The Journal of Real Estate Finance and Economics, 32:1, 41–60, 2006

How Does Appraisal Smoothing Bias Real Estate Returns Measurement?

ROBERT H. EDELSTEIN
Fisher Center for Real Estate and Urban Economics, Haas School of Business, University of California at Berkeley, 602 Faculty Bldg., MC6105, Berkeley, CA 94720-6105, USA
E-mail: edelstei@haas.berkeley.edu

DANIEL C. QUAN
Finance, Accounting and Real Estate, School of Hotel Administration, Cornell University, 442 Statler Hall, Ithaca, NY 14852, USA
E-mail: dq22@cornell.edu

Abstract

This paper examines and clarifies several related issues about real estate return indexes. Specifically, even if real estate valuation smoothing exists at the individual property level, such errors may offset in the aggregate. Using data from commercial property appraisals and corresponding transactions, appraisal smoothing errors engender an underestimation of both the first and second moments for real estate returns. After correcting for these "underestimations," real estate mean returns and the variance appear to be quite similar to those of stocks.

Key Words: appraisal smoothing, error aggregation, real estate returns

1. Introduction

This paper examines how individual property appraisal smoothing affects the accuracy of aggregate real estate performance indexes. The analysis examines and clarifies three related issues about the smoothing of real estate appraisals and the smoothing of real estate return indexes.

a) It is alleged that an index computed from a sample of "smoothed" appraisals has an artificially low second moment (variance) relative to the variance computed from the sales transactions for identical properties. If this assertion were true, it would be disquieting, since the usual metrics of risk and diversification for investment analyses are dependent crucially on measures of dispersion. An artificially smooth series will necessarily underestimate the riskiness of the real estate asset class, and may distort its correlations with returns of other assets. A wide range of research alleges that aggregate real estate rates of return indexes are smoothed because they employ smoothed individual property appraisals. We show that the use of individual property smoothed appraisals (with reasonable assumptions about errors), when "aggregated" into an average, *a priori* does not necessarily generate an aggregate index that will either over or under estimate the true return index and/or its volatility. It does not necessarily follow that the aggregate index variance is biased even if individual appraisals used to construct the index exhibit such biases, because individual property appraisals biases may offset each other in the aggregate. Thus we argue that the impact of individual property valuation smoothing upon the smoothing in an aggregate index is an empirical issue.

b) In this study, we evaluate the relationship between individual property appraisal error and the aggregate rate of return index. Using a set of plausible assumptions relating to statistical errors for appraisals and transactions (sales), we derive a statistical measure of confidence for an aggregate index based upon appraisals instead of transactions. This confidence interval can be utilized, in principle, to adjust and correct the estimators for the first and second moments of the aggregate appraisal based index. In this way, we illustrate how one could create a confidence interval around the "smoothed appraisal" generated index.
c) Combining a sample of contemporaneous sales transactions and appraisals for identical commercial real estate parcels, the implied statistical reliability of the aggregate real estate return index is computed. Two key assumptions are required to conduct the analysis. First, appraisal error over our sample period is drawn from a common random distribution. Second, the appreciation component of the aggregate index can be adequately proxied by appraisals.

Our analysis, while subject to certain limitations, provides new insights about measuring real estate performance using appraisal-based indexes, such as the National Council of Real Estate Investment Fiduciaries (NCREIF) Property Index or NPI. Three principal empirical results emerge for our sample about appraisal based aggregate return indexes:

a) The appraisal based mean rate of return for our data sample as well as the NPI index are likely to be biased downwards, thus underestimating the true rate of return for real estate.
b) The adjusted estimated rate of return for real estate for the sampe period, 1979–1984, is comparable to those achieved for CRSP and S&P stocks, but lower than those for small capitalization stocks. For the more recent period, 1996–2003, our corrected mean rate of return for the NPI index indicates that commercial real estate performed comparably with both S&P500 and the S&P Small Cap 600 indexes, and outperformed the Lehman Government Bond Index and T-Bills. The corrected volatility (second moment) for commercial real estate appears to be lower than those of both stock indices.
c) The variance for the appraisal based return index, based upon the empirical analysis of our sample, is substantially understated, perhaps by 50% to 80%. This finding is consistent with much of the existing real estate shibboleth.

To our knowledge, this is the first study to develop a statistical confidence interval with an adjustment for smoothing effects based upon individual appraisal bias. To the extent our techniques are applicable, we provide a statistical technique for correcting existing and future appraisal-based real estate return studies.

The paper is organized into six subsequent sections. The next section provides a backdrop for our approach, including a brief literature review. Section 3 discusses why individual property valuation smoothing does not imply smoothing for an aggregate rate of return index. A brief overview of the data base is presented in the next section. Section 5 creates a framework for generating measures of the aggregate index returns. Section 6, the heart of the paper, delineates our empirical findings. The last section concludes.

2. Backdrop for our approach

A substantial, respected existing literature, spanning the 1970s and 1980s, concludes that real estate provides a higher risk-adjusted return compared with other investment alternatives.[1] These studies imply that the inclusion of real estate in a portfolio of investments can substantially reduce portfolio risk; and find that real estate is a hedge against inflation. In contrast, more recent research, utilizing extended similar methodologies, but subsequent market data, has found more modest real estate performance.[2] The differing results, in part, may be attributable to cyclical sample specific effects as well as secular real estate market changes. The conclusion of both views in the literature depends upon the construction of one or more real estate return series which can be compared with similar return indexes for other investments. In the case of real estate, where market transactions are relatively infrequent, professional appraisals are used to represent market values in order to generate rate of return indexes. Indeed, this is the standard practice.

Several difficulties greatly restrict the usefulness of real estate rates of return series computed from unadjusted appraisal data. The classic criticism alleges that the aggregate real estate rate of return index is smoothed because it employs smoothed individual property appraisals.[3] A smoothed index reduces the variance of returns reported for a sample of appraisals relative to a sample of sales transactions for identical properties. Since the usual metrics of risk and diversification rely crucially upon measures of dispersion, an artificially smoothed series would necessarily underestimate the riskiness of the real estate asset class, and would distort its correlations with the rates of returns of other assets.

The notion of smoothing can be viewed as errors resulting from the calculation of a statistic using imperfect appraisal estimates of individual property values. Smoothing is defined as the deviation of an index from one which is never observed; and since it is this deviation between the series which gives rise to this problem, it is not surprising that smoothing is often demonstrated based on assumptions made about the true series and appraiser methodology and practice. The notion that appraisal-based return indexes are "too smooth" is so well entrenched in the literature that researchers have proposed methods to reverse engineer to "unsmooth" indexes, using these same assumptions [see, for examples, Barkham and Geltner (1994), Fisher et al. (1994), and Cho et al. (2003)]. Frequently, in the absence of empirical support, the existence of smoothing is largely assumed and unsmoothing techniques amounts to a contrived solution to an assumed problem.

There have however been several endeavors to model index errors via simulations [Quan and Quigley (1989); Giaccotto and Clapp (1992)]. For studies using actual transaction prices, Webb et al. (1992) construct an index using only sold properties. The discrepancy between appraisals and prices has been investigated by Cole et al. (1986), Miles et al. (1991), Webb (1994) and Graff and Young (1999) for commercial properties, and Dotzour (1988) for houses. The influence of valuation errors on the return index has not been explored in these studies. More recent studies have focused on understanding how individual appraisers form their estimates, and what factors influence the appraisal valuation [Hansz and Diaz (2001); Diaz and Wolverton (1998); McAllister et al. (2003)]. Clayton et al. (2001) estimate and test structural models of individual appraisals as well as develop methodologies to estimate a returns index, using repeated appraisals from external appraisers, which are alleged to be more accurate [see Geltner and Goetzmann (2000)]. Similarly, Brown and Matysiak (1998) used repeated appraisals to empirically investigated the stability of the smoothing parameter in smoothing models.

This paper, utilizing a sample of arms length market transactions for 71 large commercial real estate parcels between 1979 and 1984, investigates and provides empirical evidence of real estate return index smoothing. The relationship between individual parcel appraisal smoothing and aggregate index smoothing is indeterminate a priori, and is ultimately an empirical issue. To the best of our knowledge, no one has statistically modeled and quantified the effects of aggregation errors upon real estate rates of return indexes.

Our method for examining the effects of appraisal smoothing will use observed arms length sales transactions to infer the statistical reliability of the appraisal based index. In this context, the appraisal index is a random variable subject to sampling errors. For a random sample of transaction prices, appraisal errors and their distribution can be estimated.[4] Based on the sample, the error distribution for the appraisal based rate of return index can be derived. Confidence intervals can be computed and used to determine the region of reliability. This calibration creates both a "smooth-adjusted" return and volatility measures to compare real estate return performance over two sample periods, 1979–1984 (the period corresponding to our transaction prices) and 1996–2003. Since we do not have transaction prices for this latter period, we make the additional assumption that the error distribution calibrated from the earlier period is valid for this latter period.

In the next section, the need for and importance of empirical support is motivated by a general discussion about how smoothing is presented and treated in this literature. This is followed by a proposed procedure which use appraisal errors to infer the degree of reliability of indexes. Applying this methodology to our sample, we find that the mean return of the aggregate index is on average biased downwards by 9%; supporting the often held view that real estate is a high yielding asset.

3. The nature of the problem

The appraisal based return of a real estate portfolio is constructed using individual property appraisals. Since individual appraisals and their associated errors are averaged, one cannot a priori determine whether the resulting index will over- or under-estimate the true return index. Similarly, it does not follow necessarily that variance measures of portfolio performance based on appraisals are biased even if individual appraisals exhibit biases; that is, individual property appraisal biases may offset in the aggregate.[5]

Consider a simple numerical example for an appraisal based aggregate index and its standard deviation computed using 2 properties with appraisal "smoothing." In Table 1, appraisals for the two properties over 4 periods are "smoothed" relative to the true unobservable values as indicated by the standard errors. This corresponds to the case when appraisers smooth property appraisals individually, and the degree of smoothing is independent for the two properties. In the construction of the aggregate rate of return index based only on the appraised values, the returns from each property are determined and averaged to form the aggregate index values for each time period. By construction, from this simple example, even though appraisals may be "smoothed" for each individual property, the resulting index will not exhibit "smoothed" behavior. In the example, the errors offset such that the appraisal based index standard deviation is identical to that of the "true" market value based return index. This hypothetical situation offers a clear counter-example to the seemingly widely accepted view that individual property appraisal smoothing will necessarily lead to smoothing in the aggregate property rate of return index.

Smoothing in the aggregate can occur if appraisal errors are identical for each property in the portfolio in each time period. In this case, any given observed error can be viewed as an error from a representative property in an identical pool and if smoothing occurs for the representative property, then smoothing occurs for the pool. Alternatively, if all properties in a portfolio were appraised by the same appraiser, then systematic errors may exist. If systematic appraisal errors exist, this does not necessarily imply smoothing in the aggregate. Since errors are also a function of the true price movements, the true market value of properties in a portfolio may change in

Table 1. Hypothetical two parcel index calculation.

	Prices				
	t = 1	*t = 2*	*t = 3*	*t = 4*	
Property 1					
True Market Value	103	135	145	140	
Appraised Value	110	135	142	129	
Property 2					
True Market Value	120	125	162	110	
Appraised Value	119	123	150	90	
		Period Property Returns			σ
Property 1					
True Return (%)		31.07	7.41	−3.45	**18**
Appraisal Return (%)		22.73	5.19	−9.15	**16**
Property 2					
True Return (%)		4.17	29.60	−38.27	**34**
Appraisal Return (%)		3.36	21.95	−40.00	**32**
		Aggregate Index Return			σ
True Return (%)		17.62	18.50	−20.86	**22**
Appraisal Return (%)		13.04	13.57	−24.58	**22**

a manner causing the resulting errors to wash out. In practice, the possibility of systematic errors is unlikely. The 1992 NCREIF index of real estate returns is based upon the performance of 1892 commercial properties diversified by region and land use, and held by several independent institutional real estate investment funds. The appraised values are determined by "independent" appraisers chosen by each fund. It is implausible that systematic errors exist for all properties in the sample.[6]

In sum, without detailed information about how each appraiser forms his estimates *and* how the estimates differ between appraisers, the detection of smoothing of the rate of return or standard deviation smoothing at the aggregate level is an empirical issue the example is meant to provide.

The previous example is meant to provide a plausible case and counter-example to commonly accepted views about aggregate index smoothing.

A more convincing demonstration of aggregate index smoothing based on a property appraisal smoothing would be to derive a general relationship using a model of appraisal behavior. To this end, we use the Quan and Quigley (1991) (QQ) model of appraisal behavior to investigate the volatility of an index comprised of 2 properties.

QQ showed that transaction prices at time t, P_t^T, can be expressed as a linear noisy signal of the true unobserved price P_t where the noise stems from market imperfections in the trading environment.

$$P_t^T = P_t + \nu_t \tag{1}$$

Furthermore, the true prices are assumed to follow a random walk.

$$P_t = P_{t-1} + \eta_t \tag{2}$$

where $v_t \sim N(0, \sigma_v^2)$, $\eta_t \sim N(0, \sigma^2)$ and $E(v_t v_{t-j}) = E(n_t n_{t-j}) = 0$ for all $j \neq 0$.

Based on their model and the assumptions, the optimal appraiser's (linear) updating rule is

$$P_t^* = KP_t^T + (1-K)P_{t-1}^* \tag{3}$$

where P_t^* is the appraiser's estimate of P_t and $K = \sigma^2/(\sigma^2 - \sigma_v^2)$ is the "signal-to-noise" ratio. If returns x_t^* are represented as first differences of the prices such that $x_t^* = (P_t^* - P_{t-1}^*)$, then the updated returns will follow a first order autoregressive process of the form:

$$x_t^* = (1-K)x_{t-1}^* + K\eta_t \tag{4}$$

In order to demonstrate the effects of aggregation, consider an aggregate index which is constructed from two independent appraisals for 2 properties with the following returns:

$$x_{1t}^* = (1-K_1)x_{1t-1}^* + K_1\eta_t \tag{5}$$

$$x_{2t}^* = (1-K_2)x_{2t-1}^* + K_2\eta_t \tag{6}$$

where $K_i = \sigma^2/(\sigma^2 - \sigma_i^2)$ for $i = 1, 2$. We assume that the true prices for both properties are the same but that the appraisals differ because of differences in v_i, the idiosyncratic noise facing each appraiser. The assumption of the same true price for both properties simplifies our analysis and focuses our attention on offsetting errors arising from different appraisers being exposed to different market noise. This assumption is not crucial to our result. We clearly do not suggest that all properties follow the same dynamic process. This specification is more restrictive than the general case of allowing both the appraisals and the true prices to vary which, depending on the relationship between the true price dynamics and the appraisals, could lead to an additional source of offsetting aggregate errors. However, if errors can offset in this simpler specification, then clearly offsetting errors in the more general framework could occur.

Due to the random walk assumption, σ^2 is the true price volatility; that is $Var\,(P_t - P_{t-1}) = Var(\eta_t) = \sigma^2$. Consider the variance of an index constructed from the average of x^*_{1t} and x^*_{2t}.[7] Thus we are interested in the time series variance of the following aggregate index R_t:

$$R_t = \frac{x^*_{1t} + x^*_{2t}}{2} \tag{7}$$

It is shown in Appendix that

$$Var(R_t) = (\sigma^2)^2 \left[\frac{1}{4(\sigma^2 + 2\sigma_1^2)} + \frac{1}{4(\sigma^2 + 2\sigma_2^2)} + \frac{2}{\sigma^2 + \sigma_1^2 + \sigma_2^2}\right] \tag{8}$$

The volatility of the aggregate index, which is the average of individual appraisals, is a function of σ^2, the true volatility of the underlying price process which we assume to be the same for both appraisers, and σ_1^2, σ_2^2, the volatility of the transaction noise contained in the transaction price observed by each appraiser in his own market. From inspection, depending on the values taken by these parameters, $Var(R_t)$ can be greater than, equal to or less than the true volatility σ^2. To easily see this, we plot (8) in Figure 1 with $\sigma_1^2 = 50$ and $\sigma_2^2 = 75$ and by letting the true variance σ^2 increase from 0

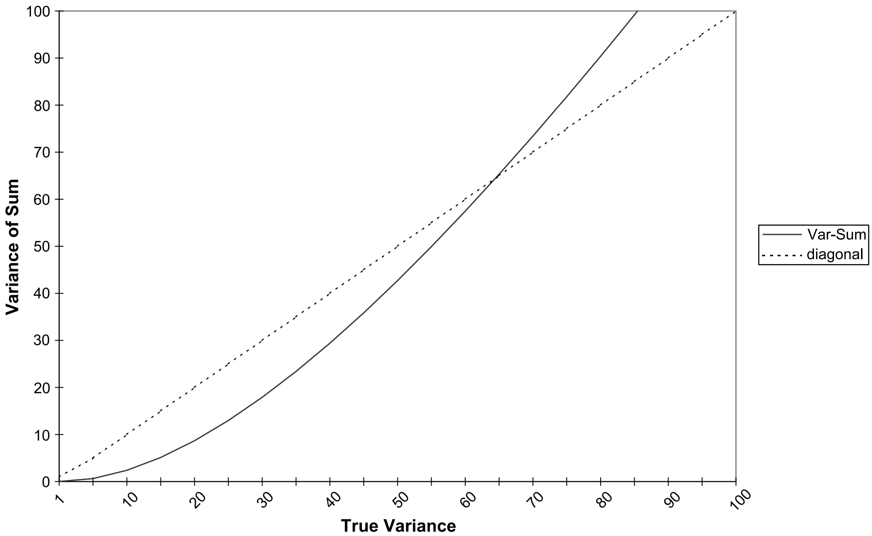

Figure 1. Hypothetical two parcel analysis of variance of sum versus true variance (sigma 1 = 50, sigma = 75).

to 100. The assumption of different individual noise variance corresponds to the case where there is more transactions price noise in the second market than there is in the first. We can see that the volatility of the aggregate index can be greater than, equal to or less than the true volatility of the index. Based on the specific values used, as long as the true volatility is less than 65, the index volatility will underestimate the true volatility; whereas at any values above this threshold, the aggregate volatility will overestimate the true volatility.[8] Thus, based on reasonable parameterization, the index volatility can be extremely unreliable and a priori one cannot determine the direction of the bias. A generalization to more than two properties is straightforward.

In sum, without detailed information about how each appraiser forms his estimates *and* how the estimates differ between appraisers, the impact of smoothing of the rate of return or the standard deviation of returns at the aggregate level is an empirical issue.

4. Understanding the data

4.1. The data base

The sample contains appraised value and arms length transactions sales data for 100% equity owed nonresidential properties between 1975 and 1984. Each of these parcels is held in one of six large commingled Real Estate Funds for Pension Trusts.[9] Within the sample, 71 properties were sold between 1981 and 1984.

A feature of the data is the similarity of the property composition to that of the NPI data base. Hence, we use our sample data to estimate the distribution of the difference between appraised value and the actual sale price; and we then use the findings from the sample to make inferences about the NPI index. To verify this similarity, a sample of 102 properties with complete appraisal information are employed to construct an overall return index as well as sub-indexes corresponding to the income and appreciation components using the NPI index methodology. Consistent with the NPI's calculations, the income and appreciation return components are calculated as follows:[10]

$$\text{Income Return}_t = \frac{Y_t}{A_{t-1} + 0.5(C_t - Y_t)} \quad \text{and}$$

$$\text{Appreciation Return}_t = \frac{A_t - A_{t-1} - C_t}{A_{t-1} + 0.5(C_t - Y_t)}$$

A_t represents the appraised value at time t, C_t is the property capital expenditure for year t, and Y_t is the net operating income (cash flow) in year t. The total index is the sum of both components. The results are provided in Table 2. An examination of Table 2 and Figure 2 (containing plots of the two indexes), demonstrates that our indexes are highly correlated ($\rho = 0.98$) with the NPI index, and that the means for teh rates of return indexes are similar. Our index appears to exhibit slightly more volatility than the NPI index, with much of this excess volatility arising from the appreciation component. In Figures 3 and 4, we display each index along with the

Table 2. Rates of return comparison for NPI and sample property indexes.

Year	*FRC index*	*Our index*	*FRC income*	*Our income*	*FRC capital*	*Our capital*	$\frac{A_t - A_{t-1}}{A_{t-1}}$
1979	20.8	26.8	9.0	8.7	11.1	18.0	18.0
1980	18.1	22.3	8.3	8.4	9.1	13.9	14.6
1981	16.8	15.9	8.1	7.8	8.3	8.1	8.9
1982	9.4	6.8	7.8	8.0	1.4	−1.3	−0.4
1983	13.2	12.2	7.8	7.4	5.1	4.8	6.4
1984	13.0	11.0	7.3	6.8	5.4	4.2	6.4
Mean	15.21	15.83	8.05	7.85	6.73	7.95	8.98
σ	4.12	7.49	0.58	0.69	3.46	7.01	6.55

corresponding income and appreciation components. The bulk of the variability originates from the appreciation component. For the overall return indexes (Figures 3 and 4 and Table 2), the bulk of the volatility originates from the appreciation component which is appraisal based. The low volatility of the income component may be related to long term tenant lease contracts that constrain income variability. Figure 5 contains the appreciation components for both indexes, which move together ($\rho = 0.97$) even though our overall index is more volatile.

4.2. The relationship between the index and appraisals

Because of our data base and the index construction methodology, we cannot isolate the pure appraisal influence from the total index (or the appreciation component). For this reason, certain simplifying assumptions about how appraisals influence the aggregate index are necessary. The percentage change in appraisals in our sample is reported in the last column of Table 2. Within our sample, the appraisals follow very closely to the appreciation component ($\rho = 0.99$). The mean real estate rate of return

Figure 2. Overall real estate rate of return: NPI index versus Our index.

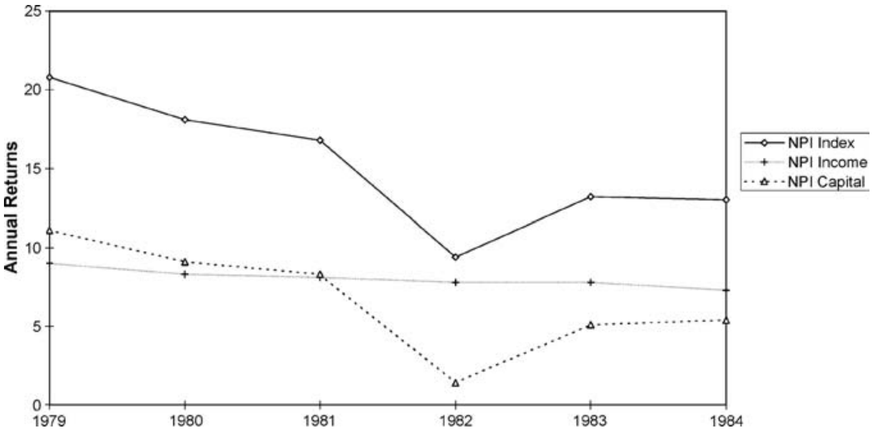

Figure 3. NPI return index and capital and income components.

for appraisals (8.98) is higher than the mean rate of return (7.95) of our appreciation component. However, the volatilities are quite close, a standard deviation of 7.01 for the appreciation component of our sample versus 6.55 for the appraisals. Because of this similarity, one approach is to substitute the appraisal return for the appreciation component in our sample to calibrate the error in our overall index. Utilizing the assumption that the appreciation component of the NPI index is also similar to our appraisal index, we then substitute our appraisal index for the NPI's appreciation component to calibrate its error.

To see further the influence of the appraisal only index on the aggregate index, both our aggregate index and the NPI aggregate index are regressed on our appraisal only return index (Table 3). In both regressions, the high R^2s suggest that the variability of both the NPI real estate return index and our constructed return index are highly

Figure 4. Our return index and capital and income components.

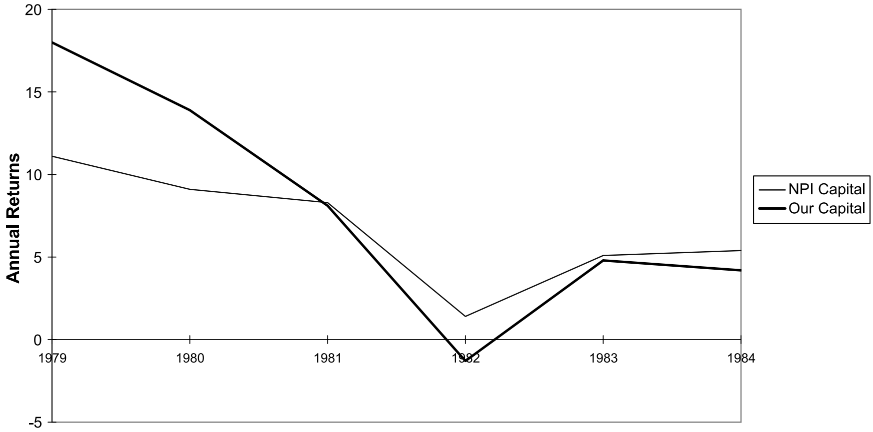

Figure 5. Capital components of NPI and Our index.

correlated with the rate of change of the appraisals (which in turn strongly correlates the appreciation component of the index[11]). Based on these results, the subsequent analysis is simplified by assuming that the aggregate return index = income return + appraisal return. By determining the bias in the appraisal return and assuming that the income component is relatively constant over our sample period, one can determine the error in the aggregate return index. The next section describes how the error in the appraisal return is determined, and computes its distribution.

5. A characterization of aggregate error

Our position is that errors in the aggregate real estate rate of return index are determined by fundamental errors in appraisals. There have been several studies which have quantified appraisal errors (Dotzour, 1988; Cole et al., 1986; Miles et al., 1991). None of these studies attempts to relate appraisal or transaction value errors to return errors.

We derive the relationship between appraisal error and the aggregate rate of return error. In order to obtain a tractable closed form expression for the "bias," several key assumptions are required about the error distribution. Let t denote the time period for which the return is defined, and i is the index for properties at t. Appraisal error is defined multiplicatively.

$$P_{it} = A_{it}\epsilon_{it} \tag{9}$$

Table 3. Regression of index on sample appraisal returns (standard errors in parenthesis).

Dependent variable	*Constant*	*Our sample's appraisal return*	R^2
Our aggregate return index	5.69	1.128 (0.092)	0.974
NPI aggregate return index	9.69	0.616 (0.064)	0.957

where P_{it} represents property i's price at t and A_{it} is its appraisal. Assume that the error term is independent and lognormally distributed:[12]

$$\epsilon_{it} \sim LN(\mu, \sigma^2) \tag{10}$$

We assume both cross-sectional and time series independence for the appraisal errors. Hence, for the purposes of estimating its distributional parameters, no distinction is made between cross-sectional and time series data. The lack of data within our sample precludes the estimation of potentially interesting time varying moments.[13] Also, assume that the appraisal estimate A_{it} is independent of its error ϵ_{it}. This condition is justifiable if the appraiser is viewed as an optimal processor of information, both private and public, in making his estimate (as described in QQ). Under such an interpretation, the appraiser forms an expectation of the random variable log(P) conditional on his information set. In these circumstances, an appraisal will be independent of his errors. Similarly, if the appraiser utilizes a regression estimated with all available information, the conditional expectation of the error term will be zero. The independence assumption may not hold if the appraisers in our sample err due to a common inappropriate appraisal methodology or other source of uniform bias, a condition unlikely to hold in our sample since our data is geographically diverse.

Since the analysis is concerned with the time series moments for periodic average returns, one needs to distinguish between expectations taken across properties and time. Consistent with the manner in which the aggregate index is constructed, and letting $E_t[\cdot]$ denote the expectations taken over time, P_t be the expected value of P_{it} taken over i for t, with A_t and ϵ_t similarly defined, the desired time series moments for our appraisal return index are:

$$E_t\left[\frac{P_t}{P_{t-1}}\right] = E_t\left[\frac{A_t\epsilon_t}{A_{t-1}\epsilon_{t-1}}\right] \tag{11}$$

We are interested in the time series moments of averages taken across different properties in each time period. Based on our independence assumption, equation (11) can be rewritten as equation (12):

$$E_t\left[\frac{P_t}{P_{t-1}}\right] = E_t\left[\frac{A_t\epsilon_t}{A_{t-1}\epsilon_{t-1}}\right] \equiv E_t[R_P] = E_t[R_A\epsilon] \tag{12}$$

For ease of exposition, delete the time subscript from the expectation, and assume that R_A is independent of ϵ to obtain the following:

$$E[R_P] = E[R_A]E[\epsilon] \tag{13}$$

By Jensen's inequality, $E[\epsilon] > 1$ thus $E[R_P] > E[R_A]$. This source of bias is strictly due to the manner in which returns are calculated, and suggests that the mean returns from an appraisal determined rate of return index will underestimate the return for a series based on transaction prices.[14]

In order to calculate the time series variance bias of an appraisal based return index, assume the independence between R_A and ϵ to obtain the following expression for σ_P^2, the time series variance of R_P:

$$\sigma_P^2 = E\left[R_A^2\right]E\left[\epsilon^2\right] - E[R_A]^2E[\epsilon]^2 \tag{14}$$

Noting that $E[R_A^2] = \sigma_A^2 + \mu_A^2$ and $E[\epsilon^2] = \sigma_\epsilon^2 + \mu_\epsilon^2$ where σ_A^2 and μ_A^2 are the variance and the mean of A_t, respectively, and σ_ϵ^2 and μ_ϵ^2 are similarly defined for the error terms, (14) simplifies to (15):

$$\sigma_P^2 = \sigma_A^2\left(\sigma_\epsilon^2 + \mu_\epsilon^2\right) + \mu_A^2\sigma_\epsilon^2 \tag{15}$$

In order to compute (15), one needs to derive the distribution for ϵ. Our assumption of lognormality and intertemporal independence allows the computation of the moments of the aggregate return errors. If ϵ_{it} is lognormally distributed with mean μ and variance σ^2, then $\overline{\epsilon}_{it} \equiv \log\epsilon_{it} \sim N(\overline{\mu}, \overline{\sigma}^2)$ such that

$$\mu = e^{\left(\overline{\mu}+\frac{1}{2}\overline{\sigma}^2\right)} \text{ and } \sigma^2 = e^{\left(2\overline{\mu}+\overline{\sigma}^2\right)}\left(e^{\overline{\sigma}^2} - 1\right) \tag{16}$$

Using the above results, it can be shown that the ratio of i.i.d. lognormal random variables will also be lognormally distributed with the following moments:

$$\epsilon \equiv \frac{\epsilon_{it}}{\epsilon_{it-1}} \sim LN\left[e^{\overline{\sigma}^2}, e^{2\overline{\sigma}^2}\left(e^{2\overline{\sigma}^2} - 1\right)\right] \tag{17}$$

Jensen's inequality will produce $E[\epsilon] = e^{\overline{\sigma}^2} > 1$ since $\overline{\sigma}^2 > 0$ and therefore implying that expected returns based on appraisals will *always* underestimate the transaction based expected return.

6. Empirical results

Since the ϵ_{it}'s are assumed to be distributed lognormally, the first two moments of the return error are estimated from the appraisal return index via (16). Using our sample of 71 properties which have both appraisals and transaction prices, the mean and standard deviation of the natural log errors are -0.0329 and 0.3897, respectively.

6.1. Mean returns

With the fitted error distribution and the simplifying assumption that the appraisal return is identical to the appreciation component, one can construct confidence intervals for the mean return for both the NPI as well as our index. Since $R_P = R_A\epsilon$ and by assumption $E[R_P] = E[R_A]\ E[\epsilon]$, the return error is derived by evaluating $E[\epsilon]$. Using the previously calculated standard deviation of log errors yields $E[\epsilon] = e^{.3897^2} = 1.164$. To derive the confidence interval, use the well known result that $\frac{\sqrt{N}(\overline{\epsilon}-\mu)}{S} \sim$ t-distribution with $N-1$ degrees of freedom where S is the sample standard deviation and $\overline{\epsilon}$ and μ are the sample and population means, respectively.

Based on these assumptions, $E[TotalIndex] = E[income] + E[R_A\epsilon]$. Also, by assumption, one can either use the mean return in our appreciation component or the mean return from the appraisal only index for determining $E[R_A]$. This expression is evaluated at the mean, $R_A = 7.95\%$ for the appreciation component and income =

7.85%. The corresponding error adjusted mean return for our overall index is 17.10% as opposed to the unadjusted index value of 15.83%. The 95% confidence interval bounds are 15.78% and 18.42%. If R_A = 8.98, the mean of the appraisal only return, the corrected mean total index return is 18.3 and its 95% confidence interval is bounded by 16.82 and 19.78. By using the same methodology to infer a confidence interval on the NPI index, the error corrected return is 15.88 as opposed to the reported 15.21 value; and its 95% confidence interval is between 14.77 and 16.99. These results imply that the adjusted returns as well as their confidence interval bounds are still relatively large.

In the more recent period from 1996–2003, the mean NPI return was 10.88% with a standard deviation of 3.2%. Using the same correction procedure as before and assuming the same error distribution as we determined with our sample of transaction prices, the corrected mean return is 11.10%. The corresponding 95% confidence interval is 9.78% to 12.42%.

6.2. Measuring variability

As in the last section, several simplifying assumptions are required to estimate the variability of our aggregate rate of return index as well as the NPI index. Since the overall index is the sum of the income and the appreciation components, a simplifying and reasonable approximation for the variability of the aggregate index is that the standard deviation is also additive:

$$std.dev.[TotalIndex] = std.dev.\ [Income] + std.dev.\ [Appreciation]$$

Using standard deviation values form Table 2 for our index, the sum of the income and appreciation component standard deviations is 7.7 as opposed to the standard deviation for the total index which is 7.49; using the NPI index values, the sum is 4.04 as opposed to the actual value of 4.12. Both of these results suggest that the approximation is reasonable and provides an adequate representation of this relationship.

By examining our index, the relationship between lognormal random variables indicates that $\sigma_\epsilon^2 = e^{2(0.3897)^2}\left(e^{2(0.3897)^2} - 1\right) = 0.4809, \mu_\epsilon^2 = \left(e^{(0.3897)^2}\right)^2 = 1.355$ and $\mu_A^2 = (0.0898)^2 = 0.0081$ if one uses the mean return from the appraisal only index. Using these values and noting that σ_A^2, the variance of the appraisal based only index, is $(0.0655)^2 = 0.0043$, using (15) we determine that $\sigma_P = 0.1086$. Adding this value to the standard deviation for the income component yields an adjusted standard deviation of 11.55 as opposed to the reported 7.49. If we were to use the mean and the standard deviation of the appreciation component as opposed to the appraisal only series, we would obtain the similar result of 11.64. Using the same methodology for the NPI index, the appraisal adjusted standard deviation is 7.22 as opposed to the 4.12 reported value based on appraisals. These results suggest that for the period 1979–1984, the true variability may be much larger than the variability measures based on indexes constructed from appraisals.

Using the same methodology but applying it to the NPI index values for the period from 1996–2003, the corrected standard deviation is 11.2%, which is substantially higher than the raw standard deviation of 3.20%. That is, over the more recent period, the volatility is likely to have been substantially underestimated.

6.3. Comparison with stocks

The results of the previous calibration for the NPI and our indexes are reported in Table 4. By inspection, the corrected mean returns from our real estate index are comparable with the mean returns for the CRSP and the Ibbotson S&P stocks over the 1979–1984 period. As expected the mean small cap stock returns are significantly larger than all reported stock returns. The 95% confidence bounds for both our sample as well as for the NPI index mean returns encompass both the CRSP and S&P mean stock returns. Using a one tailed t-test, at the 95% confidence level, the true mean returns using the appreciation, the appraisal and the NPI component are greater than 16.01, 17.07, and 14.96, respectively. Similarly, the true mean return is less than 18.19, 19.53, and 16.81, respectively, at the 95% confidence level. These results imply that, adjusting for the appraisal effect, commercial real estate returns are comparable to those of stocks. With the one-sided test, one cannot reject the null hypothesis that appraisal based real estate rates of return are not distinguishable statistically from the CRSP and S&P returns. However, one can reject the null hypothesis that real estate returns are comparable to those of small cap stocks.

The corrected volatilities of our index as well as the NPI index are much larger than those calculated values based on appraisals. The corrected standard deviations for our index using our sample are on average 61% larger than those commonly calculated using appraisals. Despite the underestimation of the variability, the corrected standard deviations for the various real estate indexes are smaller (but similar) for those calculated for the CRSP and S&P stock return indexes. However, real estate return variability appears to be much smaller than the variability of small cap stock returns. These results taken together imply that real estate performance is very similar to the risk-return characteristics of stocks over our sample period.

Table 4. Return and variability comparisons (1979–1984).

	Mean return	*Corrected mean return*	*Lower bound*	*Upper bound*	*Standard deviation*	*Corrected standard deviation*
Our sample (appreciation component)	15.83	17.10	15.78	18.42	7.49	11.64
Our sample (appraisal index)	15.83	18.30	16.82	19.78	7.49	11.55
NPI index	15.21	15.88	14.77	16.99	4.12	7.22
CRSP return[a]	16.65				13.34	
S&P return[b]	16.02				13.26	
Small cap stock return[b]	26.37				19.50	

Notes:
[a] The CRSP returns are the annual returns with dividend on a value-weighted market portfolio.
[b] These returns are similarly calculated from Ibbotson and Associates.

Table 5. Return and variability comparisons (1996–2003).

	Mean return	*Corrected mean return*	*Lower bound*	*Upper bound*	*Standard deviation*	*Corrected standard deviation*
NPI index	10.88	11.71	10.60	12.82	3.20	5.72
S&P 500 return	11.51				22.06	
S&P Small Cap 600	12.56				16.50	
Lehman gov't. bond index	6.64				5.00	
T-bills	4.14				1.79	

In the more recent period, 1996–2003, we compare in Table 5 the corrected NPI index values with the performance of the S&P 500, the S&P Small Cap 600 stock index returns, the Lehman Government bond index and T-Bills index. The raw unadjusted NPI return of 10.88% is comparable to the S&P 500 return of 11.51%, but lower than the small cap return of 12.56%. Although the corrected NPI return increased to 11.10%, it does not outperform the small cap index. However, the upper bound of the 95% confidence interval attained a value of 12.42% which is close to that of small cap stocks. As expected, both the corrected and uncorrected NPI index perform better than both bond indices. The error corrected standard deviation increases dramatically to 11.2% from the unadjusted value of 3.20%. Although the adjusted increase indicates that real estate was riskier than the unadjusted measure, the corrected value is still much lower than that of S&P 500 stocks (22.06%) and the small cap stocks (16.5%) but still higher than bonds. Taken together, using the corrected values, for 1996–2003, our estimates suggest that commercial real estate has performed better than large stocks, and perhaps better than small cap stocks on a risk-adjusted basis.

7. Conclusion

This paper examines how individual parcel appraisal smoothing affects the accuracy of aggregate real estate performance indexes. Prior studies allege that "smooth" individual parcel appraisals engender "smoothed" aggregate indexes. This conclusion does not follow necessarily since individual errors may offset in the aggregation process; and the impact of appraisal smoothing is ultimately an empirical issue.

Combining contemporaneous sales transactions and appraisals for a sample of large commercial real estate properties, we examine the statistical reliability of the aggregate appraisal generated real estate return index. Two key assumptions are required to conduct the analysis. First, appraisal errors over our sample time period are random draws from a common distribution. Second, the appreciation component of the aggregate index can be adequately proxied by appraisals.

Our analysis, while subject to certain limitations,[15] provides new insights about measuring real estate performance using appraisal indexes. Three principal results emerge from the study:

a. The appraisal based rate of return index is biased downwards, understating for our sample the true mean return for real estate by approximately 9%.

b. The rate of return for real estate is comparable to those achieved by CRSP and S&P indexes, but less than those for small cap stocks.
c. We, as others before us, detect smoothing in the aggregate appraisal based index. The variance of appraisal based return indexes is substantially understated; the variances are undervalued by 55% for our sample overall real estate return index, and 75% for the NPI overall return index. In the out-of-sample period, the volatility would appear to be underestimated substantially as well.

Appendix

We wish to derive $Var(R_t)$ where $R_r = (x^*_{1t} + x^*_{2t})/2$. In the following derivation, we suppress the asterisks for ease of presentation. Since $Var(R_t) = \frac{1}{4}Var(x^*_1) + \frac{1}{4}Var(x^*_2) + Cov(x^*_1, x^*_2)$, we require expressions for the variance and the covariance terms.

Since the appraisals have an AR(1) representation, each appraisal can be rewritten as

$$x_t = (1-K)x^*_{t-1} + K\eta_t = K\sum_{j=0}^{\infty}(1-K)^j\eta_{t-j} \tag{18}$$

Thus $Var(x_t) = E(x_t^2) = K^2\sigma^2\sum_{j=1}^{\infty}(1-K)^{2j} = K^2\sigma^2\sum_{j=1}^{\infty}\left[(1-K)^2\right]^j = \frac{\sigma^2K^2}{1-(1-K)^2}$ where the last step stems from the observation that $0 \le K \le 1$ and the general convergence result that $\sum_{j=0}^{\infty} r^j = \frac{1}{1-r}$. Simplifying the above expression further, we obtain $Var(x_t) = \frac{K\sigma^2}{2-K}$ which is expression 29 in QQ.

To calculate the covariance,

$$\begin{aligned}
Cov(x_{1t}, x_{2t}) &= K_1K_2E\left[\sum_{j=0}^{\infty}\left[(1-K_1)^j\eta_{t-j}\right]\left[(1-K_2)^j\eta_{t-j}\right]\right] \\
&= \sigma^2K_1K_2\sum_{j=0}^{\infty}(1-K_1)^j(1-K_2)^j \\
&= \sigma^2K_1K_2\sum_{j=0}^{\infty}[(1-K_1)(1-K_2)]^j \\
&= \frac{\sigma^2K_1K_2}{1-(1-K_1)(1-K_2)}
\end{aligned} \tag{19}$$

Putting these expressions together

$$Var(R_r) = \sigma^2\left[\frac{K_1}{4(2-K_1)} + \frac{K_2}{4(2-K_2)} + \frac{K_1K_2}{1-(1-K_1)(1-K_2)}\right] \tag{20}$$

Substituting in the expression $K_i = \sigma^2/(\sigma^2 - \sigma_i^2)$ for i = 1, 2 and simplifying we obtain (8)

$$Var(R_t) = (\sigma^2)^2\left[\frac{1}{4(\sigma^2+2\sigma_1^2)}+\frac{1}{4(\sigma^2+2\sigma_2^2)}+\frac{2}{\sigma^2+\sigma_1^2+\sigma_2^2}\right]$$

Acknowledgments

The authors are grateful for the comments of our discussant Charles Ward and other seminar participants at the 2004 Cambridge-Maastricht Real Estate Conference.

Notes

1. Such studies include, amoung others, Fama and Schwert (1977), Miles and McCue (1984), Ibbotson and Siegel (1984), Brueggeman et al. (1984) and Hartzell et al. (1986).
2. See, for example, Dokko et al. (1991).
3. See for example Giliberto (1988), Geltner (1989a, b, 1991) and Ross and Zisler (1991).
4. The sample of transactions prices parcels may not be representative of the universe of all parcels, sold and unsold; our analysis ignores this potential complication.
5. The notion of aggregate smoothing is distinct from intertemporal smoothing raised by Geltner (1993), who considers the possibility of smoothing as a consequence of averaging appraisals done over different time periods. This analysis focuses on the aggregation of contemporaneous appraisals.
6. A similar analysis can be applied to the standard deviation for returns generated from the appraisal values vis-a-vis the true market values.
7. Since we are expressing returns as differences, our analysis of determining the variance of average returns is equivalent to analyzing the time series returns of an index constructed from averages. The NPI index is constructed using this latter approach.
8. The possibility that the aggregate appraisal index exhibiting higher variance that the true index has also been suggested by Lai and Wang (1998).
9. The database for this study encompasses the one reported in Dokko et al. (1991) and the interested reader is referred to that article for additional data information.
10. The FRC return includes partial sales which are not available in this data sample. For this reason, this category is deleted in our index.
11. The coefficients of these regressions are biased estimates because appraisals are measured with error causing an errors in variable bias. The purpose of this exercise is to show that the bulk of the volatility of the index can be approximated by the volatility of the returns based on appraisals.
12. We acknowledge that even though most smoothing models assume autocorrelated appraisal errors; we do not incorporate such intertemporal behaviour in our calibration exercise. Empirical attempts to capture this autocorrelation will likely require us to observe repeated errors, a condition unlikely to occur in commercial real estate markets.
13. This may be particularly relevent in light of our sample which spans from 1979 to 1984, a period when property values appear to be increasing, but not uniformly over time.
14. Similar biases have been reported in studies which measure stock returns when they are constructed using the last daily traded prices (Blume and Stambaugh, 1985). Since the traded price can reflect either the bid or the ask price, the resulting return index will similarly overstate the true returns.
15. These results should be interpreted with caution. The findings for our sample period, 1979–1984, may not generalize to other time periods, and sample data limitations preclude testing for time varying moments in the error distribution. In addition, the sample database may not be indicative of other commercial real estate samples (e.g., the NPI sample) or the universe of commercial real estate.

References

Barkham, R., and D. Geltner. (1994). “Unsmoothing British Valuation-Based Returns Without Assuming an Efficient Market,” *Journal of Property Resarch* 11, 81–95.

Blume, M., and R. Stambaugh. (1985). "Biases in Computed Returns: An Application to the Size Effect," *Journal of Financial Economics* 12(3), 387–404.

Brown, G., and G. Matysiak. (1998). "Valuation Smoothing Without Temporal Aggregation," *Journal of Property Research* 15(2), 80–103.

Brueggeman, W. B., et al. (1984). "Real Estate Investment Funds: Performance and Portfolio Considerations," *AREUEA* 12, 333–354.

Cho, H., Y. Kawaguchi, and J. Shilling. (2003). "Unsmoothing Commercial Property Returns: A Revision to Fisher-Geltner-Webb's Unsmoothing Methodology," *Journal of Real Estate Finance and Economics* 27(3), 393–405.

Clayton, J., D. Geltner, and S. Hamilton. (2001). "Smoothing in Commercial Property Valuations: Evidence from Individual Appraisals," *Real Estate Economics* 29(3), 337–360.

Cole, R., D. Guilkey, and M. Miles. (1986). "Toward An Assessment of the Reliability of Commercial Appraisals," *The Appraisal Journal* (July).

Diaz, J. Jr., and M. Wolverton. (1998). "A Longitudinal Examination of the Appraisal Smoothing Hypothesis," *Real Estate Economics* 26(2), 349–358.

Dokko, Y., R. Edelstein, M. Pomer, and E. Urdang. (1991). "Determinants of the Rate of Return for Nonresidential Real Estate: Inflation Expectations and Market Adjustments Lags," *AREUEA* 19(1), 52–69.

Dotzour, M. (1988). "Quantifying Estimation Bias in Residential Appraisal," *Journal of Real Estate Research* 3(3), 1–11.

Fama, E., and G. Schwert. (1977). "Asset Returns and Inflation," *Journal of Financial Economics* 5, 115–146.

Fisher, J., D. Geltner, and R. Webb. (1994). "Value Indices of Commercial Real Estate: A Comparison of Index Construction Methods," *Journal of Real Estate Finance and Economics* 9, 137–164.

Geltner, D. (1989a). "Estimating Real Estate's Systematic Risk from Aggregate Level Appraisal-Based Returns," *AREUEA* 17(4), 463–481

Geltner, D. (1989b). "Bias in Appraisal-Based Returns," *AREUEA* 17(3), 338–352.

Geltner, D. (1991). "Smoothing in Appraisal-Based Returns," *Journal of Real Estate Finance and Economics* 4(3), 327–345.

Geltner, D. (1993). "Temporal Aggregation in Real Estate Return Indices," *AREUEA* 21(2), 141–166.

Geltner, D., and W. Goetzmann. (2000). "Two Decades of Commercial Property Returns: A Repeated-Measure Regression-Based Version of the NCREIF Index," *Journal of Real Estate Finance and Economics* 21(1), 5–21.

Giliberto, S. (1988). "A Note on the Use of Appraisal Data in Indexes of Performance Measurement," *AREUEA* 16(1), 77–83.

Giaccotto, C., and J. Clapp. (1992). "Appraisal-Based Real Estate Returns Under Alternative Market Regimes," *AREUEA Journal* 20(1), 1–24.

Graff, R., and M. Young. (1999). "The Magnitude of Random Appraisal Error in Commercial Real Estate Valuation," *Journal of Real Estate Research* 17(1/2), 33–54.

Hansz, J., and J. Diaz II. (2001). "Valuation Bias in Commercial Appraisal: A Transaction Price Feedback Experiment," *Real Estate Economics* 29(4), 553–656.

Hartzell, D., et al. (1986). "Diversification Categories in Investment Real Estate," *AREUEA* 14, 230–254.

Ibbotson, R., and L. B. Siegal. (1984). "Real Estate Returns: A Comparison with Other Investments," *AREUEA* 12, 219–242.

Lai, T., and K. Wang. (1998). "Appraisal Smoothing: The Other Side of the Story," *Real Estate Economics* 26(3), 511–535.

McAllister, P., A. Baum, N. Crosby, P. Gallimore, and A. Gray. (2003). "Appraiser Behaviour and Appraisal Smoothing: Some Qualitative and Quantitative Evidence," *Journal of Property Research* 20(3), 261–280.

Miles, M., and T. McCue. (1984). "Commercial Real Estate Returns," *AREUEA* 12, 355–377.

Miles, M., D. Guilkey, B. Webb, and K. Hunter. (1991). "An Empirical Evaluation of the Reliability of Commercial Appraisals, 1978–1990," Working Paper.

Quan, D., and J. Quigley. (1989). "Inferring an Investment Return Series for Real Estate from Observations on Sales," *AREUEA* 17(2), 218–230.

Quan, D., and J. Quigley. (1991). "Price Formation and the Appraisal Function in Real Estate," *Journal of Real Estate Finance and Economics* 4(2), 127–146.

Ross, S., and R. Zisler. (1991). “Risk and Return in Real Estate,” *Journal of Real Estate Finance and Economics* 4(2), 175–190.

Webb, B. (1994). “On the Reliability of Commercial Appraisals: An Analysis of Properties Sold from the Russell-NCREIF Index (1978–1992),” *Real Estate Finance* 11(1), 62–65.

Webb, J., M. Miles, and D. Guilkey. (1992). “Transactions-Driven Commercial Real Estate Returns: The Panacea to Asset Allocation Models?” *AREUEA* 20(2), 325–357.