database or family of information products would be required to manage this information. The primary focus here is on the investment industry and the asset market, which includes the first three items in the above list.

In addition to periodic cash flows, the information most important to a real estate research index is transaction prices and activity. Of these two, tracking asset market prices is probably the most fundamental and important, and also presents the greatest challenge. Accordingly, our discussion of the ideal asset class research index is focused on the tracking of asset price changes.

Considerations and trade-offs in deciding what information to collect

Although we would like as much information as possible, cost/benefit trade-offs exist, and they relate in part to institutional and organizational concerns as well as technical concerns. Data-vending firms already supply many types of information. The “low hanging fruit” has become more plentiful as electronic commerce and internet-based information firms (for example, CoStar and Real Capital Analytics) further develop in the real estate industry.

In addition to the type of information currently collected by NCREIF, we strongly suspect the types of information collected by the IPD in England

and the PCA in Australia could be included in a research index database in the U.S. Perhaps this could be accomplished by a joint-venture between an industry association and one or more private commercial data-vending firms. Such an index would therefore include, in addition to asset prices and other transaction information, space market and property operational data.

Total returns versus appreciation returns

Although total return information is vital for assessing investment performance, investment performance is not the only subject of interest, either for improving fundamental understanding of the real estate asset class or for supporting practical real estate decision-making. Partial information relevant to components of the total return can also be quite valuable in conducting many types of research relevant for investment performance. For example, relevant studies of volatility, systematic risk, cross-correlation, serial correlation, forecasting of returns, and analysis of lead/lag relationships all can be carried out using only the capital gain component of the periodic total return.

In short, a research index can be of considerable use and value even if it is incapable of reporting total returns. Depending on the purpose and use of the index, other types of information, such as periodic appreciation, space-market or property operational variables, and a variety of other information products can be quite useful. This is in contrast to agent evaluation benchmark indices, whose more specialized purpose requires they be able to produce reports of total returns.

Transaction prices versus appraisals

Transaction prices are not always superior to appraised values as a basis for constructing a research index for private real estate. There is much similarity and mutual influence between transaction prices and appraisals, although there are also some differences in these two types of valuation observations.

When considering the use of either transaction prices or appraised values as proxies for “true” market value, it is important to note that transaction prices and appraised values both contain errors. In this context, the word “error” refers to the difference between empirically observable prices or valuations and the underlying (but unobservable) true “market values.” However, there do tend to be differences between transaction prices and appraisals in the nature of the valuation errors. These variations, in turn, require different adjustments in the use of a research index depending on whether it is based on appraised values or transaction prices (and depending also on how the

index is constructed). For example, a research index designed to track price movements on a quarterly basis without the lagging that would typically occur in an appraisal-based index, would need to employ transaction price evidence to eliminate the temporal lag in an appraisal-based index.⁷

Equally weighted versus value-weighted

Another important consideration is whether the constituent properties in a research index should be equally-weighted or value-weighted when computing periodic returns. This depends on whether the research index is being viewed as a statistical sample, or a population census. As noted, the individual constituents of a sample should normally be equally-weighted.⁸ It is usually more efficient and useful to conceive of a research index as a statistical sample, rather than as a population census, at least for property-level research. This implies that the typical property-level research index should be equally-weighted, at least within strata or relatively homogeneous market segments.

Reporting Frequency in the Research Index

Which is best: annual, quarterly, or monthly return reporting frequency in a research index? In the case of private real estate, greater frequency does not necessarily imply a more useful index because of a fundamental trade-off. For any given level of underlying data collection, higher frequency reporting tends to result in more errors *per period*.⁹ These errors include both temporal lag bias (smoothing) and random error (noise).

Although the terms of the noise-vs-lag trade-off can be improved by the use of better statistical methods of index construction, ultimately the only way to increase reporting frequency without increasing either noise, temporal lag, or both, is to invest more in the underlying property-level valuation or data collection process. Thus, at some point the trade-off is between reporting frequency and accuracy and the budget constraint placed on the data collection function. However, the underlying sluggishness and inertia in private real estate asset markets tends to reduce the incremental benefits of higher frequency reporting. Short return periods tend to be statistically dependent on adjacent periods, so one does not gain as many useful return observations as one might first imagine, even if one ignores the valuation

⁷ This issue is discussed in detail in Geltner and Ling (2006).

⁸ This is true unless the sample is stratified so that certain components represent disproportionate shares of the underlying population, in which case the equal weighting applies within each "stratum" or "segment" of the sample.

⁹ The issues in this section are discussed in more depth in Geltner and Ling (2006).

errors in the index. Based on widespread industry practice, the ideal reporting frequency in a private real estate research index is probably quarterly.

The ideal historical time span of a research index is, of course, as long as possible. However, although some research questions require long historical time-series of data (e.g., estimates of long-term statistical characteristics, such as mean, volatility, correlation, lead-lag relationships), others do not (such as the question of whether the market “turned around” last quarter). Note also that some research questions require rather precise contemporaneous measurement of value changes (e.g., the “turnaround” question requires quarterly representations without lagging or smoothing), while other questions do not require such precision. For example, the long-run mean return of an index is not much affected by lagging and smoothing. Finally, note that the historical time span multiplied by the index periodic return reporting frequency gives the time-series sample size, which is positively related to the statistical inference power of the index.

Research Indices in Private versus Public Markets

A sample-based index is used to represent an entire market segment for research purposes. Sample-based indices are common in the public securities investment industry and are widely used for academic and industry research purposes. Perhaps the most prominent example is the S&P500 Index, which comprises a sample of 500 stocks, but which is often used for research purposes to represent all large-cap stocks. Two attributes make the S&P500 Index an acceptable passive research index for this market segment. One attribute has to do with the adjective “market.” The other has to do with the adjective “passive.”

The S&P500 Index is a “market” index because it effectively represents a well-defined segment of the equities market. However, the S&P500 Index is only a sample of U.S. large-capitalization stocks. Nevertheless, the S&P500 well represents the large-cap market segment because, statistically speaking, the S&P500 Index: (i) is a large enough sample (500 stocks), and (ii) is sufficiently representative of the large-cap market segment because its constituent stocks are chosen through a deliberate process designed to insure representativeness. As a result, the returns to the S&P500 Index are highly correlated with most reasonable definitions of the entire population of all the U.S. large-cap stocks.

Next, consider what makes the S&P500 Index a “passive” index. The S&P500 Index does not change its composition very frequently. Thus, it can

accurately reflect the performance of an essentially passive investor in the market segment or style represented by the S&P500. In the public securities industry it is possible to, in effect, hire a passive investment manager who essentially mimics a passive market index like the S&P500. Such managers have relatively low research and trading costs, and therefore relatively low expenses, as in the case of “index funds.”

The closest equivalent the private real estate investment industry could conceivably have to a passive market index like the S&P500 would be an index representing the investment performance of virtually all of the properties in a given property market segment or group of market segments. For example, an index representing all of the income-generating properties in the U.S. with a market value greater than \$20 million might be considered roughly comparable at a conceptual level to a passive market index of large-cap stocks such as the S&P500. However, adequate representation of a market does not require that all the properties in that market be included in the index database.

Definitions and Purpose of Agent Evaluation Benchmarking

We now turn our attention to the second basic type of index product: an “agent evaluation benchmark index.” Geltner and Ling (2000 and 2001) suggest there are at least three reasons why evaluation benchmarking is done in commercial real estate:

- (1) To aid communication between the principal and agent regarding investment objectives, strategy, and tactics;
- (2) To help to align the interests of the principal and agent; and
- (3) To help weed out obviously inferior investment managers from active involvement in the industry.

The first purpose is to help the principal and agent to clarify and understand their objectives, functions, and roles in their mutual relationship. By together deciding on an appropriate evaluation benchmark at the outset of the management contract, the principal and agent can make clear their mutual expectations. Moreover, by explaining his relative performance with respect to the benchmark, the agent can help the principal party understand what the agent can do (or has done), and what he cannot (or did not) do.

The second function of evaluation benchmarking is to help align the interests of the principal party and the agent. If the benchmark reflects an appropriate standard on which to base measurement of the agent’s actual contribution to achieving the principal’s objectives, and if the agent knows his reward will

be positively correlated with his performance relative to his benchmark, then benchmarking gives the agent a direct incentive to perform in the interest of the principal.

Finally, the mere existence of the benchmarking function helps to keep inferior managers out of the business. “Inferior” managers by this definition are those who would tend to consistently underperform the market on a risk-adjusted basis.

Note that, in large measure, evaluation benchmarking plays essentially an *ex ante* role. That is, by the end of the contract period when the final review of the agent’s performance is reported, the agent’s behavior and performance cannot be changed, at least for the past management contract or review period. The only actions that can then be influenced are those in a subsequent contract or review period (if any). Therefore, problems associated with drawing accurate inferences from measurements of noisy *ex post* returns do not necessarily undercut the beneficial *ex ante* functions of benchmarking described here (Geltner and Ling, 2000 & 2001).

Using evaluation benchmarking to identify superior managers

Because statistical “noise” in the measurement of *ex post* investment returns makes it difficult to distinguish superior performance from luck, principals should avoid relying on quantitative performance measures in the evaluation of agents. Instead, principals should focus attention on the agents’ explanations of their results with respect to their benchmark. These explanations will involve detailed discussions of strategy, tactics, and procedures, including both the acquisition/disposition and operational management functions. Such discussions aid communications and help with interest alignment.

Characteristics of the Ideal Private Real Estate Evaluation Benchmark Index

Geltner and Ling (2000 and 2001) argue that an ideal performance evaluation benchmark index is defined by six fundamental criteria:

- (1) The benchmark’s return can be calculated over the time span of the investment management contract;
- (2) The investor (principal) can invest directly in the benchmark index as an alternative to hiring the investment manager (agent);
- (3) The investment manager will never be forced to place a bet against the benchmark index;

- (4) The benchmark should reflect the investment style or specialization of the manager;
- (5) The manager should not be able to influence the performance of the benchmark to any significant degree; and
- (6) The benchmark should be mutually agreed upon by the investor and manager at the outset of the investment management contract.

The necessity for the first criterion is obvious, because a primary purpose of evaluation benchmarking is the quantitative measurement of investment performance.

The second of the six criteria, that the investor can invest directly in the benchmark, derives from the fundamental precept that evaluations should be based on the incremental value added by the manager over and above what the client could do without hiring the manager. Incremental evaluation is a fundamental construct in economics and the decision sciences.

The third criterion, the no-forced-bet against the benchmark criterion, is based on the widely accepted management evaluation principle that responsibility should be equated with authority. If the manager is going to be held responsible for his performance relative to a benchmark, then the manager should have the authority to control that relative performance as much as possible. This suggests that the manager should be able to decide whether, how, and when to deviate from the agreed upon benchmark. Note that the no-forced-bet criterion overlaps the second criterion at a practical level. If the principal can invest in the benchmark directly (the second criterion), so to can the manager, thereby avoiding a bet against the benchmark (the third criterion). Thus, a common practical characteristic implied by both the second and third fundamental criteria is that the ideal evaluation benchmark should be investable by both the principal and the manager.

The basic management principle of equating responsibility with authority also underlies the fourth criterion: that the benchmark should reflect the manager's style or specialization. This is sometimes referred to as the *appropriateness* criterion and is important for achieving the objective of fairness in the evaluation of the manager and for promoting the usefulness of the benchmarking process for implementing the investor's overall investment strategy. Typically, an investment manager is hired to implement a component of an overall strategic investment plan. For example, a manager may have particular expertise in picking small-cap value stocks, or in finding, acquiring, and managing apartment properties in the western U.S. If so, the manager's benchmark should reflect this specialization. In this way, the

manager is not tempted to deviate from his area of specialized expertise in order to avoid placing a bet against his benchmark. This encourages “style-pure” and specialized expertise-based investment management.

The fifth criterion, that the manager should not be able to directly influence the performance of the benchmark index, is necessary in order for the process to be meaningful and to provide the *ex ante* incentives described above. In the absence of this criterion, the manager could conceivably manipulate the performance evaluation to his advantage.

Finally, the sixth criterion of an ideal evaluation benchmark is that both parties to the management contract should explicitly agree upon the benchmark index in advance. Mid-contract changes can be made, but only with the agreement of both parties. As with the fourth criterion, this requirement helps to insure that the interests and objectives of both parties are well understood and aligned, and it helps the manager to plan and carry out an investment policy in the client’s interest.

Peer universe index versus a passive market index

Two types of evaluation benchmark indices are often used in the investment management industry: peer universe indices and passive market indices. A “peer universe index” is designed to reflect the performance of all managers that are effectively competitors of the subject manager; that is, all managers who have the same style or specialization as the subject manager. In the case of core institutional real estate investments, the NCREIF Index is an example of a property-level peer universe index in the U.S. As such, the NCREIF Index can be viewed as either: (i) directly representing the property-level performance of peer universes among the NCREIF data-contributing members (largely investment managers in the tax-exempt fiduciary branch of the industry); or (ii) indirectly representing the property-level performance of all the tax-exempt funds (trustees) who either are themselves NCREIF data-contributing members or who have hired NCREIF data-contributing members as their investment agents. Similarly, the IPD Index in Great Britain, the PCA Index in Australia, and the ROZ Index in the Netherlands can be viewed as institutional real estate peer universe indices in their respective countries.

A passive market index, on the other hand, is constructed to represent the performance of all the assets in a given market segment or group of segments. In the public securities investment industry, so-called passive investment management is possible and widely practiced. Passive management in this context typically refers to the replication or mimicking of a specified passive market index, thereby eliminating the individual stock

selection (active trading) function of the manager, with a resulting minimization of management expenses. In the private real estate investment industry, truly passive investment is not possible for three reasons: (1) the need for asset operational management; (2) the need for specialized local expertise; and (3) the impossibility of replicating a passive market index due to the trading of unique, whole assets.

A fundamental characteristic of private real estate investment is that property owners are responsible for the operational management of their asset holdings. Because properties are held for long periods of time (in part due to high transaction costs), and typically provide most of their total return in the form of periodic income rather than capital gain, property operational management can have a significant impact on the long-term investment performance of private real estate. Operational management is certainly a type of “active” investment management, in the sense that it is a major (and costly) activity for which the investment manager is responsible (and for which he is compensated). It is virtually impossible to invest in private real estate assets without the use of a manager who is “active” at least in the sense of being responsible (directly or indirectly) for operational asset management at the property level.

In contrast, active management in the public securities investment industry consists purely of asset picking; that is, the attempt to find superior-performing individual assets. This is trading-oriented active management, as distinguished from operational management-oriented active management. Short-term, opportunistic trading of assets in private real estate markets may be self-defeating, due to the much higher transaction costs (as a fraction of asset value).¹⁰ In fact, the most “active” private real estate investment managers (in terms of individual asset trading) probably trade assets less frequently than the typical passive manager in the public securities industry. Moreover, even a passive real estate manager holding a fixed set of properties may be very “active” in terms of property operational management.

Furthermore, the high degree of operational management required in private real estate markets makes it difficult empirically to rigorously distinguish manager value added attributable to asset picking and trading from manager value added attributable to ongoing operational management. It would be as if the roles of the investment manager picking IBM stock and that of IBM corporate management were somehow merged into one role.

¹⁰ The results of Brounen et al. (2007) suggest that active management strategies do not, on average, add value on a risk-adjusted basis, to investor portfolios.

Closely related to this point is the fact that private real estate investment requires greater levels of specialized local market expertise than is required in the public securities industry. This is not only because unique, whole assets are traded in illiquid, private markets, but because the market in which these assets are traded is less informationally efficient than the public securities markets. This lack of informational efficiency means investors do not have the power of a highly efficient, competitive marketplace protecting them from doing “bad deals,” in the sense of overpaying for purchases or selling at prices below market value. In contrast, when you buy shares in IBM, you can be relatively certain, *ex ante*, that you are not paying more than market value. Hundreds of investors are buying and selling IBM at publicly-quoted prices at or very near the price you pay, in a public exchange environment designed to maximize public information dissemination and minimize the abuse of insider private information. However, it requires specialized knowledge and expertise to be sure you are not overpaying for a property located at 1000 North Main Street.

Finally, in the private real estate investment industry, it is often virtually impossible to even define an appropriate and measurable passive market index. Moreover, even if such an index could be defined, and its investment performance measured, a passive index in private real estate would not be investible. This is because unique, whole assets are traded in the private property market. Because the assets are already be owned by someone else, it is impossible for any one investor (or any one investment manager) to hold the assets that would comprise any conceivable passive market index. This eliminates the potential in private real estate to implement passive investment management defined as the replication of a passive market index by the manager. In fact, real estate managers would be forced to “bet against” any passive market index, at least at the individual property level, and investors could not directly invest in the index as an alternative to hiring a manager. As a result, significant tracking error between the investor’s portfolio and the benchmark cannot be avoided even when the benchmark is a passive market index.¹¹

In summary, there are important differences between the public securities and private real estate investment industries regarding the relative appropriateness and meaningfulness of passive versus peer universe based benchmarks. However, just as a review of *Consumer Reports* is a reasonable and prudent exercise when making consumer product purchase

¹¹ “Tracking error” refers to deviations in periodic returns between a subject portfolio and a benchmark. Because in private real estate the benchmark itself is not investible, idiosyncratic differences in returns between the manager’s properties and the benchmark properties would make tracking error inevitable.

decisions, so to is evaluation benchmarking based on a peer universe index a common-sense approach to real estate investment management decisions.¹²

Coverage and Information Considerations in the Ideal Evaluation Benchmark Index

Ideally, simple averages of periodic returns should be used for performance evaluation benchmarking when the agent does not have effective control over the timing of investment capital into, or out of, the investments he manages. This is usually the case in traditional portfolio management. In contrast, the manager should be evaluated on the basis of the (dollar-weighted) IRR he achieves for the client if he does have control over investment timing. In real estate, this is most commonly the case for opportunistic and value-added type funds, including funds involving staged development or “turnaround” projects. For example, a fund that effectively gives the manager a “line of credit” with broad discretion as to when to draw down the line and when to liquidate assets and pay back funds to the investor, should be benchmarked using the IRR achieved by the manager for the client.

Equally-weighting or value-weighting the constituent assets

Whether the constituent assets’ returns should be equally-weighted or value-weighted depends on whether the benchmark index is intended to be a statistical sample or a population census.¹³ Consistent with the ideal agent evaluation benchmark index being based on a peer universe, such indices should generally be viewed as population censuses, consisting of all the peers of the subject manager. This reasoning suggests the ideal benchmark index should usually be value-weighted, measuring the return achieved by the entire portfolio of the subject agent’s peers. However, it may be argued that the population census requirement is extreme in many situations; for example, when the agent is being benchmarked only against average peer performance, and provided that a statistical sample could provide a highly accurate indication of that average. As a practical rule, as large a fraction as possible (ideally at least well over half) of the peer universe population should be included in the benchmark index.

¹² Survivorship bias is always a potential problem in the construction of any index. However, survivorship bias should not be a more serious problem in peer universe based indices than in passive market indices. Any index can avoid the survivorship bias problem by retaining the historical data from funds or assets that have gone out of business. This can be done as a matter of policy in any index. Peer universe based indices built on property level data (such as the NCREIF, IPD, ROZ and PCA Indices, for example) have such policies.

¹³ See Geltner (2000).

Appraised values versus transaction prices

As noted earlier, appraisal-based indices are usually more appropriate than transaction price based indices for evaluation benchmarking. One reason concerns random error. In a typical institutional property portfolio, appraisals occur much more frequently than transactions. The typical property will often be appraised once per year or even more frequently in some cases. In contrast, the holding period for investment property is five years or more. Thus, appraisal data tends to provide larger empirical samples of valuation observations, for a given number of properties and length of historical observation. Thus, appraisal-based indices tend to be less “noisy.”

Benchmarking generally requires the comparison of a benchmark average return to the manager’s portfolio return over a predetermined period of time. A manager’s portfolio is likely to contain a relatively small number of properties; thus, purely random valuation noise tends to present the most serious problem in small samples. Appraisal-based valuations typically reduce random error because the sample is typically larger than would be possible with transaction price data alone. There may also be less random cross-sectional dispersion in appraised values than in transaction prices. Therefore, the use of appraisal-based valuations to compute the manager’s time-weighted return is usually preferred in the benchmarking process. For consistency, appraisal-based valuations should also be used in constructing the agent’s benchmark index.

It is important to note that the benchmarking process is focused on time-weighted mean return across a multi-period historical sample (say, three to five years). The type of measurement error most prevalent in appraisal-based valuations, temporal lagging bias, is precisely the type of error that tends to matter little in a comparison of multi-period time-weighted mean returns. Furthermore, it is likely that the lagging and smoothing will be similar in magnitude between the manager’s portfolio and the benchmark index. If so, the errors will largely cancel out in the benchmarking comparison.

An additional argument for the use of appraisal-based indices in evaluation benchmarking is that an index used to evaluate an agent should be auditable down to the individual property level, rather than based on statistical inference. A formal, direct transaction based price index typically requires the use of regression-based statistical techniques such as hedonic valuation models (HVM) or repeated-measures regression (RMR).¹⁴ With HVMs,

¹⁴ See, for example, Gatzlaff and Geltner (1998), Geltner and Goetzman (2000), Fisher et al. (2006), and Marcato (2005).

there is never any one specification of the model that is clearly “correct,” so no uniquely correct return can be computed for either the agent’s portfolio or the benchmark index. RMR-type transaction-based indices, in contrast, require the properties on which they are based to be sold at least twice over the sample period. However, the vast majority of properties that still remain in the agent’s portfolio or in the benchmark index have typically been sold only once (when the property was acquired). As a result, the data on which transaction price indices are based is only a sample of the underlying portfolio’s properties, and generally not a representative or random sample. This may introduce bias into the transaction-based index that is deemed unacceptable in the context of agent evaluation.

Data requirements of the ideal evaluation benchmark index

The optimal frequency and historical span for the ideal agent evaluation benchmark index is generally different than for the ideal research index. Evaluation benchmarking is oriented primarily toward comparisons of mean returns over moderately long periods, typically 3 to 5 years. Shorter evaluation periods do not provide a long enough sample of time from which to draw valid conclusions about the manager’s performance. Moreover, second moment statistics (such as volatility) have relatively little use in agent evaluation benchmarking. For both of these reasons, high-frequency periodic returns are of relatively little use in evaluation benchmarking.

At a minimum, in order to produce a performance index of periodic total returns, data-contributors from the appropriate peer universe of agents must contribute periodic information regarding asset values and cash flows, as well as detailed property description information (such as type, location, size, age, etc). Particularly at the property-level, such data requirements overlap considerably with the information requirements for an ideal research index.¹⁵

Considerations in the Appropriate Use of Evaluation Benchmark Indices

Four types of considerations can be important in defining an appropriate peer universe, or in understanding the appropriate use of an existing peer universe index:

¹⁵ In addition, it may be argued that the ideal agent evaluation benchmark index should also provide property operational-level performance attribution analysis services. This requires that more detailed property operational-level data be collected, such as information on operating expense and capital expenditure breakdowns, information on leases, and appraisal inputs and assumptions.

- (1) Property segment specialization,
- (2) Risk control,
- (3) Appraisal policy, and
- (4) Fee policy.

Managers can be characterized by these four attributes. Ideally, a given agent should be benchmarked against a peer universe that includes only other agents similar to the subject agent in all four dimensions. For example, managers specializing in apartments should be benchmarked against other managers specializing in apartments.

Property market segment refers to the type and geographic location of the properties the agent is expected to consider for inclusion in the subject portfolio. Real estate market segments are defined largely by property usage types (e.g., office, industrial, retail, apartment, etc) and geographical area (usually metropolitan areas within broader geographic regions). Specialized expertise and resources are typically required to invest in a given property segment, and the acquisition of such expertise and resources involves substantial start-up cost and risk. The principal party generally does not want to encourage his investment agent to consider market segments with which the agent is not already familiar, a result that could occur if the benchmark index included such segments.

Risk control in benchmarking is required to assure that the agent cannot beat his benchmark simply by taking on more risk than is present in the benchmark. In theory, risk can be controlled for by basing the comparison on risk-adjusted return measures (such as a "Treyner Ratio"), rather than the raw returns.¹⁶ In practice, however, such an approach does not work well for private real estate. One problem is that returns must be adjusted for risk in a manner consistent with how the capital market prices risk and this, unfortunately, is not well understood in the case of private real estate. For example, how should we define "beta" within the private real estate asset class, such that beta can be reliably quantified for individual portfolios of properties, and such that expected returns are clearly a positive linear function of such a beta?

A more practical approach to controlling for risk in private real estate markets is to include only portfolios of similar risk to the subject portfolio in the peer universe, and to restrict the agent from pursuing investment strategies that deviate significantly from the investment strategies of the peer universe. In this way, discrete categories of investment attributes and styles are used as proxies for risk categories. For example, leverage limits,

¹⁶ The Treynor ratio is the portfolio's excess return over T-bills divided by its systematic risk as represented, for example, by its "beta."

development project exposure limits, occupancy limits, and style descriptors (such as “opportunistic,” “value-added,” “global,” and “core”) may be used to characterize both the subject agent and his peer universe. In practice, this approach requires considerable specification and elaboration between the agent and the principal party at the outset of the management contract, and often on an on-going basis during that period.

Appraisal policy should also be considered to make sure the subject agent’s policy does not differ from that used to appraise properties in his peer universe benchmark in any way that could significantly bias comparisons. Finally, it is necessary to understand differences in fee policies between the subject manager and the peer universe benchmark in order to adjust for the effect of fees.

Defining appropriate segment or style weights

Suppose a manager has the discretion to allocate the principal’s investment capital across more than one property type or geographic market segment. Suppose further that the peer universe benchmark index includes managers that, taken together, encompass all of the possible segments in which the subject manager might invest. What segment weights should be employed in the manager’s evaluation benchmark index? If the peer universe index can be disaggregated (as, for example, the NCREIF Index by property type and geographical region), then the manager’s benchmark can be defined using the pure-segment sub-indices of the peer universe. The weights should reflect the client’s objectives and the role the subject manager is hired to play in the client’s overall investment strategy.

For example, if the manager is hired to invest in industrial and office properties, a 50/50 weighting of office and industrial property in the manager’s benchmark would be a reasonable weighting to employ in the benchmark index. If the client wishes to skew the manager toward a particular strategy or allocation (for example, if the client thinks that industrial properties are likely to outperform office properties), then the weighting could reflect such an objective (e.g., 75% industrial and 25% office weighting in the benchmark).

Another approach to segment weighting is to attempt to estimate the approximate relative magnitudes of the target market segments, and weight the benchmark proportionately based on these market weights. For example, if there is twice as much office as industrial property in the market (by market value), the benchmark weights might be set at 67% office and 33% industrial. The rationale for this approach is that it may be twice as easy for the manager to find suitable acquisitions in the office market than in the

industrial market. In any case, the aggregate peer universe index is only used to provide the within-segment property performance for the benchmark. Weights agreed upon by the client and manager are used to construct the benchmark index returns based on the peer universe within-segment returns.

Yet a third approach would be to define the benchmark weights as the weights that happen to exist in the peer universe. In this case, the peer universe index would not be customized, but simply used directly as the benchmark index.

It is important to emphasize that it is generally not advisable to set the evaluation benchmark segment weights equal to the weights in the portfolio constructed by the agent for the principal. Such an approach mixes the effect of the agent's within-segment property-level performance with the effect of the agent's portfolio-level allocation decision across segments. This makes it impossible for the agent to either excel or fall short of the benchmark on the basis of her segment allocation decisions. As a result, the agent's performance relative to her benchmark is thus determined purely by her within-segment property-level performance. This removes any incentive the agent would have to improve her performance by means of superior allocation across segments.

Note, however, that if the manager does *not* have discretion over her allocation weights, then it is appropriate to neutralize as far as possible the effect of segment allocation in the manager's relative performance differential with respect to her benchmark. Use of manager segment weights in her benchmark is an effective way to accomplish this objective. However, the magnitude (as opposed to the sign) of the manager's differential performance with respect to her benchmark will still be influenced by the segment allocation weights to which she is subject.

The ideal peer universe and multiple investment styles

There are various sectors or branches of the U.S. private real estate investment industry, including the tax-exempt fiduciary branch, the REIT branch, the insurance company general account branch, other financial institutions, foreign investors, and wealthy private individual investors.¹⁷ Moreover, the investors in each of these sectors have different objectives, styles, and constraints in their real estate investing. Because a basic principle of evaluation benchmarking is that agents be benchmarked against other agents of similar style and specialization, separate benchmark indices

¹⁷ Note that REITs in this context are investors in the private property market, and thus a category of direct property owners.

should generally be constructed for each sector (and perhaps subsector) of the private real estate investment industry.¹⁸

The Downside to Agent Evaluation Benchmarking

Although evaluation benchmarking can be useful when applied carefully, there are dangers. As noted, benchmarking affects the communication between the principal and the agent, and the incentives faced by the agent. Inappropriate benchmarking can cause problems in both of these important functions. Two problems are most common:

Use of an inappropriate evaluation benchmark index

If the benchmark index does not well represent the style or specialization for which the principal has hired the agent, misleading or meaningless comparisons will tend to result. Moreover, perverse incentives may face the agent. For example, if an agent is hired to manage apartment investments because of her specialized expertise in this segment, the aggregate NCREIF Index would be an inappropriate benchmark. The manager might outperform (or underperform) the NCREIF benchmark because of events beyond her control, including the relative performance of apartments versus the other property types in the NCREIF Index). In some cases, a manager may be tempted by an inappropriate benchmark to stray beyond her specialized expertise and competency. A more appropriate benchmark for such a manager would be the NCREIF Apartment sub-index, or a geographical component of that sub-index.

Excessive reliance on quantitative comparisons

There is measurement noise in any benchmark index as well as in any agent's portfolio returns. Thus, there is even more noise in any comparison of the differential between an agent's portfolio and a benchmark. Excessive use of purely quantitative comparisons can have perverse results. For example, management incentive fee structures that provide large, discrete "jumps" in the manager's fee as a function of performance relative to a benchmark index, and incentive fee structures that are asymmetric between the upside and downside, may give the manager an incentive to look for

¹⁸ There are some specialized uses of evaluation benchmark indices for which it would sometimes be appropriate to extend the peer universe beyond one branch or style. For example, property operational level performance attributions and comparisons may be considered to be of interest with respect to all properties of a given type, no matter what type of investor owns the property. Nevertheless, comparisons among styles or types of investment vehicles is more aptly characterized as broad asset class research than specific agent evaluation benchmarking.

excessively risky investments, or may discourage the truly best managers from even wanting to risk working for the client.

Performance measurement difficulties in private real estate do not negate the potential usefulness of evaluation benchmarking. However, they do make evaluation benchmarking a “blunter” tool in the private real estate investment industry than it is in the public securities industry. As discussed in detail by Geltner and Ling (2006), there are three sources of randomness or “noise” in measurements of realized investment returns: true cross-sectional dispersion, true longitudinal volatility, and measurement error. The first two of these exist in the public securities investment industry just as much as (or perhaps more than) in the private real estate investment industry. The third source of randomness, the existence of return measurement errors in private real estate, adds additional noise, and often certain types of bias, that does not exist in public securities benchmarking. This makes the *ex post* use of benchmarking to distinguish superior managers on the basis of their realized performance even more difficult in private real estate than it is in the public securities industry.

Nevertheless, recall that benchmarking serves other purposes that are *ex ante* in nature, and that are probably more important than the *ex post* quantitative comparison function. Neither the benchmark index, nor the agent’s performance, need to be measured perfectly in order for benchmarking to improve communication and incentive alignment between the principal and agent. As long as the agent significantly improves his chances of beating his benchmark (*ex ante*) by adopting behavior that is in the principal party’s best interest, then the benchmarking process will be serving its interest-alignment purpose.

Summary and Conclusion

The primary purpose of the paper is to address the following question: “What are the characteristics of the ideal index for supporting investment decision making and improving the role of the private real estate asset class in the efficient functioning of the overall U.S. capital market?” Geltner and Ling (2000, 2001) concluded that there is not one answer to this question, but two. In the investment and information environment of the 21st century, the real estate investment industry’s needs for performance measurement, research and decision support are too diverse to be optimally met by a single type of index product. In order to realistically and rationally consider optimal index design, we still believe it is essential to distinguish two types of index products: (1) a real estate asset class research index, and (2) an agent evaluation benchmark index.

The primary function of an asset class research index is to improve understanding of the real estate asset class by supporting diverse and fundamental research on private real estate investment performance, carried out in both the academic and industry research communities (i.e., a “CRSP for real estate”). The primary function of an agent evaluation benchmark index is to support performance evaluation the private real estate investment management industry.

Despite the arguments put forth by Geltner (2000) and Geltner and Ling (2000, 2001), both the asset class research index and the agent evaluation benchmark index functions in the U.S. continue to be served and supported primarily by a single index, the NCREIF Index. NCREIF has made significant progress during the last five years by adding operational and valuation data and expanding their database and products beyond those based on fully leased, institutional quality, properties. Nevertheless, progress toward a publicly available index optimized for asset class research has been slow, perhaps because the industry still tends to think in terms of a need for a single, all-encompassing “flagship” index. However, in 2006 the MIT Center for Real Estate launched the publication of their Transactions-Based Index of Institutional Commercial Property Investment Performance (TBI) in cooperation with NCREIF (and based on NCREIF data). As noted, the TBI measures market movements and returns on investment based on transaction prices of properties sold from the NCREIF Index database, and is publicly available.

This paper broadens and deepens the analysis of Geltner and Ling (2000, 2001) of the ideal benchmarking and asset class research indices for private commercial real estate. The purpose of the ideal research index, the optimal depth and breadth of market coverage, and the type of information included in the research index database are discussed in detail. This discussion is followed by an analysis of the ideal agent evaluation benchmark index, including the definition, purpose, and uses of evaluation benchmarking, the characteristics of the ideal benchmark index, market coverage and information considerations, and a comparison of benchmark indices in the private real estate market to comparable indices used to benchmark public securities. This analysis may be especially timely given that, with the publication of MIT’s transaction-based index, researchers now have access to a return index that has been designed specifically to fill some gaps in the data available to researchers. A related paper (Geltner and Ling, 2006) focuses on the technical considerations associated with the design and construction of real estate research indices.

References

- Brounen, Dirk, Piet Eichholtz, and David C. Ling (2007). Trading intensity and real estate performance, *Journal of Real Estate Finance and Economics*, forthcoming.
- Fisher, Jeffrey, David Geltner, and Henry Pollakowski (2006). MIT/CREDL 'Commercial Real Estate Data Laboratory': A Quarterly Transaction-Based Index of Institutional Real Estate Investment Performance and Movements in Supply and Demand, Working Paper, MIT Center for Real Estate, January.
- Gatzlaff, Dean and David Geltner, A. (1998). Transaction-based index of commercial property and its comparison to the NCREIF index, *Real Estate Finance*, 15, 1, 7-22.
- Geltner, D. (2000). Benchmarking manager performance within the private real estate investment industry, *Real Estate Finance*, 17, 1, 23-34.
- Geltner, David and William N. Goetzmann (2000). Two decades of commercial property returns: A repeated-measures regression-based version of the NCREIF index, *Journal of Real Estate Finance and Economics*, 21, 1, 5-21.
- Geltner, David and David C. Ling (2000). Benchmarks & index needs in the U.S. private real estate investment industry: Trying to close the gap. A RERI Study for the Pension Real Estate Association (published by RERI and PREA, at www.reri.org and www.prea.org).
- Geltner, David and David C. Ling (2001). Ideal research and benchmark indexes in private real estate: Some conclusions from the RERI/PREA technical report, *Real Estate Finance*, 17, 4, 1-12.
- Geltner, David and David C. Ling (2006). Considerations in the design and construction of investment real estate research indices, *Journal of Real Estate Research*, 28, 4, 411-444.
- Geltner, David and Norman G. Miller (2001). *Commercial Real Estate Analysis and Investments*, South-Western Publishing.
- Marcato, Gianluca (2005). Repeated measures regression and the construction of real estate indexes with thin information, Working Paper.
- Shapiro, Carl and Hal Varian (1999). *Information Rules*, Harvard Business School Press, Boston.