

**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

By

David Geltner, PhD
University of Cincinnati

David Ling, PhD
University of Florida

Part I:
Executive Report:
How Do We Close the Gap?

Final Report
October 17, 2000

Disclaimer:

All statements and assertions in this report are the opinions of the authors. These opinions are not necessarily shared by the Real Estate Research Institute or the Pension Real Estate Association.

Abstract

This Executive Report represents *Part I* of a two-part report prepared by the authors, at the request of the Real Estate Research Institute (RERI), on behalf of the Pension Real Estate Association (PREA). This Executive Report summarizes and reflects our major conclusions from the Technical Report (*Part II* of the overall report), and also presents our personal suggestions regarding some of the major institutional questions and issues the U.S. private real estate investment industry is currently facing regarding its needs for indexing and benchmarking. The purpose of the overall study has been to survey and bring to the table information that can help the U.S. industry to address the indexing and benchmarking question from a broad perspective and in a rational and informed manner. In preparing this report we have considered:

- What an “*ideal index*” would look like;
- How the major existing indices compare to that ideal; and
- What are some important issues and guidelines in trying to close the gap between the ideal and the existing reality in the U.S. today.

In summary, we argue that the NCREIF Index is ready to evolve into two more specialized successor families of index products: one tailored for fundamental asset class research support, and the other tailored for investment performance evaluation benchmarking and performance attribution. There are a number of challenging and exciting issues regarding how this evolution should occur, and there are likely to be positive and constructive roles for all of the major existing national and international players in this process, as well as perhaps some new actors. In this report we raise some specific considerations that seem important to us in this process, and we suggest some guidelines that we hope will help the industry to move forward. The U.S. industry now faces an exciting and important historic opportunity. We hope this report can play a role in helping the industry to begin to grasp that opportunity.

Contents of the Executive Report

1. Introduction

2. An *Ideal* Index

3. A Gap Analysis of the Existing NCREIF Index Relative to the Ideal

4. Trying to Close the Gap: Some Institutional Considerations

5. Looking for Consensus: A Suggested 10-Plank Platform

Appendix A: “S.W.O.T.” Analysis of the Current U.S. Index & Benchmarking System

Appendix B: Table of Contents of Part II: The Technical Report

Appendix C: Acknowledgments

Appendix D: Biographical Sketches of the Authors

1. Introduction:

With the establishment of NCREIF in 1982, the U.S. real estate investment industry positioned itself as a worldwide leader in investment performance measurement. NCREIF has since made numerous important advancements and contributions. Nevertheless, the NCREIF Index and investment performance measurement in the U.S. has slipped behind the best comparable products and procedures available in some other countries, notably Britain and Australia. However, in the past decade the U.S. has led the world in the development of technological and information advances that hold great potential for moving commercial real estate performance measurement and investment information availability to a new level of usefulness for the industry. In our view, it is an opportune time for the U.S. industry to take a broad and bold look at where it stands, and where it can potentially go, in this new environment.

The RERI/PREA Benchmarks & Index Needs Study has attempted to survey and present information that can help the U.S. industry to begin this process in a rational and informed manner. At PREA's request, the focus of the study has been to consider:

- What an “*ideal*” return index would look like;
- How the major existing indices compare to that ideal; and
- What are some important issues and guidelines in trying to close the gap between the ideal and the existing indices in the U.S. today.

The first two questions are addressed in depth in an accompanying two-volume Technical Report that represents “*Part II*” of the overall RERI/PREA study report. The third question, *how do we close the gap*, is the primary subject of this Executive Report (“*Part I*” of the RERI/PREA study). Although this Executive Report is based on, and summarizes, the Technical Report, it also goes beyond the Technical Report in some respects, to focus more specifically on the broad questions that now face the U.S. private real estate investment industry.

The remainder of this Executive Report is organized into four sections: (1) a brief overview description of an “*ideal index*”; (2) a “*gap analysis*” that considers how the existing NCREIF Index measures up against this ideal; (3) a description of some major institutional considerations and steps involved in trying to *close the gap*; and (4) a concluding section in which we put forth our own personal suggestions for some *action guidelines* the industry might use as a starting point for further discussion. In addition, we include as appendices a brief “*S.W.O.T.*” analysis of real estate indexing in the U.S., and a summary table of contents of the *Part II* report of the RERI/PREA Benchmarks & Index Needs Study.

2. An “*Ideal Index*”

What would an *ideal index* look like, for purposes of 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?

It takes two...

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 believe it is essential to distinguish two types of index products:

- A Real Estate Asset Class Research Index, and
- An Agent Evaluation Benchmark Index.

An ideal research 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

Why does real estate “get no respect”?...
Real estate’s low allocation in the institutional core is often blamed on a “Rodney Dangerfield syndrome”. Real estate suffers from a lack of respect and credibility relative to stocks and bonds. If this is true, then one reason is surely the vast difference in the quantity and quality of data availability between the asset classes. Good data availability does more than just provide short-run benefits for industry research. It attracts academic research, which over the longer run builds up a knowledge base. The existence of such a knowledge base, and the prevalence of asset class research at the most prestigious universities, probably tends in itself to add to the credibility and prestige of an asset class, and to the comfort level of mainstream investors. We believe that commercial real estate is on the verge of being able to command significantly more attention at top universities, as evidenced by a growing number of articles published in top mainstream academic journals, and serious research attention at schools such as MIT, Yale, NYU, Wharton, Michigan, Berkeley and Stanford. But the pressure is strong to improve the academic quality and rigor of commercial real estate databases and index products, because the quality of data in the competing asset classes is constantly improving. If real estate can’t keep up (or ideally even improve its relative standing), it risks slipping farther down the credibility ladder relative to other asset classes. Real estate cannot produce investment performance series of the same nature as the University of Chicago’s CRSP or the NYSE’s TAQ databases, but if commercial real estate had a flagship research index that was based on scientific property samples and transaction price data, and on the type of state-of-the-art statistical procedures that are used by the top econometric scholars at the most prestigious universities, we think real estate would attract more interest and research activity in the academic mainstream, and that this would surely have a long-term beneficial effect for the asset class and the investment industry.

estate investment performance, carried out in both the academic and industry research communities. The research index should support: **(i)** strategic investment analysis; **(ii)** tactical investment analysis; **(iii)** asset market analysis and forecasting; and **(iv)** fundamental asset class research. Examples of the types of questions an ideal research index would support range from questions about the risk and return differences and lead/lag relationships between real estate and stocks and bonds, to questions about liquidity or about linkages between

asset and space markets, to more narrow questions about whether a given asset market “turned around” last quarter, or how far the market has fallen since a given peak (or risen since a given trough).

In supporting this type of broad and basic research function, an ideal asset class research index should be of sufficient technical and statistical quality so that, over the long run, the existence of the index in itself helps to improve the credibility and respectability of the private real estate asset class within the broad investment community. To do this, the ideal real estate research index probably needs to command the interest and respect of the top level of the academic research community. This, in turn, requires that such an index be based on statistical principles and methods that are widely accepted and employed in the academic research community (See box.)

An ideal benchmark index...

The primary function of an agent **evaluation benchmark index** is to support performance evaluation and investment decision-making in the private real estate investment management industry. The benchmark index should allow investors to quantify the historical performance of a given investment agent relative to that of a *peer universe* consisting of essentially all of the competing agents that have a similar specialization and style.

In supporting this function, the ideal benchmark index should also include sufficient depth and detail of portfolio and property operational-level information so as to assist

What makes benchmark indexing a “commercial product”?

A “commercial product” is one that can generate sufficient profit to attract a private firm to produce it without subsidy from a collective entity. We are not sure that benchmarking services are in themselves a commercial product in North America. However, from our discussions with IPD, it seems that a key part of their commercial success in other countries has been the coupling of custom-benchmarking services with extensive property-level performance attribution analysis. Thus, it is likely that the complementary combination of evaluation indexing and performance attribution makes a viable commercial product.

managers and investors in diagnosing and understanding the sources and causes of realized investment performance, what is often referred to as “**performance attribution**”. To accomplish this, the benchmark index should include entire populations of properties and owners that represent effective peer universes relevant for performance evaluation, and it should include property operational-level data such

as leasing and expenditure breakdowns. Moreover, the ideal evaluation benchmark index should reflect valuation and performance measurement procedures that are standardized, auditable, and widely accepted in the industry.

Why we need two index families...

To date, both the asset class research index and the agent evaluation benchmark index functions in the U.S. have been served and supported primarily by a single index, the NCREIF Index. For this reason, perhaps, the industry tends to think in terms of a need for a single, all-encompassing “flagship” index. However, from a technical perspective, a

single all-purpose index product cannot be optimal for both functions. We believe it is time for the NCREIF Index to evolve into two separate families of index products: one focused on the asset class research support role, the other focused on the evaluation benchmarking and performance attribution support role.

The reasons two separate families of index products are needed are both technical and institutional in nature. The technical reason is that the optimal types and sources of raw input data, as well as the optimal index construction methodologies, differ between the two types of indices. Consider a couple of examples of this dichotomy at the technical level. It is important, in principle, for an evaluation benchmark index to include the entire *population* (that is, a “census” or “universe”) of competing investment managers of a given type or style. In contrast, a research index can often make better use of stratified *samples* that are optimized for drawing statistical inferences. A modern research index will often require the use of modern regression-based statistical techniques in index construction, as well as the use of transaction price data (as distinguished from appraised values). Top-level academic researchers in the U.S. will probably always widely shun the use of purely appraisal-based indices. (Transaction-based housing and REIT indices currently receive much more academic research attention than appraisal-based indices. *) On the other hand, real estate investment managers are skeptical of being formally evaluated based on samples and statistical procedures. Because of differences such as these, research indices can often make greater use of public and commercially available data sources (e.g., for transaction and rental price data), whereas evaluation benchmark indices must rely more heavily on proprietary data contributed by a pool of subscribers (e.g., for property-level appraisal and income data).

Some institutional & economic considerations...

It may make sense for different entities or organizations to specialize in the production and dissemination of one type of index or the other. This is made plausible by the technical differences noted above: different types of data sources and index construction methodologies (although there is some overlap in the use of input data).

Separate production is also suggested by the different nature of the roles and usage of the two types of index products within the industry, and the different economics of index production. A top quality research index is essentially a “public good,” in that the long-run collective benefit it provides to the industry is greater than any short-run or private benefits it can provide. As a result, the ideal research index suffers from a serious potential “free rider” problem, and it tends not to be a viable commercial product. It requires subsidized or collective production. In contrast, the growth in demand for formal agent evaluation benchmarking and performance attribution over the past decade has probably created a viable commercial product out of benchmark indices and performance attribution services.

* If commercial property indices are not treated seriously at the top levels of the mainstream of U.S. financial economic research, then it will be difficult for real estate to improve its relative level of credibility and respect within the investment community establishment over the long run.

A natural monopoly...

Indeed, the information service business associated with real estate investment benchmarking probably tends to be a *natural monopoly* product as well. In part, this is due to what economists call *network externalities*. A benchmarking service is more valuable to each user as the number of other users increases. Products that have natural monopolies have the *potential* to generate “monopoly rents,” that is, profits in excess of what is required to attract a private firm to produce them. This has a potentially important implication: it may be possible for the industry to retain some of the monopoly rents associated with benchmark index production. A portion of these monopoly rents (profits) could be used for various purposes, including (for example) to subsidize the production of a better research index for the asset class. Apart from this, users of a natural monopoly product can often benefit from collective and strategic actions to avoid potential long-run problems associated with dependence on a single private supplier, such as “*lock in*” (the difficulty or impossibility for users to switch to an alternative supplier).*

Separate, but coordinated...

Although the ideal world would probably have two separate families of real estate index products, both should be related and well coordinated. Clearly, the uses and functions of research indices and evaluation benchmarks overlap and potentially reinforce each other to a significant degree. For example, both types of index products may be useful in the area of manager or property-owner performance attribution, a largely diagnostic function to help understand why or how a given set of properties performed the way they did in a given interval of time. Furthermore, there is considerable overlap and mutual utility in the input data and the data providers. As a result, the potential benefits from some degree of coordination are significant. This suggests that some sort of joint administration and oversight of the two index families is preferred.

Summarizing the ideal system...

The ideal index system we have been describing is depicted in Exhibit 1 (on the next page). A comprehensive benchmark index family of products, including complementary performance attribution services, would be provided by a benchmark index provider indicated on the right side of the figure. This is a self-supporting commercial service, so we would expect the benchmark index provider to be a private company. Furthermore, this firm would likely obtain a natural monopoly, giving it the potential to earn super-normal profits (monopoly rents). On the left of the exhibit, a research index provider produces the industry’s flagship research index. This provider might also be a private firm. If so, it would presumably be hired by a collective industry body because research index production, in our view, is not itself a viable commercial product.†

* For example, IPD argues that in the U.K., where most of their databank and revenue derives from a small number of large subscribers, it would be easy for these client firms to act in concert, and this provides an effective counter-weight to IPD’s market power.

† The primary role of the research index provider is to produce the research index. However, in the overall system the research index provider can also potentially provide another role. It can act as a *potential*

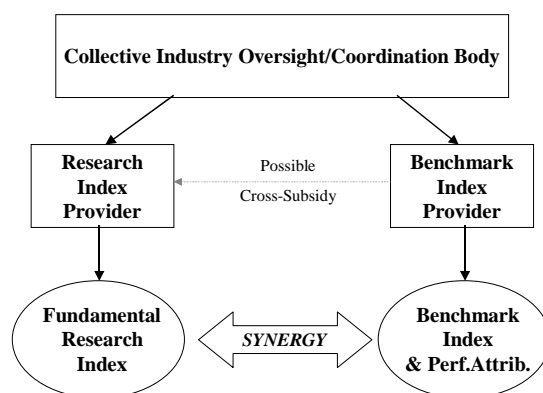


Exhibit 1: *Thoughts on an Ideal Index Production System*

The collective industry oversight and coordination body shown at the top of the exhibit has several functions:

1. On behalf of the industry, it hires the research index provider (or directly provides this function itself, as a “public good”).
2. It assists in the coordination between the research index and the benchmark index, including insuring the appropriate sharing of underlying database elements.
3. It assists the benchmark index provider in obtaining broad industry support and participation in the benchmark index.
4. It provides technical oversight for both families of index products, insuring high standards of technical quality.
5. It provides some political oversight to insure that less directly franchised interest groups, such as the ultimate investor community (e.g., pension plan sponsors) and the academic community, have a sufficient voice and role in the system, and sufficient access to it.
6. It safeguards long-run industry collective interests related to the provision of an essential industry function (benchmarking) by a private monopoly firm (the benchmark index provider). (*Note: See the Technical Report Chapter 2 Appendix B for further elaboration of this issue.*)
7. If appropriate, it provides a mechanism to retain monopoly rents earned in the benchmark index provision process for use as a cross-subsidy to support high quality research index production.

The collective industry oversight & coordination body might use various mechanisms to carry out its functions. For example, it might be important for the collective body to effectively own the proprietary database that underlies the benchmark index. The collective body might use an open RFP and bidding process to engage private firms to

competitor or back-up to the firm that has the monopoly providing the benchmark index service, thereby providing an important redundancy and balance in the system.

provide the research and benchmark index products on a contractual basis for the industry. An analogy that may be appropriate in this regard is the role of the airline industry and some European governments in establishing Airbus as a counterbalance to

What about a “natural duopoly”?

Arguably, there is no one unique “correct” or “best” way to perform evaluation benchmark indexing and complementary performance attribution services, which suggests that there might be a market for two (or even more) competing benchmark firms or entities. An analogy might be the duopoly between S&P and Moody’s in the corporate bond-rating business. However, an important difference is the degree of dependence on a complete peer universe of confidential proprietary data. While this does not rule out a duopolistic equilibrium, it may make it difficult to achieve.

Boeing, avoiding excessive reliance on a single private firm in the commercial jet manufacturing industry, an industry that might otherwise tend to be a natural monopoly. In general, the ideal system we envision retains as a large role as possible for private information firms, with the collective industry body applying as light a hand, and playing as small a role, as possible consistent with its functions noted previously.*

3. A Gap Analysis of the Existing NCREIF Index Relative to the Ideal

Currently in the U.S., the NCREIF Index is the only property-level index that can broadly serve in both the research and benchmarking roles. How close to the ideal is the NCREIF Index? In this section we present a point-by-point “gap analysis” of our perceptions about how the NCREIF Index currently stacks up against the ideal in four important dimensions: market coverage, types of data collected, index methodology, and institutional/constitutional fit.

The NCREIF Index versus the ideal research index...

NCREIF’s non-profit, collective constitution is particularly suited to the relatively non-commercial, “public good” nature of a broad asset class research index. Indeed, NCREIF’s mission statement provides a strong mandate to promote the provision of a top-quality research index for the industry:

“to foster understanding of the real estate asset class through the collection and dissemination of information, performance measurement, standards, research and education.”

* Note that the collective industry body envisioned in the ideal system described here would be broader than NCREIF or any single presently-existing U.S. real estate industry association. Ideally, such an entity would represent all major branches of the U.S. real estate investment industry, including REITs and insurance company general accounts, as well as the tax-exempt fiduciary branch of the industry, and would provide for formal and effective representation of investor and academic constituencies as well as data-contributors. It might be possible for the collective body we are suggesting here to consist of a loose network of user groups and advisory councils, as long as sufficient independence and governing authority were retained to carry out the necessary collective functions.

Nevertheless, the NCREIF Index suffers from technical and institutional limitations that render its design and functioning less than ideal as a research index. We now discuss how the NCREIF Index compares to an ideal asset class research index.

❖ **Market Coverage:** The usefulness of a sample for purposes of making inferences about a population is not measured best by the percentage of the population included in the sample, but rather by the *absolute size* and *representativeness* of the sample. With 2600 properties, the NCREIF Index is a sufficiently large sample at the aggregate level. The major problem with the NCREIF Index in terms of market coverage for a research index is the size of the sample at the detailed market-segment level, and the representativeness of the sample in general. Certain types of properties are not effectively covered, and there are too few properties in some “cells” of a scientifically stratified sample of properties for broad asset class research purposes. In a stratified sample, each property type/geographical market would carry a weight in the sample that reflects its weight in the overall population. For example, if office properties in Orlando Florida constitute X percent of the total market value of all commercial real estate in the U.S., then Orlando office properties should constitute X percent of the sample on a market weighted basis. Construction of a scientific sample would certainly require larger property representation than NCREIF now has, but it would not require a large fraction of the entire property population. As an example, to support scientific sample stratification the PCA in Australia regularly surveys all institutional property holdings (not just those in their index database) in order to provide information useful to determining appropriate weights.

- **Our grade for market coverage for a research index: “B” at the aggregate level, “C” to “F” at the detailed market-segment level.**

❖ **Types & Sources of Data:** The NCREIF Index collects sufficient property-level investment performance information from investors and managers in the tax-exempt fiduciary branch of the real estate investment industry. However, as a research index it has several limitations:

- **Operational-level data:** More property operational-level information would be useful, such as more detailed breakdowns of revenue sources, lease characteristics, operating expenses and capital expenditures.
- **Types of properties:** Expanding the property-level coverage to include additional types of properties within the tax-exempt fiduciary branch of the industry would be useful. Most notably, the inclusion of development projects and properties that are not reappraised annually would add considerably to the richness of the database and allow useful new index products to be produced.
- **Industry coverage:** Expanding beyond the existing tax-exempt fiduciary branch of the industry would probably prove more useful than expanding the coverage within the tax-exempt fiduciary branch of the industry (where coverage probably already exceeds 60%). Including properties owned by REITs, insurance company general accounts, other financial institutions, and some large private and foreign investors would enable the property sample size to be expanded threefold or

- more. Such investors might be willing to contribute data to a well-organized research index, particularly if such contributions did not require them to incur appraisal expenses.
- **Publicly-available information:** Expansion to include public and commercially-available information, such as transaction information (price and volume), occupancy, absorption/leasing, and construction information, would be very valuable for an asset class research index. Transaction price information could probably be used to construct a regularly-updated annual (or perhaps quarterly) transaction price-based “repeat-sales index” of commercial property for the U.S. This would be an international first, and would place commercial property on a par with housing in this regard.
 - **Our overall grade for data type coverage for a research index: “A-“ for appraisal-based investment performance data from the tax-exempt fiduciary branch of the industry; “D” to “F” for other types of data and other branches of the industry.**
- ❖ **Reporting Frequency & Index Construction Methodology:** The NCREIF Index’s quarterly reporting frequency is probably optimal for a research index in the U.S. given data and institutional constraints. However, NCREIF’s index construction methodology is far from ideal for a research index. In particular, NCREIF’s use of both current and stale appraisals every quarter and, indeed, its sole reliance on appraisal-based valuations (without any direct use of transaction price evidence) falls far below what would be ideal in a flagship research index. The current NCREIF Index lacks the credibility and usefulness a research index should ideally have with the top level of the academic research community, in part because of its construction methodology. NCREIF has begun to explore methodological improvements (e.g., its sold-property-based quarterly “Value Trends” report, and its repeated-measures regression-based “Current Value Indicator” - CVI). Nevertheless, the existing NCREIF Index is hampered because it is widely perceived and used both as a research index and as an agent evaluation benchmark index, and this prevents its optimization as a research index.
- **Our grade for reporting frequency and index construction methodology as a research index: “A” for reporting frequency, “C-“ or “D” for index construction methodology.**
- ❖ **Institutional Considerations & Constitutional Structure:** NCREIF’s organizational form and mission are completely consistent with research index production, as the relatively non-commercial, “public good” nature of such indices makes a non-profit, collective entity an appropriate general type of organization to produce or oversee such an index. NCREIF also has developed as a high quality and well-functioning industry technical forum (e.g., its role in developing and overseeing the Real Estate Information Standards, and its informal links with the professional accounting, appraisal, and academic research communities). Nevertheless, an obvious shortcoming is that NCREIF is limited to the tax-exempt fiduciary branch of the

overall real estate investment industry. An asset class research index ideally requires broad market coverage (including, e.g., REITs and insurance company general accounts). Another weakness is that, although it has informal links with academia and industry research communities, NCREIF lacks a formal role for these communities in index oversight.

- **Our grade for institutional considerations and constitutional structure as a research index: “B” within the scope of its traditional function representing only the tax-exempt fiduciary branch of the industry, “D” from a broader perspective.**

In summary, the NCREIF Index is usable as a research index. However, it falls far short of the ideal, and we believe far short of what the industry is capable of producing.

- **Our overall grade for the NCREIF Index as a research index: “C.”**

The NCREIF Index versus the ideal evaluation benchmark index...

NCREIF’s mission statement and the original motivation for the NCREIF Index are oriented toward the production of an asset class research index. However, many of the technical and market-coverage characteristics of the NCREIF Index are closer to optimal for an evaluation benchmark index. What follows is our gap analysis of the NCREIF Index as an evaluation benchmark index.

- ❖ **Market Coverage:** The NCREIF Index currently includes some \$83 billion worth of properties. Although this is a small share of the total U.S. commercial property market, it is a good coverage of the core portfolio investments of the tax-exempt fiduciary branch of the real estate investment industry. For purposes of investment agent performance evaluation benchmarking (as distinct from some broader research purposes), market coverage should be measured against the relevant *peer universe*, defined as the population of real estate investment managers or property owners that compete within a common style or specialization. Thus, we need to break our gap analysis of the NCREIF Index market coverage into three different levels of focus:
 - **Core investments:** The NCREIF Index covers at least 70% of the relevant peer universe, possibly as much as 80%. Although not ideal (the optimum would be 100% coverage), it seems adequate for benchmarking core investments.
 - **Non-core investments:** The official criteria of the NCREIF Index tend to prevent it from including most properties that would be classified as “non-core.” This term refers to development and rehab projects, raw land investments, and other types of properties that would be common in so-called “opportunistic” and “value-added” funds. Also, some managers of such funds are not NCREIF members. As far as we know, nothing would prevent NCREIF from relaxing the criteria for properties to be included in the NCREIF Index, so as to begin collecting data on more non-core properties. Currently NCREIF criteria require

- 60% occupancy and annual reappraisals. (Note that once a property is in the NCREIF Index it generally remains in the index even if it subsequently dips below these criteria.) On the other hand, many non-core investments should not be benchmarked using time-weighted returns, but rather using IRR inception-date cohorts in a manner more comparable to the standard benchmarking practice in venture capital and other private equity asset classes. It should be noted that NCREIF is in the process of purchasing a fund-level database and is apparently developing a fund-level index that should provide some benchmarking capabilities for non-core investments.
- **Performance attribution:** Some operational-level performance attribution analyses relevant for diagnostic purposes ideally require a valid statistical sample of the relevant property market segment, rather than a population census of a given investment peer universe of property owners. Ideal market coverage for such purposes is more akin to that described previously for detailed market segment coverage in a research index.
 - **Our grade for market coverage as a benchmark index: “A-“ for core properties; “F” or “I” (for *incomplete*) for non-core properties (with some remedial credit possible depending on how the fund-level index turns out); and “C” for coverage relevant for performance attribution.**
 - ❖ **Types & Sources of Data:** The NCREIF Index is closer to the ideal evaluation benchmark index than to the ideal research index in the type of data it currently obtains. This is particularly true if we include the fund-level database that NCREIF is purchasing, presuming that NCREIF will successfully implement the collection of fund-level data going forward. Nevertheless, in order to provide more useful performance attribution information, the NCREIF Index should collect more detailed operational-level data (such as rents, occupancy, and lease information, as well as operating expense and capital expenditure breakdowns). It should also expand its coverage to other branches of the investment industry beyond the tax-exempt fiduciary branch.
 - **Our grade for data type coverage as a benchmark index: “A-“ for evaluation benchmarking, “C” for performance attribution.**
 - ❖ **Reporting Frequency & Index Construction Methodology:** The NCREIF Index would better suit evaluation benchmarking purposes if it concentrated on producing an annual end-of-year rather than quarterly index. Nevertheless, the stale appraisal effect is a much less serious problem in an evaluation benchmark index than in a research index. This is because performance evaluation should span relatively long historical periods in which the stale appraisal effect is relatively minor, and because any such effect may largely cancel out in the comparison between the subject agent and the benchmark. Moreover, sole reliance on appraisal-based valuations and failure to use modern statistical techniques to incorporate direct transaction price evidence or reduce temporal lag bias are not, in general, fatal flaws in an evaluation benchmark index (as contrasted to the situation for a research index).

- **Our grade for reporting frequency and index construction methodology as a benchmark index: “A-“ for both frequency and methodology.**
- ❖ **Institutional Considerations & Constitutional Structure:**
 - On the positive side, we can list three strengths of NCREIF as an evaluation benchmark index provider:
 - Its data-contributing members represent the majority of the property holdings of the tax-exempt fiduciary branch of the investment industry, an appropriate peer universe for agent evaluation.
 - To the extent that evaluation benchmark index production is a function that is both essential to the industry and leads to a natural monopoly, there is arguably some need for a vehicle to facilitate collective action by users of such an index product. NCREIF is constitutionally in a position to act as such a vehicle, at least for the fiduciary branch of the industry.
 - NCREIF’s technical strengths in the functioning of its professional committees, as evidenced by its active role in the Real Estate Information Standards, make NCREIF an effective forum for addressing technical issues relevant to evaluation benchmark index construction.
 - On the negative side, we also list three weaknesses:
 - NCREIF’s non-profit, industry association structure is somewhat cumbersome and inefficient for the production of a type of index that is essentially a commercial product.
 - NCREIF lacks a formal and sufficiently powerful role for the investor client community (the pension plan sponsors) in its governance and index oversight functions, even though this is the primary ultimate user group for evaluation benchmark indices.
 - NCREIF represents only one branch of the overall real estate investment industry (the tax-exempt fiduciary branch), even though there are some benchmarking functions and services that can be better served with broader representation.
- **Our grade for institutional considerations and constitutional structure as a benchmark index: “C-“ or “D,” as the weaknesses probably slightly outweigh the strengths at the present time in the U.S.**

To summarize our gap analysis of the NCREIF Index as an evaluation benchmark index, we believe that the NCREIF Index currently provides a very serviceable and useful product for core investment performance evaluation within the tax-exempt fiduciary peer universe. It falls short in supporting property operational-level performance attribution and in failing to go beyond its one constituent branch of the industry. Also, NCREIF should provide a stronger voice to the ultimate users of performance evaluation information, the plan sponsor investor community.

- **Our overall average grade for the NCREIF Index as a benchmark index: “C+” or “B-.”**

Overall summary of the gap...

Overall, NCREIF’s current index and benchmark services provide a valuable service to the industry. However, we feel there is significant room for improvement. It is interesting that, in our view, the NCREIF Index scores better as an evaluation benchmark index than as an asset class research index. As we noted, NCREIF’s mission and non-profit status are arguably more appropriate for the production and promotion of a broad asset class research index. However, NCREIF’s orientation to working only with its own database of proprietary information, and its conservatism regarding use of modern statistical procedures, severely handicap it in the research index role. NCREIF also suffers in both the research and benchmarking roles because it represents only one branch of the real estate investment industry. In any case, the NCREIF Index is strategically hampered by having to serve as both a broad and fundamental research index and also as an evaluation benchmark index, when it is technically impossible to optimize any single index for both of these functions. Clearly, NCREIF has made important advancements and contributions since its founding two decades ago, and it has recently begun to address some of the weaknesses noted above. Nevertheless, from a broader perspective, we believe it is possible for the industry to advance much further.

<i>Gap Analysis of the NCREIF Index: Summary Grades</i>		
Consideration	Grade as Research Index	Grade as Benchmark Index
Market coverage	<i>B</i> aggregate level <i>C to F</i> mkt segment	<i>A-</i> core <i>F(I)</i> non-core <i>C</i> perf.attrib.
Types & sources of data	<i>A-</i> apprsl fiduciary <i>D to F</i> other	<i>A-</i> eval.benchmk <i>C</i> perf.attrib.
Reporting frequency & index construction methodology	<i>A</i> frequency <i>C- to D</i> method	<i>A-</i> frequency <i>A-</i> method
Institutional considerations & constitutional structure	<i>B</i> fiduciary <i>D</i> broader	<i>C- to D</i>
Overall	<i>C</i>	<i>C+ to B-</i>

4. Trying to Close the Gap: Some Institutional Considerations

We argued in Section 2 that an ideal index system would include two separate families of index products, one oriented toward broad and fundamental asset class research, and the other oriented more specifically toward evaluation benchmarking and performance attribution diagnostics. The former tends to be a public good that requires subsidized production. The latter tends to be a commercial product that has the characteristics of a natural monopoly. In Section 3, we suggested that the existing NCREIF Index attempts to serve both functions largely as a public good provided by a non-profit industry association. However, it is not technically possible for a single index product to be optimized for both functions, and the result, though serviceable, is less than satisfactory.

Although technical questions and opportunities are vital in the design and implementation of specific attempts to improve or expand the index system in the U.S., institutional issues can dominate. And, in some respects, institutional questions are more difficult to resolve. To help the industry address such questions in a rational and informed manner, we now focus on the question of industry institutional structure for index service provision. In particular, we consider how index services are currently provided in other countries, and we discuss some of the key issues relevant to industry structure, as we currently perceive them.

Overview: Primary focus on evaluation benchmarking...

Arguably, the main focus of index production systems in all countries is now evaluation benchmarking.* We therefore focus in this section on evaluation benchmark indices (the right-hand branch in the scheme depicted earlier in Exhibit 1 on page 6). However, we also attempt to keep the big picture and long-run objective of the overall ideal system in mind.

Three models of institutional structure...

Full-scale custom benchmarking and performance attribution services are currently provided in several countries outside North America. To our knowledge, in all cases but one these services are provided by one firm, the Investment Property Databank (IPD),

* As we described in Section 3, NCREIF's mission is oriented toward research index promotion, but the actual NCREIF Index is designed and produced closer to the optimal as an evaluation benchmark index than as the type of scientific and statistically sophisticated product that would be necessary to capture the interest and respect of the mainstream of the academic research community in the U.S. The Property Council of Australia produces separate index products, their "Investment Performance Index" (IPI) is oriented more toward asset class research while their "Benchmark Clearinghouse" is oriented more toward evaluation benchmarking. But the IPI is similar to the NCREIF Index in its technical design and construction. The IPD has databases in several countries that might lend themselves to research index production along the lines of what was described in Section 2, but to date IPD index products have been more oriented towards, and optimized for, peer universe-based evaluation benchmarking (and related performance attribution).

headquartered in London.* The exception is Australia, where the benchmark index and performance attribution services are provided by a non-profit industry association, the Property Council of Australia (PCA). A survey of the institutional structures used in these other countries reveals that, at a broad-brush level, there are three different models, as described below.

Model “A”: Direct IPD service provision...

In this model, the IPD and/or its local subsidiary works directly for individual data-contributing subscribers. This is the model originally pioneered by IPD in the U.K., and has more recently been adopted, more or less, in Germany and France. This structure is schematically depicted in Exhibit 2A.

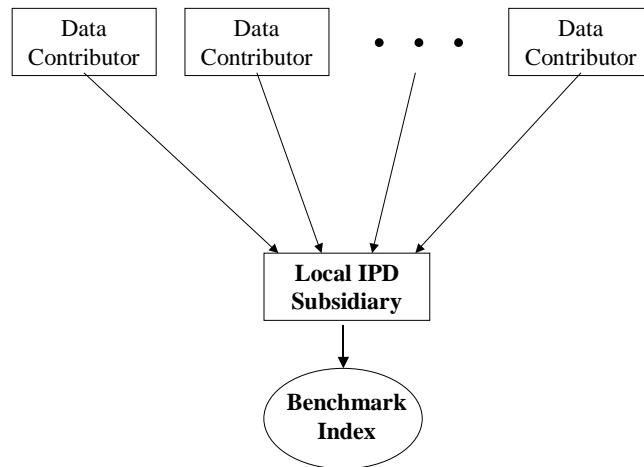


Exhibit 2A: *Institutional Structure “A” (German/French Model)*

* In the past, at least one other firm, Frank Russell Co., has been involved in index production in North America and Australia, and there are closely-related index products provided by several consulting and brokerage firms in a number of countries.

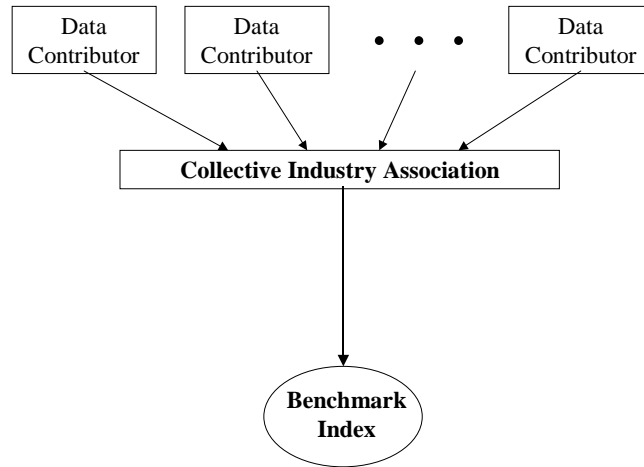


Exhibit 2B: *Institutional Structure “B” (Australian Model)*

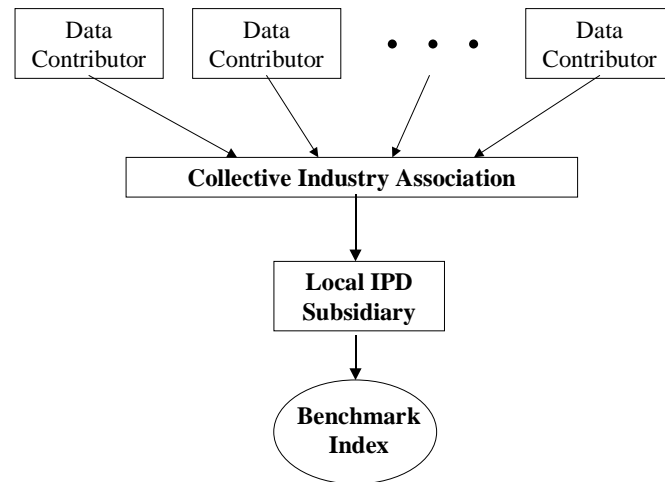


Exhibit 2C: *Institutional Structure “C” (Dutch/Swedish Model)*

***Model “B”:* Provision by an industry association...**

In this model, benchmarking and performance attribution services are provided directly by a non-profit industry association rather than by a private firm. In essence, this is the Australian model, as indicated in Exhibit 2B.

***Model “C”:* Collective industry engagement of IPD ...**

This model incorporates some aspects of both of the two previous models. In this model the local industry collectively engages IPD (or conceivably another private firm) to produce the benchmarking services through a contractual or joint venture arrangement with the local industry association. The local industry association helps IPD to obtain

broad industry participation, and it provides a vehicle to mediate and represent the interests of the local industry with regard to IPD. This model is well represented by the situation in the Netherlands and Sweden, and is depicted schematically in Exhibit 2C (on the preceding page).*

Issues & considerations in deciding on institutional structure...

A number of considerations are important in deciding what general type of industry institutional structure is best for providing benchmark indexing services. The relevant issues can be examined from economic, political, and legal/administrative perspectives. We will not attempt a comprehensive or definitive review of all the relevant considerations and perspectives. However, in the interest of generating discussion, we now highlight several important issues.

- **Oversight/governance issues:** It is important for the industry to think carefully about how best to structure the oversight and governance of benchmark index products and service provision. Although such products tend to be largely commercial in nature, they represent a key element in the efficient functioning of the entire real estate investment industry, and as such have collective ramifications for all participants in that industry. Of particular concern is the need (noted in Section 2) to insure that less directly franchised interest groups and contributors, such as investors and academics, have a sufficient voice in the oversight and governance system. Also important is both the reality and perception that the system is free from excessive influence or control by any one interest group or vested interest.
- **Relevant pre-existing products & structures:** In developing any new system for index service provision it is important to be sensitive to relevant patrimony that may already exist. Clearly, serious costs would be incurred by the industry in any major disruption of the NCREIF index because it is already well-established in the benchmarking and research roles. Moreover, the natural monopoly nature of the evaluation benchmark index product implies that the industry, through NCREIF, has a collective stake in a potentially profitable business. The industry should carefully consider how to best use this stake in any reconfiguration of the overall institutional structure. Another consideration in this regard is the existing institutional “landscape” in the country: What are the various branches of the institutional real estate investment industry, how do the various industry branches and interest groups get along together, and to what extent do existing industry organizations effectively represent the various branches and interest groups.

* It should be noted that in both the Australian model and the Dutch/Swedish model, the relevant collective industry association has considerably broader industry representation than does NCREIF or any single current U.S. real estate industry organization. For example, the PCA in Australia and the ROZ in the Netherlands represent not just pension funds but virtually all major branches of the local institutional real estate investment industry. However, it should be kept in mind that the institutional real estate investment industries in most other countries tend to be smaller and much less fragmented than in the U.S.

- **Appropriate roles for private firms vs collective entities:** As noted in Section 2, custom evaluation benchmark index service provision tends to be a commercial product, at least when coupled with complementary performance attribution services. Thus, it is important to consider what are the appropriate roles for private firms versus non-profit collective bodies in the provision of such services.
- **Research index considerations:** As we described in Section 2, the ideal index system contains a separate research index family of products that are optimized specifically for the broad asset class research role. Because such an index product has many of the characteristics of a public good, including “free-rider” problems, its production will likely require an industry subsidy. This required subsidy has clear implications for index service provision in the industry, most obviously, the need to set up a mechanism to fund and oversee research index production, as well as to coordinate the research index with the evaluation benchmark index at some level.
- **Cultural considerations:** Business institutions operate within national cultural frameworks. Different countries have differing cultures regarding the roles and functions of private firms, collective agencies, government, and so forth. There also may be differences in the way real estate as an asset class, and the real estate investment industry, is perceived by the investment establishment and the mainstream investment community. These considerations should influence the design of the institutional structure put in place for index service provision.
- **International compatibility:** In the modern global investment industry it is helpful for local institutions and procedures to be as compatible as possible with those in other countries.

Summary implications...

Taken together, the above considerations do not suggest that any one model of industry institutional structure clearly dominates. Different structures have various relative strengths and weaknesses, the importance of which will vary across time and across countries. As a result, different institutional structures have been adopted in different countries. In fact, there are three models currently operative outside North America.

Where we come out personally...

As suggested by our discussion in Section 2, we tend to think that Model “C” (the Dutch/Swedish type model that includes major roles for both a collective industry body and a private provider firm) may hold the most promise as a general guide for the current U.S. situation. We conclude this for five reasons:

1. A key objective of index development in the U.S. should be to significantly broaden the coverage of the various branches of the real estate investment industry beyond the tax-exempt fiduciary branch that is currently well represented

in NCREIF. Although the entrepreneurial drive of a private, for-profit, benchmarking firm may be valuable in helping to broaden industry participation, we think a collective industry effort is likely to be required, as this promises to be a difficult project. Furthermore, to the extent that a major obstacle to bringing in additional branches of the industry may be their reluctance to regularly appraise their property holdings, a technical way around this obstacle is presented by the modern regression-based statistical techniques (such as the repeated-measures regression - RMR) that allow use of infrequent appraisals or direct transaction price evidence. But such statistical techniques are more appropriate in a research index than in an evaluation benchmark index, and a research index tends to be more of a “public good” that likely would require subsidized development by a collective entity.

2. Extending the preceding point, we feel it is important that the U.S. real estate industry develop an index optimized for fundamental research purposes, as described in Section 2 (an index that can improve the credibility of real estate research within the academic community). Such an index will likely require subsidized production by the industry, and should be coordinated with an evolving and improving evaluation benchmark index. We think it likely that this type of development will require some sort of broad industry collective oversight and coordination.
3. Unlike the situation in Europe and other countries where IPD has commenced operations, but similar to the situation in Australia, the U.S. already has a well-established index product and supporting institutional structure. It is important that any new developments in the U.S. be sensitive to this reality and patrimony. The NCREIF Index, and NCREIF itself, represent a valuable foundation that can and should be used in the development and construction of a broader and deeper index production system in North America. Although we feel that a broader industry association will ultimately be necessary, NCREIF may provide a useful starting point. In any case it would seem important that NCREIF and its core products should not be inadvertently undermined or destroyed.
4. The political and academic environment in which an index system must operate in the U.S. is particularly rigorous. The real estate and investment industries in the U.S. tend to be more pluralistic and fragmented than in most countries, with a tendency to operate in distinct, relatively aggressive and competitive interest groups (investors, client-consultants, private asset managers of various different styles, publicly-traded firms, and so forth). In addition, the academic community in the U.S. is probably larger, generally more sophisticated, and tends to be more actively involved in the investment industry than in most other countries. Therefore, the academic community is potentially more influential in the long run, either favorably or unfavorably, in raising the credibility and respectability of real estate in the broader investment community. The academic community will tend to look for a system in which accessibility and openness to in-depth scrutiny and constructive criticism is promoted. In such an environment, a private

benchmarking firm could benefit greatly from the support that could be provided by an appropriately broad industry collective association.

5. The concerns noted previously regarding reliance on a single private firm with a natural monopoly to provide essential benchmarking services strongly suggest the need for some collective industry representation and governance.

5. Looking for Consensus: A Suggested 10-Plank Platform

In this concluding section, we suggest a 10-plank platform that we hope will encourage constructive industry discussion, and perhaps even serve as a first draft of the action items required to move the U.S. industry to a new level of index production and benchmarking. We have divided our planks into two categories, technical and institutional, each containing five planks. We first present the technical planks, as they are in some sense more fundamental. We conclude with the institutional planks, as they relate most directly to the initial action steps we believe would be helpful to move the U.S. industry forward.

Technical planks...

Plank #1: “Two ideals”. Real estate indexing/benchmarking information products serve two distinct broad functions: (i) Asset class research (e.g., comparisons between R.E. vs stocks&bonds, asset market analysis & forecasting, fundamental R&D, strategic & tactical investment policy studies, etc); and (ii) Investment performance evaluation of agents (“benchmarking” of agents’ portfolios per se). These two functions, though related, are distinct. As a result, any realistically-achievable index that is “ideal” (or “technically optimal”) for asset class research cannot also be so for agent performance evaluation.

Plank #2: “Two coordinated families of index products”. Given Plank #1, the industry should move toward developing two distinct sets or families of real estate index products: One should be oriented primarily toward fundamental asset class research; the other primarily toward performance evaluation. However, the industry should seek to maximize the mutual relevance and synergy between the two families of products.

Plank #3: “Peer universes are OK”. Appraisal-based peer universe indices are, in principle, valid and useful tools for investment performance evaluation benchmarking in the private real estate investment industry. (*Note: See Section 2.2 in Chapter 2 of the Technical Report for a detailed definition of a “peer universe index”, and an elaboration of the reasoning underlying this plank. It should also be noted that both the NCREIF Index in the U.S. and the IPD Index in the U.K. are examples of peer universe-based indices.*)

Plank #4: “State-of-the-art statistical tools”. Real estate indices that are developed or improved going forward should strive to reflect relevant modern standard statistical

methodologies as these are widely accepted in the academic research community. This should include consideration of such techniques as repeated-measures regression (RMR), mass-level appraisal, and the use of transaction price data as well as individual property-level appraisals, to address such issues as temporal lag bias and return reporting frequency as it relates to valuation frequency. Some of these considerations may be more appropriate for research purposes than for evaluation benchmarking purposes.

Plank #5: “Expand the sample”. The industry should place a high priority on expanding the property universe coverage in any real estate indices going forward. Particular objectives include: (i) Expanding property-level coverage to branches of the commercial real estate equity investment industry beyond the tax-exempt fiduciary branch, such as REITs, insurance companies, and (if possible) other financial institutions, foreign investors, and wealthy individual investors; (ii) Improving the existing coverage of the tax-exempt fiduciary branch of the real estate investment industry to include types of properties not currently included in the NCREIF Index, such as properties appraised less frequently than annually and properties with less than 60% occupancy; and (iii) Possible use of stratified statistical sampling techniques to facilitate the most efficient and effective expansion of universe coverage in any indices.

Institutional planks...

Plank #6: “Broader collective industry organization”. The U.S. private real estate investment industry should seriously consider the establishment of a broad industry "umbrella" association to address the index and benchmarking issues. This association should encompass the tax-exempt fiduciary branch of the industry (as NCREIF and PREA already do), but also the REITs and insurance companies (at a minimum, and possibly also banks, foreign investors and wealthy private individuals). At least at first, such an association might be ad hoc in nature (e.g., a “task force” or “inter-organizational committee”), and it might initially grow out of NCREIF or PREA (or a combination of the two, or alternatively, a private firm might take the initiative), but it should ultimately include representation from a number of other industry organizations. (Examples would include NAREIT, ACLI, AFIRE, ULI, NAIOP, ICSC, NAR, SIOR, The Real Estate Roundtable, and so forth.) The objective of this "umbrella of umbrellas" would be to promote and facilitate the production of high-credibility indexing and benchmarking information products for the U.S. private real estate equity investment industry.

Plank #7: “A formal role for academia”. The U.S. private real estate investment industry should consider creating a formal and permanent role for the academic community in the oversight and direction of the indexing and benchmarking products developed for the industry. As an example, this might take the form of a formal "Advisory Council" or a "Technical Advisory Council" attached to the industry umbrella organization suggested in Plank #6. Membership on such council might be delegated to appointees from such organizations as AREUEA, ARES, or RERI. Industry technical bodies such as AIMR and the Appraisal Institute should also have formal representation on any such councils or boards. A major purpose of a formal academic and technical role would be to insure that state-of-the-art statistical procedures are used in index

construction (see Plank #4). The broader purpose of such a formal academic role and broad technical role would be to insure that private real estate indexing products would have the highest possible level of credibility and respect, thereby helping in the long run to improve real estate's credibility and viability as an asset class within the broader mainstream investment community dominated by the public securities markets.

Plank #8: “*Harness the private sector*”. In the promotion and development of future index products, the U.S. real estate investment industry should seek to tap the energy and entrepreneurship of private, for-profit information companies. Such commercial firms can be useful in the implementation and efficient production and distribution of indexes and/or index-related information products and services. Moreover, they can provide a stimulus for innovation in both broad asset class research index products and manager evaluation and attribution benchmarking products (particularly the latter, probably).

Plank #9: “*No black boxes*”. The U.S. real estate investment industry should guard carefully against the use of, or reliance on, "black boxes". To increase the credibility of the asset class, and to insure continued improvement of the index products, it is vitally important that proprietary production techniques and methodologies used in the development and production of index products be fully available to public and academic scrutiny and criticism (while nevertheless protecting the confidentiality of proprietary data).

Plank #10: “*Manage private monopoly*”. In the use of private for-profit information firms, the industry should guard carefully against becoming excessively reliant on any single firm, so as to avoid problems associated with "lock-in" and natural monopoly. It is important to recognize that some of the key and major index information products of most useful to the real estate investment industry may tend to be natural monopolies, while other such products tend to be "public goods", that is, there is a collective benefit to the industry that outweighs the private benefit of any one firm.

Appendix A:
A “S.W.O.T.” Analysis of the Current U.S. Index & Benchmarking System

Strengths:

- **NCREIF History:** The NCREIF Index provides a sufficiently long and high quality historical time series of private real estate investment performance information to be useful for many research and investment performance evaluation purposes.
- **Technical-level Institutional Infrastructure:** NCREIF, AIMR, the Appraisal Institute, NAREIT, ACLI, and other organizations provide a proven capability to develop and successfully promulgate technical standards and guidelines that improve the quality and reliability of real estate investment performance information, such as the Real Estate Information Standards (REIS).
- **Academic Infrastructure:** The U.S. has a large and advanced academic community actively involved in research relevant to investment in general and real estate in particular, with growing involvement in activities that provide support to the professional investment community, including through organizations such as the Real Estate Research Institute (RERI).
- **Breadth and Depth of Industry Research Talent:** The U.S. has broad and deep industry research community dealing with institutional real estate investment.
- **Scale of Geography and Real Estate Market:** The U.S. has a larger stock of commercial real estate than any other country. This size potentially provides more data, more observations, and higher statistical precision.
- **Scale of Resources:** The sheer magnitude of the U.S. private real estate investment industry provides the *potential* for greater resources to be focused on information problems.

Weaknesses:

- **Industry fragmentation:** The U.S. real estate investment industry is characterized by a relatively large number of firms and distinct branches of the industry, and, relative to the size of the economy, a relatively small penetration of the commercial property market by large-scale investment and financial institutions. This makes it relatively more difficult than in many other countries to organize a large segment of the industry and of the property market to contribute to information pools and adopt standardized information procedures, a vital first step in broad-based index development.

- **Collective Institutional Fractionalization:** Due to industry fragmentation, the U.S. industry is characterized by a large number of institutions and industry associations each representing a relatively small or narrow a segment of the industry or underlying property population. As a result, there is currently no single entity that can encompass and energize all the segments of the industry that are necessary to move the industry forward with regard to investment performance information needs that tend to be “public” or “collective” in nature, such as return indices useful for performance evaluation or research purposes.
- **Relatively little use of formal appraisals:** Related to the previous points, and in comparison to most other Anglo-Saxon countries, the U.S. commercial property industry probably makes less use of formal appraisals. Regular and frequent marking-to-market is practiced in fewer branches of the industry and fewer types of property investment ownership vehicles (or in a smaller share of the overall commercial property market) in the U.S.
- **More complicated properties:** Institutional-size properties in the U.S. tend to be larger, with more tenants, and shorter, less-standardized leases, than in many other countries. This tends to make real estate a less homogeneous product in the U.S., which, in turn, makes precise valuation and appraisal more difficult.
- **Technical Problems in the NCREIF Index:** Reflecting the above-listed weaknesses, the NCREIF Index suffers from several shortcomings, including:
 - small coverage of the underlying property population,
 - lagging and stale valuations in the index,
 - limited property operational level data for performance attribution,
 - lack of fund or manager-level data, and
 - a narrow investment style focus.
- **Lack of Real Estate Credibility in the Mainstream Investment Industry Establishment:** This is a long-standing problem, present in both the professional and academic communities. Real estate is often stereotyped by the mainstream investment community as being a “backwater” that lacks both sufficient quantity and quality of data and efficient centralized access to data. This perception, while based on a kernel of truth, exaggerates the actual degree of real estate investment information analysis and decision-related problems relative to those in the mainstream securities investment industry.
- **Lack of Depth and Intensity of Academic/Industry Interaction:** Relative to the mainstream securities investment industry, both the academic and industry real estate research communities are small, as is the degree of interaction and cross-fertilization, particularly in commercial real estate investment areas. This weakness is simultaneously one cause of, and also in part caused by, the previously-noted weakness of lack of credibility.

Opportunities:

- **The Web & E-Commerce:** The development of the internet and electronic transaction, research, and information services and products opens up new prospects for centralized, easily-accessible data sources relevant to commercial real estate investment performance.
- **Statistical Methodological & Computer Technological Developments:** Since the founding of NCREIF two decades ago statistical methodologies useful for investment performance index construction have advanced dramatically, notably including developments such as the repeated-measures regression (RMR) and related noise-filtering techniques. Similarly, computer hardware and software technological development has been mind-boggling during this period. Together, these two developments offer capabilities for “mass appraisal” and other tools that could be very useful for improving real estate indices, both for manager performance evaluation and for broader research purposes, but especially for the latter.
- **The IPD Initiative:** The interest and energy put forth by the IPD from England in starting up operations in the U.S. brings a possibly unique opportunity. A private, for-profit corporation may have capabilities that non-profit industry associations cannot duplicate. Although any organizational form has weaknesses as well as strengths, the IPD initiative may bring a new type of player and a new type of product that offers useful possibilities.
- **New Openness & Initiatives at NCREIF:** Possibly stimulated in part by potential competition from IPD, NCREIF has moved with more energy and creativity in the past year than in many previous years. NCREIF has developed, at least at the prototype stage and in some cases already in regular production, products that are advanced in some cases beyond anything available in other countries, including NCREIF’s inter-active website for constructing custom-benchmarks, new types of indicators of market direction based on advanced statistical techniques (such as the RMR), and new products based purely on transaction price information or based on fund-level information including non-traditional management styles.

Threats:

- **Self-destruction Due to Narrow Private or Partisan Interests:** Individual firms or industry associations representing parts of the overall industry may see at least short-run benefit in protecting private information or sources of information, or in disseminating negative criticism of private real estate information products without corresponding positive or constructive information helpful to investors and information users. This feeds into and reinforces real estate’s credibility problem with the mainstream investment industry establishment.

- **Ignorance:** Lack of understanding about the nature and capabilities of some of the new information opportunities, especially those that involve new levels of technical sophistication or new forms of electronic interaction, may slow down or even torpedo the progress that could otherwise be made.

Appendix B:
Table of Contents of Part II:
Technical Report

Volume 1: The Ideal Index (and its Ideal Use)

Introduction: Overview and Definition of Index Types

Chapter 1: The Ideal Asset Class Research Index for Private Commercial Real Estate

Chapter 2: The Ideal Agent Evaluation Benchmark Index for Private Commercial Real Estate

Chapter 3: Performance Attribution in Private Real Estate Investment Management

Chapter 4: Property Sampling Issues in the Ideal Research Index

Chapter 5: Basic Asset Valuation and Statistical Considerations for Real Estate Indices

Chapter 6: Property Revaluation Frequency, Index Return Reporting Frequency, and the Ideal Index Construction Methodology

Chapter 7: Benchmarking & Indexing Investment Performance for Real Estate Investments that Are Not Regularly Marked-to-Market

Volume 2: A Review of Some Existing Indexes

Chapter 8: The NCREIF Index in the U.S.

Chapter 9: The IPD Index in the U.K.: Some Issues Relevant to the U.S. Case

Chapter 10: The PCA Index in Australia: Some Issues Relevant to the U.S. Case

Appendix C: Acknowledgments

The RERI/PREA benchmarking study was designed from the outset to be developed using an inclusive, team-based process. The authors would like to thank a number of participants for their valuable input, feedback, and time in this process. Special appreciation goes to the “official” task force that was set up at the outset under RERI auspices to provide guidance throughout the study. This task force included:

Brian Abrams (DuPont Pension Fund Investment)
Jacques Gordon (JL-LaSalle, PREA Research Committee Liaison)
Wiley Greig (RREEF Funds)
Mike Grupe (NAREIT)
Peter Katseff (Tennessee Consolidated Retirement System)
John McClelland (Los Angeles County Employee Retirement Assoc.)
Will McIntosh (Prudential, RERI President)
Tim Riddiough (MIT, RERI Study Task Force Chairman)
John Seckman (Delta Airlines, PREA Plan Sponsor Council Liaison)

In addition, a number of other individuals contributed resource papers or feedback and suggestions on working drafts and components of the study. These included:

Richard Barkham (Grosvenor Ltd)
Blake Eagle (MIT, NCREIF)
Jeff Fisher (Indiana University, NCREIF)
Mary Ellen Grant (State Teachers Retirement System of Ohio)
Aart Hordijk (ROZ – Dutch Real Estate Council)
Susan Hudson-Wilson (Property & Portfolio Research)
Tony Key (IPD)
Stephen Lee (Reading University)
Colin Lizieri (Reading University)
Mike Miles (Real Estate Finance)
Jeanne Murphy (Watson Wyatt Investment Consulting)
Neil Myer (Townsend Group, Cleveland State University)
Rupert Nabarro (IPD)
Douglas Poutasse (AEW Capital Management)
Charles Ward (Reading University)
R. Brian Webb (UBS Brinson, NCREIF President)
Karen Whiley (Property Council of Australia)
Mike Young (RREEF Funds)

Of course, the authors take full responsibility for any errors in this report, which expresses the authors’ opinions, not necessarily shared by any of the contributors noted above.

Appendix D: Biographical Sketches of the Authors

David Geltner is the Greg A. Power Professor of Real Estate in the Finance Department of the College of Business Administration at the University of Cincinnati. He has been teaching at UC at both the undergraduate and graduate levels since 1989. He also serves as a Visiting Professor in the Department of Land Management of the Faculty of Urban and Regional Science at the University of Reading (England), a research appointment. Dr Geltner received his PhD in 1989 from the Massachusetts Institute of Technology. He also has degrees in urban studies from Carnegie-Mellon University and the University of Michigan. He served during 1994-99 as a Managing Editor of Real Estate Finance, and presently is a co-editor of Real Estate Economics (the leading academic real estate journal and the official journal of the American Real Estate & Urban Economics Association). Dr Geltner also serves on the editorial boards of the Journal of Real Estate Finance & Economics, the Journal of Real Estate Research, and the Journal of Property Research. He served during 1994-2000 on the RERI Advisory Board and Board of Directors. Dr Geltner is a Fellow of the Homer Hoyt Institute, and serves as Academic Advisor to NCREIF (a non-paid voluntary position). Dr Geltner is the External Academic Member of the Real Estate Investment Committee of the State Teachers Retirement System of Ohio (a pension plan sponsor with over \$6 billion of directly managed real estate holdings). Dr Geltner has published extensively in leading academic journals in the area of real estate economics, investment analysis, and performance measurement. Dr Geltner is co-author Commercial Real Estate Analysis & Investments (Prentice-Hall, 2001), a new graduate-level real estate investments textbook.

David Ling is the William D. Hussey Professor of Finance and the Director of the Center for Real Estate Studies in the Warrington College of Business at the University of Florida. Professor Ling received an MBA (1977) in finance and a Ph.D. (1984) in finance and real estate from The Ohio State University. Prior to moving to the University of Florida in 1989, Ling was a professor of real estate at Southern Methodist University. Professor Ling is the current President of the American Real Estate & Urban Economics Association (AREUEA), the leading academic real estate professional association in the U.S. Ling is also a Fellow of the Homer Hoyt Institute and a member of the Board of Directors of the Real Estate Research Institute. Professor Ling is a co-editor of Real Estate Economics and serves on numerous other academic journal editorial boards including the Journal of Housing Economics, The Journal of Real Estate Research, and Real Estate Finance. Professor Ling has provided research and consulting services to several state and national organizations including the Federal National Mortgage Association, the National Association of Home Builders, the National Association of Realtors, the Florida Association of Realtors, and the Florida Real Estate Commission. His publications have been in the areas of real estate investments and investment performance measurement indices, REITs and real estate capital markets, residential and commercial mortgage markets, housing markets, home ownership, and tax policy. Dr Ling is a co-author of Real Estate Perspectives: An Introduction to Real Estate, a leading real estate principles text published by Irwin-McGraw-Hill.

**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

By

David Geltner, PhD
University of Cincinnati

David Ling, PhD
University of Florida

**Part II:
Technical Report:
*A Two Volume Review in Q&A Format***

Final Report
October 17, 2000

**Volume 1: The Ideal Index (& its Ideal Use)
Volume 2: A Review of Some Existing Indices**

Disclaimer:

All statements and assertions in this report are the opinions of the authors. These opinions are not necessarily shared by the Real Estate Research Institute or the Pension Real Estate Association.

Chapter contents of both volumes...

Volume 1: The “Ideal Index” (and its Ideal Use)

Introduction: Overview and Definition of Index Types (*p.1*)

Chapter 1: The Ideal Asset Class Research Index for Private Commercial Real Estate (*p.4*)

Chapter 2: The Ideal Agent Evaluation Benchmark Index for Private Commercial Real Estate (*p.17*)

Chapter 3: Performance Attribution in Private Real Estate Investment Management (*p.43*)

Chapter 4: Property Sampling Issues in the Ideal Research Index (*p.49*)

Chapter 5: Basic Asset Valuation and Statistical Considerations for Real Estate Indices (*p. 52*)

Chapter 6: Property Revaluation Frequency, Index Return Reporting Frequency, and the Ideal Index Construction Methodology (*p.59*)

Chapter 7: Benchmarking & Indexing Investment Performance for Real Estate Investments that Are Not Regularly Marked-to-Market (*p. 72*)

Volume 2: A Review of Some Existing Indexes

(Note: Chapter numbers begin with 8 so as not to duplicate question numbers in Volume 1.)

Chapter 8: The NCREIF Index in the U.S. (*p. 80*)

Chapter 9: The IPD Index in the U.K.: Some Issues Relevant to the U.S. Case (*p. 94*)

Chapter 10: The PCA Index in Australia: Some Issues Relevant to the U.S. Case (*p. 105*)

Detailed contents of both volumes...

Volume 1: The Ideal Index (& its Ideal Use)

Introduction:

Overview and Definition of Index Types

- 0.1 What is meant by “indexing” or “benchmarking” in the private real estate investment industry, and what types of index products and information are required? (p.1)
 - 0.2 Why is it necessary to distinguish between evaluation benchmark indices and research indices? (p.2)
 - 0.3 But isn’t there much overlap and synergy between the two types of index products? (p.2)
 - 0.4 How does the NCREIF Index fit into this scheme? (p.3)
 - 0.5 Which types of indices are discussed in this volume, and where? (p.3)
-

Chapter 1:

The Ideal Asset Class Research Index for Private Commercial Real Estate

Section 1.1:

Definitions & Purpose of the Ideal Real Estate Research Index

- 1.1.1 What is the primary purpose of an asset class research index? (p.4)
- 1.1.2 Is the asset class research index also important for maintaining or elevating the credibility of the real estate asset class in the investment community? (p.4)
- 1.1.3 What characteristics are required of a research index to support this credibility-enhancing role for the industry? (p.5)
- 1.1.4 Do these purposes make the ideal real estate research index a “*public good*”, and what are the institutional implications for index production? (p.5)
- 1.1.5 How does the purpose of a research index differ from that of an evaluation benchmark index? (p.5)
- 1.1.6 What is the relationship between a research index and a passive market index? (p.6)
- 1.1.7 What is the relationship between a research index and a peer universe based index? (p.6)
- 1.1.8 Can you provide examples of the types of questions a commercial real estate research index should be designed to help to answer? (p.6)
- 1.1.9 What is the role of the research index in answering such questions? (p.6)

Section 1.2:

Market Coverage

- 1.2.1 What is meant by breadth of market coverage in a research index? (p.7)
- 1.2.2 What is the required breadth of market coverage for the ideal research index? (p.7)
- 1.2.3 What trade-offs and considerations are involved in defining the market coverage of the ideal research index? (p.7)
- 1.2.4 As a starting point, what practical considerations can guide the index market coverage decision? (p.8)
- 1.2.5 How can statistical methodologies of index construction help expand index market coverage? (p.8)

Section 1.3:

The Level & Type of Information Included in the Research Index

- 1.3.1 What types of information should be in the ideal asset class research index, and what is the focus of the present report in this regard? (p.9)
- 1.3.2 What are some general considerations and trade-offs in deciding what information to collect in an ideal research index? (p.9)
- 1.3.3 Should a research index be focused primarily on the property level or on the fund (manager) level? (p.10)
- 1.3.4 Must a good research index report periodic returns, or is an IRR-based index potentially useful for research purposes? (p.10)
- 1.3.5 Must a good research index report total returns? (p.11)
- 1.3.6 Must the property level valuations in the ideal research index be based on transaction prices instead of appraisals? (p.11)
- 1.3.7. How does the answer to the appraisal-vs-transaction price question differ between research indices and evaluation benchmark indices? (p.12)

Section 1.4:

Sampling & Property Composition Issues in the Research Index

- 1.4.1 Is it necessary for a real estate research index to include *all* the properties in the subject market segment? (p.12)
- 1.4.2 Is it necessary for a real estate research index to be limited to a fixed and constant set of properties? (p.12)
- 1.4.3 Should the constituent assets in a research index be equally-weighted or value-weighted in computing the index periodic returns? (p.13)

Section 1.5:

Return Reporting Frequency in the Research Index

- 1.5.1 What is the ideal periodic return reporting frequency for a research index? (p.13)
- 1.5.2 What is the ideal historical span of the ideal research index? (p.14)

Chapter 1 Appendix A:

Research Indices in Private Real Estate and Public Securities: Some Comparisons

- 1.A.1 What is a “*sample-based*” index, and are such indices widely used for research purposes in the public securities industry? (p.14)
- 1.A.2 Why is the S&P500 a good example of a passive market index and a research index in the public securities investment industry? (p.14)
- 1.A.3 How does the S&P500 fall short as a research index? (p.15)
- 1.A.4 Does the lack of a perfect “one-size-fits-all” research or benchmark index distinguish private real estate from the public securities investment environment? (p.16)
- 1.A.5 What would be a conceptual example of a passive market index in private real estate, a conceptual equivalent to the S&P500? (p.16)

Chapter 2:
The Ideal Agent Evaluation Benchmark Index for Private Commercial Real Estate

Section 2.1:
Definitions & Purpose of Evaluation Benchmarking

- 2.1.1 What is meant by an “agent evaluation benchmark index”? (p.17)
- 2.1.2 Who uses evaluation benchmarks, in what role? (p.17)
- 2.1.3 What is the purpose of evaluation benchmarking? (p.18)
- 2.1.4 How is evaluation benchmarking used in an *ex post* manner to try to distinguish *superior* managers? (p.18)

Section 2.2:
The Basis and Fundamental Characteristics of the Ideal Private Real Estate Evaluation Benchmark Index

- 2.2.1 What fundamental criteria define an ideal evaluation benchmark index? (p.19)
- 2.2.2. What is the logical basis for the six fundamental criteria of the ideal performance evaluation benchmark index? (p.20)
- 2.2.3 What is the difference between a peer universe index and a passive market index? (p.21)
- 2.2.4 Which is better as an evaluation benchmark: a peer universe index or a passive market index? (p.21)
- 2.2.5 How do peer universe based indices implement the ideal evaluation benchmark index criteria in the private real estate investment industry? (p.22)
- 2.2.6 What about the problem of survivorship bias in peer universe based indices? (p.24)
- 2.2.7 Are there ever circumstances where a passive market index could be superior to a peer universe based index for evaluating manager performance in the private real estate investment industry? (p.24)

Section 2.3:
Coverage and Information Considerations in the Ideal Evaluation Benchmark Index

- 2.3.1 Should evaluation benchmark indices measure property-level or portfolio-level returns? (p.25)
- 2.3.2 Should portfolio-level benchmark indices measure returns net or gross of fees? (p.25)
- 2.3.3 Should time-weighted returns (HPRs) or money-weighted returns (IRRs) be used in evaluation benchmarking? (p.25)
- 2.3.4 Should the constituent assets in an evaluation benchmark index be equally-weighted or value-weighted in computing the benchmark periodic returns? (p.26)
- 2.3.5 How much of the underlying asset class or market population of properties should be included in a property level evaluation benchmark index? (p.27)
- 2.3.6 How much of the peer universe population should be included in the ideal evaluation benchmark index? (p.27)
- 2.3.7 Why are appraised values usually better than transaction prices for evaluation benchmarking and for constructing evaluation benchmark indices? (p.27)
- 2.3.8 What is the required frequency and historical span of the ideal evaluation benchmark index? (p.29)
- 2.3.9 What type of data should the ideal evaluation benchmark index collect from data-contributors? (p.29)

Section 2.4:

Considerations in the Appropriate Use of Evaluation Benchmark Indices

- 2.4.1 What considerations are important in defining an appropriate peer universe? *(p.30)*
- 2.4.2 What is meant by property market segment specialization and why is it important to control for segments in an appropriate evaluation benchmark index? *(p.30)*
- 2.4.3 How can risk be controlled for in evaluation benchmarking in private real estate investment? *(p.31)*
- 2.4.4 Why is appraisal policy important in evaluation benchmarking? *(p.32)*
- 2.4.5 Why is fee policy important in evaluation benchmarking? *(p.32)*
- 2.4.6 What considerations are important in defining appropriate segment or style weights in the evaluation benchmark index? *(p.32)*
- 2.4.7 What's wrong with defining the evaluation benchmark segment weights to be whatever weights the agent uses in her portfolio? *(p.33)*
- 2.4.8 Is there ever a circumstance where the use of manager weights in the benchmark is appropriate? *(p.34)*
- 2.4.9 Should the ideal peer universe extend beyond a single branch or style of the investment industry? *(p.34)*
- 2.4.10 Is there a potential downside to evaluation benchmarking? *(p.35)*
- 2.4.11 Do performance measurement difficulties in private real estate negate the potential usefulness of evaluation benchmarking? *(p.35)*

Chapter 2 Appendix A:

Evaluation Benchmark Indices in Private Real Estate and Public Securities: Some Comparisons

- 2.A.1 What are the "Bailey Criteria" and how are they related to the six fundamental criteria of the ideal performance evaluation benchmark index? *(p.36)*
- 2.A.2 How do the Bailey Criteria relate to the ideal benchmark index in the private real estate investment industry? *(p.37)*
- 2.A.3 Why is direct passive investment not possible in the private real estate investment industry? *(p.37)*
- 2.A.4 What is the difference between active management in private real estate as compared to public securities investment? *(p.38)*

Chapter 2 Appendix B:

Some Information Economics Considerations Relevant to the Ideal Evaluation Benchmark Index

- 2.B.1 Is an agent evaluation benchmark index a "public good" for the real estate industry, similar to the flagship research index described in Section 1.1 of Chapter 1? *(p.39)*
- 2.B.2 Is an agent evaluation index a "natural monopoly"? *(p.40)*
- 2.B.3 What about a "natural duopoly"? *(p.40)*
- 2.B.4 If the ideal evaluation benchmark index is a natural monopoly, does this imply that it should be produced by an industry association rather than a private firm? *(p.41)*
- 2.B.5 What are the two phases of production, and two types of finished products, in the evaluation benchmark index business? *(p.41)*
- 2.B.6 Is vertical integration technically necessary between the underlying database compilation process and the finished information products? *(p.42)*

Chapter 3:
Performance Attribution in Private Real Estate Investment Management

- 3.1 What is meant by “*performance attribution*” in private real estate investment management? (p.43)
- 3.2 Who uses performance attribution, and for what purpose? (p.43)
- 3.3 What are the two major levels of performance attribution? (p.43)
- 3.4 What is the relation between performance attribution and either the ideal research index or the ideal evaluation benchmark index? (p.44)
- 3.5 Is performance attribution analysis as an information product a “public good” or a commercial type of product? (p.44)
- 3.6 How is performance attribution done in general? (p.44)
- 3.7 How is performance attribution done at the portfolio level? (p.45)
- 3.8 How is performance attribution done at the property level? (p.45)
- 3.9 Is it necessary that performance attributions components all add up to the total performance differential at each level? (p.46)
- 3.10 How useful is performance attribution? (p.47)
- 3.11 What are the technical questions about the usefulness of performance attribution? (p.47)
- 3.12 How is performance attribution more useful from an informal, management perspective? (p.48)

Chapter 4:
Property Sampling Issues in the Ideal Research Index

- 4.1 What determines the usefulness of a sample for purposes of defining a real estate investment research index or database? (p.49)
- 4.2 What are the implications of sampling theory for efficient sampling of large, diverse populations? (p.49)
- 4.3 What are the implications of scientific sampling theory for an ideal research index of the U.S. commercial property market? (p.49)
- 4.4 How big a property sample would be necessary for a passive market index of U.S. commercial property investment performance? (p.50)
- 4.5 Can you give a simplified numerical example of how the necessary cell sample size is determined for a return index useful for research purposes? (p.50)
- 4.6 Is sampling theory also relevant for evaluation benchmark indices, in addition to research indices? (p.51)

Chapter 5:
Basic Asset Valuation and Statistical Considerations for Real Estate Indices

- 5.1 What is the conceptual difference between transaction price, market value, and appraised value? (p.52)
- 5.2 What is meant by transaction price dispersion? (p.52)
- 5.3 What is the difference between appraisal error and transaction price error? (p.53)
- 5.4 Is the evidence for temporal lag bias in appraisals uncontroversial and widely accepted? (p.53)
- 5.5 Can you give a simple numerical example of why or how appraisal temporal lag bias occurs? (p.54)

- 5.6 Why does there tend to be a trade-off between random error and temporal lag bias in property valuation estimation? (p.55)
- 5.7 What are the implications of these different types of errors (appraisal error vs transaction price error) for indices based on different types of valuations and used for different purposes (research vs evaluation benchmarking)? (p.55)
- 5.8 Can you give a numerical example of the difference between transaction price, market value, and appraised value, showing the nature of purely random cross-sectional valuation “error”? (p.56)
- 5.9 What is the relative importance of temporal lag bias vs random error in an asset class research index and in an evaluation benchmark index? (p.57)
- 5.10 Are there other sources of randomness in real estate indices besides measurement errors? (p.57)
- 5.11 Why is randomness of any source an important consideration particularly in evaluation benchmarking? (p.57)
- 5.12 Can you elaborate on the nature of the three sources of randomness in *ex post* return measurement in real estate? (p.58)

Chapter 6:
Property Revaluation Frequency, Index Return Reporting Frequency, and the Ideal Index Construction Methodology

Section 6.1:
Overview and Basic Definitions

- 6.1.1 What determines the statistical qualities of a real estate return index? (p.59)
- 6.1.2 What is meant by index return reporting frequency? (p.59)
- 6.1.3 What is meant by property revaluation frequency? (p.59)
- 6.1.4 What is meant by number of properties in the index? (p.60)
- 6.1.5 What is meant by index construction methodology? (p.60)
- 6.1.6 In general, how do these attributes affect the dynamic statistical quality of the index? (p.60)
- 6.1.7 How does the index frequency and historical span relate to the statistical power of the index? (p.61)
- 6.1.8 How do the statistical quality determinants interact in an appraisal-based index? (p.61)
- 6.1.9 How do the index statistical quality determinants interact in a pure transaction-based index? (p.61)
- 6.1.10 Is this chapter about appraisal-based indices only? (p.62)

Section 6.2:
Index Reporting Frequency and Property Revaluation Frequency: The Basics for Appraisal-based Indices

- 6.2.1 Must the index return reporting frequency equal the property revaluation frequency of the constituent properties in the appraisal-based index? (p.62)
- 6.2.2 What are the major trade-offs and considerations regarding the relationship between index return reporting frequency and property-level reappraisal frequency? (p.62)
- 6.2.3 But why can't the property-level data-contributors update their property valuation reports between major annual reappraisals? (p.63)
- 6.2.4 How expensive is it to conduct temporally independent property-level appraisals more frequently than once per year? (p.63)
- 6.2.5 Why can't the problem of stale appraisals be solved by just requiring all index data-contributors to fully reappraise all their properties every period? (p.64)

Section 6.3:

Modern Index Statistical Tools and Basic Appraisal Technology & Economics

- 6.3.1 What is the nature of scale economies in appraisals for index construction? (p.65)
- 6.3.2 Are there scale economies in increasing property-level appraisal frequency, and what are the implications for research indices vs evaluation benchmark indices? (p.65)
- 6.3.3 Are there no economies of scale in appraisal that can be useful in research index construction? What about “*mass appraisal*”? (p.65)
- 6.3.4 Is mass appraisal used in practice? (p.65)
- 6.3.5 How does mass appraisal work? (p.66)
- 6.3.6 Does mass appraisal require formal statistical procedures such as regression analysis? (p.66)
- 6.3.7 Is it possible to produce an index whose return reporting frequency is greater than its constituent properties’ reappraisal frequencies without suffering from the lagging caused by the stale appraisal effect? (p.66)
- 6.3.8 How does the RMR work, and is it useful for transaction based indices? (p.67)
- 6.3.9 Are there any particular considerations in the use of the RMR procedure in pure transaction price based indices? (p.67)

Section 6.4:

Practical Considerations for Index Construction With Modern Statistical Tools

- 6.4.1 Will the investment industry accept an index that is based on a *regression model*? (p.67)
- 6.4.2 Are there any major technical problems in regression-based index construction methodologies? (p.68)
- 6.4.3 Why don’t traditional appraisal-based indices suffer from the “noise” problem? (p.68)
- 6.4.4 Can anything be done about the noise problem in regression-based real estate research indices? (p.69)
- 6.4.5 Can regression-based research indices suffer from stale appraisals or temporal lag bias? (p.69)

Section 6.5:

Use of Less Formal Techniques in Index Construction

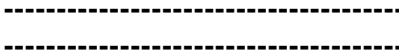
- 6.5.1 Are there other methods besides statistical regression to allow an index to have greater return reporting frequency than its average constituent property revaluation cycle? (p.70)
- 6.5.2 How does such a simple, common sense sampling approach compare to more formal regression-based procedures such as the RMR? (p.70)
- 6.5.3 What about a more common sense approach that tries to use all the relevant information? (p.70)
- 6.5.4 What are the pros and cons of the ad hoc informal approach as compared to the formal regression-based approach? (p.71)
- 6.5.5 Are the two approaches, regression and informal analysis, mutually exclusive? (p.71)

Chapter 7:

Benchmarking & Indexing Investment Performance for Real Estate Investments that Are Not Regularly Marked-to-Market

- 7.1 What is meant by “marking-to-market”? (p.72)
- 7.2 How does real estate differ from other private investments in marking-to-market? (p.72)
- 7.3 How does real estate differ from public securities investments in marking-to-market? (p.72)

- 7.4 Why is marking-to-market important in benchmarking and indexing investment performance? (p.72)
- 7.5 Why is the calculation of TWRRs important for an investment asset class? (p.73)
- 7.6 What is the difference between the time-weighted return and the IRR? (p.73)
- 7.7 Can you give a numerical example of the difference between the TWRR and the IRR? (p.74)
- 7.8 Which real estate investments are marked-to-market regularly? (p.75)
- 7.9 Which real estate investments are *not* marked-to-market regularly? (p.75)
- 7.10 What are the implications for expanding market coverage of commercial property indices? (p.75)
- 7.11 What about “*opportunistic*” and “*value-added*” real estate funds? (p.75)
- 7.12 What is the standard procedure for benchmarking private investments that are not regularly marked-to-market? (p.76)
- 7.13 How does IRR-based benchmarking work? (p.76)
- 7.14 How is the IRR of the benchmark inception cohort calculated? (p.76)
- 7.15 How is the value of residual assets handled in the calculation of since-inception IRRs? (p.77)
- 7.16 Are there any major concerns regarding residual asset valuation in the IRR-based benchmarking of fund performance? (p.77)
- 7.17 Is it valid to combine IRR-based and TWRR-based returns in an all-encompassing index that represents both opportunistic funds and traditional core investments? (p.77)
- 7.18 What is the benchmarking standard in the venture capital industry? (p.78)



Volume 2: A Review of Some Existing Indices

(Note: Chapter numbers begin with 8 so as not to duplicate question numbers in Volume 1.)

Chapter 8: The NCREIF Index in the U.S.

Section 8.1: Institutional Overview and Setting

- 8.1.1 What types of real estate investment performance index and benchmarking products are currently available in the U.S.? (p.80)
- 8.1.2 What has been the role of the NCREIF Index, and what types of information products and functions have been the traditional primary focus of NCREIF? (p.80)
- 8.1.3 What is a fund-level index, and does NCREIF produce one? (p.81)
- 8.1.4 Does NCREIF provide customized benchmarking and performance attribution analyses for its members? (p.82)
- 8.1.5 Is the NCREIF Index primarily a manager evaluation benchmark or a broad research index? (p.82)
- 8.1.6 Why does the NPI come closer to the ideal as an agent evaluation benchmark than as a broad asset class research index? (p.82)

Section 8.2: Technical Characteristics of the NCREIF Index, and the Appraisal Lag Problem

- 8.2.1 How is the NCREIF Property Index constructed? (p.83)
- 8.2.2 How frequently are NCREIF properties reappraised? (p.83)
- 8.2.3 How seasonal is the reappraisal of NCREIF properties? (p.83)
- 8.2.4 What is the effect of this procedure on the dynamic statistical quality of the index? (p.84)

- 8.2.5 Are the underlying property-level appraisals, even the “serious” ones, also lagged? (p.84)
- 8.2.6 What is the nature of the lag in the NCREIF Index? (p.84)
- 8.2.7 How problematical is this lag? (p.85)
- 8.2.8 Is NCREIF doing anything about the lag problem? (p.85)

Section 8.3:

Gap Analysis of the NCREIF Index vs the Ideal Index

- 8.3.1 How does the NCREIF Index compare to the ideal *research index* regarding property market coverage? (p.86)
- 8.3.2 How does the NCREIF Index compare to the ideal *evaluation benchmark index* regarding property market coverage? (p.86)
- 8.3.3 Why does NCREIF not have 100% coverage of the core component of the tax-exempt fiduciary branch of the industry? (p.87)
- 8.3.4 How does the NCREIF Index compare to the ideal *research index* regarding types and sources of data collected in the underlying database? (p.88)
- 8.3.5 How does the NCREIF Index compare to the ideal *evaluation benchmark index* regarding types and sources of data collected in the underlying database? (p.89)
- 8.3.6: How does the NCREIF Index compare to the ideal *research index* regarding index construction methodology & reporting frequency? (p.90)
- 8.3.7 How does the NCREIF Index compare to the ideal *evaluation benchmark index* regarding index construction methodology & reporting frequency? (p.91)
- 8.3.8 What is NCREIF’s constitutional structure, and how does this relate to its most appropriate role in index production? (p.91)

Chapter 9:

The IPD Index in the U.K.: Some Issues Relevant to the U.S. Case

Section 9.1:

Institutional Overview & Setting

- 9.1.1 What types of real estate investment performance index and benchmarking products are currently available in the U.K.? (p.94)
- 9.1.2 What has been the role of the IPD Index, and what types of information products and functions have been the traditional primary focus of IPD? (p.95)
- 9.1.3 How does the traditional major information source and product focus differ between IPD and NCREIF? (p.95)
- 9.1.4 What specific additional information does IPD include beyond what NCREIF includes? (p.96)
- 9.1.5 Is IPD’s methodology and performance measurement consistent with international state-of-the-art practice? (p.96)
- 9.1.6 How is the natural monopoly role of IPD addressed? (p.97)

Section 9.2:**Some Technical Characteristics of the IPD Index, with Some Comparisons to the NCREIF Index**

- 9.2.1 What indices does IPD currently produce in the U.K.? (p.98)
- 9.2.2 What is the peer universe and market coverage of the IPD indices in the U.K.? (p.98)
- 9.2.3 Are there institutional differences between the U.K. and U.S. that affect the difficulty of obtaining broad market coverage for a benchmark index? (p.100)
- 9.2.3 Do IPD indices in the UK suffer from temporal lag bias? (p.101)
- 9.2.4 Is there a difference in temporal lag between the IPD Annual and Monthly indices? (p.102)
- 9.2.5 Are IPD indices less temporally lagged than other British private real estate indices? (p.102)
- 9.2.6 What are the causes of the difference in temporal lag bias between the U.K. indices and the NCREIF Index in the U.S.? (p.103)

Chapter 10:***The PCA Index in Australia: Some Issues Relevant to the U.S. Case***

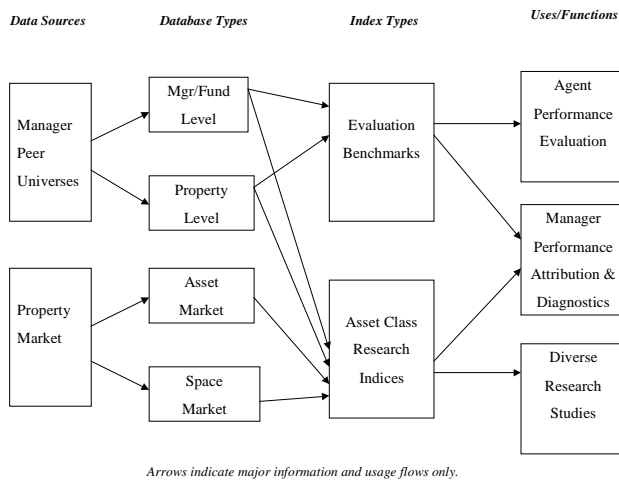
- 10.1 What is the role of the PCA Index, and what types of information products and functions does the PCA provide? (p.105)
- 10.2 What is the PCA Index's market coverage? (p.105)
- 10.3 How does the PCA Index compare to the NCREIF Index in its index construction procedures? (p.106)
- 10.4 How does the Benchmark Clearinghouse work? (p.106)
- 10.5 Are there any unique and interesting features of the PCA benchmarking methodology and information products? (p.107)
- 10.6 What data does the PCA collect directly from the space and asset markets? (p.107)

Volume 1
The Ideal Index
(& its Ideal Use)

**Introduction:
Overview and Definition of Index Types**

0.1 What is meant by “indexing” or “benchmarking” in the private real estate investment industry, and what types of index products and information are required?

The term “benchmarking” has a rather broad (and evolving) meaning, and is used in common parlance in the real estate investment industry to refer to several related, but distinct, functions. Similarly, the concept of a private real estate asset class investment performance “index” can refer to several different types of indices, used in different ways. An overview of private real estate investment information flows and products, as these relate to benchmarking and indexing, is presented in the diagram below.



In the right-hand column of the diagram, three major types of uses or functions of benchmarking and real estate indices are identified: (i) investment agent performance *evaluation*, (ii) investment agent and asset manager performance *attribution* and diagnostics, and (iii) a broader (and potentially deeper) category of uses referred to as diverse *research* studies (including, for example, asset market analysis and investment strategic and tactical research).

Each of these three types of uses or functions is ideally supported by one or more appropriate types of index products. For practical purposes, we think it makes sense to distinguish between two different types or categories of index products:

- (i) Asset class *research indices*, and
- (ii) Agent *evaluation benchmark indices*.

These two types of indices, in turn, are based on and derived from databases, which may include types of information in addition to what is used to produce the relevant investment performance index. Four types of databases are identified, though an agent evaluation benchmark can be produced using essentially only property-level and manager-level data. Finally, in the left-hand column of the diagram, two major types of data sources are identified: manager peer universes, and the underlying property market itself. The former consist of the investment management firms that own or manage property investments, typically in a fiduciary capacity, on behalf of investor clients. These firms own or manage vast amounts of raw data relevant to property and portfolio investment performance. But such firms do not own or manage all the properties in the relevant property asset markets, and they do not possess all of the relevant data. Basic information about property asset market transactions and prices, and space market rents and occupancy levels, may often be publicly available or collected by other types of data vending or brokerage firms. This is why the second box is indicated in the left-hand column.

0.2 Why is it necessary to distinguish between evaluation benchmark indices and research indices?

There are two reasons. First, ideal evaluation benchmark indices differ in important technical respects from ideal research indices. This should not be surprising, since their uses and functions are quite different. As a result, the optimal types and sources of raw input data, as well as the optimal index construction methodologies, typically differ between the two types of indices. For example, it is relatively more important for an evaluation benchmark index to include the entire *population* (aka a “census” or “universe”) of competing investment managers of a given type or style. In contrast, a research index is often better constructed using stratified *samples* optimized for drawing statistical inferences about questions that are of more research interest than evaluation interest (for example, questions about real estate market movements, volatility, or correlations). Similarly, the use of transaction price data as distinguished from appraised values, and the use of modern regression-based statistical techniques in index construction, will typically be more appropriate in research indices than in the agent evaluation benchmark role. Because trade-offs exist in the construction of any index product, it is generally impossible to technically *optimize* a single index product simultaneously for both the agent evaluation function and the broader and more diverse function of real estate asset class and asset market research. Thus, from a technical perspective we need to consider each type of index product separately if we are going to effectively describe the “ideal”.

A second reason why it makes sense to distinguish between the two types of indices is that it may be useful to consider separate production of these two types of index products. That is, it may make sense for different entities or organizations to specialize in the production and dissemination of one type of index or the other. This is made plausible by the technical differences noted above: different types of data sources and index construction methodologies (although there is some overlap in the use of input data). Separate production is also suggested by the different nature of the roles and usage of the two types of index products within the industry, and the different economics of index production. Research indices have many of the characteristics of a “public good” in that the long-run collective benefit they provide to the industry as a whole is greater than the short-run or private benefit they provide to any one firm. As a result, the ideal research index tends not to be a very good commercial product. It requires subsidized or collective production. In contrast, the growth in demand for formal agent evaluation benchmarking and performance attribution over the past decade has possibly made the production of agent evaluation benchmark indices and complementary performance attribution diagnostic services viable as a commercial product.

0.3 But isn't there much overlap and synergy between the two types of index products?

Yes. The uses and functions of research indices and evaluation benchmarks do overlap and potentially reinforce each other to a significant degree. For example, both types of index products may be useful in the area of manager or property-owner performance attribution, a largely diagnostic function to help understand why or how a given set of properties performed the way they did in a given interval of time. Furthermore, there is considerable overlap and mutual utility in the input data and the data providers. As a result, there is considerable potential

synergy between the two types of index products, and considerable potential benefit from some degree of coordination.

0.4 How does the NCREIF Index fit into this scheme?

Currently in the U.S., a single index product, the NCREIF Index, is used for both the agent evaluation and asset class research functions. However, the NCREIF Index is not optimized for either function. It was originally established as a (sort of) public good by collective industry action with the primary motivation being to foster asset class research. NCREIF's mission statement reflects this goal, which is closer to the purpose of what we are here calling the ideal research index than it is to the purpose of the ideal evaluation benchmark index. Nevertheless, during the 1990s the use of the NCREIF Index by the investment industry in the U.S. has become more heavily focused on agent evaluation benchmarking. Moreover, the technical characteristics of the NCREIF Index, the type of data and index construction methodology it employs, happen to be closer to the ideal for an agent evaluation benchmark index than for an asset class research index. But the NCREIF Index is not ideal as an evaluation benchmark either, because it is missing some key components of the evaluation function, such as fund-level data and much of the data necessary for property operational level performance attribution.

From a technical perspective, we believe that, in the more sophisticated information environment of the 21st century, it would be optimal for the NCREIF Index to evolve into two families of index products, separate, but coordinated: one focused primarily on, and optimized for, agent evaluation and performance attribution, and the other focused primarily on, and optimized for, broader asset class research purposes consistent with NCREIF's original mission.

0.5 Which types of indices are discussed in this volume, and where?

Both research indices and agent evaluation benchmarks are addressed. However, because there is overlap and mutual relevance in many issues, some questions in each part of this volume will be relevant to both types of indices. Nevertheless, in terms of major focus, Chapter 1 presents an overview of research indices, while Chapter 2 focuses on evaluation benchmarks. Subsequent chapters treat more technical topics in more depth, and are often relevant to both types of indices.

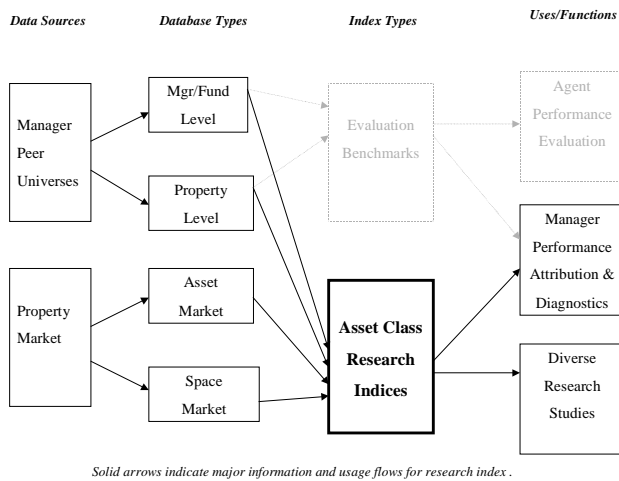
Chapter 1:
The Ideal Asset Class Research Index
for Private Commercial Real Estate

Section 1.1:

Definitions & Purpose of the Ideal Real Estate Research Index

1.1.1 What is the primary purpose of an asset class research index?

The primary purpose of a private real estate asset class research index is to support research relevant to improving understanding and decision-making relating to direct investment in commercial property. As indicated in the figure below, research indices are useful both for supporting diverse and fundamental asset class research studies and, in some cases, for supporting at least some types of performance attribution analyses useful for investment



performance diagnostic purposes. 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, than to speak only in terms of a single “research index”. Nevertheless, in the present context, we are focusing particularly on the index-oriented applications of such a database, so we shall use the term “research index” widely in this volume, keeping in mind that this may encompass a “family” of indices and index-related

information products derived from a supporting database.

1.1.2 Is the asset class research index also important for maintaining or elevating the credibility of the real estate asset class in the investment community?

Yes. *If the index is of sufficient quality*, it can help to improve the *image* and *investment utility* of real estate as an asset class within the broader investment community. By increasing the quantity and quality of information and understanding about real estate as an investment, a high-quality asset class research index will tend, over the long run, to increase the *credibility* of direct investment in privately-traded commercial real estate as a “mainstream” investment.

1.1.3 What characteristics are required of a research index to support this credibility-enhancing role for the industry?

A “flagship” research index must command and retain the respect of the mainstream academic research community, in addition to the industry research community. To succeed in this regard, a research index must be:

- Based on data sources and index construction methodologies that reflect the best practice state-of-the-art for academic and industry statistical research;
- Compiled in an unbiased, neutral manner not susceptible to manipulation by any interested party;
- Compiled in an open manner subject to public and academic scrutiny and criticism;
- Accessible widely to industry and academic researchers at reasonable cost.

1.1.4 Do these purposes make the ideal real estate research index a “public good”, and what are the institutional implications for index production?

If a commercial real estate research index accomplishes the two purposes noted above (fundamental research support and asset class credibility enhancement), it may be considered a “public good”. This means that such an index provides a collective benefit to the entire real estate investment industry, indeed, to the entire broad investment industry (not just the real estate community). This represents a contribution to the efficient functioning of the overall capital market. This contribution exceeds the aggregate benefit to the subscribing or sponsoring firms or organizations directly involved in the production and distribution of the index. This public good characteristic does not necessarily imply that the ideal research index must be produced by a public or collective institution. It is possible that a private index-producing firm might be perceived as more independent, in some respects, than an industry association. However, the public good aspect of a flagship research index does suggest that the industry has a major collective stake in providing input and helping to oversee the quality of any such index. Based on the considerations in Question 1.1.3, it is important that any such input and oversight include strong representation from the investor and research communities within the industry.

1.1.5 How does the purpose of a research index differ from that of an evaluation benchmark index?

As elaborated upon in Chapter 2, an agent *evaluation benchmark index* is one that is used specifically for purposes of evaluating the investment performance of managers by comparing the manager’s total return to an appropriately defined benchmark index’s total return over a moderately long (preferably at least 3-5 year) time period.

In contrast, an asset class *research index* in the private real estate investment industry serves a variety of specific uses, and is explicitly not necessarily optimized for agent evaluation benchmarking purposes. This is not to say that the research index may never be used in such evaluation, but that it is not designed with that purpose uniquely or primarily in mind.

1.1.6 What is the relationship between a research index and a passive market index?

The concept of a passive market index refers generally to an index that tracks a relatively static set of assets within a given asset market segment or style. An ideal passive market index in real estate would be a property-level index consisting of a fixed constant set of properties representing an entire property asset market segment or group of segments. This is generally the type of index that is most appropriate for research purposes at the property level. However, for research purposes adequate representation of a market does not require that all the properties in that market be included in the index database. (See Chapter 4, especially Questions 4.4 & 4.5.)

1.1.7 What is the relationship between a research index and a peer universe based index?

The concept of a peer universe based index refers to a database or index representing the returns to all of the investment managers of a given style or specialization. In real estate, an ideal peer universe based index would include, at the property level, all of the property holdings of all the investment managers specializing in a given style of investment and type of property. This is generally the type of index that is most appropriate for researching the performance of particular managers or for carrying out manager performance evaluations. As a result, in the private real estate investment industry peer universe based indices tend to be more appropriate as agent evaluation benchmark indices (of the type addressed in Chapter 2) than as asset class research indices.

1.1.8 Can you provide examples of the types of questions a commercial real estate research index should be designed to help to answer?

- 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 real estate asset 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 various different types of locations?
- 7) Did the real estate market (or a given market segment) “turn around” (change the direction in asset price movement) last quarter?

1.1.9 What is the role of the research index in answering such questions?

A research index is not expected to supply definitive answers to research questions (including the questions listed above). The research index provides a type of “raw material” that is a vital input

for *analysis* that can, in turn, provide answers to a range of research questions. The ability of such an analysis to supply useful answers depends, in part, on the quality and quantity of other supporting data in the research database (apart from the index per se), and also on the quality of the analysis applied to the data. Part of the quality and success of the analysis depends on an understanding of the nature of the research index and its supporting database, including its strengths and weaknesses. The ultimate objectives are to improve fundamental understanding and to improve practical decision-making.

Section 1.2: Market Coverage

1.2.1 What is meant by breadth of market coverage in a research index?

Breadth of market coverage refers to the number of different property asset market segments represented by the index. For example, should the 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 metro areas, or should it extend to all 200+ metro areas plus rural zones? Should the index only include large (“institutional size”) properties (e.g., \$10 million +)?

1.2.2 What is the required breadth of market coverage for the ideal research index?

In terms of an “ideal” index we would like as much breadth of market coverage as possible. From a realistic and practical perspective, however, the question is how difficult (and expensive) it is to extend the breadth of index coverage. The other side of the equation is how much additional usefulness does one get from the index by adding breadth. Related to this issue is the question of how different, from an investment performance and decision-making perspective, are the different market segments and levels, the additional increments of market coverage. Neither the academic nor the industry research literature provides definitive answers to these questions.

1.2.3 What trade-offs and considerations are involved in defining the market coverage of the ideal research index?

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 segments that are beyond the current institutional holdings (and therefore remain uncovered by peer universe-based indices such as the ideal evaluation benchmark index addressed in Chapter 2) could improve the ability of institutions to invest in such new market segments. This would increase the investment choice set available to institutions. A relevant question is: *to what extent would this significantly push out the efficient frontier, i.e., how much additional diversification benefit or improved risk/return opportunities would be provided?...*

Another benefit of extended market coverage in a research index, to the extent this increases the institutional investment choice set, is that the size of the pool of physical assets that could be called upon to absorb the flow of financial capital provided by institutional investment would be increased. This would presumably help to increase liquidity and reduce price distortions in up-markets (when the net capital flow is into real estate). A relevant question is: *To what extent is such up-market liquidity and price distortion a problem?...*

We do not currently have even preliminary answers to these questions. Perhaps theoretical analysis or small-scale empirical pilot studies could be used to explore these questions.

1.2.4 As a starting point, what practical considerations can guide the index market coverage decision?

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 the market segments that can be relatively easily and inexpensively covered. The relevant “ease” and “expense” includes both technical considerations and institutional or organizational considerations. Through the end of the 1990s, there was essentially no commercial real estate investment performance index coverage in the U.S. beyond the existing holdings of the tax-exempt fiduciary investor branch of the real estate investment industry. Yet other countries have been able to include other major branches of the industry in their indexes. For example in Britain, the Netherlands and Australia the local equivalents of REIT holdings, insurance company general account holdings, and some wealthy private investor holdings are included in the major index products (IPD, ROZ-IPD, and PCA respectively on those three countries). In the U.S. there is some hope that statistical methodologies developed in recent years, combined with the electronic data availability possibilities arising at the turn of the 21st century, may allow the scope of market coverage to be considerably broadened.

1.2.5 How can statistical methodologies of index construction help expand index market coverage?

Advances in statistical methodology and computer technology during the 1990s make it possible to use empirical valuation observations (either appraised values or transaction prices) more efficiently and flexibly in the construction of capital return or periodic market price-change indices. In effect, properties can be reappraised much less frequently than the return reporting frequency of the index, and/or indices can be based directly on transaction price evidence without need for individual property-level appraisals. This reduces the appraisal expense that index data-contributors or data-gatherers would otherwise face. This makes it easier and more practical to expand the property market coverage of a research index either by (i) enlisting as index data-contributors members of branches of the real estate investment industry that do not regularly or frequently mark-to-market based on appraisals (such as the REIT and insurance company general account branches of the industry), and/or by (ii) basing the index at least partly on transaction price data that can be acquired through publicly-available primary or secondary information sources without the need for data-contributing membership.

Section 1.3:

The Level & Type of Information Included in the Research Index

1.3.1 What types of information should be in the ideal asset class research index, and what is the focus of the present report in this regard?

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

1. *Asset market prices*: Both the levels and changes in property asset market prices;
2. *Asset market activity*: The nature and volume of transactions & capital flows;
3. *Investment cash flows*: Net operating income and capital expenditures, including break-downs by source;
4. *Space market prices*: Rents & leasing activity;
5. *Development industry information*: Construction costs & activity levels, land & location values.

Obviously, not all of this information can be presented in a single “index”. As noted, an entire database or family of information products would be required to encompass all of these types of information. The primary focus of the present report is on the investment industry and the asset market, which includes the first three items in the above list. Within these three types of information, the third (investment property cash flows) may often be well-addressed in the context of an ideal agent evaluation benchmark index of the type described in Chapter 2 (see also question 1.3.2 below). Hence, the information most important to a real estate research index (as distinct from an evaluation benchmark index) is the asset market information, notably prices and transaction activity. Of these two, tracking asset market prices is probably the most fundamental and important, and also presents the greatest challenge. Accordingly, our treatment of the ideal asset class research index in this report is focused on the tracking of market prices and price-changes. (See also Question 2.3.9 in Chapter 2.)

1.3.2 What are some general considerations and trade-offs in deciding what information to collect in an ideal research index?

What types of information and variables should be included in the ideal research index database? For example, should the index collect space market information (e.g., occupancy, absorption, construction, etc) as well as asset market information? Should it collect transaction volume and capital flow data as well as asset pricing and valuation data? How deep should the index drill into the property operating level (e.g., lease information, operating expense and capital expenditure information, appraisal inputs and assumptions, etc)?

Needless to say, for an “ideal” index, we would like to have as much information as possible. In practice, however, cost/benefit trade-offs exist, and they relate in part to institutional and organizational concerns as well as technical concerns. Commercial information and data-vending firms already naturally supply some types of information. Other information tends to be more “public” in nature, in that the collective benefits of its production and distribution tend to be large relative to the private benefits. (See also question 1.1.4.) Industry-wide institutional efforts at index production should logically focus on such public or quasi-public types of information,

which may include asset market price levels and price changes. At a general level, much of the discussion in the answers to the Questions 1.2.3 & 1.2.4 is also relevant here. We suspect the “low hanging fruit” will become more plentiful as electronic commerce and internet-based information firms further develop in the real estate industry. In addition to the type of information that is currently collected by NCREIF, we strongly suspect that the types of information collected by the IPD in England and the PCA in Australia could be included in a realistic research index database here in the U.S. Perhaps this could be accomplished by some sort of 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 market price and transaction information, the space market and property operational level variables noted in the preceding paragraph, among others.

1.3.3 Should a research index be focused primarily on the property level or on the fund (manager) level?

There are important investment research issues and questions at both the property level (sometimes referred to as the standing portfolio or market level) and at the portfolio or fund level (or manager level, sometimes referred as the investment level). However, in a research index more attention tends to focus on the property-level, because this is where many of the most fundamental research questions lie, and many of the questions that distinguish real estate uniquely from other investment asset classes (such as the interface between the asset market and space market). In contrast, an evaluation benchmark index would focus more attention on the fund or manager level. However, there are also important research topics at the fund or manager level, particularly regarding financial economics questions.

1.3.4 Must a good research index report periodic returns, or is an IRR-based index potentially useful for research purposes?

Periodic total returns and the time-weighted averages that can be computed from them are very important for many basic types of research questions. For example, questions relating to volatility, systematic risk, correlations, lead/lag relationships, asset market efficiency and forecasting, all require time-series of periodic returns. Time-weighted average returns (TWRR) are also considered to be the best measure for evaluating investment manager performance when the manager does not control the timing of capital flow. (See Questions 2.3.3 in Chapter 2 and 7.5 & 7.6 in Chapter 7.) However, research indices may be used for a variety of purposes, and some of these purposes lend themselves to an IRR-based analysis. For example, property-level operational performance attribution and analysis can be well accomplished with the IRR (e.g., since property acquisition: see Question 3.8 in Chapter 3). In addition, research indices may be relevant for analyzing funds and investments in which the manager controls capital flow timing and where the assets are not periodically marked to market, for which periodic returns cannot be computed and may not even be relevant. (See additional discussion in Chapter 7.) In short, IRR-based research indices can be quite useful and appropriate for some purposes.

1.3.5 Must a good research index report total returns?

The concept of “total return” encompasses both capital value change and periodic or intermediate income distribution during the investor’s holding of the asset. In this sense, both the simple holding period return (HPR) and the IRR are total return measures. The former is periodic, while the latter is a dollar-weighted multi-period return measure.

Although total return information is vital for assessing investment performance, investment performance is not the only subject of research 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 studies relevant for investment performance. For example, studies of investment risk, volatility and questions of systematic risk, cross-correlation, serial correlation, forecasting of returns, and analysis of lead/lag relationships all can usually be carried out using only the periodic appreciation return (aka the “capital return”) component of the HPR, or even using asset market periodic price-change percentages. (The latter differ from the appreciation returns in that market price-changes do not subtract out the cost of capital expenditures, and they may not hold constant the sample of properties from one period to the next, so that the effect of pure aging at the property level is not reflected in the time series of market price-change percentages.)

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 for the index, other types of information, such as periodic appreciation or price-changes, space-market or property operational variables, and a variety of other information products can be quite useful. Note that this is in contrast to agent evaluation benchmark indices (the subject of Chapter 2), whose more specialized purpose of manager performance evaluation requires that they be able to produce total return reports.

1.3.6 Must the property level valuations in the ideal research index be based on transaction prices instead of appraisals?

Transaction prices are not necessarily either better or worse than appraised values as a basis for constructing an ideal research index for private real estate. In reality, appraised valuations are based on transaction price evidence, and the market participants who are responsible for determining transaction prices typically consider information from appraisal estimates of value. There is much similarity and mutual influence between transaction prices and appraisals, though there are also some differences in these two types of valuation observations.

In considering the use of either transaction prices or appraised values in a real estate research index, it is important to note is that transaction prices and appraised values both contain “errors”. In this context, the word “error” does not imply that anyone has done anything “wrong”. It simply implies that there is likely to be a difference between empirically observable prices or valuations, and the underlying conceptual true “market values”. (For an explanation of the term “market value” in this context, and an elaboration of this point, see the answer to Questions 5.1 and 5.8 in Chapter 5.) As neither transaction prices nor appraised valuations are without error,

neither is inherently or necessarily superior to the other for purposes of research index construction.

However, there do tend to be differences between transaction prices and appraised valuations in the *nature* of the errors. These differences, in turn, require different types of 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 precisely track asset market movements on a quarterly basis without the lagging that would typically occur in an appraisal-based index, would need to employ transaction price evidence either directly in an appropriately specified regression-based index, or indirectly to eliminate the lag in an appraisal-based index. (See Chapter 5 for further discussion.)

1.3.7. How does the answer to the appraisal-vs-transaction price question differ between research indices and evaluation benchmark indices?

In general, the preceding answer (1.3.6) applies to agent evaluation benchmark indices as well as to research indices. However, in practice, appraised values will usually be better than pure transaction prices for the construction of agent evaluation benchmark indices of the type described in Chapter 2. (See Question 2.3.7 in Chapter 2 and Questions 5.7 and 5.9 in Chapter 5.)

Section 1.4: Sampling & Property Composition Issues in the Research Index

1.4.1 Is it necessary for a real estate research index to include *all* the properties in the subject market segment?

No. The S&P500 does not include *all* U.S. large-cap stocks, yet it is still considered to be an effective passive market index and research index for that asset market segment. The science of *statistical inference* and *sampling* tells us that it is not necessary to include an entire *population* in a sample database in order to draw statistically valid inferences about the overall population. If a population is very large, it is usually much more efficient to employ a scientifically based *sample* rather than to attempt a *census* of the entire population. (See Chapter 4 for more in-depth treatment of the sampling question in the research index.)

1.4.2 Is it necessary for a real estate research index to be limited to a fixed and constant set of properties?

For research purposes there are advantages to “*static portfolios*”, that is, fixed sets of properties. However, the S&P500 Index is not a constant set of securities, and this does not prevent its use as an effective passive market index and research index for many purposes in the securities industry. Furthermore, it should normally be possible to define within a property index database a constant sub-set of properties over a given span of time, even though the overall index database is a dynamically changing set. It should also be noted that real estate property populations within

index databases such as NCREIF or IPD tend to change relatively slowly, due to the long holding periods associated with most real estate investments.

1.4.3 Should the constituent assets in a research index be equally-weighted or value-weighted in computing the index periodic returns?

As noted in the answer to Question 2.3.4, the answer to this question 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* (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). 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 usually be equal-weighted, at least within strata or relatively homogeneous market segments. For manager-level or fund-level research indices there is a stronger argument to try to include the entire population of relevant managers. (See Chapter 2.)

Section 1.5: Return Reporting Frequency in the Research Index

1.5.1 What is the ideal periodic return reporting frequency for a research index?

Which is best: annual, quarterly, or monthly return reporting frequency in a research index? You may think that more frequent is always better. However, 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 empirical valuation data collection, greater frequency of index reporting tends to result in more error per period in the index periodic returns. This error includes both temporal lag bias and random error. While 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 frequency without increasing either noise or lag (or both) is to spend more money on the underlying property-level valuation or data collection process. Thus, at some point the trade-off is between index reporting frequency and accuracy versus the budget constraint on the property-level valuation and data collection function. However, the underlying sluggishness and inertia in private real estate asset markets reduces the incremental benefits from greater frequency. Short return periods tend to be statistically dependent on adjacent periods, so one does not gain so many useful return observations as one might first imagine, even apart from valuation errors in the index. (See Chapters 5 and 6 for additional relevant discussion.) As a rule of thumb based on widespread industry practice, the ideal reporting frequency in a private real estate research index is probably quarterly, versus annual for evaluation benchmark indices (see Question 2.3.8).

1.5.2 What is the ideal historical span of the ideal research index?

The simple answer, of course, is: as long as possible. However, some research questions require very long historical time-series of data (e.g., estimates of long-term statistical characteristics, such as mean, volatility, correlation, lead-lag relationships), while 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 asset market value changes (e.g., the “turnaround” question requires quarterly representations without lagging or smoothing), while other questions do not require such precision (e.g., the long-run or complete-cycle mean return is not much affected by lagging and smoothing). Finally, note that the historical span multiplied times the index periodic return reporting frequency gives the time-series sample size, that is, the number of return observations in the index, which in turn is positively related to the statistical inference power of the historical index, as discussed in Question 6.1.7 in Chapter 6.

Chapter 1 Appendix A: Research Indices in Private Real Estate and Public Securities: Some Comparisons

1.A.1 What is a “*sample-based*” index, and are such indices widely used for research purposes in the public securities industry?

A sample-based index is an index used to represent an entire asset market segment for research purposes based on detailed up-to-date price information for only a sample of the securities in that market segment. Sample-based indices are common in the public securities investment industry and are widely used for research purposes in the mainstream of the academic and industry research relating to public securities. The most prominent example is probably the S&P500 Index, which comprises a sample of 500 stocks, but is often used for research purposes to represent all large-cap stocks. Some widely used corporate bond indices are based on stratified samples of duration and credit-risk “cells”. More sophisticated and in-depth research often uses customized indices based on scientifically stratified samples, defining cells based on careful statistical analysis of the attributes important for a given study.

1.A.2 Why is the S&P500 a good example of a passive market index and a research index in the public securities investment industry?

The S&P500 Index is a passive market index whose use is widely-accepted both for evaluation benchmarking and for broader research purposes in the public securities industry. The S&P500 is considered a good passive market index that reflects the large-capitalization domestic equity market segment (or “investment style”). Two attributes make the S&P500 Index a relatively good passive market index and research index for this market segment. One attribute has to do with the adjective “market”. The other has to do with the adjective “passive”. (Both adjectives modifying the noun “index”.)

The S&P500 Index is a “market” index because it effectively *represents* a reasonably well-defined segment of the equities market, namely, the large-cap domestic stock market. However, the S&P500 Index does not include *all* U.S. large-capitalization stocks. It is only a *sample*. 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. The sample size and its representativeness *together* insure that there is relatively little random tracking error and relatively little systematic bias between the S&P500 Index and the population of all 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 well reflect the performance of an essentially *passive* investor in the market segment or style represented by the S&P500. A passive investor is one who does not frequently trade individual stocks, and whose allocation among individual stocks thus changes primarily only in response to the stocks’ relative price changes (as in the index). 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”.

1.A.3 How does the S&P500 fall short as a research index?

Although it is widely used for research purposes, the S&P500 has some limitations that prevent it from being used (or make it sub-optimal for use) in some important types of research. For example, the S&P500 does not include small and mid-cap stocks, nor does it include all large-cap stocks. It also includes stocks from more than one major stock exchange, but not all of the stocks from any one stock exchange. And the S&P500 is not a fixed sample of stocks. For all of these reasons and others, the S&P500 is inappropriate or sub-optimal for addressing some important research questions. For example, market micro-structure questions comparing NYSE vs NASDAQ performance cannot be well addressed using the S&P500. Questions about the effect of size (market capitalization) on stock performance cannot be optimally addressed using the S&P500. Questions about the CAPM-theoretical “market portfolio” cannot be investigated using the S&P500, because the CAPM-theoretical “market portfolio” is much broader than the S&P500. The S&P500 may have a “survivorship bias” problem, because stocks that fall in value may be dropped out of the index. To address questions and problems such as these, other stock market indices must be used or developed. Much academic and industry research in the public securities investment industry today uses custom-tailored indices constructed from databases such as University of Chicago Center for Research in Security Pricing (CRSP), The New York Stock Exchange’s Trade-And-Quotes (TAQ), and others, rather than the S&P500.

1.A.4 Does the lack of a perfect “one-size-fits-all” research or benchmark index distinguish private real estate from the public securities investment environment?

No. As pointed out in the preceding answer to Question 1.A.3, even such a widely-acclaimed “flagship” index as the S&P500 is far from perfect or optimal for all research or benchmarking purposes. Therefore, the fact that the private real estate investment industry lacks a perfect research or benchmark index, and needs multiple types of indices for multiple purposes, does not distinguish the private real estate investment industry from the public securities investment industry. To be sure, both the quantity and quality of data in the public securities industry exceeds that available in private real estate, but the same general types of data problems face both of the two investment industries

1.A.5 What would be a conceptual example of a passive market index in private real estate, a conceptual equivalent to the S&P500?

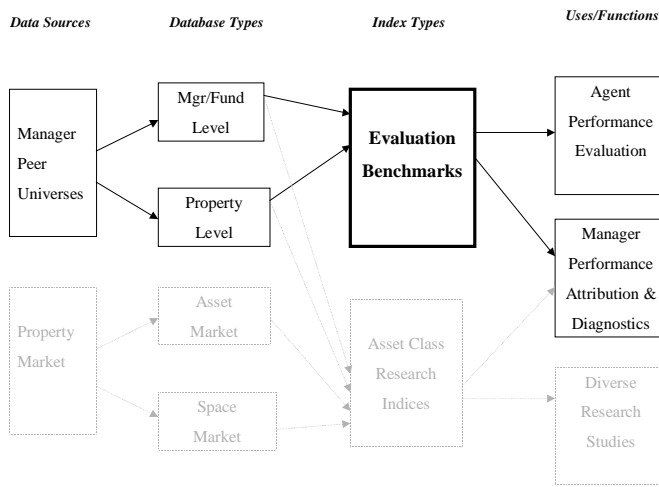
The closest equivalent the private real estate investment industry could conceivably have to a passive market index would be an index representing the investment performance of virtually *all* of the properties in a given property market segment or group of market segments. For example, an index representing all of the income-generating properties in the U.S. with a market value greater than \$10 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 to be included in the index database. (See Questions 1.4.1 above and 4.2 & 4.4 in Chapter 4.)

Chapter 2:
***The Ideal Agent Evaluation Benchmark Index
for Private Commercial Real Estate***

Section 2.1:
Definitions & Purpose of Evaluation Benchmarking

2.1.1 What is meant by an “agent evaluation benchmark index”?

As we are using the term in this report, an “agent evaluation benchmark index” (or simply “evaluation benchmark” for short) is a *reference point* that can be used as a *standard* to *quantify*



Solid arrows indicate major information and usage flows for evaluation benchmark index.

the relative performance of an investment manager (the “agent”) on behalf of an investor (the “principal”). The nature and role of an evaluation benchmark index as distinguished from an asset class research index as described in the previous chapter is indicated in the figure. (See also Chapter 1, Section 1.1.)

2.1.2 Who uses evaluation benchmarks, and in what role?

The investment performance evaluation may be of a professional

investment manager evaluated by or on behalf of a pension fund trustee, or it may be an evaluation of a pension fund trustee evaluated by or on behalf of the pension plan beneficiaries or plan sponsor. In some cases, the evaluation may be of an investment manager’s asset management or portfolio management employees by the investment manager’s CEO level within the organization. No matter what the specific principal-agent or superior-subordinate relationship, the practice of evaluation benchmarking refers to the formal comparison of the agent’s performance with that of one or more evaluation benchmarks. The agent is expected to explain or address performance differentials relative to the benchmark. The agent’s performance relative to the benchmark, in combination with her explanation of that differential, will typically be important (though not necessarily determinative) in the principal party’s decision to re-hire or extend the agent’s investment management contract.

2.1.3 What is the purpose of evaluation benchmarking?

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. *What does the client want the manager to do for him? How do both sides expect this function to be accomplished?* By together deciding on an appropriate evaluation benchmark at the outset of the management contract, the principal and agent can make more clear and concrete 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 for that agent, and if the agent knows that 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 obviously "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, all three of these purposes of evaluation benchmarking operate *before* the final reckoning of the manager's performance relative to the benchmark at the end of the management contract period. That is, these purposes occur, or are effected and accomplished, at the outset of, and during the course of, the execution of the management contract period. In this sense, evaluation benchmarking plays essentially what may be described as an *ex ante* role. By the end of the contract period, when the final review and reckoning of the agent's performance with respect to his benchmark is tallied and reported, the water is over the dam, at least for the past contract or review period. The only actions that can then be influenced are those in a subsequent contract or review period (if any). This is important, because it means that problems associated with drawing accurate inferences from measurements of *ex post* investment returns do not necessarily undercut the *ex ante* functions of benchmarking described here.

2.1.4 How is evaluation benchmarking used in an *ex post* manner to try to distinguish superior managers?

In practice, it is typically more difficult to rigorously identify *superior* managers *ex post* than it is to discourage and weed out obviously *inferior* managers *ex ante*. "Superior" managers are those who can *consistently* beat the market on a risk-adjusted basis and (therefore) consistently beat the

majority of their peers. As inferior managers are constantly being replaced, superior managers must consistently beat a moving target of an improving average. Superior managers are therefore rare, by definition, and when they do exist, their performance over long periods of time is likely to be only slightly better than the average of their peers. Statistical “noise” in the measurement of ex post investment returns makes it difficult to rigorously distinguish such superior performance using ex post quantitative return comparisons alone (see Question 5.11 in Chapter 5 for further explanation). Thus, wise principal parties try to avoid relying purely or excessively on quantitative performance measures in the evaluation of their agents. In this context, a key part of

Purposes of evaluation benchmarking...

- Ex ante purposes (most important):
 - Communication (client ↔ manager)
 - Interest-alignment (client → ← manager)
 - Weed out inferior managers
- Ex post purposes (less important or less feasible):
 - Identify superior managers (in combination with qualitative information)

the effective use of evaluation benchmarking is the consideration of agents’ *explanations* of their results with respect to their benchmark. These explanations involve descriptions of strategy, tactics, and procedures, including both the acquisition/disposition and operational management functions. They involve analyses of space and asset markets. In short, the *combination* of quantitative performance results and non-quantitative reasoning (in some cases supported by performance attribution analysis) provides a more powerful and

useful indication of an agent’s true abilities and effort than mere quantitative total return performance comparisons alone. The preceding box summarizes the discussion in this section of the purpose of evaluation benchmarking.

**Section 2.2:
The Basis and Fundamental Characteristics of the Ideal Private Real Estate Evaluation Benchmark Index**

2.2.1 What fundamental criteria define an ideal evaluation benchmark index?

An ideal performance evaluation benchmark index for *any* asset class 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 directly 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.

(For a discussion of how these six criteria relate to the “Bailey Criteria” prevalent in the public securities investment industry, see Question 2.A.2 in Chapter 2 Appendix A at the end of this chapter.)

2.2.2. What is the logical basis for the six fundamental criteria of the ideal performance evaluation benchmark index?

The necessity for the first criterion is obvious, because the purpose of evaluation benchmarking is to be a 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 evaluation 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 decision science. If the agent achieves x , but the principal party could have achieved y without hiring the agent, then the relevant performance achieved by the agent for the principal is not x , but $x-y$. The second criterion simply says that the benchmark should equal y .

The third of the six criteria, the *no-forced-bet* criterion, is based on the widely accepted management evaluation principle that *responsibility should be equated with authority*. This criterion is of particular concern to the manager, from the perspective of the fairness of the performance evaluation. If the manager is going to be held responsible for (that is, evaluated based on) his/her performance relative to a benchmark, then the manager should have the authority to control that relative performance as much as possible (without influencing the benchmark performance itself). This suggests that the manager should be able to control the decision of whether, how, and when to deviate from his/her benchmark. This third criterion overlaps the second criterion at a practical level. If the investor can invest in the benchmark directly (the second criterion), then presumably the manager also could invest in the benchmark, 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 simply *investability*. The ideal evaluation benchmark should be easily investable by both the client (principal) and the manager (agent).

The basic management principle of equating responsibility with authority also underlies the fourth criterion of the ideal evaluation benchmark: that the benchmark should reflect the manager’s style or specialization. This is sometimes referred to as the *appropriateness* criterion. This criterion is important not only for the objective of fairness in the evaluation of the manager, but also for the objective of promoting the usefulness of the benchmarking process for implementing the investor’s overall investment strategy. Typically, an investment manager is hired specifically to implement a component of an overall strategic investment plan, because the manager presumably has particular expertise for that component of the plan. For example, a manager may have particular expertise in picking small-cap value stocks, or in finding and managing apartment properties in the western U.S. The 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 their 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 for the benchmark comparison process to be meaningful and to provide the *ex ante* incentives described in Section 2.1. In the absence of this criterion, the manager could conceivably manipulate the performance evaluation to his advantage against the interest of the client.

Finally, the sixth criterion 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 criterion helps the benchmark to serve the investment objectives of the client. It helps both parties to insure that their interests and objectives are well understood and aligned, and it helps the manager to plan and carry out an investment policy in the client's interest. In short, it is "common sense" in a management contractual relationship.

2.2.3 What is the difference between a peer universe index and a passive market index?

Two different types of evaluation benchmark indices are typically defined in the public securities investment industry, based on the source of investment performance information in the index. A "*peer universe* index" reflects the performance of all the managers that are effectively competitors of the subject manager, that is, all managers who have the same style or specialization as the subject manager. A "*passive market* index", on the other hand, would represent the performance of *all the assets* in a given asset market segment or group of segments.

In the case of private real estate institutional core 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 essentially as institutional real estate investment peer universe indices in their respective countries. (We are not claiming here that any of these indices are "ideal" as such, but they are relevant examples of real estate peer universe-based indices.)

2.2.4 Which is better as an evaluation benchmark: a peer universe index or a passive market index?

In the public securities investment industry there is some argument, based largely on the Bailey Criteria (see Question 2.A.1 in Chapter 2 Appendix A), that passive market indices are preferred over peer universes as benchmarks for manager performance evaluation (see Question 2.A.2). Nevertheless, although this argument has been propounded for almost a decade, peer universes are still widely used.

In the private real estate investment industry this question lacks the same meaning and importance, for at least three reasons:

- First, at a practical level it is often virtually impossible to define an *appropriate* and measurable passive market index for real estate, so a peer universe index may be the only choice if benchmarking is to be done at all.
- Second, even if a passive market index can be defined and its investment performance accurately measured, a passive market index in private real estate would not be *investable* because whole assets are traded in private real estate. Thus, assets in the market index would already be owned by someone else. As a result, passive market indices do not implement the fundamental ideal benchmark criteria noted in Question 2.2.1 (or the Bailey Criteria) in the private real estate investment industry in the same way that they do in the public securities investment industry. For example, 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. Said differently, significant *tracking error* between the investor’s portfolio and the benchmark cannot be avoided even when the benchmark is a passive market index. (“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 investable, idiosyncratic differences in returns between the manager’s properties and the benchmark properties would make tracking error *inevitable*.)
- Third, unlike public securities investors, private real estate investors cannot choose between hiring *active* versus *passive* investment managers. Virtually *all* private real estate investment must be accomplished via investment managers that are “*active*” in a meaningful sense (see Question 2.A.3, in the Chapter 2 Appendix A). Because private real estate investors cannot hire passive investment managers and cannot themselves directly implement passive investment strategies, a passive market index loses much of its fundamental meaning and purpose as a benchmark.

In summary, there are important differences between the public securities and private real estate investment industries regarding the relative appropriateness and meaningfulness of passive market versus peer universe based benchmarks. Peer universe indices make at least relatively more sense as (“realistically ideal”) evaluation benchmarks in the private real estate investment industry as compared to the public securities industry. Moreover, it can be argued that peer universe based indices have considerable rationale in the private real estate investment industry based on the six fundamental criteria presented in Question 2.2.1. (See Question 2.2.5 immediately below.)

2.2.5 How do peer universe based indices implement the ideal evaluation benchmark index criteria in the private real estate investment industry?

Although no benchmark index in the real world can be perfect, peer universe based indices can serve as valid and useful benchmark indices in the private real estate investment industry. Consider how such indices relate to the six fundamental evaluation benchmark index criteria presented in Question 2.2.1.

(1) Measurability: Peer universe returns can be measured (e.g., in the U.S. as prescribed by the Real Estate Information Standards promulgated by NCREIF, PREA and NCREIM, consistent with AIMR performance measurement standards).

(2) Client investability: If the investor (that is, the principal party doing the evaluation) is a pension fund trustee/CIO and the manager (that is, the agent being evaluated) is a professional real estate investment management firm, then the investor can, in principle, hire any one (or several) of the subject manager's competitors as an alternative to using the subject manager. Thus, although the investor cannot exactly duplicate the performance of a given manager's entire peer universe, the investor could expect to obtain the average performance of the manager's peer universe, on an *ex ante* (beforehand) basis. In this sense the peer universe index is "investable" from an *ex ante* expectational perspective. Even if the principal party doing the evaluation is a plan sponsor or beneficiary (evaluating the pension fund trustees or CIO as agent), the peer universe index may be viewed as essentially similar to a "*Consumer Reports*" of the investment industry. It tells you how a particular agent (brand) compared with other similar agents (brands of the same type of product) over a given historical period (e.g., in a given test or survey of the product's performance). This is a plausible way of implementing the client investability criterion within the scope of what is realistically possible in the private real estate asset market.

(3) Manager's no forced bet: Although in private real estate the subject agent cannot exactly duplicate the property holdings of her peers, she should be able to find properties similar to those held by her peers, given that the peers are defined as being agents of a similar style and specialization as the subject agent. Thus, the agent is not forced to place a bet against an appropriately-defined peer universe benchmark, in the sense that the agent can invest in properties that have the same general characteristics as those held by her peers (provided of course that the peer universe index reports sufficient data about the nature of the properties in the index).

(4) Appropriateness: Unless the agent's specialization is unique, a peer universe based index can theoretically be defined based on the other agents with the same (or similar) specialization.

(5) Non-manipulatability: This criterion can be implemented with a peer universe benchmark index simply by removing the subject agent's own properties from the benchmark. This is easily done with any peer universe index.

(6) Agreement in advance: This criterion can be implemented with any regularly-published index, whether peer universe based or otherwise.

In summary, just as a review of "*Consumer Reports*" is a reasonable and prudent exercise in making consumer product purchase decisions, so evaluation benchmarking based on a peer universe index provides a common-sense based reasonable and prudent exercise in making investment management decisions. Peer universe based indices are, in principle, reasonable and useful tools in the evaluation of investment performance of agents in the private real estate investment industry. Their proper use as such will be addressed further in Section 2.4 below.

2.2.6 What about the problem of survivorship bias in peer universe based indices?

Survivorship bias refers to the fact that managers or funds that perform poorly tend to go out of business and therefore drop out of the peer universe based index. Indices that report the historical performance of only the *currently existing* managers or funds will then present an average historical performance that is biased on the high side.

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.

2.2.7 Are there ever circumstances where a passive market index could be superior to a peer universe based index for evaluating manager performance in the private real estate investment industry?

If the principal party explicitly wants his agent to deviate from her peers, it will be impossible to use a peer universe based index to benchmark the agent's performance. In this case, a passive market index would theoretically offer the only possibility to benchmark the agent. In private real estate, such a "passive market index" would be best represented by an ideal asset class research index of the type described in Chapter 1. (See in particular Questions 1.1.6, 1.1.7, and 1.2.3.) Because the passive market index is supposed to represent the entire market or asset class, it presumably can represent the performance of whatever assets the agent is investing in, even though that type of property is not held by any of the agent's peers. Thus, an appropriate component of a sufficiently complete passive market index could be used to benchmark the agent's unique investments.

In this circumstance, however, if the agent is being benchmarked at all, then she is being evaluated against something that none of her peers (no competing agent) provided as an alternative. By definition, the subject agent has no peers against which to be compared with respect to the performance of the types of properties that are in the passive market index but not in any peer universe index. As a result, the meaningfulness of the relative comparison of the subject agent versus the benchmark is less clear than normally. The principal party could not have invested directly in the benchmark properties without the aid of an agent. Would another agent have done better or worse than the one selected to manage this unique type of investment? Highly unique investment strategies are inherently difficult to benchmark.

Section 2.3:

Coverage and Information Considerations in the Ideal Evaluation Benchmark Index

2.3.1 Should evaluation benchmark indices measure property-level or portfolio-level returns?

Both property level and portfolio (or fund) level returns are of interest, and therefore either or both may be used for performance evaluation benchmarking. Nevertheless, the portfolio or fund level is in principle more relevant than the property level. This is because the returns actually realized by the investor are a more direct function of the portfolio level than of the property level. However, the property level is the underlying source of most of the portfolio-level returns, and the property level generally represents the purpose for which the manager was hired. Property level performance analysis is therefore particularly relevant for diagnostic and analytical purposes. (See Chapter 3 below.)

The difference between the property and portfolio levels will vary greatly depending on the type or style of manager fund or portfolio being considered. Traditional, unlevered “core” funds that place all capital relatively quickly into fully-operational assets will have portfolio-level performance very similar to property-level performance. At the other extreme, opportunistic or value-added funds that engage in development or redevelopment activities, make intensive use of leverage, and/or experience rapid turnaround of individual property asset holdings, will have portfolio-level returns that could differ substantially from any property-level returns that might be calculated (although accurate measurement of property-level returns might be difficult in such circumstances). In many cases, such funds may be better benchmarked by the use of indices based on inception-time-cohort IRRs rather than time-weighted average returns. (See Question 2.3.3 below.)

2.3.2 Should portfolio-level benchmark indices measure returns net or gross of fees?

In principle, returns net of all fees matter to investors. Therefore, the ideal portfolio-level performance evaluation benchmark index would measure returns net of fees, and such a benchmark would be compared to the subject manager’s net-of-fee returns for the client. In practice, it is often very difficult to compile net-of-fee benchmark indices, so the direct comparison between the manager and his benchmark may be conducted at the gross-of-fee level, with the impact of fees separately considered.

2.3.3 Should time-weighted returns (HPRs) or money-weighted returns (IRRs) be used in evaluation benchmarking?

(Note, for an explanation of the difference between time-weighted and money-weighted returns, and the definition of the HPR and IRR, see Questions 7.4, 7.6 and 7.7.)

Ideally, time-weighted returns should be used for performance evaluation benchmarking when the agent does not have much effective control over the timing of capital flow into or out of his management. This is usually the case in traditional portfolio management. In contrast, the manager should be evaluated on the basis of the IRR he achieves for the client when the manager

does have effective control over capital flow 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 which 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. As described in Chapter 7, IRR-based benchmarking must be based on equivalent inception-date cohorts, and should also consider how much capital remains with the agent, and how accurately the remaining assets have been appraised.

Even if the agent’s performance is measured by the IRR, in some cases his benchmark should be measured by a time-weighted return. Whether the performance evaluation benchmark index is measured using time-weighted return or IRR depends on the style and nature of the manager and the type of assets being acquired, and (related to that) on the appropriate peer universe defined for the manager. If traditional core assets are being acquired and the manager differs from traditional core managers only in the degree of discretion over cash flow timing, then the benchmark might well consist of the time-weighted return to a traditional peer universe based real estate index. In this way, the benchmark index will be measured by a return that is neutral with respect to cash flow timing. On the other hand, if the manager is being hired for more opportunistic type investment, the appropriate benchmark might also be IRR-based, along the lines of the inception-cohort indices used in the venture capital industry. (See Chapter 7.)

2.3.4 Should the constituent assets in an evaluation benchmark index be equally-weighted or value-weighted in computing the benchmark periodic returns?

The answer to this question depends on whether the benchmark index is intended to be a statistical sample or a population census. In contrast to a census, a sample does not attempt to include all members of the underlying population. In general, the individual constituents of a *sample* should be *equally-weighted* (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 “cell” of the sample, as described in Chapter 4). In contrast, the individual constituents of a *population* census should be *value-weighted*, so that the index return replicates the return achieved by a hypothetical portfolio consisting of the entire population of assets.

Consistent with the ideal evaluation benchmark index being based on a peer universe (see Question 2.2.5), such indices should generally be viewed as population censuses, consisting of all the peers of the subject manager. This reasoning suggests that 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 a bit extreme or legalistic in some circumstances, for example, where the agent is being benchmarked only against the average (mean) peer performance, and provided that a statistical sample could provide a highly accurate indication of that average (see answer to Question 2.3.6 below).

2.3.5 How much of the underlying asset class or market population of properties should be included in a property level evaluation benchmark index?

Consistent with the ideal evaluation benchmark index being based on a peer universe (see Question 2.2.5), such indices need not represent the entire asset class or market population of properties, but only the properties held by the subject agent's peers. For example, the theoretical population to be represented in a typical institutional core manager's benchmark would be all the properties held by other institutional core managers. Other types of properties, and even similar types of properties not held by the manager's peers, need not be represented in a peer universe index.

2.3.6 How much of the peer universe population should be included in the ideal evaluation benchmark index?

As a practical rule, as large a fraction as possible of the peer universe population should normally be included in the benchmark index. In part, this is because it is usually not possible to insure that smaller samples are random or representative of the entire peer universe population. Therefore, to minimize bias in the benchmark index, it is desirable to include a very large fraction (ideally at least well over half) of the peer universe population in the index. In part, this reflects a somewhat legalistic concern that agent evaluation should be based on the most accurate measure possible, rather than on a measure that is merely "acceptably accurate" in terms of traditional statistical confidence bounds.

For some benchmarking purposes, it also is useful to know the entire distribution of performance (e.g., the upper and lower quartiles) of the peer universe. This requires a very high percentage coverage of the peer universe population in the index. However, if a representative (unbiased) sample can be defined, and if the statistic of interest is simply the average (mean) return of the peer universe, and provided a statistical inference perspective rather than a legalistic perspective of measurement accuracy is adopted, then a large percentage coverage of the peer universe population may not be necessary in the evaluation benchmark index (see Chapter 4, relating to asset class research indices).

2.3.7 Why are appraised values usually better than transaction prices for evaluation benchmarking and for constructing evaluation benchmark indices?

As noted in Chapter 1, appraisal-based indices are usually more appropriate than direct (pure) transaction price based indices for evaluation benchmarking, though this conclusion does not necessarily hold for research indices of the type described in Chapter 1 (see Question 1.3.6). There are three reasons why appraisals tend to be more appropriate in evaluation benchmarks: two technical and one legalistic. The two technical reasons are based on a consideration of the two major types of errors that can exist in portfolio valuations and index returns: purely random error and temporal lag bias. The legalistic reason involves concerns about auditability in an agent evaluation framework. All three of these reasons are discussed below.

The essential consideration regarding evaluation benchmarking and purely random valuation error is that, in the typical institutional property portfolio, appraisals are conducted much more

frequently than transactions. For example, the typical property will be seriously appraised at least once every three years (ERISA) and often once per year or even more frequently in some cases. In contrast, the typical property is held for about five years or more between transactions. As a result, appraisal data tends to be “denser”, providing larger empirical samples of valuation observations, for a given number of properties and length of historical observation. This causes appraisal-based indices to be less “noisy”, in terms of purely random valuation error, than transaction price based indices.

Evaluation benchmarking requires the comparison of a benchmark index time-weighted average return to the subject agent’s portfolio over a given historical period of time. An individual portfolio of a single agent is likely to contain a relatively small number of properties, and in small samples purely random valuation noise tends to present the most serious problem. This is because the “Square Root of n Rule” of statistics (see Question 4.1 in Chapter 4) is working with an n (sample size) too small to sufficiently reduce the random error.

Valuation techniques that minimize purely random error are most at a premium in working with small portfolios, such as the agent’s portfolio in the evaluation process. Appraisal-based valuation typically minimizes purely random error because the sample of appraisal observations is typically larger than would be possible with transaction price data alone, and also, perhaps, because there may be less random cross-sectional dispersion in appraised values than in transaction prices (see Questions 5.2 & 5.3 in Chapter 5). Because we therefore want to use appraisal-based valuation to compute the agent’s time-weighted return in the benchmarking process, we also want to use appraisal-based valuation in the agent’s benchmark index, in order to compare “apples-vs-apples”. Indeed, if it is deemed desirable that the evaluation benchmark index should be composed as the sum of all the individual agents’ portfolios in the relevant peer universe, then the evaluation benchmark index would by mathematical necessity also be appraisal based.

The second reason to use appraisal-based valuations instead of pure transaction price based indices in evaluation benchmarking is because the benchmarking process concerns itself almost entirely with the time-weighted mean return across a multi-period historical sample of time (ideally at least three to five years). The major type of error in which appraisal-based valuations tend to be worse than pure transaction price observations is precisely the type of error that tends to not matter much in a comparison of multi-period time-weighted mean returns, namely, lagging and smoothing bias. Lagging and smoothing bias does not much affect multi-period mean returns (and the longer the time sample the better in this regard). Furthermore, it is possible that the lagging and smoothing will be similar in magnitude between the manager’s portfolio and the benchmark index, in which case the effect of appraisal-based lagging and smoothing error will largely cancel out in the benchmarking comparison. (See also Question 2.4.11.)

The third reason why appraisal-based indices tend to be more appropriate for evaluation benchmarking is legalistic in nature, related to the notion that an index that is being used to evaluate an agent should be auditable down to the individual property level, rather than based on statistical inference. A formal, direct transaction price based index typically requires the use of regression-based statistical techniques such as hedonic valuation models (HVM) or repeated-measures regression (RMR). (See Chapter 6, Section 6.3.) HVMs raise the problem that 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 have the problem that they require the properties on which they are based to be sold twice. But all the properties that still remain in the agent’s portfolio or in the benchmark index have typically only been sold 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, and may be frowned upon from a legalistic perspective in which performance differentials based on samples as opposed to entire populations is viewed unfavorably.

2.3.8 What is the required frequency and historical span of the ideal evaluation benchmark index?

The answer in this case is somewhat different than for the ideal research index as described in Chapter 1 (see Section 1.5). As noted in the present chapter, evaluation benchmarking is oriented primarily toward comparisons of time-weighted mean total returns over moderately long periods, typically 3 to 5 years. Shorter evaluation periods tend to be too “noisy”, not providing a long enough sample of time to draw valid or useful conclusions, especially in the more legalistic context of agent evaluation. (See the previous question.) Moreover, return temporal second moment statistics (such as volatility) have relatively little use in evaluation benchmarking (see in particular Question 2.4.3 below). For both these reasons, high-frequency periodic returns per se are of relatively little use in evaluation benchmarking. However, the higher the index reporting frequency, the more flexibility the benchmark index provides for matching any given evaluation time frame. Nevertheless, there is probably little added value to a return reporting frequency greater than annual for the purpose of agent evaluation benchmarking.

2.3.9 What types of data should the ideal evaluation benchmark index collect from data-contributors?

At a minimum, in order to produce an investment 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 as described in Questions 1.3.1 and 1.3.2 in Chapter 1. In addition, it may be argued that the ideal agent evaluation benchmark index should also provide property operational-level performance attribution analysis services. (See Chapter 3, especially Questions 3.4 and 3.12.) 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.

Section 2.4:

Considerations in the Appropriate Use of Evaluation Benchmark Indices

2.4.1 What considerations are important in defining an appropriate peer universe?

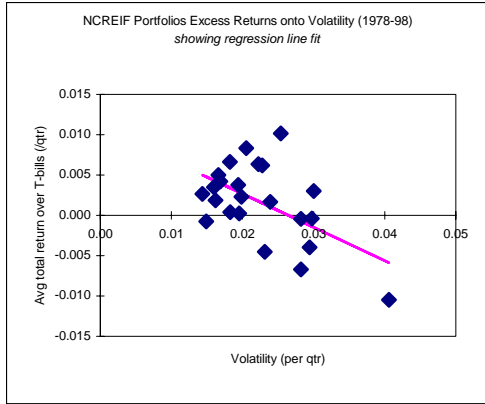
Four types of considerations can be important in defining an appropriate peer universe, or in understanding the appropriate use of a given (feasible) peer universe index:

- (i) Property segment specialization,
- (ii) Risk control,
- (iii) Appraisal policy, and
- (iv) Fee policy.

Managers or funds can be characterized by these four attributes. Ideally, a given agent should be benchmarked against a peer universe that includes only other agents that are similar to the subject agent in all four dimensions. For example, managers specializing in apartments should be benchmarked against other managers specializing in apartments. Where such similarity is not possible, differences along these four dimensions must be controlled for. In some cases, this is facilitated by recognizing that some aspects of the above four considerations can be treated as analytical levels or components of overall return measures. For example, leverage affects risk, but levered returns can often be broken down into two components: the underlying unlevered returns plus the incremental effect of leverage. Thus the manager's return could be unlevered and compared against an unlevered benchmark, or an unlevered benchmark's return could be levered-up for direct comparison with the manager's return. The above four dimensions are considered in the following questions.

2.4.2 What is meant by property market segment specialization and why is it important to control for segments in an appropriate evaluation benchmark index?

Property market segment refers to the type and geographic location of the properties the agent is expected to consider for inclusion in the subject fund or portfolio. Real estate asset market segments usually correspond closely to space market segments, which are defined largely by property usage types (e.g., office, industrial, retail, apartment, etc, including sub-categories in some cases), and geographical area (usually specified metro areas within broader geographic regions). It is important for benchmarks to be appropriate in this dimension because the performance of property segments can vary substantially over a given 3 to 5 year period. 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 would generally not want to encourage his investment agent to delve into market segments the agent was not already familiar with, a result that could occur if the benchmark index included such segments (otherwise the agent would be placing a "forced bet" against his benchmark).



2.4.3 How can risk be controlled for in evaluation benchmarking in private real estate investment?

Risk control refers to the need to assure that the agent cannot expect to beat his benchmark simply by taking on more risk than is present in the benchmark, relying on the market’s normal risk/return trade-off to achieve higher average returns (unadjusted for risk). In theory, risk can be controlled for by using risk-adjusted return measures in both the benchmark and the agent’s portfolio, and basing the evaluation comparison on the adjusted measures rather than the raw returns. For

example, comparison on the basis of a “Treynor Ratio” makes good theoretical sense in such a role. (The Treynor Ratio is the portfolio’s excess return over T-bills divided by its systematic risk as represented, for example, by its “beta”.) However, in practice such an approach does not work well for private real estate. One problem is that returns must be adjusted for risk in the way that the *capital market prices risk*, and this is not theoretically well understood, especially 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”? (This problem is illustrated in the exhibit, which shows typical real estate portfolio ex post returns plotted against their volatilities. A valid risk measure would result in a strong upward-sloping relationship between risk and return.)

The more practical approach to controlling for risk in private real estate investment is to define the peer universe to include only portfolios of similar risk to the subject portfolio, and to restrict the agent from pursuing investment strategies that significantly increase that risk. Because it is difficult to measure risk in the private real estate market in the way that risk is priced by the capital market (i.e., as reflected in expected returns), discrete categories of investment attributes and styles are used as proxies for risk categories. For example, leverage limits, 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 or review period, and often on an on-going basis during that period.

For example, a fund that allows high leverage, but only on core-type fully-operational properties, might be benchmarked either against a peer universe consisting only of other such funds, or against a core property-level index (such as the NCREIF Index) adjusted for leverage. If the latter approach is being used, normally the leverage adjustment in the benchmark index should specify at the outset a fixed percentage of leverage, not simply mimic the manager’s portfolio’s leverage, assuming that the manager has discretion over the amount of leverage in the subject portfolio. (See the answer to Question 2.4.7 below.)

2.4.4 Why is appraisal policy important in evaluation benchmarking?

Appraisal policy should also be considered to make sure that the subject agent's appraisal policy does not differ from that in his peer universe benchmark in any way that could significantly bias the comparison between the two. In some cases, any such differences might be able to be adjusted or dealt with in the comparison procedures. For example, if a subject manager appraises all properties as of the end of the year, and the peer universe has staggered appraisals throughout the year, then the peer universe will be slightly lagged behind the manager in time, as of the end of each year. Various techniques can be used to adjust for such differences (e.g., the manager's returns could be lagged to match the benchmark by applying linear interpolation of the manager's annual value levels, or the peer universe's returns could be projected forward using a forecasting model of the index).

2.4.5 Why is fee policy important in evaluation benchmarking?

Fee policy may also be important to consider in defining an appropriate benchmark or the appropriate use of a given feasible benchmark. Managers or funds with substantially different fee policies may differ in other important respects as well, such as the degree of risk in the investments. Furthermore, benchmark index performance will often have to be quantified at a before-fee level. Therefore, 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, to make an after-fee comparison of net returns realized by the investor.

2.4.6 What considerations are important in defining appropriate segment or style weights in the evaluation benchmark index?

Suppose a manager is hired with the discretion to allocate investment capital across more than one segment of investment. (Or the discretion could include different styles, or any other dimension that could affect performance, such as degree of leverage, for example.) 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 cannot be disaggregated to reveal pure-segment performance sub-indices, there is no choice. The benchmark must use the aggregate weights in the overall peer universe index. But if the peer universe index can be disaggregated (as, for example, the NCREIF Index by property type and geographical region), then a choice exists. The manager's benchmark can be defined using the pure-segment sub-indices of the peer universe and whatever weights the manager and client together agree are appropriate for the manager's benchmark. What should those weights be?

There is no single answer to this question, except to say that the weights should reflect the client's objectives and the role that the subject manager is hired to play in the client's overall investment strategy. For example, if the manager is hired to place capital into industrial and office properties, then 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. In the absence of a clear reason otherwise, a simple solution such as equal-weighting makes sense. 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 skewed weighting would reflect such an objective (e.g., 75% industrial and 25% office weighting in the benchmark. Another approach would be 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 value), then 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.

2.4.7 What's wrong with defining the evaluation benchmark segment weights to be whatever weights the agent uses in her portfolio?

The problem with this approach is that it 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 on the basis of her segment allocation decisions. The sign (+ or -) of 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.

For example, suppose a manager is hired to place capital into industrial and office properties, with the allocation between these two segments at the manager's discretion. Now suppose the manager chooses to allocate 75% of the capital under her management to the industrial segment and only 25% to the office segment. Suppose this allocation decision turns out to be a wise choice because industrial properties outperform office properties by, say, 600 basis-points. As a result, as compared to a 50/50 allocation, the manager has gained 150 basis-points for her client due to her allocation decision $[(75\% - 50\%)600 = (0.25)600 = 150]$. On the other hand, suppose that the manager performed 100 basis-points below her peer universe within each segment, that is, both her industrial and office properties provided 100 basis-points less return than the average among her peers. Then if the manager's benchmark is weighted by her own segment weights, she will fall 100 basis-points below her benchmark, even though she has in fact provided her client with a net advantage of 50 basis-points above a peer universe benchmark weighted equally between office and industrial property. (For example, if the average industrial property provided a 16% return and the manager's 15%, while the average office property provided a 10% return and the manager's 9%, then the manager would have achieved $(0.75)15\% + (0.25)9\% = 13.5\%$, while the 50/50 benchmark would have achieved $(0.5)16\% + (0.5)10\% = 13\%$.) Of course, the 50/50 weighting in the benchmark might be arbitrary in some sense, but so might a 75/25 weighting. The relevant point here is that setting benchmark segment weights equal to the

manager's actual portfolio weights removes any incentive for the manager to improve her performance via superior segment allocation decisions.

2.4.8 Is there ever a circumstance where the use of manager weights in the benchmark is appropriate?

Yes. 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.

2.4.9 Should the ideal peer universe extend beyond a single branch or style of the investment industry?

There are various distinct branches of the U.S. private real estate investment industry. Although the definition of such branches is somewhat unclear, one can rather easily distinguish among, for example, the tax-exempt fiduciary branch, the REIT branch, the insurance company general account branch, other financial institutions, foreign investors, and wealthy private individual investors. (Note, REITs in this context are *investors* in the private property market, and thus a category of direct property owners.) Arguably, each of these different branches has characteristically different concerns, objectives, and styles in their real estate investing. As described in Section 2.2 (especially Question 2.2.5), a basic principle of evaluation benchmarking requires that agents be benchmarked only with respect to other agents of similar style and specialization. Thus, in general, separate peer universe evaluation benchmark indices should be defined and used within each branch of the private real estate investment industry. For example, pension fund investment managers should not generally be benchmarked against REITs. Indeed, even within one branch there may be agents that specialize in different styles or market segments (as noted above in Question 2.4.6).

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. Moreover, comparisons between peer groups are also relevant at a broader level. For example, pension fund plan sponsors may wish to compare the performance of the tax-exempt fiduciary peer group against the REIT peer group in deciding how they want to make their real estate investments. But this level of comparison, between styles or types of investment vehicles, is more in the field of broad asset class research than specific agent evaluation benchmarking. Benchmarking at this broad asset class level is somewhat different conceptually, perhaps more appropriately addressed in the context of the ideal asset class research index (as described in Chapter 1) rather than the ideal agent evaluation benchmark index.

2.4.10 Is there a potential downside to evaluation benchmarking?

Yes. 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 general dangers 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, then misleading or meaningless comparisons will tend to result. Moreover, perverse incentives may face the agent. For example, if a manager is hired because of her specialized expertise in apartment properties to manage apartment investments, then the overall aggregate NCREIF Index would be an inappropriate benchmark. The manager might beat her benchmark (or get beaten by it) because of events beyond her control (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 the area of her specialized expertise and competency. (For example, in order to avoid placing a “bet” against her benchmark, the apartment manager might be tempted to invest in other types of properties that are in the NCREIF Index. 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.*** An agent’s differential performance relative to a benchmark is inherently “noisy”. There is noise (of all three types described in Chapter 5, Questions 5.10 & 5.12) both in any benchmark index as well as in any agent’s portfolio returns. There is therefore even more noise in any comparison of the differential between an agent’s portfolio and a benchmark, even a very appropriate benchmark. Excessive use of purely quantitative comparisons, or failing to think very carefully through any use of quantitative performance differentials, can have perverse results. For example, 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 excessively risky investments, or may discourage the truly best managers from even wanting to risk working for the client.

2.4.11 Do performance measurement difficulties in private real estate negate the potential usefulness of evaluation benchmarking?

No, but they do make evaluation benchmarking a “blunter” tool in the private real estate investment industry than it is in the public securities industry. As described in Chapter 5 (Questions 5.10-12), 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, for example, the attempt 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 also from the answer to Question 2.1.3 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 (or of discouraging truly inferior managers from entering or remaining in the business), provided the benchmarking process is carefully applied (including a recognition and consideration of the nature of the noise and errors in *ex post* measurement). For example, 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 interest, then the benchmarking process will be serving its interest-alignment purpose. In general, even a noisy index is more likely to be beaten by an agent who truly performs better than by one who truly performs worse.

Chapter 2 Appendix A: Evaluation Benchmark Indices in Private Real Estate and Public Securities: Some Comparisons

2.A.1 What are the “Bailey Criteria” and how are they related to the six fundamental criteria of the ideal performance evaluation benchmark index?

In the public securities investment industry the following list of criteria defining an ideal evaluation benchmark index are widely cited and employed:

- (1) ***Unambiguous***: Names and weights of the securities in the index are clearly specified;
- (2) ***Investable***: The option is available to forego active management and simply hold the benchmark;
- (3) ***Measurable***: The benchmark's return can be frequently calculated;
- (4) ***Appropriate***: The benchmark reflects the manager's style or specialization;
- (5) ***Knowable***: The manager can be expected to have current knowledge about the securities making up the benchmark;
- (6) ***Specified in advance***: The benchmark is constructed and mutually agreed upon between the manager and client prior to the start of the management contract.

These are sometimes referred to as the “Bailey Criteria” or the (“BRT Criteria”, for Bailey, Richards, & Tierney), after the authors who initially promoted them. These criteria are practical, common-sense measures for implementing *within the public securities investment industry* the six fundamental benchmark index criteria described in Questions 2.2.1 and 2.2.2 (or at least most of them: it's not clear how the Bailey criteria alone guarantee that the manager cannot manipulate the benchmark).

The Bailey Criteria are not necessarily “fundamental” at the same level of generality as the six criteria presented in Question 2.2.1, although there is considerable overlap between the two sets

of criteria. In fact, there is some overlap and redundancy even within the Bailey Criteria. For example, the “Unambiguous” criterion is necessarily implied by the “Investability” and “Appropriateness” criteria (hence there is no logical need to identify “unambiguousness” as a separate criterion). The “Knowable” criterion is surely implied by the “Appropriateness” criterion. (How could an index reflect a manager’s style or specialization if the manager could not be expected to have current knowledge about its constituents?) Thus, there are four *fundamental* Bailey Criteria: Measurable, Investable, Appropriate, and Specified-in-Advance. As noted in Question 2.2.2, *investability* is a practical way to implement two of the six fundamental criteria of Question 2.2.1: direct investability by the client and no forced bet by the manager. Thus, the Bailey Criteria effectively implement five of the six fundamental ideal index criteria, within the public securities investment industry. (The criterion left out by the Bailey criteria, as noted, is the fifth criterion of Question 2.2.1: “No manager influence on the benchmark performance.”)

2.A.2 How do the Bailey Criteria relate to the ideal benchmark index in the private real estate investment industry?

In contrast to the public securities environment, within the private real estate investment industry the Bailey Criteria described in 2.A.1 do not necessarily implement the six fundamental evaluation benchmark index criteria described in 2.2.1. Thus, the Bailey Criteria do not have the same meaning and relevance in the private real estate investment industry that they do in the public securities investment industry.

For example, in the public securities industry the Bailey criterion of *unambiguousness* effectively implements the *investability* criterion (which in turn implements the direct *client investability* and *no-forced-bet* criteria), because public securities are liquid, publicly-traded, homogeneous shares. If we know that an index contains 1% IBM stock, then we can go out and replicate that part of the index by allocating 1% of our portfolio to IBM stock. The IBM stock we buy will be identical to the IBM stock held in the benchmark index. If we know all the constituents of the index, *and all the constituents are liquid, homogeneous shares*, then we can replicate the entire index in our portfolio.

In contrast, the Bailey *unambiguousness* criterion would not implement the more fundamental *investability* criterion in the private real estate investment industry. Knowing the exact identities of the properties in the benchmark index does not enable either the client or the manager to replicate the benchmark, because it is not generally possible to purchase the properties in any benchmark index. In the private real estate asset market, unique, whole assets are traded privately at the discretion of their current owners, rather than homogeneous shares in a public exchange that backs up the liquidity of their listed securities as a matter of policy.

2.A.3 Why is direct passive investment not possible in the private real estate investment industry?

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 passive investment is not possible, as this concept is understood in the public securities investment industry. There are three reasons for this: the need for asset operational management; the need for specialized local expertise; and 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 investors 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 income rather than capital gain, property operational management has a large impact on the overall 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.

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 privately, 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 that the investor does not have the power of a highly efficient, competitive marketplace protecting the investor from doing “bad deals”, in the sense of overpaying for purchases or selling at prices below market value. When you buy shares in IBM, you can be relatively certain that you are not overpaying. 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. It requires some specialized expertise to be sure you are not overpaying for 1000 North Main Street.

Finally, replication of a passive market index is not possible in the private real estate investment industry. The fact that unique, whole assets are traded in the private property market makes it impossible for any one investor (or any one investment manager) to hold the assets that compose any conceivable passive market index. This makes it impossible in private real estate to implement passive investment management defined as the replication of a passive market index by the manager.

2.A.4 What is the difference between active management in private real estate as compared to public securities investment?

As noted in the preceding question, active management in private real estate is typically focused heavily on *operational management*. In contrast, active management in the public securities investment industry consists purely of *asset picking*, the attempt to find superior-performing individual assets. This is *trading-oriented* active management, as distinguished from operational management-oriented active management. Short-term trading of assets in private real estate

markets is usually self-defeating, due to the much higher transaction costs (as a fraction of asset value). 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.

Of course, real estate investment managers must, like active stock managers, pick individual assets. However, the overall long-term investment performance of real estate managers is often due as much or more to their operational management ability and effort as to their individual asset picking and trading performance. Furthermore, it is empirically very difficult to rigorously distinguish manager performance attributable to asset picking and trading as opposed to manager performance attributable to 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.

Chapter 2 Appendix B: Some Information Economics Considerations Relevant to the Ideal Evaluation Benchmark Index

2.B.1 Is an agent evaluation benchmark index a “public good” for the real estate industry, similar to the flagship research index described in Section 1.1 of Chapter 1?

Although the ideal agent evaluation benchmark index has some of the same industry public good characteristics as the ideal research index, a benchmark index is not a public good to the same degree as a flagship research index. Two main differences should be noted in this regard between the ideal benchmark index and the ideal flagship research index:

- **Private purpose:** The specific agent investment performance evaluation function is a private purpose, serving specific principals and agents to help them improve and mediate their private business relationships. Admittedly, the benchmark index’s support of the general ability within the industry to carry out such private evaluations improves the overall functioning of the industry and therefore redounds to the ultimate benefit of the asset class. Nevertheless, the direct and overriding purpose of agent evaluation per se is essentially private in nature. This is also true of the function of benchmark indices in supporting performance attribution at the individual firm level, where such attribution is used primarily for diagnostic purposes either within the agent’s firm or in the review and evaluation process between the agent and principal. In contrast, although some uses of an industry flagship research index are for private purposes, there is also much use for academic, public sector, and industry research that is of a more fundamental or collective foundation-building nature (e.g., industry “white papers”).
- **Commercial product:** There is considerably more potential for the ideal evaluation benchmark index to be produced as a commercial product than is the case for the ideal research index. Although this has yet to be done in the U.S., IPD has demonstrated that, in combination with complementary performance attribution information services, it is

possible for an agent evaluation benchmark index to be profitably produced in several other countries.

2.B.2 Is an agent evaluation index a “natural monopoly”?

The term “natural monopoly” refers to a market where the entire demand can be satisfied most efficiently by production from a single producer. (Note that the terms “natural monopoly” and “public good” are not synonyms.) Many information products in the so-called “new economy” tend to be natural monopolies. The ideal benchmark evaluation index arguably has important characteristics of a natural monopoly. These characteristics are seen in both the demand and supply sides of the market for such an index:

- **Supply side:** Once the data necessary to compile and produce the ideal benchmark index is obtained, it can be distributed at very little additional cost to an unlimited number of users. Thus, there are large economies of scale in the production and distribution processes.
- **Demand side:** Users of the benchmark index for agent evaluation purposes may derive some benefit from having a standard, uniform product for all such evaluations. Evaluation is inherently an exercise in *comparison*, and comparisons are facilitated by the use of a single common reference point. Comparisons are also most valuable when they can be made against an entire “peer universe” (see Questions 2.2.3-2.2.5, and 2.3.6). These considerations suggest that the use of a single industry-wide standard benchmark is more valuable to users than having to deal with a variety of competing benchmarks in the evaluation function. This provides the ideal evaluation benchmark index with “*network externalities*”, which impart a demand-side basis for the natural monopoly, complementing the supply-side economies of scale noted above.

The markets for products that have supply and demand side natural monopoly characteristics tend to be what economists call “tippy”. Once a single supplier captures a majority market share, the entire market tends to rapidly “tip” in favor of that supplier, often very quickly resulting in a monopoly supply situation.

2.B.3 What about a “natural duopoly”?

This is not a term we are familiar with in the technical economics literature, but it might refer to situations where a natural monopoly is managed for technical and institutional reasons as a duopoly, in which two competing private firms are fostered or cultivated, precisely to mitigate the possible dangers and problems associated with monopoly. (These dangers can include such concerns as monopolistic pricing, stagnation of innovation, and excessive dependency on a single private firm.) Arguably, an example of this type of managed duopoly is found in some of the U.S. Defense Department’s procurement policies, where there is some attempt to maintain two independent suppliers of, say, fighter jets (e.g., Lockheed and Boeing). In the case of evaluation benchmarking, a counter-argument to the demand-side natural monopoly argument described in the previous question suggests that, as there is no *unique* best or correct way to do evaluation benchmarking and performance attribution, users would in fact benefit from having competing benchmarks, such as the several benchmark indices that are used in the REIT industry. Admittedly, there is some danger that a large number of competing benchmarks in private real estate performance evaluation would be confusing, and possibly even lead

benchmark production firms to be susceptible to losing their evaluation independence as they are forced to curry favor with potential data-contributing subscribers. However, there might be an interesting argument that a small number of competing independent benchmark suppliers, such as two, might be better than either one or many. Perhaps the analogy that is most on point for evaluation purposes is the domination of the corporate bond-rating business by two independent private firms: S&P and Moodys. However, a difference with the bond-rating business is the degree of dependence on confidential proprietary information (see Question 2.B.6 below).

2.B.4 If the ideal evaluation benchmark index is a natural monopoly, does this imply that it should be produced by an industry association rather than a private firm?

Not necessarily. As noted, natural monopolies are commonplace in the new “information economy”, and almost all of these natural monopoly products are supplied by private, for-profit firms. However, users of products that tend to be natural monopolies need to be aware of certain dangers and opportunities. In particular, experts counsel methods to help users avoid the negative consequences of “lock-in”, and even to try to tap into some of the “super-normal” profits suppliers of natural monopolies can obtain. For example, users of natural monopoly products are advised to:

- Bargain hard *before you are locked in*, to obtain some form of “sweetener” or long-term protection from the monopoly supplier, such as discounts, equity participation, and contractual guarantees against monopolistic exploitation in the future.
- Keep your options as open as possible, for example, by cultivating or even subsidizing a potential competitor to the monopoly supplier, e.g., through *multiple-sourcing* as practiced by the U.S. Defense Department (see previous question).
- If possible, negotiate purchase agreements through collective user groups or cooperatives, so as to balance monopoly power with monopsony power. (Monopoly refers to one producer; monopsony refers to one purchaser.)

(Note: The term “lock-in” refers to the tendency of users to become “locked in” to a single supplier of a natural monopoly product, such as PC manufacturers getting locked in to Microsoft operating system software. The term “super-normal profits” (aka “monopoly rents”) is a technical term in the economics field that refers to profits greater than those that would be sufficient to attract producers, given the costs and risks involved in the production process. For an excellent, practical guide to dealing with the problems and opportunities facing users of information products that tend to be natural monopolies, see Information Rules, by Carl Shapiro and Hal Varian, Harvard Business School Press, Boston, 1999.)

2.B.5 What are the two phases of production, and two types of finished products, in the evaluation benchmark index business?

There are two phases in the production of evaluation benchmarking services. The first involves collecting and compiling the raw property-level and fund-level data from the peer universe whose investment performance is to be benchmarked. (For an explanation of “peer universe”, see Questions 2.2.3-2.2.5.) The second phase involves aggregating this raw data into two types of finished information products:

- A **benchmark index**, measuring the periodic investment performance of the relevant peer universe (with appropriate segment breakdowns and sub-indices); and

- **Custom benchmarking services**, comparing individual agents or portfolios to specified custom benchmarks, and possibly including complementary information products, such as portfolio-level and property operational-level performance attribution analysis relative to the benchmarks.

2.B.6 Is vertical integration technically necessary between the underlying database compilation process and the finished information products?

Vertical integration refers to the inclusion of multiple phases of a production process within a single producer. In the present context, the question is whether the database compilation process and the finished information product production process need necessarily be included in a single producing entity. The answer to this question revolves primarily around the need to assure confidentiality of proprietary information for the data-contributors who supply the raw fund-level and property-level data. This consideration is somewhat unique to the private real estate investment industry. Because of this, it is important for any entity compiling a benchmark database to be able to guarantee the necessary confidentiality to the data contributors. It is difficult to imagine how such confidentiality could be guaranteed if the database-compiling entity is going to turn around and sell the raw data to other firms that would then produce one or more finished information products from the data. Although some finished information products can be produced from masked data without revealing confidential proprietary information, the construction of an evaluation benchmark index, and the production of custom-benchmarking and performance attribution analysis services inherently require access to the entire peer universe of non-masked, individual property and individual owner data. Thus, confidentiality requirements probably make vertical integration between the database compilation process and the evaluation benchmark index production process necessary. That is, both of these phases of production need to be housed in a single producing entity. Such an entity could, however, involve an appropriately constructed joint venture between two or more independent organizations.

Chapter 3:
Performance Attribution in Private Real Estate Investment Management

3.1 What is meant by “performance attribution” in private real estate investment management?

The term performance attribution refers to the quantitative breakdown of a given historical investment performance into various additive components, so as to *attribute* the total performance to various sources. The investment performance being “attributed” may be measured by a given agent’s total return or by the differential between that agent’s total return and the total return on a given evaluation benchmark index of the type described in Chapter 2. The term “agent” here is used as defined in Chapter 2 to encompass investment managers and fiduciaries working for principal parties that could include investment management firm CEOs, investment trustee fiduciaries, or investment plan sponsors or beneficiaries.

3.2 Who uses performance attribution, and for what purpose?

Performance attribution is typically used by two main types of users, for two main purposes:

- By an investment agent himself (e.g., a manager or a trustee fiduciary) for internal diagnostic purposes to help understand the sources of his past performance;
- By an investment principal party (e.g., a trustee fiduciary or pension plan sponsor) in the context of a performance evaluation of an investment agent, for example to help understand or diagnose the nature of the agent’s performance.

Note that both uses are essentially diagnostic in nature. Performance attribution is used to diagnose the sources of ex post investment performance in a formal, quantitative sense.

3.3 What are the two major levels of performance attribution?

Performance attribution in real estate is applied broadly at two levels: the portfolio (or fund) level, and the property operational level. These are explained below:

- **Portfolio level:** At this level two major underlying sources of investment performance are traditionally recognized: *allocation* performance and *selection* performance. These correspond to the two major functions of a portfolio manager in the public securities investment industry. The allocation function is the decision of how to allocate the overall portfolio among different categories of investments. The selection function is how to select the specific individual asset holdings within each category.
- **Property operational level:** At this level several breakdowns of the total return are typically considered. For a single short period, or for multiple periods using arithmetic mean time-weighted returns, the total return can be broken into two additive components: income and appreciation. Appreciation within a single short period can be further attributed to income-change and yield-change components. If property operational data is sufficiently available, the income-change component can be further broken down into effects of revenue growth and expense reduction.

3.4 What is the relation between performance attribution and either the ideal research index or the ideal evaluation benchmark index?

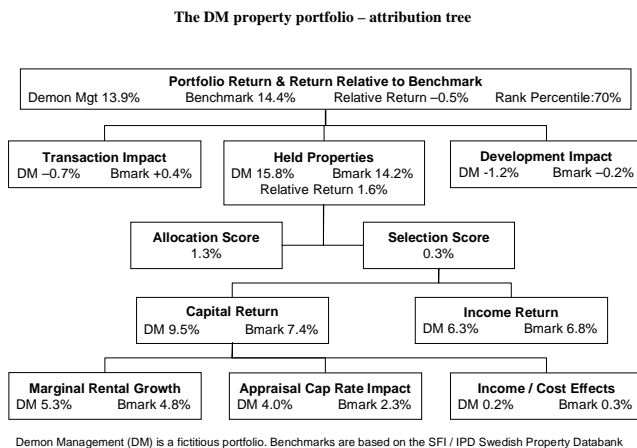
To answer this question let us first distinguish between performance attribution as a function or activity that is performed by investment agents and principals, versus performance attribution as an information product provided by an entity that produces real estate indices. Both research indices and evaluation benchmark indices can provide information that is relevant to carrying out the performance attribution activity or function broadly speaking. However, the evaluation benchmark index plays a more vital role in traditional performance attribution than the research index, because performance attribution is traditionally oriented heavily toward comparisons with an appropriate benchmark. In part, this is because performance attribution is often performed in the context of performance evaluation. Because of this close functional connection between evaluation benchmark indices and performance attribution, performance attribution analysis is often provided as a complementary product to the provision of agent evaluation benchmark indices. Indeed, from a technical production perspective, the evaluation benchmark index is a *byproduct* of the performance attribution analysis product, at least to the extent that performance attribution is defined *relative to* an appropriate benchmark. Both these products are therefore currently typically provided by a single entity that obtains data from the relevant peer universe of investment agents. Thus, performance attribution is more closely associated with the ideal agent evaluation benchmark index (as described in Chapter 2) than with the ideal research index (as described in Chapter 1). (See also Appendix B of Chapter 2.)

3.5 Is performance attribution analysis as an information product a “public good” or a commercial type of product?

Performance attribution as described in this chapter has few “public good” characteristics. It provides information relevant to individual firms and investors. If the benefits performance attribution provides to those private users are worth less than the cost of producing the performance attribution analysis information, there is little or no reason for the industry to subsidize the provision of such information. In other words, performance attribution analysis is a commercial product. It is a natural complement to the provision of agent evaluation benchmark indexing, the product that was described in Chapter 2. Indeed, it may be that the performance attribution information product per se has more profit-generation potential than the evaluation benchmark index product by itself without the complementary performance attribution. To the extent that this is true, then from a commercial perspective, performance attribution is the main product, and the evaluation benchmark index is a byproduct. As noted elsewhere, the evaluation benchmark index does have some characteristics of a public good, particularly if there is no separate asset class research index of the type described in Chapter 1, so that the evaluation benchmark index serves also as the “flagship” research index for the asset class. (See Questions 1.1.4 and 2.B.1.)

3.6 How is performance attribution done in general?

Broadly speaking, performance attribution can be done in numerous ways, and there is no single methodology that is widely recognized as being uniquely correct. Currently, the most common procedure is to break down into additive components the difference between the agent’s total



return and the total return on the benchmark portfolio within a given short period (e.g., a year or less). As an example, the exhibit at left gives a schematic overview of the way performance attribution is currently done in the SFI/IPD Swedish Property Databank.

3.7 How is performance attribution done at the portfolio level?

At the portfolio level the above-noted differential (agent’s return minus appropriate benchmark return) is

presumed to be due to allocation and/or asset selection differences between the agent’s portfolio and the benchmark portfolio. The “pure allocation” effect can be defined as the product of the benchmark return in each category of assets times the difference between the agent’s weighting minus the benchmark portfolio weighting in each category, summed across all the categories. Similarly, the “pure selection” effect is defined as the product of the benchmark weighting in each category times the difference between the agent’s return minus the benchmark return within each category, summed across all categories. These two “pure” effects do not sum to the total differential between the agent’s and the benchmark’s total return. The difference between the agent’s and benchmark’s return that is not attributed to either type of “pure” effect (allocation or selection) is often termed the “interaction effect”. In private real estate applications (unlike the typical public securities application), even the “pure selection” effect is not due solely to the quality of property selection. Rather, it may also reflect the quality of his operational management of those properties.

3.8 How is performance attribution done at the property level?

At the property level, the return in a given short period can be first broken into income and appreciation components. There are alternative methods of calculating this breakdown, depending on whether capital expenditures are subtracted from the income component or the appreciation component. Subtracting capital expenditures from the income component makes this a current cash yield measure and the appreciation return becomes a measure of capital value change between the beginning and end of the period. This is arguably most appropriate (e.g., most comparable to similar measure in the public equity markets) if capital expenditures are financed by the internal operating cash flow of the property. Subtracting capital expenditures from the appreciation return makes the income return a measure more comparable to the “earnings/price” ratio in public equities and the appreciation return then reflects only asset market-based effects on asset value. This approach is arguably most appropriate when capital expenditures are financed largely by external sources of capital, such as debt or additional equity infusions from investors. The latter method (subtracting capital expenditures from appreciation) is the most common in real estate investment performance indices worldwide. In any case, when a property is initially acquired, a relatively high income return or current yield (compared to an

otherwise comparable benchmark property) would presumably be some evidence of successful asset selection on the part of the agent (e.g., a “bargain” acquisition, other things being equal). However, there is no such implication for the income return in subsequent periods after the time of acquisition. (Relatively high income returns during the period of property ownership could be caused by relatively successful income generation or relatively unsuccessful market valuation.) For this reason, multi-period time-weighted average income return components are of questionable meaningfulness or use. (Characterization of properties as being relatively “income-oriented” versus “growth-oriented” in their investment performance during the period in question can be accomplished more directly via other means, such as comparison of time-weighted capital returns.)

For a single period’s return the appreciation component can be further broken down into the effect of changes between the beginning and end of the period in the property’s income and in its yield (that is, the property value/income multiple). If sufficient data are available, property-level performance attribution can be broken down further. For example, the income-change component can be broken down into revenue growth and expense reduction components. At this level, the analysis may provide some insights regarding property operational level management, though care must be taken in interpreting results, particularly from a single short period.

Multi-period performance attribution at the property level should arguably be based on IRRs rather than time-weighted returns. This is because, at the property level (as contrasted to the portfolio level), the agent normally has substantial control over the timing of capital flow into and out of the property (e.g., capital expenditure and leasing decisions). As noted in Questions 7.6 and 7.7 in Chapter 7, the IRR reflects the effect of capital flow timing decisions, whereas the TWRR is neutral in this regard. Thus, much of the basic benchmarking methodology and principles developed for IRR-based benchmarking in the venture capital and private investment industry (e.g., the use of inception-time cohorts as described in Questions 7.12-7.15 in Chapter 7) would make sense as an ideal benchmarking methodology for property-level performance attribution in private real estate. For example, if property-level selection and operations were benchmarked using acquisition-date IRR cohorts, then more meaningful comparisons could be conducted regarding initial (acquisition) yields, income-change effects, and yield-change effects within the overall IRR (between the agent’s properties and an appropriate benchmark of similar properties).

3.9 Is it necessary that performance attribution components all add up to the total performance differential at each level?

A characteristic of the way performance attribution analysis services are currently provided is that all the attribution components typically add up to the total differential between the subject agent’s performance and the appropriate benchmark index performance not only within any given “level” of the breakdown but vertically through all the levels as well. This enables the ultimate portfolio or fund-level performance differential to be entirely “attributed” to the various levels and components of the attribution analysis. For example, the revenue-growth, expense-change, and yield-change effects would sum to the appreciation return, which in turn would combine with the income return to sum to the total return at the standing portfolio level. This type of comprehensive vertical articulation is sometimes useful. In other cases, however, it may

be more useful to simply compare a measure that is most indicative at one level with the equivalent measure in the benchmark, without worrying about vertical and horizontal additivity. For example, to gain some insight on a given portfolio's relative performance in leasing, it may be more useful to simply compare rents or net income per square foot between the subject agent's portfolio and the benchmark portfolio of similar properties, rather than trying to use additive return components such as income returns.

3.10 How useful is performance attribution?

From a management perspective, we believe that performance attribution can be very useful as an internal diagnostic tool for property owners and managers. As a component of investment agent evaluation, performance attribution should be used much more carefully. From a technical perspective, we believe that the usefulness of performance attribution in the quantitative evaluation context remains a somewhat open question. However, if carefully used, it can probably often provide a useful input as part of a broader evaluation process.

3.11 What are the technical questions about the usefulness of performance attribution?

Performance attribution claims to be a formal, quantitative tool to help understand the causes or sources of historical investment total return results. As a purely mechanistic, formal mathematical device, this is straightforward enough if one can agree upon relevant definitions of the return breakdown components and attributes. The technical questions about the usefulness of such an exercise arise from two major considerations:

- **Questionable meaningfulness of attribute measures:** It is sometimes difficult to define meaningful, unambiguous return component breakdowns. For example, there is no way to break down the differential between a portfolio total return and a benchmark total return into two additive components one of which uniquely and purely reflects the effect of portfolio allocation decisions and the other of which uniquely and purely reflects the effect of asset selection decisions on the part of the portfolio manager. Similarly, it is difficult to define unambiguous or unique breakdowns at the property level to distinguish among the effects of initial yield, income-growth, and yield-change on the multi-year differential performance of a set of properties with respect to a benchmark.
- **Difficulty in making reliable inferences:** To the extent performance attribution is used formally in the evaluation of agent investment performance, it can be argued that users are engaging in an exercise in statistical inference. That is, they are trying to make inferences about the ability of agents to perform various types of activities or functions that affect the total return achieved by the agent relative to a benchmark. But as noted in Question 2.4.11 (and elaborated upon further in Questions 5.10-12), the inherent randomness (or "noise") present in ex post (historical) returns renders inferences about performance relative to a benchmark drawn from short samples of historical returns typically unreliable in a rigorous statistical sense. This problem is more severe in private real estate than in the public securities investment industry because empirical real estate return measures are subject to additional types of error (see Questions 5.10 and 5.12).

3.12 How is performance attribution more useful from an informal, management perspective?

The technical questions noted in the preceding answer suggest that quantitative performance attribution needs to be “taken with a grain of salt”. In particular, one should be wary of using performance attribution in an excessively mechanistic manner, without ample consideration given to other relevant qualitative or less formally quantifiable evidence. With this understood, it is nevertheless clear that performance attribution is a method of obtaining additional data, analysis, and perspective on investment performance and that as such it can certainly be used for a variety of useful purposes both by investment agents for internal diagnostic purposes and by principal parties in the process of agent performance evaluation. We note three such uses here:

- **Agent performance evaluation:** When applied in a careful, non-mechanistic manner, performance attribution can be helpful in the overall performance evaluation process. In particular, recall from Questions 2.1.3 and 2.1.4 that agent evaluation benchmarking serves several important *ex ante* functions (such as aiding communication and interest-alignment), and that use of benchmarking in *ex post* quantitative analysis should to be accompanied by additional qualitative information. Appropriate use of performance attribution can sometimes be helpful in providing such additional information and perspective.
- **Manager internal diagnostics:** Performance attribution can help managers to obtain insights and perspectives relevant to the nature and causes of their realized investment performance. Such insight need not necessarily derive directly from the formal or mechanistic use of performance attribution, but may result from lines of thought and analysis that are triggered by the performance attribution process.
- **International standardization:** Formal performance attribution as part of the benchmarking process has spread rapidly in leading overseas real estate investment markets. This type of analysis has been hampered in North America by the lack of fund-level and performance attribution information in the NCREIF and Russell-Canadian Property Indices. As the real estate investment industry becomes ever more global, it is increasingly useful to be able to apply similar procedures in all countries.

Chapter 4:
Property Sampling Issues in the Ideal Research Index

4.1 What determines the usefulness of a sample for purposes of defining a real estate investment research index or database?

Statistics theory says that the power and usefulness of a sample drawn from a large population is a function of three characteristics:

- (i) How much *random dispersion* exists in the underlying population;
- (ii) The *absolute* size of the sample (e.g., 500 in the case of the S&P500, *not* the percentage of the population included in the sample);
- (iii) The *representativeness* of the sample (i.e., the lack of *bias* in collecting the sample).

Furthermore, as a general rule, the effect of the absolute size of the sample is a function of the *square root* of the number of *independent* observations in the sample. This is known as the “***Square Root of n Rule***”. Thus, for example, a sample that is four times larger tends to only double the statistical accuracy or inference power (e.g., half the typical estimation error magnitude or range in a given confidence bound). In other words, *declining benefits to scale* exist for sample size increments, other things being equal. (Sometimes there may also be economies of scale in collecting data, at least up to a point, so that the benefit/cost ratio of expanding the sample size may be more favorable up to a point.)

4.2 What are the implications of sampling theory for efficient sampling of large, diverse populations?

Because of the declining returns to scale implied by the Square Root of n Rule, it may be a waste of effort and resources to attempt to include all or even a large fraction of a large population in a statistical sample useful for research purposes. Rather, if the population is diverse and complex, it may be more efficient to *stratify the sample*. Sample stratification refers to the identification of “strata” or “cells” consisting of sub-populations or segments of the overall population that are relatively homogeneous within each cell. For example, one cell could be office properties in Florida. It is necessary to sample from each cell, but within each cell the Square Root of n Rule applies.

4.3 What are the implications of scientific sampling theory for an ideal research index of the U.S. commercial property market?

Suppose the total population of U.S commercial properties consists of 1,000,000 properties. Now suppose these properties can be usefully classified into, say, 1,200 relatively homogeneous market segments or cells (e.g., 50 MSAs X 6 property types X 4 size or sub-type categories = $50 \times 6 \times 4 = 1200$). The Square Root of n Rule could be invoked to establish that a sufficient sample size within each cell is, say, $n=50$. Then the research index need only include $1200 \times 50 = 60,000$ properties, only 6% of the underlying population. Such an index can be designed to be far more useful than an index with, say, 200,000 properties that were drawn haphazardly from the population, with over-representation from some cells and no representation at all from others.

As we move up the strata (to higher aggregates of the sub-indices), to include for example all properties in a given MSA, the relevant sample size increases, leading to greater statistical power. (In our previous example the typical MSA would include $6 \times 4 \times 50 = 1200$ properties, which would give $\sqrt{1200/50} = 5$ times the statistical power of each 50-property cell.) In order to aggregate properly, however, we need to know the relative sizes of each cell as a proportion of the total population. Thus, scientific sample stratification requires not only a determination of the minimum sample size in each cell, but the relative magnitude of the population within each cell. For example, we would like to know the relative market value of each MSA, each property type within each MSA, and each sub-type or size category within each type. In general, it is considered acceptable to estimate the relative cell sizes without recourse to an entire population census. Nevertheless, some method of surveying or otherwise estimating the relative cell magnitudes would need to be applied.

4.4 How big a property sample would be necessary for a passive market index of U.S. commercial property investment performance?

The answer to this question depends on a number of considerations. At a general level, the most important of these considerations include:

1. What types of research questions will the index be optimized to address?
2. How much random dispersion is in the underlying population regarding the variables that are important for addressing the target research questions?
3. How much standard error (e.g., in terms of a 95% confidence interval around the best estimate) is acceptable in the index reports?

In the case of an index whose primary purpose is to report estimates of the periodic returns or percentage price-changes in a commercial property market, there is some evidence that cell sample sizes of a few dozen repeat-observations per index period are sufficient to produce useful return indices, based on the repeated-measures regression (RMR) methodology combined with noise-filtering techniques such as the Bayesian use of ridge regression. (See the article by Gatzlaff & Geltner in the Spring 1998 issue of *Real Estate Finance*, and the article by Geltner & Goetzmann in the July 2000 issue of *The Journal of Real Estate Finance & Economics*.)

4.5 Can you give a simplified numerical example of how the necessary cell sample size is determined for a return index useful for research purposes?

Suppose property income flow is observed each period without error. Suppose *each* property is valued *independently* at the end of *each* period of time. Suppose each individual property valuation observation (whether this be based on an appraised value or a transaction price) has a standard error of 5% of the property value. (That is, 5% of the property value is the standard deviation of the cross-sectional dispersion in the observable price distribution around the theoretical true market value of the property, as of any point in time when the property is valued or a value observation is made.) Suppose further that we want the raw index (without any Bayesian noise filtration) to have a standard error in each periodic index return report equal to one-half the true periodic return volatility. Finally, assume the true periodic volatility is 2%. This means that we want each individual periodic return report in the index to have a standard error of

1% (100 basis-points). Assuming normally distributed errors, each periodic return reported by the index would therefore have about a two-thirds chance of being within 100 basis-points of the true return that period.

With independent valuations, the index value level standard error is 5% divided by the square root of n . With longitudinally independent errors (errors independent across time), the index return standard error is the square root of twice the square of the index value level standard error. (Index returns are essentially like the arithmetic differences in the consecutive value levels. This is precisely true working in log-levels.) Thus, the index return standard error is $\sqrt{2(5\%/\sqrt{n})^2}$. Setting this equal to 1% and solving for n , we find:

$$n = \frac{(5\%)^2}{(1\%)^2 / 2} = 2(5^2) = 2(25) = 50.$$

That is, the index requires a cell sample size of 50 properties (valued independently each period). This result holds no matter how many properties are in the underlying population within the cell. (This assumes the cell population is large. If the cell population is very small, then we might get the necessary 1% standard error with a smaller sample size.)

4.6 Is sampling theory also relevant for evaluation benchmark indices, in addition to research indices?

In principle, yes. However, as noted in Section 2.2 of Chapter 2, evaluation benchmark indices are based on the peer universe concept, and for the reasons noted in Section 2.3 of that chapter it is traditionally considered desirable to base an index that is used to evaluate agents on as close to a population census as possible, rather than on a statistical sample (see especially Question 2.3.6).

Chapter 5:
Basic Asset Valuation and Statistical Considerations for Real Estate Indices

5.1 What is the conceptual difference between transaction price, market value, and appraised value?

Transaction prices and appraised valuations are empirically observable values. They occur (or *exist*) only when and if a property transacts or is appraised.

Market values are theoretical or conceptual constructs. They always exist for every property, though any given property's (or a portfolio of properties') market value generally changes continuously through time. Market values change continuously because news arrives continually relevant to asset values (e.g., news about the economy and the capital markets affects the values of all capital assets in various ways, as well as news specifically about any given asset).

The market value of a property as of a given point in time is frequently defined in real estate markets as the "most likely" (or the *expected*) transaction price of the property as of that point in time. It is therefore the *mean* of the ex ante transaction price probability distribution as of the stated time. It is thus the opportunity cost of holding onto the property rather than selling it.

Market value defined in this way also closely approximates the actual transaction price at which we would expect to empirically observe transactions in a highly liquid, dense market where homogeneous assets are bought and sold very frequently by numerous buyers and sellers, such as the trading of equity shares in the stock market. Market values represent "market clearing prices" at which the number of buyers equals the number of sellers for homogeneous assets.

In private real estate (much more so than in the public securities business) asset market values are *not empirically observable*, unlike transaction prices or appraised values. This is because in private real estate whole assets are traded, and (like whole assets anywhere as distinct from shares) these assets are each unique, and traded infrequently. This does not imply that market values do not exist. Market value is a well-defined and useful construct for practical decision-making. As noted, market value equals opportunity cost, what the property owner is giving up by *not* selling the property, and what a potential buyer can expect to have to pay to obtain it.

5.2 What is meant by transaction price dispersion?

By definition, observable transaction prices are individual draws from underlying probability distributions that are centered around the unobservable true market values of the properties being transacted. Transaction prices at any given point in time therefore exhibit ***cross-sectional dispersion*** around the underlying (unobservable) true market values as of that point in time. In other words, any given transaction price is likely to differ from the (unobservable true) market value of the property being transacted. This difference between the observable transaction price and the unobservable true value is often referred to as "*transaction price noise*", or "*transaction price error*" (but note that this does not imply that anyone has done anything wrong). Cross-sectional dispersion occurs because no one can observe the true market value of the property being traded, so neither side in the negotiation knows exactly at what price the property "should"

trade. Each side seeks its best deal. In a given transaction, one side or the other will typically end up getting the better deal (though it will usually be impossible to know which side it is). One side may have a little better information, or be under a little less pressure to close the deal, or simply more skillful at negotiation.

5.3 What is the difference between appraisal error and transaction price error?

Similar to transaction prices, appraised valuations are also dispersed cross-sectionally around true market values as of any given point in time. If you hire any two appraisers to appraise the same property as of the same point in time, and prevent them from communicating with each other, they will almost certainly not arrive at exactly the same estimation of market value. At least one of them must be “wrong”, in the sense that their valuation differs from the true market value of the subject property, because conceptually there is only one true market value for a given property as of a given point in time. (Most likely *both* appraisers are “wrong” in this sense.) The difference between a given empirical appraised value and the (unobservable true) market value as of the same point in time is called “*appraisal error*”, even though there is no implication that the appraiser has exhibited any incompetence, negligence, or impropriety.

Although appraised values are, like transaction prices, dispersed around the underlying true values, unlike transaction price dispersion, the appraisal value dispersion is not necessarily *centered* on the true value. This implies that appraised values may be *biased*. That is, the ex ante mean (or expected) appraised value may not equal the true market value. Such bias may result from very rational behavior on the part of the appraiser given the nature of the empirical information available in the real estate market (or in any case, the bias may result from very *human* behavior). A growing body of academic literature is studying, documenting, and describing the nature of this appraisal bias. One of the major types of bias that is most likely is that the appraised value tends to be lagged temporally behind the true contemporaneous market value. (See, for example, the article by Chinloy, Cho & Megbolugbe in the January 1997 issue of *The Journal of Real Estate Finance & Economics*, the article by Diaz & Wolverton in the Summer 1998 issue of *Real Estate Economics*, the article by Hamilton & Clayton in the Fall 1999 issue of *Real Estate Finance*, and the article by Fisher & Geltner Spring 2000 issue of *Real Estate Finance*.) This is referred to as *temporal lag bias*.

5.4 Is the evidence for temporal lag bias in appraisals uncontroversial and widely accepted?

In our opinion, the case for temporal lag bias in appraisals is strong, but this is not a closed question in the academic research literature. Appraisal lag bias is still a hypothesis that is being actively investigated, both empirically and clinically. Several studies have been published in the academic literature finding evidence of the existence of temporal lag bias in appraisals. These studies cover widely different types of appraisals and different types of empirical or clinical evidence, and use a variety of different analytical methodologies. In contrast, we know of no empirical or clinical research that has been published that claims to find to the contrary, that temporal lag bias does not exist. Theoretical normative and behavioral models of appraisal have also been put forth in the academic literature consistent with the temporal lag bias hypothesis (e.g., the article by Quan & Quigley in the June 1991 issue of *The Journal of Real Estate Finance & Economics*). Although the appraisal lag bias hypothesis seems to be widely accepted

in the academic community, a few academics do challenge it (see notably the article by Lai & Wang in the Fall 1998 issue of *Real Estate Economics*).

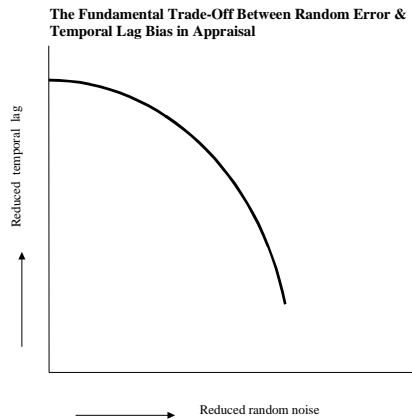
5.5 Can you give a simple numerical example of why or how appraisal temporal lag bias occurs?

There may be more than one source or type of appraisal temporal lag bias. The following illustrates one major hypothesis. Suppose you hire an appraiser who faces the following choice. He can give you a value estimate based on “Method A” which will be unbiased (meaning that the most likely and expected value of the appraisal equals the true market value), but this value estimate has only a 50% chance of being within 10% of the true market value. Alternatively, the appraiser can give you a value estimate based on “Method B” which will have a little bias. In particular, the expected value of the Method B appraisal equals the true market value from six months ago, or the value halfway between the current true market value and the true market value a year ago. For example, if true market values were 2% lower a year ago, then the Method B appraisal has a 1% bias on the low side. However, Method B has less purely random error. There is a 75% chance that the Method B appraisal will be within, say, 5% of the current true market value of your property. Furthermore, Method B will provide you with more solid historical evidence explicitly documenting the estimated value (e.g., more “comps”). The appraiser says he will charge you the same price for either method, and the choice is yours. Which method would you prefer?

The most typical answer is Method B, because it will probably give you a more useful estimate of value, *provided* you are concerned about *documenting* the value of the *single individual* property that is the subject of the appraisal. You are willing to accept a little bit of temporal lag bias in order to get a more precise and well-documented estimate of value. Of course, if you own the property and are in the process of negotiating a sale price, you would probably hesitate to sell for a price equal to the appraisal estimate because you would probably be aware that market values have been increasing over the past year (in this example). Furthermore, if you are not concerned about the value of that single, subject property, but rather you care only about the overall *aggregate* value of a large portfolio of properties, and if you are particularly concerned to obtain an *up-to-date* estimated value of the portfolio, then you should prefer the appraiser’s offer of Method A for the individual property in question. After all, the purely random error in Method A will diversify out of your overall aggregate portfolio valuation, thanks to the Square Root of n Rule. Nevertheless, the standard practices and procedures of the appraisal profession in the U.S., and the typical method of employing and using appraisers in the institutional real estate investment industry in the U.S., probably results in appraisal methods more similar to B than A.

5.6 Why does there tend to be a trade-off between random error and temporal lag bias in property valuation estimation?

Because in order to reduce purely random error the appraiser needs to increase the sample size of transaction price observations he considers in estimating value. (This trade-off is depicted in the



figure, and it is yet another example of the Square Root of n Rule.) In order to increase his sample size of transaction price observations the appraiser has to reach farther back in time to find sales of comparable properties. This introduces bias whenever market values have been changing over time (which is usually), because the appraiser can only imperfectly observe and adjust for such longitudinal market changes, and both theoretical and practical considerations generally dictate that the appraiser (differently from the market participant) make a conservative

adjustment for market value changes. Thus, although appraised values may have less purely random error than transaction prices (narrower cross-sectional dispersions), they exhibit temporal lag bias.

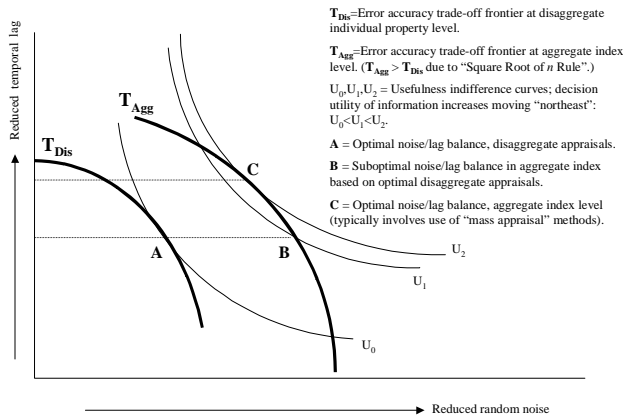
5.7 What are the implications of these different types of errors (appraisal error vs transaction price error) for indices based on different types of valuations and used for different purposes (research vs evaluation benchmarking)?

In an index or portfolio, where many individual property valuations are summed together, purely random error tends to diversify away (via the operation of the Square Root of n Rule), resulting in less noisy valuations at the index or portfolio level than at the individual property level. (This is one reason why the NCREIF Index appears so smooth across time.) However, appraisal bias is *systematic*, operating similarly on all individual property appraisals as of a given point in time. Therefore, temporal lag bias does not dissipate in the aggregation of individual appraisals into a portfolio or index valuation. As a result, transaction based indices often exhibit less temporal lag bias than appraisal-based indices, though this depends on exactly how the transaction price index is constructed (see Question 6.4.5 in Chapter 6).

In contrast, appraisal-based indices tend to exhibit less “noise” (or *purely random error*) than transaction price based indices. In part this may be because appraised values exhibit less cross-sectional dispersion than transaction prices at the property level. In addition, appraisal-based indices may sometimes benefit from the fact that appraisals are normally much more frequent than transactions within a given set of institutionally-owned properties. This causes appraisal-based valuation data to be much more plentiful than transaction price data, resulting in a larger effective sample size of valuation observations for indices representing a given peer group of institutional owners.

Methods can be devised to either adjust appraisal-based indices to correct for temporal lag or to adjust transaction-based indices to filter out random noise. (See for example the article by Fisher & Geltner in the Spring 2000 issue of *Real Estate Finance*.) Such adjustments often make sense because the optimal trade-off between bias and purely random error may be different at the

Fundamental appraisal methodology and economic considerations in real estate index construction: *The interaction of statistics and economics...*



aggregate (index) versus disaggregate (individual property) levels of valuation (see exhibit), as well as for different types of indices. In particular, random error is relatively more problematical in evaluation benchmark indices than in research indices (see 2.3.7). In contrast, temporal lag bias is often a greater problem than random noise in research indices, at least for addressing certain types of research questions (see Questions 1.1.8, 1.3.7, and 1.5.2 in Chapter 1). Studies relating to real estate return time-series second moments (volatility, systematic risk,

correlations, beta), and studies relating to current market movements, lead/lag relationships, and forecasting of asset market returns, are especially sensitive to temporal lag bias, but often are less affected by random noise. Thus, optimal index design is typically different for research indices as contrasted to evaluation benchmark indices.

5.8 Can you give a numerical example of the difference between transaction price, market value, and appraised value, showing the nature of purely random cross-sectional valuation "error"?

Assume properties A, B, and C are very similar physically, similarly located, and have very similar tenants and lease profiles. They have the same true market value. Assume also that each has a true market value of \$10,000,000.

Now suppose all three properties happen to sell as of the same date. Is it likely that each will sell for exactly the same price? No. Is it necessarily true that any one (or more) of these three sales transactions will occur at exactly the \$10,000,000 true market value? No. Is it very likely even that the arithmetic average of the three sales prices will *exactly* equal the true market value of \$10,000,000? No. For example, Property A might sell for \$9,900,000, Property B for \$10,500,000, and Property C for \$9,750,000 (or whatever). (Note that the average of these three prices is \$10,050,000.)

Suppose that, just prior to these three transactions, three appraisers were hired to independently appraise each of these three properties. Is it likely that all three will appraise each property at exactly the same value? No. Is it necessarily true that any one (or more) of these three appraised values will exactly equal the \$10,000,000 true market value? No. Is it very likely that the

arithmetic average of the three sales prices will exactly equal the true market value of \$10,000,000? No. For example, Property A might be appraised at \$9,800,000, Property B at \$10,100,000, and Property C at \$10,050,000. (Note that the average of the above three appraisals is \$9,983,333.)

In this example, is the true market value of \$10,000,000 (for each of the three properties) empirically observable by any appraiser or any market participant? No.

5.9 What is the relative importance of temporal lag bias vs random error in an asset class research index and in an evaluation benchmark index?

The key determinant of the relative importance of different types of valuation errors in an index is the purpose for which the index is being used, and the resultant types of statistics that need to be accurately measured. Temporal lag bias is most problematic in trying to answer certain questions that tend often to be important for a research index to be able to address, but which are typically not important for an evaluation benchmark index. Temporal lag bias can greatly affect the analysis of questions related to, for example: current market directions (e.g., Did the market turn around last quarter?); the relative magnitude of market value change since a turning point; the nature of lead/lag relationships; correlations; forecastability (including questions of asset market informational efficiency); and possibly the quantification of asset class risk measures such as beta or volatility. Therefore, appraisal lag bias or smoothing has to be recognized in the usage of a real estate research index. In contrast, as noted in the answer to Question 2.3.7, appraisal lagging tends to be less important in the evaluation benchmarking process, and in the construction and use of evaluation benchmark indices.

5.10 Are there other sources of randomness in real estate indices besides measurement errors?

Measurement errors caused by appraisal value dispersion or transaction price dispersion (or both) are an important source of randomness in real estate investment performance indices. But they are not the only sources of randomness. Including measurement error, there are three sources of randomness in realized investment performance results:

- 1) *Cross-sectional* randomness in the true performance of individual properties;
- 2) *Longitudinal* randomness in the true aggregate asset market performance over time; and
- 3) *Measurement error* in both individual property and market aggregate performance.

Of these three, the first two are present in indices of public securities returns, while the third source, measurement error, is a major problem only in real estate indices.

5.11 Why is randomness of any source an important consideration particularly in evaluation benchmarking?

Each of the sources of randomness noted in the preceding question can cause realized (ex post) investment results to differ from valid expectations (ex ante). This makes it difficult in practice to use ex post return data to rigorously identify superior agents. Even if agents' true relative abilities were constant over time, three to five years of performance data are rarely enough to

overcome the three sources of randomness noted in Question 5.10. The result is that, from a rigorous statistical perspective, it is very difficult to distinguish superior performance due to skill from superior performance due to luck. Yet it is only the former type of superior performance that is likely to repeat consistently over a long period of time. Therefore, as noted in Question 2.1.4, quantitative benchmarking should be used carefully and in conjunction with other information in making judgments about agents' abilities. This point is important enough that at least one major national industry association, the Property Council of Australia (PCA), officially issues the following "health warning" in their *Investment Performance Measurement & Presentation Standards* that govern their benchmarking service:

"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."

5.12 Can you elaborate on the nature of the three sources of randomness in *ex post* return measurement in real estate?

Cross-sectional randomness reflects the idiosyncratic or specific risk unique to any given property. No two properties will provide the same *ex post* investment performance, even if they are very similar physically and in similar types of locations. At least part of the idiosyncratic differences in *ex post* performances across properties is random in the sense that it could not be reasonably predicted in advance and may be due effectively to investment "luck" (either good or bad).

Longitudinal randomness reflects the volatility in *ex post* returns even for well-diversified aggregate market segments or styles. Not all of this volatility is reasonably predictable in advance, nor could any such volatility be avoided by a "style-pure" or segment-constrained investment manager.

Measurement error reflects the difficulty of defining and accurately measuring the true asset market value-based total return performance for a given real estate portfolio or market segment, and even more so for any given individual property. Such measurement difficulties are particularly acute in the private real estate investment field because unique, whole assets are traded infrequently and privately in relatively illiquid markets, making the observation of market values an inexact science that can only be treated rigorously as an exercise in statistical inference.

Chapter 6:
***Property Revaluation Frequency, Index Return Reporting Frequency,
and the Ideal Index Construction Methodology***

Section 6.1:
Overview and Basic Definitions

6.1.1 What determines the statistical qualities of a real estate return index?

The dynamic statistical quality of a return index refers to the type of periodic return time-series statistics that can be computed from the index, as well as the quality of those statistics. For example, how frequently can return statistics be calculated? Does the index tend to be very “noisy”? To what degree do its periodic returns tend to exhibit temporal lag bias? The four major attributes (or dimensions) of the index which interact together to determine its dynamic statistical quality include:

1. Return reporting frequency;
2. Frequency of revaluation observations per property;
3. Number of properties used in index construction;
4. Methodology used to construct the index from the underlying disaggregate valuation observations.

6.1.2 What is meant by index return reporting frequency?

The index return reporting frequency refers to the frequency with which the index reports investment returns (or price-changes); for example, annually, quarterly, or monthly. Higher frequency implies shorter individual return periods.

6.1.3 What is meant by property revaluation frequency?

In an appraisal-based index, revaluation frequency refers to how often the properties composing the index are reappraised. In practice, “reappraisal” has multiple definitions, ranging from superficial “desk-top” appraisal “updates” to full independent appraisals. For the purpose of determining or understanding the important dynamic characteristics of an index, superficial reappraisals do not have much effect. As a result, in this chapter we generally use the term revaluation or reappraisal frequency to refer to “serious” reappraisals, which generally means more than just a cursory update of value. (See Question 6.2.3 below.)

In a transaction-based index, property revaluation frequency might refer to the average number of transaction price observations per index reporting period, a good measure of the effective average sample size for drawing statistical inferences about periodic returns. In the case of a repeated-measures regression (RMR) based index, the valuation observations would be repeat-valuation observations.

6.1.4 What is meant by number of properties in the index?

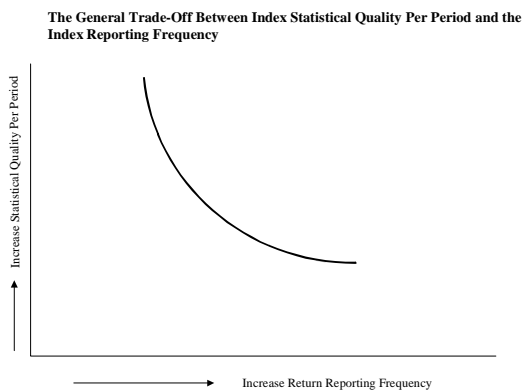
In an appraisal-based peer universe index, the number of properties simply refers to the number of properties whose values are regularly reported by the data-contributing members of the index.

In a transaction price based index the number of properties might refer to the total pooled number of transaction price observations in the index estimation database, including all the transacting properties across all the history of the index. In the case of a repeated-measures regression (RMR) based index, the number of properties in the index might refer to the cross-sectional number of *different* properties (each of which transacts at least twice, providing at least one repeat-transaction value observation).

6.1.5 What is meant by index construction methodology?

Index construction methodology refers to the procedure by which individual disaggregate valuation observations are aggregated to compute the reported index return each period. In the case of an appraisal-based index, this may be as simple as the summing of all the constituent properties' current reported values and comparing that with the sum of the same properties' valuations the preceding period. However, modern statistics provides more sophisticated procedures for producing transaction price based indices and for dealing with problems such as stale appraisal reports in appraisal-based indices.

6.1.6 In general, how do these attributes affect the dynamic statistical quality of the index?



For a given index reporting frequency, statistical quality improves *per index reporting period* the greater the frequency of property revaluation, the greater the number of properties used, and the more statistically efficient and effective the index construction methodology employed. On the other hand, holding these other characteristics constant, the greater the index return reporting frequency, the lower the dynamic statistical quality *per reporting period*, but the more frequently the return statistics can be calculated. Thus, broadly speaking, there is a general trade-off between statistical

quality per period and the frequency of index reporting, holding constant the overall quantity and quality of raw valuation data and index construction methodology. This trade-off is depicted in the figure.

6.1.7 How does the index frequency and historical span relate to the statistical power of the index?

We always come back to the “Square Root of n Rule”. The power of statistical analysis is roughly proportional to the square root of the number of *independent* observations in the database. Much (but not all) investment-oriented research focuses on *time-series* analyses (that is, analysis *across time*, rather than across properties within a single period of time). For such time-series analyses, the number of observations equals the number of periods of time covered by the index. For example, a 20-year quarterly index has 80 return observations, while a 10-year annual index has 10. However, it is important to note that real estate asset markets tend to have *inertia* because they lack informational efficiency compared to public securities markets. This is usually so even apart from, or in addition to, any temporal lag bias in appraisals, as described in Questions 5.3-5.7 in Chapter 5. As a result, individual periodic return observations are not statistically independent. Rather, they are positively correlated with the preceding and subsequent returns. This positive correlation (lack of statistical independence) is greater the shorter is the frequency of the periodic returns in the index. Correlation may be increased or decreased depending on the nature of the valuation process and the index construction methodology, as described below (see especially Questions 6.3.7, 6.3.9, and 6.4.5 below). Thus, a 20-year quarterly index does not provide 4 times the number of *independent* observations, and therefore does not provide 2 (that is, $\sqrt{4}$) times the statistical power of a 20-year annual index.

6.1.8 How do the index statistical quality determinants interact in an appraisal-based index?

In principle, all four of the attributes listed in Question 6.1.1 can be varied independently in an appraisal-based index. That is, the index producer can separately determine the index return reporting frequency, the constituent property revaluation frequency, the number of properties in the index, and the methodology used to construct the aggregate index from the underlying individual property valuation observations. In practice, of course, budget and institutional constraints restrict the range of choice in all four dimensions. In an appraisal-based index, either the index producer or the data-contributors directly or indirectly bear the cost of property-level appraisals. As a result, greater frequency of revaluation observation per property (greater appraisal frequency) increases index production costs per property included in the index.

6.1.9 How do the statistical quality determinants interact in a pure transaction-based index?

The frequency of revaluation observations per property is generally beyond the control of the index producer in a direct transaction price based index. This is because properties are sold at the discretion of their owners, not for the sake of providing input data for an index. As a result, in a transaction-based index only three of the four quality-determining attributes can be manipulated: return reporting frequency, number of properties (sample size), and index construction methodology. On the other hand, there are no appraisal costs associated with a pure transaction price based index.

6.1.10 Is this chapter about appraisal-based indices only?

The primary focus of the present chapter is on appraisal-based indices. However, some of what is addressed here is also relevant to transaction-based indices. We shall endeavor to note these points of relevance.

Section 6.2:

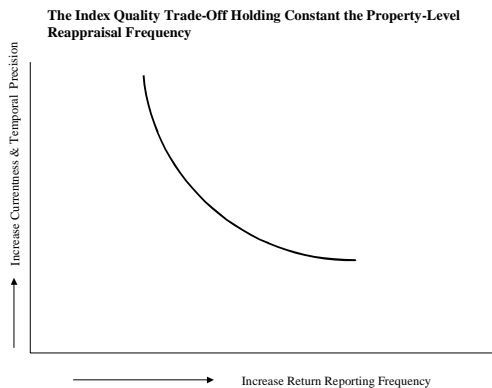
Index Reporting Frequency and Property Revaluation Frequency: The Basics for Appraisal-based Indices

6.2.1 Must the index return reporting frequency equal the property revaluation frequency of the constituent properties in the appraisal-based index?

Not necessarily. In some conditions an index can report returns more frequently than its constituent properties' revaluation cycles. For example, an index could report quarterly returns when its properties are being revaluated annually. However, there are some important trade-offs and methodological issues that should be considered in addressing this question.

6.2.2 What are the major trade-offs and considerations regarding the relationship between index return reporting frequency and property-level reappraisal frequency?

Suppose you have a population of properties each of which is reappraised once per year. If all properties are reappraised *as of the same time every year* (e.g., as of December 31), you can produce an *annual frequency* index that will report *completely up-to-date* annual returns each year based on your entire population of property revaluations. That is, there will be no lagging or smoothing in the index relative to contemporaneous appraised values. (However, the possibility



that the appraised values themselves are lagged behind purely contemporaneous transaction price-based evidence is another question, not addressed here. See Questions 5.3-5.7 in Chapter 5.) With this annual-common-date procedure you won't be able to produce an index with any greater frequency than annual, because you will have no valuation observations as of intermediate times within the year. Suppose, on the other hand, that the appraisals are still once-per-year at the individual property level, but staggered throughout the year in the population of properties. For example,

some properties are on a January reappraisal cycle, some are on a February reappraisal cycle, and so forth. Then you can produce a quarterly or even monthly frequency index. However, if you insist on constructing such an index as a simple aggregate of *all* the properties' reported values each quarter (or month), then the index will suffer from a temporal lag bias due to a large

proportion of effectively “stale” valuation reports in the index each quarter (or month). Properties last fully reappraised in January will still be entering the index the following December at values equal to, or not much different from, their previous January valuation. This *stale appraisal effect* results in temporal lag bias in the index even if the individual appraisals have no lag bias when they are conducted on each property. The lag bias in the index is greater, measured in average number of reporting periods of lag, the greater the reporting frequency. For example, with annual property-level reappraisals, the average lag in an annual index would be one-half period, 2 periods in a quarterly index, and 6 periods in a monthly index. The nature of this trade-off is depicted in the exhibit.

6.2.3 But why can’t the property-level data-contributors update their property valuation reports between major annual reappraisals?

They can, and may, make some effort to do so. But there are two fundamental problems in any such effort:

- The first is the *expense* of conducting an update that is *temporally independent*. In order for an appraisal update to be “temporally independent”, it must not rely *at all* on the previous value report (or on any prior value report) for the subject property. This is not merely an “independent” value report in the legalistic sense that an external fee appraiser signs off on the update. Temporal independence for the purpose of avoiding a stale appraisal effect in a real estate index requires that the person estimating the updated value be entirely unaware of what any previous value report was for the subject property. Superficial updating of the value reports (so-called “desktop appraisals”) between the annual full (or “clean-slate”) reappraisals will be inadequate, because they will inevitably rely heavily on prior valuation reports for the subject property. (See Questions 5.3-5.7 in Chapter 5.)
- The second problem is that there is often *too little purely contemporaneous information*, within the *current* quarter or month, for the appraiser to get a sufficiently precise indication of a given individual subject property’s value. Unless the property is of a very common or homogeneous type that trades very frequently, there may be little or no transaction price evidence available purely from the current period. This is a bigger problem the shorter is the current period (e.g., a bigger problem for monthly reappraisal than for quarterly, and a bigger problem for quarterly reappraisal than for annual). This makes it difficult for the appraiser to avoid the use of stale information.

6.2.4 How expensive is it to conduct temporally independent property-level appraisals more frequently than once per year?

A “full” (or “clean-slate”) appraisal at the individual property level (i.e., when conducted separately for each property) typically costs about 2/10000 of the value of the property for “institutional size” properties in the U.S., or about two basis-points. For example, the appraisal fee for a typical \$30 million property might be about \$5,000, with perhaps another \$1,000 administrative and clerical costs associated with the appraisal born directly by the property owner in addition to the appraiser’s fee. To place this cost in an appropriate relative scale of magnitude, consider that investment manager fees in the private real estate industry (which are of course expenses to their client investors) run typically in the range of 65-100 basis-points per

year. With full appraisals conducted on each property once per year, appraisal costs are about 3% of total management expenses (e.g., 2 basis-points for annual appraisals as a fraction of 75 basis-points total management costs including the manager's profit). If full appraisals were conducted *every quarter on every property*, the annual appraisal costs would be multiplied by four, to eight basis-points, or over 10% of the total investment management costs.

It should be noted that the additional appraisal costs would not necessarily be born entirely by the investment manager (agent), or entirely by the investor (principal). The point is that these are real costs (a "deadweight burden" in economic jargon) that will have to come out of some combination of either the manager's profits or the investor's return, the latter to the extent managers are able to pass the extra cost on to their clients. Comparing incremental appraisal costs to the magnitude of investment management expense is simply a good way of conceptualizing the scale of these costs, relative to other recurring fund-level (or portfolio-level) administrative costs of property investment.

6.2.5 Why can't the problem of stale appraisals be solved by just requiring all index data-contributors to fully reappraise all their properties every period?

This might be referred to as the "*brute force*" technique. The answers to the two preceding questions directly relate to this question. Another way of quantifying the potential costs involved is to note that if full independent appraisals were done quarterly at the micro-level on all individual properties in the NCREIF Index, NCREIF data-contributing members (and/or their clients) would incur approximately \$50 million per year additional appraisal expense, some four-times the existing annual appraisal costs within this peer group. In practice, the brute-force technique seems to work fairly well at the annual frequency, for example, in Britain with the IPD Annual Index. But there is evidence that the brute force technique has trouble working at quarterly or greater frequencies. For example, the IPD Monthly Index in Great Britain shows signs of temporal lag bias of the type we would expect from a stale appraisal effect. (The IPD Monthly Index shows some evidence of lagging behind the Annual Index.)

As noted in the preceding questions, there are two major reasons why the brute-force technique has trouble working at quarterly or monthly frequencies: expense, and lack of sufficient current real estate market information. Mass appraisal techniques (see Section 6.3 below) are probably required to break through this barrier. However, mass appraisal requires the *index constructor* to get rather directly involved in the valuation process, at least at the aggregate level or in a coordinating role, rather than relying purely on the data-contributors to independently supply individual property-level valuation information each period. Such active involvement by the index constructing entity may be viewed by some data-contributors as acceptable in the context of a research index (as described in Chapter 1) but not in the context of an evaluation benchmark index (as described in Chapter 2). Nevertheless, if there is sufficient desire for a quarterly-frequency evaluation benchmark index that does not require much additional property-level appraisal expense and yet does not suffer from a stale-appraisal temporal lag bias, then the industry may have to consider such an evolution of the benchmark index construction procedures.

Section 6.3:

Modern Index Statistical Tools and Basic Appraisal Technology & Economics

6.3.1 What is the nature of scale economies in appraisals for index construction?

Economies of scale in appraisal derive from sharing of common information across multiple appraisal reports or multiple properties. When such information sharing occurs *across time* (as in quarterly or monthly “updates” of prior annual appraisals), such scale economies may be referred to as *longitudinal* scale economies. When the information is shared *across properties* cross-sectionally (as of the same point in time), the term “*mass appraisal*” may be employed.

6.3.2 Are there scale economies in increasing property-level appraisal frequency, and what are the implications for research indices vs evaluation benchmark indices?

Scale economies arising from increased appraisal frequency result from appraising *the same property* more frequently *across time*. The only way appraisals of individual properties can obtain such longitudinal economies of scale is by relying on prior appraisal information. This would cause such appraisals to be *temporally dependent* (as defined in Question 6.2.3). Thus, the only way to achieve scale economies from more frequent property-level appraisals is to accept an increase in the temporal lag bias in the average appraisal (see also Question 5.5 in Chapter 5). Such bias does not diversify away at the aggregate level and so would bias the index returns, causing them to lag the true market (see Question 5.7). This is a serious concern for research indices (see Questions 1.3.6 and 5.7). However, as noted in Chapter 2, temporal lag bias is less of a concern in evaluation benchmark indices (see Question 2.3.7). This makes more frequent property-level valuations less of a problem in evaluation benchmark indices than in research indices. Thus optimal index construction methodology differs in this regard between research indices and evaluation benchmark indices.

6.3.3 Are there no economies of scale in appraisal that can be useful in research index construction? What about “*mass appraisal*”?

Yes there are. *Cross-sectional* economies of scale, derived from the simultaneous appraisal of numerous properties, can provide appraisal cost economies in an index without necessarily causing the temporal lag bias that is particularly problematical for research indices. The term used for appraisal methods that seek to take advantage of economies of scale in the simultaneous appraisal of numerous properties is called “*mass appraisal*”.

6.3.4 Is mass appraisal used in practice?

Yes, “mass appraisal” (the “batch” appraising of a large number of individual properties all at once) is widely used in the residential appraisal industry, particularly by institutions that must appraise large numbers of properties on a regular basis. For example, mass appraisal is widely used by property tax assessment authorities. To date, mass appraisal has not been widely applied to commercial real estate in the U.S. However, there have been some academic and industry research studies that have applied mass appraisal techniques to commercial properties in the U.S.

6.3.5 How does mass appraisal work?

Formal mass appraisal techniques are based on statistical regression analysis, and the most widely used mass appraisal techniques rely heavily on such formal analysis. A statistical “value model” is created, based on analyzing hundreds of transaction prices. In the typical property-tax assessment application, the property owners have the right to appeal the valuation assigned to their property by the statistical regression model. The overall result is an informal combination of formal statistical methods and case-by-case corrections by human intervention.

6.3.6 Does mass appraisal require formal statistical procedures such as regression analysis?

In principle, mass appraisal need not necessarily use regression analysis or any formal statistical procedure. Whenever the same underlying information base is used to simultaneously estimate the values of large numbers of individual properties in a coordinated manner, a type of mass appraisal is occurring. Arguably, the IPD makes, or has made, some use of mass appraisal techniques in making sure that all constituent property value reports are updated in a consistent manner as of the end of each year in the annual IPD-UK Index. Indeed, when the vast majority of the properties owned by a given peer group are being reappraised as of the same time in coordination with the data submission requirements of the peer group’s evaluation benchmarking entity, a type of informal, de facto mass appraisal would tend to happen rather naturally. From this perspective, the IPD’s Annual UK Index may represent the largest commercial property mass appraisal operation in the world.

6.3.7 Is it possible to produce an index whose return reporting frequency is greater than its constituent properties’ reappraisal frequencies without suffering from the lagging caused by the stale appraisal effect?

Yes. The key is to exclude the stale valuations from the index return calculation methodology. Most generally and fundamentally, this requires that the index construction methodology employ some sort of *sampling*, and this is most effectively done if the index employs some sort of mass appraisal procedure. In effect, sampling is used to *infer* the values of all the properties in the index using current period information from only some properties in the index. The type of mass appraisal procedure has to be specified carefully to avoid inducing a temporal lag bias in the index. This can be done, in principle, using several possible methods. The most widely-used approach in academic research, and in industry practice in the housing sector, is the repeated-measures regression (RMR). This technique works without dropping out any properties or serious reappraisals from the index return calculations. In some respects, such techniques can allow the index to “have its cake and eat it too”. For example, you could have all properties on annual (or even less frequent) reappraisal cycles yet the index could be reported on a quarterly basis and not suffer from a stale appraisal effect, as long as each quarter contains a sufficient number of serious reappraisals. (Typically a few dozen repeat-valuation observations per index reporting period is sufficient.)

6.3.8 How does the RMR work, and is it useful for transaction based indices?

The RMR may be applied to transaction-based indices as well as to appraisal-based indices, or even to a combination of the two. The repeated-measures regression can be viewed in the present context as a mass appraisal technique that uses repeat-observations of individual asset values (the “full” annual appraisals in an appraisal-based index, or actual transaction prices in a transaction-based index) to *infer* the change in the entire population value (the capital return or price-change) in each index reporting period. Essentially, the RMR can be viewed as a method of *sampling*. However, the RMR is more sophisticated than simply using a sample of properties that might happen to have been fully reappraised or transacted at both the beginning and end of a given period (consecutive reappraisals or transactions). The RMR does not require that *any* properties be reappraised or transact in consecutive periods. It is able to use the information contained in non-consecutive value observations to infer what happened to the population as a whole within each period. Under some traditional statistical assumptions, it can be shown that the RMR is the most powerful method for making such inferences.

6.3.9 Are there any particular considerations in the use of the RMR procedure in pure transaction price based indices?

As properties typically only transact on average every 5 years or so, relatively long periods of historical data must be accumulated before there is enough data to estimate a purely transaction-based RMR index. Also, the relative scarcity of transaction data compared to appraisal data for a given set of properties (at least if they are owned by institutions that regularly mark-to-market), causes the estimation sample size to be smaller for a transaction-based index than for an appraisal-based index. These differences give the index a tendency to be more “noisy”, which must be dealt with using certain technical specification and procedural methods in the index construction. Nevertheless, RMR-based transaction price indices of commercial property have been successfully estimated for the states of Florida and California using publicly-available data, and for the properties sold from the NCREIF Index using NCREIF data.

Section 6.4:

Practical Considerations for Index Construction With Modern Statistical Tools

6.4.1 Will the investment industry accept an index that is based on a *regression model*?

As noted previously, regression-based mass appraisal techniques similar to the RMR are widely employed by property tax assessing authorities, and recognized by the courts that administer such taxes. The RMR is widely used in academic research in the housing sector, and has been so for well over a decade. (The procedure was first developed in 1963.) It is also increasingly used in industry and government research on housing. For almost a decade, FNMA and FHLMC have published a quarterly RMR-based index of housing values by metropolitan area, and the *Wall Street Journal* often reports the Case-Shiller-Weiss (CSW) Housing Index, which is also a RMR-based index. The fixed-income securities investment industry uses sample-based indices, particularly in the construction of passive bond index funds.

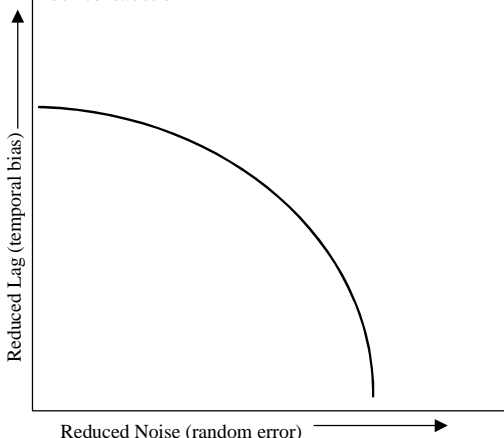
The RMR procedure is *objective, open, and replicable*. It is fully described, explicated and analyzed in the academic and practitioner literature. In summary, it seems that use of the RMR and/or other such regression-based index construction methodologies should certainly be acceptable to the industry for the purpose of constructing commercial real estate *research indices*. Use of such techniques in the construction of agent evaluation benchmark indices is more problematic, and would probably be less acceptable in the industry, in part for the reasons described in Chapter 2 regarding the need for evaluation benchmark indices to be based on populations rather than samples. (See Question 2.3.7.)

6.4.2 Are there any major technical problems in regression-based index construction methodologies?

Yes. The most important problem is the possibility of “*noise*”, or random estimation error, in the index returns. Fundamentally, such noise is caused by random cross-sectional dispersion in individual property valuation observations (whether appraisals or transaction prices, see Questions 5.1-3, & 5.8 in Chapter 5). Such noise appears in the index as excess volatility and spurious periodic return “spikes”. Regression-based index noise is subject to the “Square Root of *n* Rule” of statistics. The magnitude of the noise declines as a function of the square root of the number of valuation observations per index reporting period in the index database. However, keep in mind that periodic returns are “small numbers”. For example, the quarterly appreciation return in unlevered commercial property investment is typically less than 100 basis-points, or less than 1% of the property value, and typical quarterly volatility may be only a little more than this. Thus, even if the magnitude of appraisal dispersion (e.g., the range or standard deviation of appraised values around the average appraised value as of a given point in time) is a small percentage of property *value*, it will be a large percentage of the *periodic return*. For example, if the typical appraisal can be seen as falling within 5% of the true property value, then that appraisal falls in a range of 500% around the typical *quarterly change* in the property value (because the quarterly change in property value is typically only 1% of the magnitude of the property value). Thus, even a rather large sample size (a large “*n*” in the Square Root of *n* Rule) can potentially still allow quite a bit of noise in the RMR-based index.

6.4.3 Why don’t traditional appraisal-based indices suffer from the “noise” problem?

The inherent trade-off between random noise vs temporal lag bias in index construction

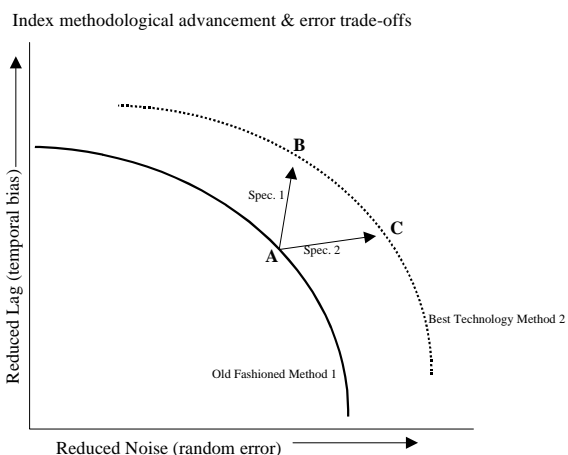


They do. But the traditional appraisal-based index is based on significantly more valuation observations each period. This is because it uses valuation reports from the *entire population* of properties in the index *each period* (whether or not each property was fully reappraised each period). In effect, the traditional approach makes a *trade-off*. By using the entire population of properties every period (instead of just a sample), *n* is increased and noise in the index is thereby reduced. But the other side of the trade-off is that temporal lag bias is introduced into the index,

because most of the property valuation reports used each quarter are not fully up-to-date reappraisals (the “stale appraisal effect”). This trade-off between noise and temporal lag bias is illustrated graphically in the figure. The graph in the exhibit shows the general relationship between the amount of random noise error and the amount of temporal lag bias in an index, holding constant both the index return reporting frequency and the number of valuation observations available in the index database.

6.4.4 Can anything be done about the noise problem in regression-based real estate research indices?

Yes. There are a variety of statistical procedures designed to “filter out” noise in regression-based estimates. In the context of a returns time series index, most of these procedures would tend to interject temporal lagging into the index, which would be undesirable from the perspective of using the index to indicate fully contemporaneous, up-to-date market and investment performance information. However, some noise-filtering techniques do not introduce a temporal lag into the index. In general, these techniques impose some sort of conditions on the index. This is known as a “Bayesian” statistical approach. The index constructor uses some prior knowledge or theory about real estate markets to impose certain conditions that help the index to have characteristics that seem reasonable for a real estate index. These techniques have been



found to be very effective in filtering out noise, without obfuscating real market movements, even with sample sizes as small as a few dozen repeat-valuation observations per period. In essence, by using the best available statistical techniques to improve the construction of real estate indices, we can “push out” the noise-vs-lag trade-off frontier, as indicated in the exhibit. Furthermore, the exhibit reveals that different specifications of the regression procedure and noise filter will tend to provide different choices on the noise-lag trade-off frontier. For some research purposes, noise may be less of a problem than

temporal bias. For other purposes, temporal bias may be more of a concern than random noise.

6.4.5 Can regression-based research indices suffer from stale appraisals or temporal lag bias?

Yes. Use of regression-based procedures such as the RMR does not automatically eliminate the danger of temporal lag or a stale appraisal effect. In general, improvements in index construction methodology cannot eliminate the fundamental trade-off between noise and lag, they can only improve the terms of the trade-off by pushing out the frontier. For example, if the appraisals or valuation reports used in the RMR have dates spread uniformly throughout each period, then the resulting index value will reflect the *average value level* of the properties in each period, that is, averaged across the period, rather than the value as of the end of the period. This is a type of “temporal aggregation” which results in some smoothing and lagging in the return index

(constructed based on the differences in the index value levels each period). However, it is possible to construct the RMR so as to minimize or eliminate this problem. For example, the RMR can be estimated for short periods (such as quarters or months), or the regression can be specified so as to effectively weight valuation observations between periods (e.g., a valuation as of the middle of the period is attributed one-half to the current period and one-half to the previous period). Nevertheless, nothing can make the RMR or any other procedure immune to the fundamental noise-vs-lag trade-off. Methods to reduce temporal lag tend to bring greater random noise. This makes use of noise filters such as the Bayesian techniques noted in the previous question all the more important. In effect, they are techniques to push out the index accuracy trade-off frontier.

Section 6.5: Use of Less Formal Techniques in Index Construction

6.5.1 Are there other methods besides statistical regression to allow an index to have greater return reporting frequency than its average constituent property revaluation cycle?

Yes. We have already noted that mass appraisal need not be limited to formal statistical techniques. For example, suppose the index population of properties consists of a “mixed bag” in the sense that some properties are fully reappraised every period, others every other period, others once per year, and still others not even that frequently. (For example, in the NCREIF Index about 20% of the properties are typically “seriously” reappraised in both the current and previous quarter.) Each period, you could use the sub-sample of properties that happened to be fully reappraised at both the beginning and end of that period, as an indicator of what happened to the entire population that period. You could then apply that implied appreciation percentage to all the other properties in the index, perhaps making adjustments for property market segments.

6.5.2 How does such a simple, common sense sampling approach compare to more formal regression-based procedures such as the RMR?

The simple procedure would be less efficient and less effective than the regression-based approach. It would, in effect, “waste data”, because it would ignore the information that can be validly used for inference purposes from other properties that did not happen to have serious reappraisals in consecutive periods. The simple sampling approach would also be more subject to bias than the regression-based approach because the sub-sample of properties that are consecutively reappraised in a given period might not be typical or representative of the population as a whole. A professional statistician would probably refer to the simple sampling approach as the “Sub-Optimal Sampling Method”.

6.5.3 What about a more common sense approach that tries to use all the relevant information?

Suppose we have the “mixed-bag” of property reappraisal cycles described in Question 6.5.1, and we use the consecutively-reappraised sub-sample to get some indication of current period

market movements. In addition, we also consider the particular characteristics of the properties in that sub-sample, and we also examine the appraisals and the property characteristics of other properties that were fully reappraised in the current period but not in the previous period. Next, combining all this current value level and current period appreciation information, we turn our attention onto the remaining properties, those that were not fully reappraised this period. We try to estimate how they would (or should) have been reappraised, if they had been fully reappraised this period, based on what we see in the sub-sample that was actually fully reappraised. For example, some of the superficially-updated or non-reappraised properties might have been fully reappraised just last period. We could perhaps apply the current period appreciation return estimated from the sub-sample of consecutively fully reappraised properties to try to estimate what the properties last reappraised last period would be worth today. A good name for this type of approach would be “Ad Hoc” or “Informal” Mass Appraisal. The approach described here is less “formal” than the previously-described regression-based procedures in the sense that it is less amenable to formal mathematical specification, and is not derived rigorously based on statistical theory.

6.5.4 What are the pros and cons of the ad hoc informal approach as compared to the formal regression-based approach?

The major advantages of the regression-based approach relative to the informal approach are the following:

- **Objective & transparent:** The regression-based approach is based on procedures that are well-known and that can be precisely expressed mathematically, making it easily replicable by anyone with access to the underlying data;
- **Optimal & efficient:** Because it can be precisely expressed mathematically, the regression approach can be, and has been, analyzed from a formal statistical perspective by the academic community. From this perspective, the regression-based approach meets certain standards of “optimality” or “efficiency”, as these are widely accepted in the statistics profession.

The major advantages of the informal approach relative to the regression approach are the following:

- **Less mechanistic:** The informal approach is more flexible than regression, because it allows for ad hoc adjustments and subjective judgments to be made. It allows more “human intervention”;
- **Less “nerdy”:** The informal approach may be complicated and messy, but it is based on common sense. It can be explained to, and probably understood even at the “nuts & bolts” level by people who do not have special technical training.

6.5.5 Are the two approaches, regression and informal analysis, mutually exclusive?

No. This is not an “either/or” proposition. Regression-based analysis can be formally imbedded in the informal ad hoc procedure as one component or “step” in such a procedure, or regression analysis can simply be used informally by those individuals charged with carrying out the informal method of index construction. Going the other way, it is also possible to incorporate some of the ad hoc or “human” inputs of the informal procedure into formal regression-based models, particularly through the use of Bayesian statistical procedures.

Chapter 7:
Benchmarking & Indexing Investment Performance for Real Estate Investments that Are Not Regularly Marked-to-Market

7.1 What is meant by “marking-to-market”?

Marking-to-market refers to the estimation of the current market value of the assets in a fund or portfolio, and recording or carrying forward the values of those assets at their estimated market values. In the case of private assets, this requires some sort of “appraisal” process.

7.2 How does real estate differ from other private investments in marking-to-market?

A well-functioning asset market exists for commercial property. There are numerous competing buyers and sellers constantly in the overall market for commercial properties, and transactions are virtually always occurring, at least somewhere. Commercial properties are relatively standard, easily understood assets about which a wealth of information is typically publicly available in general, and either privately or publicly available for specific assets that are in the market. This makes it generally possible to estimate the current market value of a given commercial property with reasonable accuracy. To serve this value estimation function, the profession of real estate appraisal is a long-standing and well-established profession with detailed and widely accepted standards of practice. In contrast, there is virtually no market for many other types of private assets as such. The typical underlying assets in the venture capital asset class are much less liquid than the typical commercial property, and it is usually more difficult to precisely estimate the current market values of such assets in a manner that can be well-documented and widely accepted. In short, marking-to-market in private real estate markets tends to be easier, more accurate, and more able to be documented in a standard, widely-acceptable manner, than is the case in many other types of private investments.

7.3 How does real estate differ from public securities investments in marking-to-market?

The real estate asset market is characterized by the private trading of unique, whole assets. That is, each transaction typically involves one complete underlying asset, with the deal being privately negotiated between the buying and selling parties. The private real estate market is relatively “thin”, in that each underlying asset typically trades only rarely and irregularly through time. In contrast, public securities markets are much “denser”, with large quantities of homogeneous shares of (unique underlying) assets typically being simultaneously bought and sold continuously at publicly quoted prices in stock exchanges that promote liquidity and information availability. This makes it easier to precisely estimate and document the current market value of publicly traded securities, which makes marking-to-market for those securities easier, more accurate, and less expensive than for private real estate.

7.4 Why is marking-to-market important in benchmarking and indexing investment performance?

Regular, periodic marking-to-market of assets is required to calculate periodic holding period returns (HPRs). Time-series of periodic HPRs are used for a number of investment performance

analysis purposes, and they are the required input to calculate the time-weighted rate of return (TWRR). The TWRR is the average of the periodic HPRs across a given historical span of time. For example, if an investment produces periodic returns of 1%, 1%, and 4% in three consecutive periods, then its TWRR is 2% on an arithmetic mean basis $[(1+1+4)/3=2]$, or 1.99% on a geometric mean basis $[(1.01*1.01*1.04)^{(1/3)}-1=0.0199]$.

7.5 Why is the calculation of TWRRs important for an investment asset class?

There are two major reasons:

- **Comparability to the core public securities asset classes:** The vast majority of performance data available on the mainstream asset classes that compose the core of institutional investment portfolios (stocks and bonds) are calculated using TWRRs and their underlying periodic HPR time-series. Thus, the use of TWRRs and their underlying periodic returns series allows the statistical properties of commercial real estate, such as mean returns, standard deviations, and correlations, to be compared directly with the mainstream core asset classes.
- **Appropriateness for agent performance evaluation:** In most (but not all) professional investment management situations, the investment agent (e.g., the manager or fiduciary) does not effectively *control the timing* of capital into and out of her management. Therefore, fairness requires the measurement of the agent's investment performance using a return measure that is not sensitive to capital flow timing. (Recall the underlying management principle of equating responsibility with authority.) The TWRR effectively removes the impact on calculated returns of capital inflows and outflows at the fund or portfolio level. In this respect, the time-weighted methodology produces an average historical return that is neutral with respect to capital flow timing, reflecting only the performance through time of the underlying assets that are held in the fund or represented by the index. As a result, TWRR-based evaluation does not reward (or punish) an agent for having more of the principal's money invested when the market segment the agent is responsible for happens to be performing relatively well (or poorly).

7.6 What is the difference between the time-weighted return and the IRR?

Time-weighted returns are the averages across time of simple periodic HPRs ("holding period returns"). The HPR in each period is, in turn, the difference between the ending and beginning capital value, plus any net cash distributions during the period, all as a percent of the capital value at the beginning of the period. Computation of periodic HPRs (and hence of time-weighted returns) requires asset valuations as of the end of each return-reporting period (e.g., marking-to-market at the end of each day, quarter, or year). Time-weighted returns are neutral with respect to the timing of capital flow into and out of the manager's control.

The IRR is the single rate that discounts the sum of all the net cash flows (and remaining capital value) to the initial capital investment amount. In contrast to the TWRR, the IRR is a dollar-weighted (or "money-weighted") return. It is more heavily influenced by the rate of return achieved during periods when the investor had more capital invested. The IRR is thus sensitive to the timing of capital flow into and out of the manager's control. Computation of the IRR does

not require asset marking-to-market at the end of each period, but merely requires a record of the magnitude and timing of all net cash flows to and from the investment (e.g., the fund), as well as a valuation of any assets remaining in the investment (fund) as of the ending date when the performance evaluation is taking place.

7.7 Can you give a numerical example of the difference between the TWRR and the IRR?

Consider a stock investment that currently sells for \$50 and is paying a constant annual dividend of \$2, paid at the end of each year. Assume you purchase one share of the stock now and an additional share at the end of the first year for \$53. At the end of year two you sell both shares for \$54. Total cash outlays are:

Time	Outlay
0	\$50 to purchase first share
1	\$53 to purchase second share a year later
<hr/>	
	Proceeds
1	\$2 dividend from initially purchased share
2	\$4 dividend from the two shares held in the second year, plus \$108 received from selling both shares at \$54 each

To calculate the internal rate of return (IRR) experienced by the investor, we solve for the discount rate that equates the present values of cash inflows and outflows:

$$50 + \frac{53}{(1+r)} = \frac{2}{(1+r)} + \frac{112}{(1+r)^2}$$

The unique solution to this equation is $r = 0.0718$, giving an IRR of 7.18%.

Note that the IRR is a multi-period average return (covering two periods of investment in this case). It is also a “dollar weighted” return because the stock’s performance in the second year, when two shares of stock were held, has a greater influence on the IRR than the first year return, when only one share is held.

In contrast to the IRR, the TWRR ignores the number of shares of stock held in each period. The stock’s simple HPR in the first year was 10% (\$2 in dividends plus \$3 in capital gains, divided by the \$50 purchase price). In the second year, the periodic return was 5.66% (\$2 in dividends plus \$1 in capital gains, divided by the \$53 purchase price). The multi-period TWRR is calculated either as the simple (arithmetic) average of 10% and 5.66%, which is equal to 7.83%, or the geometric average, which is equal to 7.81%. In either case this result differs from the IRR of 7.18%. The TWRR considers only the period-by-period return on an investment of \$1 without regard to the amounts invested in the stock during each period.

Note that in this example the dollar-weighted return (IRR) is less than the TWRR. This is because the stock fared relatively poorly in the second year when the investor was holding more shares. The greater weight that the IRR places on the second-year return results in a lower

measure of average investment performance. In general, IRR and TWRR returns will differ, with the sign and magnitude of the difference depending on the sequence of the periodic returns and the composition of the portfolio.

7.8 Which real estate investments are marked-to-market regularly?

Pretty much the *only* commercial property investments that are regularly market-to-market are those held by investment fiduciaries. ERISA guidelines require serious reappraisals at least once every three years for pension funds, while many co-mingled funds perform external fee appraisals annually (with some updating more frequently). NCREIF requires all properties in its index to be reappraised at least once per year.

7.9 Which real estate investments are *not* marked-to-market regularly?

Most commercial real estate investments are not regularly appraised and marked-to-market. GAAP accounting requirements do not generally require marking-to-market of real estate assets, even in REITs. The cost of independent fee appraisals typically ranges from around 10 basis-points of property value for small properties to as low as 1 or 2 basis-points for very large, “institutional size” properties. In the absence of regular, frequent marking-to-market, it is impossible to calculate periodic HPRs for the subject assets directly without recourse to mass appraisal statistical techniques such as those described in Chapter 6.

7.10 What are the implications for expanding market coverage of commercial property indices?

In order to expand the property-level coverage of commercial real estate indices beyond the tax-exempt fiduciary peer universe currently covered by the NCREIF Index, it would be necessary to either: (i) convince the other major industry segments (e.g., REITs, insurance company general accounts, foreign investors and wealthy private investors) to begin regular, frequent marking-to-market; or (ii) employ index construction methods that take advantage of modern statistical mass appraisal and transaction price indexing techniques of the type described in Chapter 6.

7.11 What about “*opportunistic*” and “*value-added*” real estate funds?

During the 1990s there was a great expansion in so-called “opportunistic” and “value-added” real estate funds. These funds typically invest in properties or projects that require considerable capital expenditure over time on development or rehabilitation of the properties. Two typical characteristics of such funds are relevant to the indexing and benchmarking question:

- Many of these funds do not regularly appraise and mark-to-market the assets in the fund. Instead, properties are carried at historical cost until the projects are complete and/or the completed buildings are sold. It often makes little sense to try to appraise and mark-to-market buildings that are effectively under construction or rehabilitation, as it is difficult to obtain accurate estimates of value for such assets. Raw land is also often relatively difficult to appraise.
- Opportunistic and value-added funds often give their managers greater discretion than traditional core property funds in controlling capital flow timing into and out of the fund.

The investor client may make a commitment up front for a specified total dollar amount which the fund manager draws down in amounts and timing largely at his discretion. Then the fund liquidates gradually as projects are completed and properties sold, again largely at the manager's discretion.

Both of these differences between opportunistic funds and traditional core real estate investment vehicles that specialize in acquiring fully-operational leased-up properties have implications for performance measurement and benchmarking. In effect, these differences make the opportunistic real estate funds more comparable to such private investment asset classes as venture capital, for which IRR-based benchmarking is the standard procedure, instead of TWRR-based benchmarking. (See also Question 2.3.3 in Chapter 2.)

7.12 What is the standard procedure for benchmarking private investments that are not regularly marked-to-market?

Private investments that are not regularly marked-to-market using widely-accepted appraisal techniques such as those of core real estate assets are usually evaluated using IRR-based benchmarks rather than TWRR-based benchmarks. Although AIMR recognizes the appropriateness of the TWRR in most institutional investment situations, it recommends that practitioners measuring the performance of individual private investments employ the IRR from the inception of those investments. Furthermore, AIMR acknowledges that there are inherent problems in applying a time-weighted return methodology to assets that trade in private markets. IRR-based benchmarking is most appropriate for real estate opportunistic and value-added funds that are more similar in style to such other aggressive "alternative investment" asset classes as venture capital and LBO funds.

7.13 How does IRR-based benchmarking work?

IRR-based benchmarking uses the annualized IRR since inception of the subject fund being evaluated, known as the "*since-inception IRR*". This requires the use of inception-date cohorts in the evaluation benchmark. It is not valid to compare the since-inception IRR of the subject fund with the since-inception IRR of other funds that began at different times from the subject fund. Performance evaluation comparisons should always be made over identical historical periods. For example, a subject fund that began in 1995 would be benchmarked against the average IRR of other (similar style) funds that also began in 1995.

7.14 How is the IRR of the benchmark inception cohort calculated?

There are a variety of ways to calculate the average IRR across the subject's peer universe cohort of funds represented in the benchmark, including equal-weighted average, value-weighted average, median, and the pooled IRR. The most widely employed method is to use the pooled IRR. In this method the net cash flows from all the funds in the cohort are pooled within each time period, as if they were all one giant fund, and the IRR of those pooled cash flows is then determined. If the benchmark index were viewed as a statistical sample and each fund in it were considered to be an equally-representative example of the performance of the underlying population, then it might be argued that an equal-weighted average of the individual fund IRRs across all the funds in the index would be a better way to construct the benchmark. However,

evaluation benchmark indices in the private investment industry are typically viewed not as statistical samples, but as peer universe populations. From this perspective, the pooling method makes considerable sense. It is also less susceptible than the equal-weighted average to the influence of a few outliers.

7.15 How is the value of residual assets handled in the calculation of since-inception IRRs?

Residual assets are those remaining in the fund at the time when the since-inception IRR is being calculated. The current value of these assets must be estimated. Then, the annualized since-inception IRR of the fund can be calculated using the fund's actual net cash inflows and outflows from/to its investors within each historical period and assuming that the residual assets are sold in the most recent period at their current estimated market values.

7.16 Are there any major concerns regarding residual asset valuation in the IRR-based benchmarking of fund performance?

Yes. For one thing, it is often difficult to obtain accurate, well-documented, and widely acceptable estimates of value of the residual assets. Yet these value estimates can greatly affect the fund's since-inception IRR calculation, particularly if the fund is young or still has a large amount of invested capital. Also, it is a somewhat "apples vs oranges" comparison to compare the IRR of two funds that have very different amounts of residual assets left in them, even when the two funds have the same inception date.

7.17 Is it valid to combine IRR-based and TWRR-based returns in an all-encompassing index that represents both opportunistic funds and traditional core investments?

The traditional answer to this question is "no". There is no conceptually "perfect" (apples-to-apples) way to combine dollar-weighted and time-weighted average returns. AIMR recommends that private equity managers disclose that such comparisons, when used, can be misleading. To the extent that the overall real estate investment universe includes some types of investments that are best benchmarked using IRR-based measures and other types of investments that are best benchmarked using TWRR-based measures, the conservative position holds that the overall universe must remain a "salad" of asset class returns rather than a "melting pot" of a single return or single index representing the entire universe. However, it has been suggested that repeated-measures regression (RMR) based index construction techniques could, in principle, allow a periodic total returns time-series to be constructed from fund-level IRR data, and such periodic returns time-series could then be combined in some manner with the corresponding time-series of periodic returns from other types of assets that are regularly marked-to-market. (See the article by Geltner & Goetzmann in the July 2000 issue of *The Journal of Real Estate Finance & Economics* for a description of the methodology in question.) We are not prepared at this time to pass judgment on the validity of such a procedure for the use suggested in the above question.

7.18 What is the benchmarking standard in the venture capital industry?

The National Venture Capital Association (NVCA) and AIMR recognize that when a management contract calls for a series of investments, spread out over time at the discretion of the manager, TWRR-based performance measurement and evaluation are not appropriate. The NVCA and AIMR deem the IRR to be the most appropriate measure of returns for venture funds and other such private equity investments. Benchmarking is applied using the since-inception cohort approach described previously. (See Question 7.13.)

It should be noted that the venture capital industry in the U.S. does, in effect, attempt to mark-to-market asset values at the end of each quarter. Limited partners in venture capital funds compute their IRRs based on their cash-on-cash returns, plus their share of the residual value of the venture capital fund's holdings of cash and investments in (typically new) companies that have yet to be distributed to the limited partners. These residual values are recorded for the end of each quarter. However, the residual value is added to the periodic cash flows in the calculation of the IRR only at the end of the holding period for which the IRR is being calculated. Similar to the commercial real estate investment industry, interim (i.e. quarterly) asset valuation is the most problematic component of cash flow analysis for venture capital funds. In fact, estimating the unliquidated, or "residual," value of a venture fund's assets is probably more difficult than valuing investment grade commercial properties. (See Question 7.2.)

In 1993, AIMR proposed performance measurement guidelines that recommended a time-weighted approach to presenting fund performance. After some concerns were expressed by investors and fund managers, a special subcommittee of private equity industry investors and experts appointed by AIMR studied the applicability of time-weighted returns to the private equity industry. They recommended that fund managers and intermediaries present their private equity performance results on a dollar-weighted (IRR) basis. However, over the past six years the leading venture capital benchmarking firm, Venture Economics, has been reporting both dollar-weighted and time-weighted returns to their clients.

Volume 2:
A Review of Some Existing Indexes

*Chapter 8:
The NCREIF Index in the U.S.*

**Section 8.1:
Institutional Overview and Setting**

8.1.1 What types of real estate investment performance index and benchmarking products are currently available in the U.S.?

Regularly produced indexing or benchmarking products for commercial property investment performance in the U.S. currently fall into three major categories:

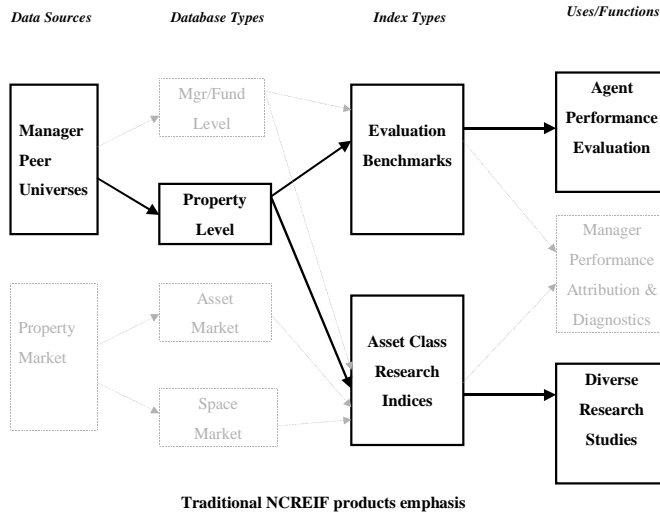
- **The NCREIF Index:** The NCREIF Property Index (NPI) is a property-level index that reports quarterly, appraisal-based investment performance periodic returns (“holding period returns”, or HPRs) on an unleveraged basis. It reports total returns broken down by income and appreciation components, at both the aggregate level as well as numerous sub-indices defined by property type and geographic location. The NPI is what might be called a “semi-public” information product in that both the index itself and the methodology used to produce it are made widely available at minimum cost in keeping with NCREIF’s mission.
- **Proprietary fund-level indices:** Proprietary fund-level historical periodic investment return indices are produced by several commercial consulting firms, primarily for the purpose of assisting their clients, who are typically tax-exempt fiduciaries in the business of engaging professional real estate investment management firms to handle their direct private property investments. These index products are typically based on quarterly financial reports of multi-investor (“co-mingled”) real estate funds, and represent the investment performance realized by investors in the constituent multi-investor funds. It is unlikely that any one consulting firm’s index includes complete coverage of the entire peer universe of co-mingled funds (though they probably all include wide coverage). Separate account investments not included in co-mingled funds are not generally represented in these indices. As of September 2000 NCREIF was in the process of purchasing one of these proprietary fund-level indices and its supporting historical database.
- **REIT Indices:** Indices of publicly-traded REIT equity performance are published by several sources, including the National Association of REITs (NAREIT). The NAREIT Index includes all publicly-traded REITs (with sub-indices of various types).

In addition, there are several sources of information on prices per square foot and/or cap rates (such as ACLI, NREI, Korpacz), but these do not provide periodic or time-weighted total return information.

8.1.2 What has been the role of the NCREIF Index, and what types of information products and functions have been the traditional primary focus of NCREIF?

The NCREIF Index has been perceived as, and used as, both an asset class research index and an agent evaluation benchmark index in the U.S. private real estate investment

industry. However, the design of the NCREIF Index is not optimized for either of these two types of index products, as these products were defined and described in Volume 1 of this Report (see Questions 0.1, 1.1.1, & 2.1.1). NCREIF’s founding purpose, as reflected in its mission statement, is oriented (among other things) toward the development and promulgation of a research index for the asset class. The property-level focus of the



NCREIF Index is clearly consistent with such an objective (see Question 8.3.8 below), and evidence that NCREIF takes this objective seriously. This is represented by the bottom branch in the figure. Moreover, NCREIF’s non-profit, collective constitution is particularly suited to the relatively non-commercial, “public good” nature of a broad asset class research index. (See Questions 1.1.4 & 1.1.5 in Volume 1.) However, many

of the technical and market-coverage characteristics of the NCREIF Index are closer to optimal for an evaluation benchmark index than for a broad asset class research index (see Question 8.1.6 below). Perhaps because of this, during the 1990s the NCREIF Index began to be widely used in the benchmarking process. This use of the NCREIF Index is represented by the top branch in the figure. Yet, through the 1990s NCREIF did not provide either a fund-level performance index or a customized performance attribution analysis service, both prominent features of agent evaluation benchmark indices and benchmarking services outside of North America. This is represented by the fact that these elements have been shaded out in the figure. However, NCREIF has begun to develop and provide some elements of these latter types of information and services (see Questions 8.1.3 & 8.1.4 below).

8.1.3 What is a fund-level index, and does NCREIF produce one?

In addition to the performance of underlying properties held in the funds, fund-level performance reflects the effect of fund-level policies, such as leverage, fund-level administrative expenses, and the return on fund holdings in cash or other non-real estate assets. NCREIF is in the process of beginning production of a quarterly fund-level index. In 2000 NCREIF purchased such an index and its supporting historical database from IPC, a private consulting firm, consistent with a new NCREIF policy of commencing production of a NCREIF fund-level index with some historical background.

8.1.4 Does NCREIF provide customized benchmarking and performance attribution analyses for its members?

NCREIF does not provide such products as part of its regular services to its members. Moreover, NCREIF has not to date collected some of the more detailed property management and operational information that would be necessary to provide some of the finer breakdowns of property operational level performance attribution analysis that are provided, for example, by IPD in England and PCA in Australia. However, NCREIF does provide at nominal charge customized indices suitable for benchmarking (e.g., a market segment sub-index with a given member's properties removed from the index). Also, NCREIF has developed and implemented an interactive website that enables members to construct custom disaggregations of the NCREIF Index. This website also allows members to conduct queries of the database that make it relatively easy for members to perform their own performance attribution analyses at the standing portfolio level and at some of the higher property operational levels. This allows some of the diagnostic function of performance attribution to be supported, but it is less oriented toward supporting the agent evaluation function.

8.1.5 Is the NCREIF Index primarily a manager evaluation benchmark or a broad research index?

There is no simple answer to this question. As noted in Question 8.1.2 above, the NCREIF Index is used in both roles. There is a need for both a manager evaluation benchmark and for a broad asset class research index, and the NPI is essentially all that is available for either purpose at present for direct private real estate investments (with the exception of the proprietary fund-level indices that are used by consultants for manager evaluation). The history and design of the NPI suggests that both the evaluation benchmark and the research index roles have always been considered at NCREIF (not necessarily clearly distinguished). However, one can make an argument that the NPI is closer to the optimal as an agent evaluation benchmark (for core institutional property holdings) than as a broad asset class research index.

8.1.6 Why does the NPI come closer to the ideal as an agent evaluation benchmark than as a broad asset class research index?

There are three major reasons for this conclusion:

- The NCREIF Index includes a large fraction of the direct private property holdings of the tax-exempt fiduciary branch of the real estate investment industry, a reasonable definition of a relevant "peer universe". (See Questions 2.2.4 and 2.4.9 in Volume 1, and Question 8.3.2 below.) In contrast, the NPI includes a much smaller fraction of the entire real estate asset class or market, and that fraction is not constructed as a scientifically stratified random or representative sample of the overall commercial property market. (See Questions 1.1.6, 1.2.2, and 4.3 in Volume 1, and 8.3.1 below.)
- The NCREIF Index is value-weighted rather than equally-weighted (although NCREIF now makes an equal-weighted version available on its web site),

suggesting that the NCREIF property base is viewed as a population (a peer universe) rather than a statistical sample (e.g., of the asset class as a whole). (See Questions 1.4.3 and 2.3.4 in Volume 1.)

- The fact that the NCREIF Index is appraisal based is not as serious a problem for the agent evaluation benchmarking function as it is for certain research functions. (See Questions 2.3.7, 5.7 and 5.9 in Volume 1.)

Section 8.2:

Technical Characteristics of the NCREIF Index, and the Appraisal Lag Problem

8.2.1 How is the NCREIF Property Index constructed?

NCREIF's members who own fully operational properties (at least 60% leased) that are reappraised at least annually for tax-exempt investors make quarterly reports to NCREIF regarding the values of all such properties, as well as the net operating income, capital improvement expenditures, and partial sales receipts associated with each property (along with other information primarily of a property-descriptive nature). NCREIF computes the quarterly holding period total return (HPR) for each property, and then constructs the index as the value-weighted average across all the constituent properties' returns.

8.2.2 How frequently are NCREIF properties reappraised?

NCREIF rules require each property to be fully appraised at least once per year. Empirical evidence suggests that well over 90% of the properties included in the NCREIF Index are indeed seriously reappraised at least annually (either by an external fee appraiser or by an equivalently detailed in-house appraisal). Here, the term "serious" reappraisal is defined as a value report that differs (by more than just the amount of capital improvement expenditures) from the previous quarter's value report, or else a value report that the data-contributor explicitly noted was based on an external appraisal. A few (typically smaller) properties are not always seriously reappraised every year, but at the other extreme about 20% of properties are typically rather seriously reappraised from one quarter to the next (3-month interval). The most common frequency of serious reappraisal is annual, with about half of all NCREIF properties being seriously reappraised at that frequency.

8.2.3 How seasonal is the reappraisal of NCREIF properties?

As noted, the most common reappraisal frequency for NCREIF properties is annual. Historically, about 40% of all serious reappraisals have occurred in the fourth calendar quarter, with about 20% in each of the other three quarters.

8.2.4 What is the effect of this procedure on the dynamic statistical quality of the index?

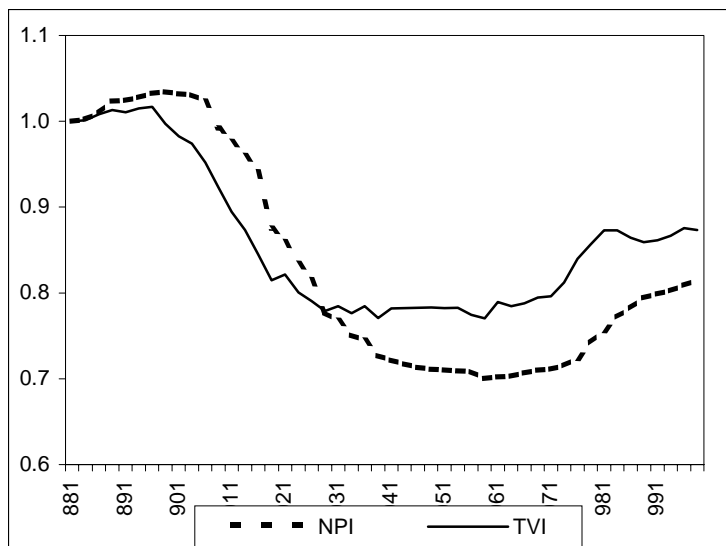
The above-described procedure cause the NCREIF Index to exhibit a “stale appraisal effect”. Largely annual property-level reappraisals in a quarterly-reported index causes the index to lag an average of about two quarters behind completely contemporaneous appraised values. (See Question 6.2.1 in Volume 1.) The lag is slightly less in the 4th-quarter index value, and slightly greater in the other three quarters. That is, the index “catches up” a little bit at the end of each calendar year, but is never fully up-to-date. There is also a slight seasonal tendency, characterized by 4th quarter return “spikes” when the property market tracked by the index has been trending in one direction for more than a year. (The spikes are positive if the market has been increasing in value, negative if it has been losing value.) The spurious 4th quarter spikes inject extra volatility at the quarterly frequency, while the general stale appraisal effect dampens quarterly volatility.

8.2.5 Are the underlying property-level appraisals, even the “serious” ones, also lagged?

This is a somewhat controversial question. We believe that the best current evidence strongly suggests that even the so-called “serious” property-level appraisals in the NCREIF Index tend to lag behind the purely contemporaneous transaction price evidence, most likely by about 3 quarters, on average. (See Questions 5.3 - 5.5 in Chapter 5 of Volume 1.) Thus, the total average lag, including both the stale appraisal effect described in the previous question and the property-level appraisal lag would be around five quarters, on average.

8.2.6 What is the nature of the lag in the NCREIF Index?

Suppose the average overall lag is five quarters, as suggested in the previous answer. This does not mean that the NCREIF Index simply equals the true contemporaneous transaction price level five quarters previous. Rather, the NCREIF Index is a moving average that tends to spread out the underlying true market movements across time. If the market moves, say, +10% in quarter 2000.1, then a small part of this 10% movement will indeed be reflected in the NCREIF Index in 2000.1, and another part of it will show up in 2000.2, and so on in subsequent quarters. By 2001.2 (five quarters from 2000.1) about half the original 2000.1 market movement of 10% is likely to be reflected in the index. The original 2000.1 market movement might not fully incorporated into the index until perhaps early 2002 or thereabouts. Meanwhile, of course, market movements in other quarters before and after 2000.1 are also entering the index in a similarly spread-out manner. The point is that if the average lag is 5 quarters, then there would still be some incorporation of old information several quarters beyond that average lag, while some traces of new, up-to-date information would also appear contemporaneously in each quarter’s index. The net effect of this is a tendency to average out and obfuscate brief movements in the market, and to cause turning points in the market to appear smoothed and lagged in time. Index volatility may be slightly dampened (though this may depend on how one defines “true market volatility”). However, the more important effect is



probably the temporal lag bias in the index. An illustrative example of how temporal lag bias may effect the NCREIF Index appreciation levels is shown in the exhibit. The dashed line is the NCREIF Index, and the solid line is an estimate of the corresponding contemporaneous transaction price levels, quarterly from 1988 through 1999. (This chart is from an article by Fisher & Geltner in the Spring 2000 issue of *Real*

Estate Finance.) Note that the transaction price index “leads” the NCREIF Index, both in down markets and in stabilizing and improving markets.

8.2.7 How problematical is this lag?

This depends on the purpose for which the NCREIF Index is being used. In Volume 1 we explained why appraisal lag bias does not cause major problems for some uses, such as estimating long-run time-weighted mean returns, while it can pose very serious problems for other uses, such as indicating which way the market moved last quarter, or measuring the lead-lag relationship between private real estate and REITs. In general, we suggested in Volume 1 that appraisal lag bias tends to be a more serious problem for some important asset class research uses of the index than for the typical evaluation benchmark use of the index (see Questions 2.3.7, 5.7, and 5.9).

8.2.8 Is NCREIF doing anything about the lag problem?

NCREIF is studying various methods to construct alternative indicators of contemporaneous, quarterly market movements, including repeated-measures regression (RMR) based approaches and use of transaction price evidence from properties sold from the NCREIF Index. (See Questions 6.3.7 and 6.3.8 in Volume 1.) These methods, which are still under study and development at NCREIF, are not being contemplated as replacements for the traditional NCREIF Index, but might be made available as supplementary information products. As noted, the lag in the traditional NCREIF Index is not necessarily a prohibitive problem for some typical uses of the NCREIF Index, especially in the benchmarking function.

Section 8.3:

Gap Analysis of the NCREIF Index vs the Ideal Index

8.3.1 How does the NCREIF Index compare to the ideal *research index* regarding property market coverage?

The NCREIF Index currently includes just over 2,600 properties. There are probably well over 1,000,000 commercial properties in the U.S. (for example, there are approximately 125,000 within the MSAs in the State of Florida alone, and Florida contains less than 6% of the U.S. population). At least 100,000 (or 10%) are of “institutional size” (\$10,000,000 or greater in value), probably considerably more than that. Thus, the NCREIF Index is a sample of less than one percent of the entire U.S. commercial property population and less than 3% of the institutional size property population, measured by number of properties. However, as described in Chapter 4 of Volume 1, the validity and usefulness of a sample for purposes of drawing statistical inferences about the population is not measured best by the percentage of the population covered by the sample, but rather by the *absolute size* and *representativeness* of the sample. The “Square Root of n Rule” is relevant in this regard (see Question 4.1 in Volume 1). This suggests that the major problem with the NCREIF Index in terms of market coverage for a research index is not the size of the sample at the aggregate level ($n = 2600$ is good enough at that level), but rather the size of the sample at the detailed market segment level, and the representativeness of the sample in general. The NCREIF Index is not a *scientifically stratified sample*. (See Questions 4.2 - 4.4 in Chapter 4 of Volume 1.) Certain types of properties are not effectively covered, and there are too few properties in some “cells” of an ideal stratified sample of properties for broad asset class research purposes. While these sampling deficiencies do not render the NCREIF Index worthless for research purposes, they do leave it far below the ideal, particularly regarding the needs of more rigorous and sophisticated academic research.

8.3.2 How does the NCREIF Index compare to the ideal *evaluation benchmark index* regarding property market coverage?

It is probably most instructive to break this answer into three parts: traditional institutional “core” investments, “non-core” investments, and certain performance attribution analysis considerations that are complementary to benchmarking.

- **Institutional core investments:** In the first quarter of 2000 the NCREIF Index included over \$83 billion worth of direct private property holdings within the tax-exempt fiduciary branch of the real estate investment industry. Although a precise tabulation is difficult, this NCREIF coverage almost certainly exceeds 70% of the entire value of direct private property holdings of traditional “core” institutional real estate investments (fully operational properties) within this branch of the industry, and it may be nearer to 80% coverage. In January 2000 *Pension & Investments* reported total real estate private equity investment of the top 200 pension funds was \$83 billion, some of which would have been levered, which suggests that the underlying direct property holdings of this peer group were probably around or slightly more than \$100 billion. However, a part of this would

be in “opportunistic” and “value-added” funds not characteristic of the institutional “core”. Keeping in mind that evaluation benchmark comparisons should, in principle, be limited to peer universes populated by investment agents of similar style and specialization, the NCREIF Index probably represents at least 70% of the most relevant peer universe for evaluation benchmark purposes within the core investment style, and possibly closer to 80%. Although less than the ideal of 100% coverage, it seems to us to be reasonably adequate for benchmarking core investments, given the purpose of evaluation benchmarking as described in Chapter 2 of Volume 1 (see especially Questions 2.1.3, 2.4.9, and 2.4.11).

- **Non-core investments:** The official criteria of the NCREIF Index tend to prevent it from including most properties that would be classified as “non-core”, where this term is understood to refer to development and rehab projects, raw land investments, and other such investments, the types of properties that would be common in so-called “opportunistic” and “value-added” funds. Also, some major managers of such funds are not NCREIF members. As far as we know, nothing would prevent NCREIF from relaxing the criteria for properties to be included in the NCREIF Index, so as to begin collecting data on more non-core properties. Currently NCREIF criteria require 60% occupancy and annual reappraisals. (Note that once a property is in the NCREIF Index it generally remains in the index even if it subsequently dips below these criteria.) On the other hand, there is a good argument that many non-core investments should not be benchmarked using time-weighted returns, but rather using IRR inception-date cohorts in a manner more comparable to the way benchmarking is done in the venture capital and other private equity asset classes (see Question 2.3.3 and all of Chapter 7 in Volume 1).
- **Performance attribution:** Some property operational-level performance attribution analyses relevant for diagnostic purposes ideally would require a valid statistical sample of the relevant property market segment for benchmarking purposes, rather than a population census of a given investment peer universe of property owners. Ideal market coverage for such purposes is more akin to that described in the answer to Question 8.3.1 regarding research indices. (See Questions 3.8 and 3.12 in Chapter 3 of Volume 1 for further elaboration regarding an ideal index in this regard.)

8.3.3 Why does NCREIF not have 100% coverage of the core component of the tax-exempt fiduciary branch of the industry?

Two main reasons have been suggested to us, and they may be somewhat related:

1. NCREIF requires properties in the index to be appraised at least once per year. Some investment managers and property owners within the tax-exempt fiduciary branch of the industry prefer not to go to the expense of appraising properties that frequently. ERISA requires only once every three years at a minimum, and tax-exempt endowment funds are under no legal constraints at all in this regard.
2. Not all of the investment managers and/or property owners in the relevant peer universe are members of NCREIF. Although NCREIF membership dues are modest, for various reasons some potential NCREIF data-contributing members

have chosen not to join. One reason may be the point noted above that they do not wish to incur the expense and trouble of appraising all their properties at least once per year. Another reason may have to do with the “public good” nature of the NCREIF Index, particularly to the extent that the NPI has traditionally been perceived as primarily a research index rather than an evaluation benchmark index, and because NCREIF provides relatively little customized benchmarking and performance attribution information services (see Question 8.1.4 above). In economics, the problem of “*free riders*” is well known as a classic potential problem with public goods. As long as the good is provided collectively, it may be possible for some to benefit from its provision without paying any share of the cost or otherwise contributing to the production of the public good.

8.3.4 How does the NCREIF Index compare to the ideal *research index* regarding types and sources of data collected in the underlying database?

The NCREIF Index does a good job of collecting property-level investment performance information from the peer universe of investors and managers in the tax-exempt fiduciary branch of the real estate investment industry. However, we feel that four additional types of information and data sources should be tapped in an ideal research index. These are listed below:

1. ***Additional operational-level data:*** More property operational-level information would be useful in the database, such as more detailed breakdowns of revenue sources, lease characteristics, operating expenses and capital expenditures. This would be useful especially for certain performance attribution analyses useful for diagnostic purposes.
2. ***Additional types of properties:*** Expanding the property-level coverage to more different types of properties within the tax-exempt fiduciary branch of the industry would be useful. Most notably, expansion to include properties that are not “fully operational” (i.e., to include properties with less than 60% occupancy) and to include those that are reappraised less frequently than annually, would add considerably to the richness of the database and allow useful new index products to be produced. (Infrequent reappraisal can in principle be addressed using statistical techniques such as RMR, as described in Questions 6.3.7 and 6.3.8 in Volume 1.)
3. ***Additional branches of the industry:*** Expanding the data-contributor information source pool beyond the existing tax-exempt fiduciary branch of the industry would potentially enable the property sample size to be expanded threefold or more from the existing NCREIF Index. REITs, insurance company general accounts, other financial institutions, and possibly foreign investors and wealthy private investors are types of investors that might be willing to contribute data to a well-organized research index, particularly if such contribution did not require them to incur additional appraisal expenses. (Once again Question 6.3.7 is relevant.)
4. ***Publicly-available information:*** Expansion to include publicly-available and commercially-available information from the space markets and asset market, such as occupancy, absorption/leasing, construction, and transaction (price and

volume) information, would be truly path-breaking and valuable from the perspective of an asset class research database. Public or commercially available transaction price information could probably be used to construct a regularly-updated annual (or perhaps quarterly) purely transaction price-based national “repeat-sales index” of commercial property for the U.S. This would be an international first, and would place commercial property on a par with housing in this regard.* A scientifically stratified sample might be designed for use with the repeat-sales index based on surveys of institutional property holdings, along the lines of the survey conducted by the PCA in Australia.

The above list is presented in approximate order of increasing difficulty of obtaining the incremental data as well as increasing usefulness of the data. A possible exception to this order is item (4). It might well be easier to obtain public and commercially available asset market and space market data than to expand the data-contributing subscribers to include other branches of the industry. Indeed, such expansion of data-contributing subscribers would probably be impossible within the current constitutional structure of NCREIF (see Question 8.3.8 below). However, while (4) might be politically easier than (3), it would cost some money to locate, collect, and in some cases pay for public and commercially-available data on the asset and space markets.

8.3.5 How does the NCREIF Index compare to the ideal *evaluation benchmark index* regarding types and sources of data collected in the underlying database?

The NCREIF Index is closer to the ideal evaluation benchmark index than to the ideal research index in the type of data it currently obtains. This is particularly true if we include the fund-level database that NCREIF is in the process of purchasing, assuming that NCREIF will successfully implement the collection of fund-level data going forward. In addition, items (1) and (2) in the preceding question would also be quite relevant and useful for evaluation benchmarking purposes. That is, NCREIF would more closely approach the ideal evaluation benchmark index if it would collect more property operational-level detailed information, and if it would expand the range of properties it includes from within the holdings of the tax-exempt fiduciary branch of the industry.

Expansion beyond the fiduciary branch of the industry would also be useful for evaluation benchmarking in two ways. First, it could support certain types of property operational-level performance attribution analyses that can benefit from benchmarking against a broader statistical sample of properties regardless of the type of owner. (See Questions 3.8 and 3.12 in Chapter 3 of Volume 1 for further elaboration regarding an ideal index in this regard.) Second, it could enable additional peer groups to be benchmarked, that is, the additional branches of the investment industry that might be

* Already, pilot projects have produced pure transaction price-based repeat-sales indices for: (i) all commercial property in Florida at the annual frequency (see Gatzlaff & Geltner, REF, Spring 1998); (ii) institutional size commercial property in California at the annual frequency based on Comps.com data (see Young Chai’s working paper presented at the April 2000 ARES meeting); and (iii) properties sold from the NCREIF Index at a quarterly frequency (see Fisher & Geltner, REF, Spring 2000).

included (such as REITs and insurance company general accounts). However, this type of expansion might not be possible under NCREIF's existing constitution, as it is our understanding that the organization is presently effectively limited to the fiduciary branch of the industry.

8.3.6: How does the NCREIF Index compare to the ideal *research index* regarding index construction methodology & reporting frequency?

The NCREIF Index's quarterly reporting frequency is probably about as good as can be realistically expected in a research index in the U.S. Even the FNMA/FHLMC repeat-sale index that is based on millions of sales observations of houses and apartments is reported only at the quarterly frequency. (See Question 1.5.1 in Volume 1.)

On the other hand, the NCREIF Index's construction methodology falls rather farther below the ideal mark. The primary research problem posed by the existing NCREIF Index methodology is the temporal lag bias that results from that methodology. While this is not critical for some types of important research questions (e.g., comparisons of long-run average investment performance across asset classes), for other types of questions it can be a serious problem (e.g., studies of lead-lag relationships and contemporary correlations). (See also Questions 8.2.6 & 8.2.7 above.) Admittedly, it is possible to deal with lag bias in several ways (e.g., use of "unsmoothing" techniques or bounded analysis taking REIT index returns as the opposite end of the range of reasonable temporal indication of real estate performance). But these solutions are imprecise, and the temporal lag problem is one major reason that the current NCREIF Index lacks the credibility and usefulness a research index should ideally have with the mainstream of the academic research community.

For research purposes in the U.S., there is no "quick-fix", nor probably any single solution, to the temporal lag bias problem. There are fundamental trade-offs that come into play in any solutions. For example, either data collection costs or random "noise" error tends to increase when one tries to reduce temporal lag bias holding index reporting frequency constant (see Question 5.6 and Sections 6.1 and 6.2 in Chapters 5 & 6 of Volume 1). Therefore, it seems to us that multiple responses should be pursued through a family of index products produced under the auspices of a high quality research index.

Two general approaches should be seriously considered:

1. **Direct use of transaction price evidence:** While appraised valuation information is a vital and important source of information for any real estate research index, it can be quite useful to supplement appraisal-based data with direct transaction price data, much of which can be obtained from public and commercial sources (in addition to the transaction price data that is in the NCREIF database from properties sold by data-contributors). (See Questions 1.3.6, 1.3.7, 5.7, 5.9 and 6.4.5 in Chapters 1, 5 & 6 in Volume 1.)
2. **Use of more sophisticated statistical techniques:** In conjunction with the direct use of transaction price data, as well as to help address the "stale appraisal" problem in the use of appraisal-based valuation data, we believe

modern statistical techniques should be used to help reduce temporal lag bias and improve the terms of the “noise-vs-lag” trade-off frontier (see Sections 6.3 & 6.4 in Volume 1). Techniques such as “mass appraisal” and *repeated-measures regression* (RMR) can be very helpful in improving the accuracy of real estate research indices.

As noted in Question 8.2.8 above, NCREIF has begun to explore both of the above suggestions. However, the existing NCREIF Index is hampered in this regard because it is widely perceived and used not only as a research index but also as an agent evaluation benchmark index. As explained in Volume 1 of this report, the optimal index construction methodology (including the importance of addressing the temporal lag problem) differs between an ideal research index and an ideal evaluation benchmark index. (See Question 8.3.7 below.)

8.3.7 How does the NCREIF Index compare to the ideal *evaluation benchmark index* regarding index construction methodology & reporting frequency?

Arguably, the NCREIF Index attempts to report returns a little more frequently than is optimal for an evaluation benchmark index in the U.S. As suggested in Volume 1, the ideal reporting frequency for an evaluation index is probably annual rather than quarterly. (See Question 2.3.8.)

In other respects, the NCREIF Index probably comes closer to the ideal evaluation benchmark index than it does to the ideal research index. We believe reliance on appraised values is generally appropriate for evaluation benchmark indices (see Question 2.3.7 in Volume 1). As noted in Question 8.3.5, fund-level performance is important for some evaluation benchmarking purposes, and it is our understanding that NCREIF has recently begun development of a fund-level index.

8.3.8 What is NCREIF’s constitutional structure, and how does this relate to its most appropriate role in index production?

NCREIF is a non-profit industry association representing the tax-exempt fiduciary branch of the real estate investment industry. NCREIF was established in 1982 with a mission to “*foster understanding of the real estate asset class through the collection and dissemination of information, performance measurement, standards, research and education.*” NCREIF is governed by its Board of Directors, in which professional real estate investment management firms have a prominent, but not unanimous, role. NCREIF’s membership is dominated by investment management firms and professional firms that service the institutional investment management industry. However, NCREIF membership also includes some pension plan sponsors and academic members. Within the industry, NCREIF is known as a primarily “technical level” association.

Regarding the index production role, NCREIF’s constitution and mission have several implications, which are somewhat different for the two types of ideal indices defined in Volume 1: broad asset class research indices and agent evaluation benchmark indices.

Our perception of NCREIF's main constitutional strengths and weaknesses for production of both of these types of indices are noted below:

- **Research indices:**
 - (i) *Strengths:* First, NCREIF's organizational form and mission are completely consistent with research index production. The relatively non-commercial nature of such indices makes a non-profit, collective entity an appropriate general type of organization for research index production (see Question 1.1.4 in Volume 1). Second, a strength of NCREIF in such role is the quality and extent of its technical-level functioning, including its role in overseeing the Real Estate Information Standards and the informal link between NCREIF's Research Committee and the academic real estate investment research community.
 - (ii) *Weaknesses:* First, NCREIF is limited to one branch of the real estate investment industry, the tax-exempt fiduciary branch, while a broad asset class research index requires broad market coverage and is relevant to all branches of the industry (including, e.g., REITs and insurance company general accounts). Second, apart from an informal position of "Academic Advisor", NCREIF lacks a formal role for the academic community in its governance or index oversight, even though academic credibility is important for the long-run success of a broad asset class research index (see Question 1.1.3 in Volume 1).
- **Evaluation benchmark indices:**
 - (i) *Strengths:* First, NCREIF's data-contributing members represent the majority of the property holdings of an entire relevant peer universe for agent evaluation, namely the tax-exempt fiduciary branch of the investment industry. (See Question 2.4.9 in Volume 1.) Second, because evaluation benchmark index production is a natural monopoly, there is a need for a vehicle such as NCREIF to facilitate collective action by users of such an index product (see Questions 2.B.2-4 in Appendix B of Chapter 2 in Volume 1). As an industry association, NCREIF can provide a vehicle for such collective action. Third, another strength is NCREIF's technical committees, and NCREIF's highly active role in the Real Estate Information Standards, making NCREIF an effective forum for addressing technical issues relevant for evaluation benchmark index construction.
 - (ii) *Weaknesses:* First, NCREIF's non-profit, industry association structure is cumbersome and inefficient for the production of a type of index that is essentially a commercial product (see Question 2.B.1 in Appendix B of Chapter 2 in Volume 1). Second, NCREIF lacks a formal and sufficiently powerful role for the investor client community (the pension plan sponsors) in its governance and index oversight functions, even though this is the primary ultimate user group for evaluation benchmark indices. Third, NCREIF's representation of only one branch of the overall real estate investment industry (the tax-exempt fiduciary branch), while it is a relatively complete, relevant peer universe, is nevertheless only one branch of the industry. There are some benchmarking functions and

services that can be better served with broader representation (see Question 8.3.2 above, and 2.4.9 in Volume 1).

*Chapter 9:
The IPD Index in the U.K.:
Some Issues Relevant to the U.S. Case*

**Section 9.1:
Institutional Overview & Setting**

9.1.1 What types of real estate investment performance index and benchmarking products are currently available in the U.K.?

It is our understanding that commercial real estate equity investment performance indices in the UK are of three main types:

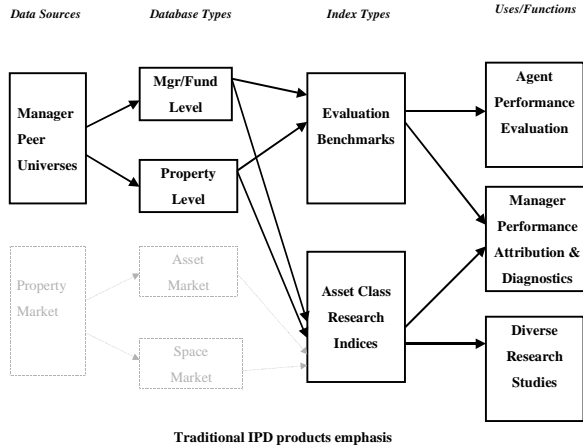
- Property-level indices produced by individual firms of agents/brokers which are close to the transaction data on the properties that are offered for sale, based on appraised values;
- Fund-level indices, produced by actuarial consultants such as WM, Watson-Wyatt and CAPS, are constructed from data from institutional investment clients, and;
- The IPD Index, produced by the firm of that name, an independent company initially financed by a group of surveying firms and agents. This index reports at both the fund (“investment” or “portfolio”) level and the property (“market” or “standing portfolio”) level.

Examples of the first type include the Jones Lang LaSalle (JLL) Index and the Richard Ellis Index (REI). These indices are based uniquely on the properties they manage on behalf of clients. As such, they cover relatively small property samples (e.g., the JLL recently included 179 properties worth less than US\$ 1 billion, while the REI was based on 331 properties worth about US\$3.75 billion). However, the JLL is useful because it extends as far back as 1967 on an annual basis and provides quarterly-frequency return reporting; the REI because it provides monthly returns over a longer history than the IPD Monthly Index.

The IPD collects data directly from large professional investors (including institutions, property companies, and open-ended investment funds), and produces indices for monthly and annual performance. Since it was established in 1985, IPD has aimed to reflect the entire market in the UK. Our understanding is that IPD estimates that as of 1999 their property database included approximately 77% of all the large, professionally-owned commercial properties in the U.K. (not counting corporate real estate and most residential income properties), some 18,000 properties. Their largest index, the IPD Annual Index, represented at that time approximately 60% of this same property population by value (some 14,000 properties in some 230 portfolios), with an aggregate value of approximately US\$120 billion (using a \$1.5/£1 exchange rate). (See Question 9.2.2 below for more details).

9.1.2 What has been the role of the IPD Index, and what types of information products and functions have been the traditional primary focus of IPD?

It is our understanding that the IPD Index is the most widely used in the U.K., both as a research index and agent evaluation benchmark index. This is no doubt due both to its breadth of market coverage and its provision of custom-benchmarking and performance attribution analysis services that complement the evaluation benchmarking function and provide useful diagnostic information to subscribers. Also, because of the size and extent of its constituent portfolio, IPD can produce more sub-indices than the proprietary agents' indices noted in the preceding question. The major

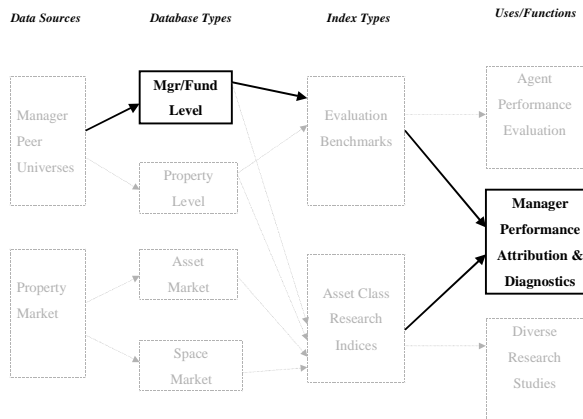


Traditional IPD products emphasis

focus of IPD's information gathering and information products to date is schematically represented in the exhibit.

9.1.3 How does the traditional major information source and product focus differ between IPD and NCREIF?

From its inception, IPD has had a broader information focus than NCREIF, particularly with regard to the collection and reporting of fund-level performance and the provision of custom-benchmarking and performance attribution services. Indeed, IPD has been an international pioneer in the development of evaluation benchmark indices and



Traditional IPD products emphasis NOT traditionally emphasized by NCREIF.

complementary performance attribution analysis services as a commercial information product. The difference in information scope and focus between IPD and NCREIF is summarized in the lower exhibit, which highlights the elements traditionally included by IPD and excluded in NCREIF. (Note, however, that NCREIF has recently begun developing some of these products, as described in Questions 8.1.3 & 4 above.) At least in the U.K. and some other countries, IPD's coupling of fund-level and

custom-benchmarking and performance attribution services to the basic investment performance index product has resulted in a successful for-profit commercial product.

(See Chapter 3 in Volume 1 for a general description of the types of information typically envisioned in real estate investment performance attribution analysis services.)

9.1.4 What specific additional information does IPD include beyond what NCREIF includes?

This question is aimed at a moving target, as both IPD and NCREIF update and extend the information they collect and report. Traditionally, however, IPD has collected and reported substantially more information than NCREIF, as suggested in the answer to the previous question and the exhibits on the previous page. To be more specific, IPD has from the outset included fund-level information and has collected in its property databank information on properties that were under construction or rehabilitation. A basic IPD tenet is to collect information on *all* properties managed by each subscriber. In addition, IPD has collected considerably greater depth of property appraisal and operational-level information, including appraisal assumptions and inputs, lease, rental and occupancy information, and more detailed breakdowns of operating expenses. Moreover, IPD compiles and packages its information into a regular custom-benchmarking and performance attribution analysis product that it provides on a commercial basis to its data subscribers (at fees considerably higher than NCREIF dues, it should be noted). Although NCREIF enables its members to perform some performance attribution analyses and custom-queries of the NCREIF database via its members' interactive website (see Question 8.1.4), IPD provides this service for its subscribers and includes greater levels of depth in the property operational level of the performance attribution analysis. Indeed, IPD has standardized the benchmarking process and refined it to a high degree as an articulated hierarchy of additive attribution components of the differential between the agent's return and the benchmark return (see questions 3.8 & 3.9 in Chapter 3 of Volume 1). By regularly providing this to all their subscribers, IPD probably has significant economies of scale and scope advantages. Furthermore, by having the analysis done by a "3rd party" (IPD) rather than by the agent himself, some additional credibility may possibly be gained for the analysis, which may make it more useful for the agent for external purposes.

9.1.5 Is IPD's methodology and performance measurement consistent with international state-of-the-art practice?

The broad answer to this question is clearly "yes". IPD helped to pioneer, and has subsequently been closely involved in the continuing evolution of, the state-of-the-art methodology and practice outside of North America regarding custom-benchmarking and performance attribution analysis in the private real estate investment industry. Of course, this state-of-the-art is constantly evolving and developing, particularly regarding specific technical-level issues, and some traditional IPD measures or information products may have occasionally been superceded by newer procedures. It is our understanding, however, that IPD is oriented primarily toward providing the information products that its subscribers want, which tend to reflect the evolving state-of-the-art. Thus, we would expect IPD, as a commercial firm, to be relatively flexible in keeping up with advances in the relevant state-of-the-art.

9.1.6 How is the natural monopoly role of IPD addressed?

In contrast to NCREIF, which is a non-profit industry association, IPD is a private, for-profit firm whose equity is controlled by the managing principals in the firm. This institutional difference should be considered in light of the fact that evaluation benchmark indexing tends to be a natural monopoly as well as an information product that can be essential to the efficient functioning of the real estate investment industry (see Appendix B to Chapter 2 in Volume 1). Although other publicly-available indices do exist in the U.K., IPD clearly dominates in the property-level custom benchmarking and performance attribution businesses. Indeed, to our knowledge, there is no effective competing firm providing these types of services in the U.K.

Although the enjoyment of a natural monopoly no doubt provides certain advantages to IPD as a private firm, the successful long-run dominance of the market for property-level benchmarking services also requires that any firm or entity providing such services must exhibit a high level of responsibility, integrity and sensitivity to certain issues. For example, the ability to maintain confidential proprietary information from data-contributing members is the “1st Commandment” in this business. Similarly, it is necessary to maintain a sufficiently “arms-length” relationship from the data-contributors, while also obtaining something like industry-wide consensus on certain key technical issues and standards. It is also necessary to make index products and methodology widely-available and accessible at reasonable cost to industry and academic researchers, including being open to scrutiny and criticism. This is particularly important where the evaluation benchmark index overlaps with the broader “public good” functions of a “flagship” research index for the real estate asset class. We conclude that IPD must have an adequate track record in carrying out such responsibilities or it could not have been as successful as it has been, both in the U.K. and in an expanding list of foreign countries.

In some countries where IPD has recently begun operations, it has done so by working directly with individual data-contributing firms, much as it does in the U.K. For example, this is essentially the case in Germany and France. However, in other countries (e.g., Netherlands and Sweden), IPD operates through contracts with local industry associations that collectively represent the local data-contributing firms and broader industry considerations. This provides a collective mechanism to address at least some of the concerns associated with a private firm holding a natural monopoly on a product that is essential to the industry (see Question 2.B.4 in Volume 1).

In addition, IPD has always provided for industry and client representation in the governance of their system through several devices:

- Several surveyor firms hold minority shares in IPD;
- Investor/advisor clients are represented as non-executive Board members;
- Consultative groups are established for both the index and benchmarking services, with the former including actuaries, the Bank of England, the UK Government, the actuarial consultant firm WM, academics, and major users.

It should also be noted that there is some scope to balance “monopoly” power with “monopsony” power in the British system. (As described in Question 2.B.4 in Appendix B of Chapter 2, “monopoly” refers to one supplier, while “monopsony” refers to one customer.) IPD depends for a large fraction of its database (its vital supply of “raw material”), and its revenue, on a small number of major data-contributing firms. This makes it potentially relatively easy for IPD’s users to form a collective “user group”, and IPD is aware of this. IPD has stated that they target profit margins of 20%-25% of revenues, and as a LTD company their annual accounts are available for public inspection.

Section 9.2:

Some Technical Characteristics of the IPD Index, with Some Comparisons to the NCREIF Index

9.2.1 What indices does IPD currently produce in the U.K.?

IPD’s original and “flagship” index is its Annual Index of U.K. investment performance. This index begins in 1981, with an historical research series that extends back to 1971. This is an index of both fund-level (what IPD calls “investment” or “portfolio” level) and property-level (what IPD calls “market standing” level) total returns. This is combined with a full custom-benchmarking and performance attribution service capable of breaking down total returns into several layers of components. In 1990 IPD developed a Monthly Index, based on unit-funds and other sources that report external property valuations on a monthly basis. The Monthly Index starts in 1987 and has about one-fifth the number of properties and about one-tenth the aggregate valuation of the Annual Index. Currently, IPD is developing a quarterly index (and benchmarking service), based on a US\$50 billion sub-sample of their Annual Index properties that are valued quarterly.

9.2.2 What are the peer universes and market coverage of the IPD indices in the U.K.?

IPD attempts to include all types of professionally-owned investment real estate in the U.K., not limited either by type of investment vehicle or by “branch” of the industry (type of owner). The major types of professional investment *owners* are:

- Pension funds,
- Insurance companies (general accounts), and
- Property companies (publicly-traded real estate companies similar in some respects to REITs and REOCs, with no special tax treatment).

IPD also includes some of the major landed estates in Britain (some of which are organized as privately-held property companies). The major types of investment *vehicles* include:

- Separate (or “segregated”) accounts (the dominant vehicle for the pension funds),

- Commingled and pooled funds of various types (including unit trusts and limited partnerships, which are popular among smaller pension funds),
- Publicly-traded shares in property companies, and
- Joint-venture structures of various types (including limited partnerships).

IPD estimates that their property databank includes about 77%, by value, of all large commercial properties owned by professional investors (approximately 18,000 properties). About three-quarters of this databank (some 14,000 properties) is represented in the IPD Annual Index, in particular, all the properties valued as of December each year. Thus, the IPD Annual Index represents about 60% of the large, professionally-owned commercial properties in the U.K. Some of the non-December-valued properties are represented in other IPD indices, and IPD is working on developing a single index that would reflect all 18,000 properties, no matter what their valuation cycle. (To put this in some perspective, the US\$120 billion value of the IPD Annual Index property base is some 1.5 times the value of the NCREIF Index property base in a national economy with a GDP about 1/8th that of the U.S., which suggests a relative market coverage ratio some 12-times that of NCREIF, based on the ratio to GDP.)

As of 1999, the IPD Annual Index property ownership was composed as follows by value:

- 44% Insurance company funds
- 23% Pension funds (largely separate accounts)
- 13% Short-term funds (largely unit trusts and pooled LPs)
- 20% Other funds (including property company and estate holdings)

It should be noted that the IPD databank includes a much smaller representation of residential income property than other types of income property. This is because much residential income property in the U.K. is owned either by public sector entities or by unlisted specialized investors. In effect, it seems that much of the residential sector is viewed as a separate branch of the real estate investment industry that IPD has not yet incorporated in its databank. If all large residential income properties in the UK were included in the denominator, then the 77% “market” coverage percentage noted above would be reduced.

It should also be noted that IPD’s coverage of the publicly-traded property company sector, though much more substantial than its coverage of the residential sector, appears to be lower than its coverage of fiduciary institutions and unit funds. There are over 100 publicly-traded property companies in the UK, with some 80 firms in the Financial Times Property Index, while only eight firms are currently IPD data-contributing subscribers (according to Tony Key, IPD’s Research Director). However, these eight are among the largest property companies, and as the sector is quite concentrated, it is likely that they include over half of all the property holdings in the sector, by value. One reason that more property companies are not included may be that (similar to REITs in the U.S., if we understand correctly) such companies may not be required by law to regularly and frequently conduct full mark-to-market appraisals of their property holdings. Nevertheless, it does appear that, either out of legal requirement or accounting standard

practice, most of the *largest* British property companies do conduct regular and frequent appraisals of all of their property holdings (in this respect differing from most U.S. REITs). However, in many cases these firms do not use a December-ending fiscal year, so even if their properties are in the IPD databank, they are not well represented in the IPD Annual Index. Our understanding is that the gross asset value of the property holdings of the publicly-traded property company sector in the UK is probably on the order of £70 billion, while less than £12 billion of this is represented in the IPD Annual Index.* The establishment of IPD's new quarterly index should improve property company representation in an IPD index, as it will enable properties appraised on fiscal year-ends other than the calendar year-end to be included.

9.2.3 Are there institutional differences between the U.K. and U.S. that affect the difficulty of obtaining broad market coverage for a real estate index?

There are three major differences between the U.S. and U.K. that tend to make it more difficult to obtain broad market coverage in a real estate index in the U.S.

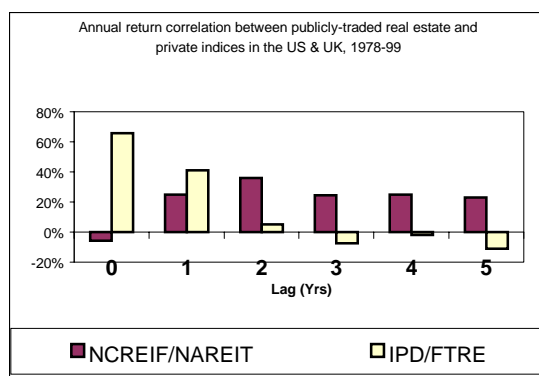
1. **Fragmented industry:** The institutional real estate investment industry in the U.S. is more fragmented than in the U.K. In the U.K. a small number of very large life insurance companies and pension funds own a larger proportion of all institutional property holdings than in the U.S. Institutional investors in the U.S. also tend to fragment their investments across multiple managers and advisory firms to a greater extent than in Britain, further complicating the institutional structure. It also seems likely that the U.S. may have a somewhat greater diversity of types or styles of investment vehicles and managers. For example, “opportunistic funds” in the U.S. are relatively rare in the U.K. These sources of greater fragmentation and diversity tend to increase the potential “free rider problem” and other obstacles to rapidly obtaining large market coverage in an index in the U.S.
2. **Fragmented standards:** There is probably less uniformity of standards in the U.S. regarding valuation practices and property investment performance reporting across the various types of investment vehicles and branches of the investment industry. For example, insurance company general accounts in the U.K. must regularly mark-to-market. In addition, the Royal Institution of Chartered Surveyors (RICS) exerts a very strong influence over the appraisal methods employed in all branches of the industry. Because a successful benchmark index requires broad agreement on valuation and performance measurement standards and the use of common standards across all index participants, this type of fragmentation also makes obtaining broad index market coverage more difficult in the U.S.
3. **Smaller institutional role:** The entire institutional ownership sector in the U.S. represents a significantly smaller fraction of all commercial property ownership than in the U.K. In part, this is because institutions in the U.S. tend to limit themselves to larger properties than in the U.K. For example, the average property

* This is an example of the rigor or IPD's index construction methodology in avoiding “stale appraisals” by only including in the index properties appraised as of the end of the calendar period covered by the index return report.

in the IPD Annual Index is worth about US\$10 million, and the average property in the IPD Monthly Index is worth only about \$5 million. In the U.S. \$10 million is often considered to be the *minimum* size for institutional investment. In fact, the average NCREIF property is worth over \$30 million. Many individual retail shops in British towns are owned by institutional investors, whereas most such properties in the U.S. are owned by individual investors and operating owner-occupants. In part, this is because insurance companies are more heavily invested in real estate in England than in the U.S. Also, the publicly-traded property company sector in the U.K. has traditionally been relatively larger than the REIT sector in the U.S. A rough estimate of the value of all property owned by U.S. institutions that roughly correspond to the institutions covered by IPD in the U.K. (essentially: pension funds, insurance company general accounts, and REITs), is \$500-\$600 billion.* This is less than three times the value of the peer universe IPD tries to cover in the U.K. (If the US\$120 billion IPD Annual Index is 60% of the overall U.K. professional investor universe, this implies that the British target population has a value of about US\$200 billion, with about 77% of this currently in the IPD databank.) Yet, the U.S. economy is approximately eight times the size of the British economy, as measured by GDP. The implication is that, even if a U.S. index achieved a penetration of the institutional property ownership comparable to what IPD has in the U.K., this would still represent a much smaller fraction of the overall commercial property population in the U.S. (In other words, there may be as much as \$2000 billion worth of “large” commercial properties owned by professional investors of one sort or another in the U.S.; 75% coverage of the major institutional holdings would be only some \$420 billion worth, or about 20% of the underlying property population by value.)

9.2.3 Do IPD indices in the UK suffer from temporal lag bias?

There appears to be much less temporal lag bias in U.K. IPD indices than in the NCREIF Index. Indeed, we find this to be one of the most impressive features of the IPD Indices and of the U.K. property market.



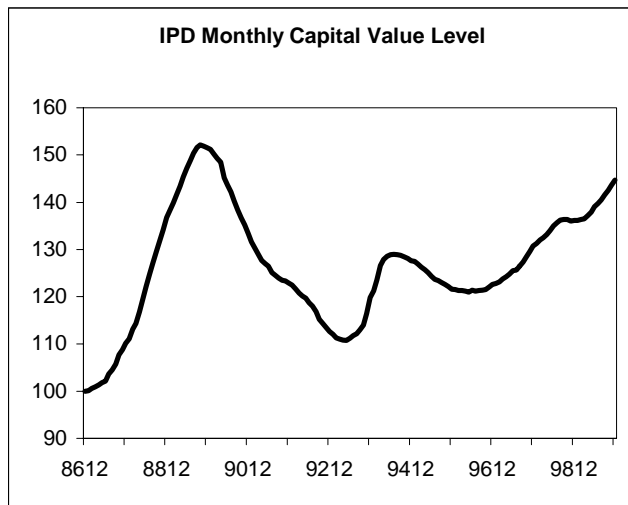
One way to quantify this difference is indicated in the exhibit. Assume share price movements by publicly-traded property companies or REITs as a basic “benchmark” of informational efficiency in the incorporation of relevant news into contemporaneous valuations. In other words, assume the London and New York Stock Exchanges are essentially equally efficient in an informational sense: when value-relevant news arrives, LSE-traded property company share prices reflect such

* This estimate is based on conversations with industry researchers, and reflects a rough breakdown of \$150 billion in pension funds, and \$300 billion in REITs, and \$100 billion in insurance companies.

news as quickly and accurately as REIT share prices traded on the NYSE. Then, the temporal lag between the public share price movements and the corresponding movements in private property index values would give some indication of the magnitude of temporal lag. This can be evidenced by the amount of lagged positive correlation between publicly-traded real estate share price returns and private real estate index returns later in time. As indicated in the exhibit, the average lag between the Financial Time Real Estate Index (FTRE) and the IPD Annual Index is less than one year (as indicated by the average lag with positive correlation). In contrast, the average lag between the NAREIT and NCREIF indices is over two years.

9.2.4 Is there a difference in temporal lag between the IPD Annual and Monthly indices?

There is some evidence that the IPD Monthly Index is slightly lagged behind the IPD Annual Index. This would occur if a significant proportion of the monthly valuations that underpin the Monthly Index are less than full, temporally independent reappraisals. (See 6.2.3 in Chapter 6 of Volume 1 for the meaning of the term “temporally independent” appraisals and an explanation of the importance of this issue for avoiding temporal lag bias in real estate indices.) For example, if the average property in the Monthly Index is



fully reappraised only once per year and not substantially updated in the intervening months, then the Monthly Index would lag about six months behind the Annual Index as of the end of each year (assuming all properties in the Annual Index are fully reappraised as of the end of the year). Although the evidence is not definitive, there is some empirical indication that the Monthly Index lags on average two to five months behind the Annual Index. For example, the widely-perceived sharp but temporary downturn in the property market in the third quarter

of 1998 (reflecting the global equity retreat triggered by the Russian financial crisis) was not reflected until the fourth-quarter in the IPD Monthly Index, and only then in a leveling-off rather than a sharp downtick. Nevertheless, even the IPD Monthly Index appears to be much less lagged than the NCREIF Index.

9.2.5 Are IPD indices less temporally lagged than other British private real estate indices?

Apparently not. There is no evidence that IPD indices lead other British private real estate indices of similar reporting frequencies. For example, the Richard Ellis and IPD Monthly Indices seem to be highly contemporaneous. In general, the contemporary

correlation between the IPD indices and other equal-frequency British indices is around 90% or more. (See Question 9.1.1 for a description of the Richard Ellis and JLL indices.)

9.2.6 What are the causes of the difference in temporal lag bias between the U.K. indices and the NCREIF Index in the U.S.?

There are three sources of the difference in temporal lag bias between the two countries. From most to least fundamental, these sources are: (1) the nature of the property asset markets, (2) the nature of the appraisal process at the individual property level, and (3) differences in the way the indices are constructed. Our perception of each of these sources of the temporal lag difference is described below.

- 1) **Property asset market differences:** There are important differences between the institutional commercial property asset markets in the U.K. and U.S., which probably cause the U.K. markets to be more informationally efficient than the U.S. markets. These differences probably make it easier for market participants to value the typical property in the U.K. as compared to the U.S., and to more accurately incorporate the effect of the latest news and information in such valuation. This would cause property markets to be less “sluggish” in the U.K., resulting in little (if any) temporal lag in the property market valuations themselves. The major differences between the U.S. and U.K. in this regard include:
 - a) *Property and leasehold differences:* Institutional commercial property in the U.K. tends to make greater use of much longer-term leases than is the case in the U.S. This gives private property equity investment more “bond-like” characteristics than in the U.S. This also makes property a more *homogeneous* asset, and therefore easier to value from a capital market perspective. In effect, some of what is unique and risky in the typical U.S. property investment is in the U.K. context passed on to the tenant who, in the traditional arrangement, typically signed a 25-year lease with 5-year upward-only rent reviews. Also, individual properties in the U.K. tend to be smaller than in the U.S., typically resulting in fewer numbers of tenants and leases per property, which again simplifies the valuation process.
 - b) *Geographic differences:* The economic geography of the U.K. is simpler than that in the U.S. There are fewer regions, with less economic diversity, and the institutional commercial property market is dominated by a single metropolitan area (London). Geographic simplicity also tends to make real estate more homogeneous and therefore easier to value.
 - c) *Role of publicly-traded property companies:* Throughout most of recent history a significantly greater share of British institutional commercial property was owned by publicly-traded property companies. In combination with the greater geographical and property-level homogeneity noted above, this makes it relatively easy for property “*price discovery*” to occur in the more efficient public stock exchange rather than in the private property market. British market participants and appraisers can more easily and directly glean information relevant to the valuation of privately-held properties simply by observing property company share price movements in the stock exchange.

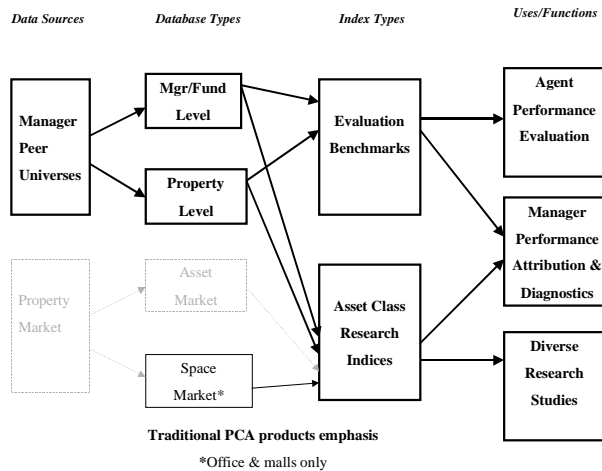
- 2) **Appraisal practice differences:** Both the U.S. and U.K. have old and well-respected property appraisal professions with long-standing traditions and formally-instituted standards of practice. Although broadly similar, there are nevertheless important differences. The RICS dominates the appraisal industry in the U.K. to perhaps a greater degree than the Appraisal Institute does in the U.S. Also, related to the property market differences noted in (1) above, RICS procedures may be developed to a greater degree of rigor and standardization. We have the personal impression that the appraisal profession is treated somewhat more seriously in Britain, and that the appraisal process in the U.K. relies more heavily on public capital market informational inputs. Perhaps as a result of this (as well as other cultural differences), we get the impression that property market participants in England pay more attention to the opinions of appraisers in the U.K. This would result in causality flowing more equally in both directions between appraisals and transaction prices in the U.K., whereas in the U.S. the flow of causality is more dominantly from transaction prices to (subsequent) appraisal estimates of value. Such differences would result in less (or no) temporal lag between property market values and appraised values in the U.K., while such a lag would tend to exist in the U.S.
- 3) **Index construction differences:** The NCREIF Index is constructed as a quarterly index from properties that are largely valued annually on a rolling basis throughout the year. The result is a “stale appraisal effect” in the NCREIF Index, which is one cause of temporal lag bias in that index. (See Questions 8.2.1-4 in the previous chapter.) The stale appraisal effect is largely absent from the British indices. Most notably, in the IPD Annual Index all properties are seriously reappraised as of the end of each calendar year, contemporaneously with the reporting interval of the index. At this frequency (and given the property market and appraisal practice differences noted above), these appraisals may well be “temporally independent” to an impressive degree. Certainly, the simultaneous reappraisal of so many properties (and a significant proportion of *all* institutionally-held properties, as noted in Question 9.2.2) must inevitably have an effect on the nature of the appraisal process. Even though a value is ultimately pegged onto each individual property, some elements of a “*mass appraisal*” process likely come into play in the British system, particularly as this annual end-of-year appraisal process is now coordinated to some extent (at least informally) by IPD (e.g., if nothing else then by virtue of the fact that most of the appraisals will be submitted into the IPD databank and hence subject to IPD review). As described in section 6.3 of Chapter 6 (in Volume 1), mass appraisal offers economies of scale and improves the terms of the “*noise-vs-lag trade-off*” in the appraisal process, helping to reduce temporal lag bias.

In summary, the difference between the U.S. and U.K. private real estate indices regarding temporal lag bias reflects only in part differences in index construction methodology between NCREIF and IPD (or other British indices). It seems likely that the greater part of the difference in temporal lag bias reflects more fundamental differences in the property asset markets and appraisal practices. On the other hand, it is important to note that IPD (and/or the system of annual benchmarking developed by IPD) has had some influence on the nature and functioning of the British markets and valuation process.

Chapter 10:
The PCA Index in Australia:
Some Issues Relevant to the U.S. Case

10.1 What is the role of the PCA Index, and what types of information products and functions does the PCA provide?

The Property Council of Australia (PCA, formerly the Building Owners & Managers Association - BOMA) is a multi-purpose, non-profit industry association that represents virtually the entire spectrum of the commercial property investment industry in Australia. Their membership encompasses all leading institutional investors in Australia, including pension funds, property trusts (roughly equivalent to REITs), and financial organizations, as well as private investors and developers. Compared to both NCREIF in the U.S. and IPD in England, the PCA is a relative newcomer to the indexing and benchmark service provision business. PCA began in 1990 with a property-level index very much like the



NCREIF Index, which they produced in association with Frank Russell Co. This index is still produced, now independent from Russell, under the name: *Investment Performance Index (IPI)*. In 1995, PCA extended their information services to include a fund-level custom-benchmarking and performance attribution service broadly similar to the IPD service in England, which they call the *Benchmark Clearinghouse: Fund Managers Information Service*

(“the Clearinghouse”, for short). Thus, the current information product coverage of the PCA in Australia is similar to that of the IPD in England, and broader than that of NCREIF in the U.S., as indicated in the exhibit. The Clearinghouse Service and its associated benchmark index sits above the Investment Performance Index in the value chain and draws heavily on the information from the Index. Not all data-contributors to the IPI have as yet elected to join the Clearinghouse Service, which has higher fees. On the other hand, some types of properties and investments that are included in the Clearinghouse (e.g., development projects, indirect property investments) are not included in the IPI.

10.2 What is the PCA Index’s market coverage?

The PCA Investment Performance Index includes over 80% of the institutionally-owned properties in the reported sectors, which include office, retail and industrial. As of June 2000 the index included 619 properties worth A\$42 billion (US\$27 billion). This is about one-third the value of the NCREIF Index property base, in an economy with less than 1/20th the GDP of the U.S., for a relative coverage ratio almost seven-times that of

NCREIF. The properties not included are largely those that are not re-valued at least annually. The types of owners covered include virtually all institutional investors in Australia. In addition to pension funds and insurance companies, the covered investors include Listed Property Trusts (LPTs, the Australian equivalent of REITs), which in Australia are required to annually mark-to-market their property holdings. (PCA includes A\$24 billion worth of the total pool of A\$29 billion LPT property holdings.) The PCA annually surveys members in order to quantify the total market base by location and property sector, including properties that are not part of the index.

10.3 How does the PCA Index compare to the NCREIF Index in its index construction procedures?

The scope of information products and service provision of the PCA is more similar to that of IPD than NCREIF. However, its Investment Performance Index has more technical index construction characteristics in common with the NCREIF Index than with the IPD Annual Index. In particular, the PCA Index is quarterly, but based on a property population that is largely reappraised annually at various times throughout the year. (More than half the PCA properties are revalued annually, several contributors value their properties twice per year, and a small minority value every quarter.) As a result, the PCA Index experiences a “stale appraisal effect” similar to that in the NCREIF Index. (The PCA Index was biannual from 1984 through 1994, moving to quarterly in 1995.)

10.4 How does the Benchmark Clearinghouse work?

The Clearinghouse currently includes 454 properties from 41 funds, worth A\$32 billion in aggregate. The Clearinghouse was established to report on fund level performance rather than the direct property level focus of the Investment Performance Index, and to provide a custom-benchmarking and performance attribution service at the property operational level. Clearinghouse subscribers pay an additional fee and must contribute some additional data beyond what is required for the IPI. At present the Clearinghouse reports on a semi-annual basis and does not include all of the data-contributors in the IPI. Within the Clearinghouse the IPI is used to benchmark certain property market segments that have too few properties within the Benchmark Clearinghouse to meet the minimum criteria for publication. The additional functionality that the Benchmark Clearinghouse adds beyond PCA’s original IPI contribution is as follows:

- Covers all direct property sectors in investor portfolios (development sites, agriculture, car-parking, hotels and tourism, non metropolitan office etc, as well as the core office, retail and industrial sectors included in the IPI).
- Benchmarks the performance at the fund-level, including investments other than direct property (e.g., listed and unlisted property trusts, cash etc).
- Measures the performance of properties held in non-transparent JV’s, which are analysed on the basis of their underlying property performance and the performance of the JV structure.

- Provides attribution analysis at the portfolio and property operational levels (including separate peer group benchmarking of cash and indirect property investments).
- Measures lease expiry profile by income and area of the direct property portfolio
- Allows comparison by peer group in addition to the market comparisons available through the IPI
- Includes all property information including partial quarters of ownership and impact on the fund
- Measures direct property returns both pre and post transaction costs. Transaction costs are excluded from comparisons in the IPI

10.5 Are there any unique and interesting features of the PCA benchmarking methodology and information products?

We have not had time to attempt a detailed and comprehensive review of all the benchmarking information services provided by PCA. We have, however, been impressed with the technical committee structure and procedures the PCA has instituted with the objective of insuring the PCA's information products reflect the latest international state-of-the-art practices, and that the procedures adopted by the PCA are well publicized and explained to the industry. This system seems to work effectively. For example, to our knowledge, the PCA was the first benchmarking agency to implement the 3-part breakdown of portfolio-level performance attribution defined on "pure" allocation and selection components plus an "interaction" effect, an approach that has some advantages in clarity of attribution and academic acceptance. Also, as a multi-purpose real estate association, the PCA is not limited to collecting data from the peer universe of data-contributors involved in the IPI and Benchmark Clearinghouse. As a result, the PCA collects additional data directly from the space and asset markets.

10.6 What data does the PCA collect directly from the space and asset markets?

As a multi-purpose industry association with very broad representation across the Australian property industry, the PCA is able to tap a number of different types of sources of data, in addition to the IPI and Clearinghouse data contributed by subscribers and participants in those indices. Much of this information is potentially useful for research purposes. As examples, PCA:

- has maintained an extensive database on the office space market for some 15 years. This tracks on an individual property basis such items as quality grade, vacancy (direct, sub-lease and total), gross and net leaseable areas, and status of space (withdrawn, new supply, partial or full refurbishment).
- surveys new developments for future supply coming online over a period of 3 years, obtained from real estate agencies and property owners. (This data is a component of the official Australian index of leading economic indicators.)
- maintains a database of every shopping centre in Australia, including information on size, turnover, pedestrian estimates, major tenants and tenant areas, number of specialities and a range of other key statistics on each centre.

- is in the process of developing a complete inventory of Australian industrial property, testing a satellite imaging system against government cadastral maps to generate a complete stock list.
- conducts an annual survey of all their members to track the magnitude of institutional holdings in all markets, whether or not properties are included in the IPI or Clearinghouse databanks (e.g., due to lack of sufficient reappraisal frequency).
- collects on an annual basis information on office and retail properties entire chart of accounts (rental income, salaries, water rates, lifts, air conditioning, and so forth).