Benchmarking Local Private Commercial Real Estate Returns:
Statistics Meets Economics

Executive Summary

Liang Peng

September 2016

It is well known that real estate is traded infrequently; as a result, observations of property values and returns are very scarce. For private commercial real estate, the problem of data scarcity is even more severe, which makes it very challenging to benchmark local returns of private commercial real estate.

This paper proposes a parameter-reduction approach to overcome extreme data scarcity and to estimate metro level total return indices of private commercial real estate. This approach uses a small number of parameters to characterize return heterogeneity across metro areas, so one needs to estimate much fewer parameters for the estimation of local indices.

Using property level NCREIF data, I estimate national total return indices for each of the four main property types, as well as two parameters that characterize how each metro-level index differs from a national index: the average excess return per quarter and the sensitivity of the metro index to the national index. The national indices and the two metro-specific parameters allow me to construct apartment total return indices for 34 metro areas, industrial total return indices for 41 metro areas, office total return indices for 31 metro areas, and retail total return indices for 21 metro areas.

I evaluate the economic merits of the metro indices using an in-sample test, a placebo test, and an out-of-sample test. I find that (1) metro-level indices provide incremental explanatory power for local property returns compared to the national index, (2) indices from random wrong metro areas do not provide any explanatory power, and (3) metro indices have significant out-of-sample explanatory power.

This paper makes two main contributions. First, it is the first to propose the parameter-reduction approach to overcome extreme data scarcity, and the first to apply this approach to estimate metro-level total return indices of private commercial real estate. Second, this paper is the first to propose and use the three tests to assess the economic merits of local indices. Both the parameter-reduction approach and the tests can be easily applied to other non-traded assets.

**Benchmarking Local Private Commercial Real Estate Returns:
Statistics Meets Economics**

Liang Peng*

Smeal College of Business
The Pennsylvania State University
University Park, PA 16802
Phone: (814) 863 1046
Fax: (814) 865 6284
Email: liang.peng@psu.edu

September 2016

Abstract

This paper uses a parameter-reduction approach to overcome the challenge of extreme data scarcity and estimates total return indices at the metro level for apartment, industrial, office, and retail properties in the U.S. This approach uses a small number of parameters to characterize return dynamics of local properties. Estimates of such parameters and national indices allow the construction of metro indices using very small samples of local property returns. Using three novel tests, an in-sample incremental explanatory power test, a placebo test, and an out-of-sample incremental explanatory power test, I find that metro indices estimated using this approach have significant economic merits, in the sense that they successfully capture unique local return dynamics and help explain local property returns both in- and out-of sample.

JEL classification: C13, C43, C58, G12

Key words: commercial real estate, total return, index, and parameter-reduction

* I thank Real Estate Research Institute for a research grant and thank the National Council of Real Estate Investment Fiduciaries for providing the data. This paper greatly benefits from constructive comments from William Maher, Andrew McCulloch, Timothy Bellman, and participants of the 2017 RERI conference. All errors are mine.

I. Introduction

Benchmarks of asset investment returns are crucial for economists, investors, and policy makers. Well-known benchmarks such as the S&P indices for stocks, the Barclays indices for bonds, and the S&P Case/Shiller indices for single-family homes are widely used by researchers and practitioners. The private commercial real estate is a trillion-dollar asset class (see, e.g. Plazzi, Torous and Valkanov (2008) and Peng (2016)). While there are indices that track their investment performance at the national level (see, e.g. Geltner and Goetzmann (2000)),[1] the commercial real estate market is likely segmented due to differences in both investment risk and income growth across metro areas as well as property types. Therefore, it is important to construct benchmark indices for local property returns.

The construction of local private commercial real estate return indices faces a significant statistical challenge: *extreme* data scarcity. It is well known that real estate is traded infrequently; as a result, observations of property values and returns are very scarce. To overcome this problem, economists have developed a large literature on index-estimation methods for non-traded assets. Such methods, particularly the repeat sales regression (Bailey, Muth and Nourse (1963)) as well as its many modifications and improvements (see, e.g. Case and Shiller (1989), Goetzmann (1992), Gatzlaff and Haurin (1997), Clapp (2004), Fisher, Gartzlaff, Geltner and Haurin (2003), Goetzmann and Peng (2006), Peng (2012), among many others), have been mostly applied to the housing market. The S&P Case/Shiller city-level house price indices are successful examples of such applications.

For commercial real estate, however, the problem of data scarcity seems much more severe, to the extent that the repeat sales regression alone is unable to overcome it. For instance, the sample used in this paper has an average of about 15 apartment properties, 23 industrial properties, 19 office properties, and 6 retail properties per metro area. The average number of properties is even smaller than the number of quarters the indices need to cover!

---

[1] The National Council of Real Estate Investment Fiduciaries (NCREIF) provides quarterly income return, capital appreciation, and total return indices for commercial real estate, which are constructed from appraised values.

This paper uses a parameter-reduction approach to mitigate the problem of extreme data scarcity. This approach works under two assumptions. First, returns of properties in different metro areas are correlated; therefore, property returns from one metro area may contain useful information to help construct return indices for other metro areas. Second, differences in return dynamics across metros are not completely random or arbitrary. As a result, one can use a small number of parameters to characterize at least some return heterogeneity across metro areas. These two assumptions help reduce the number of parameters needed to construct the index for each metro area from the number of time periods into the few characterizing parameters.

To apply this approach, in this paper I jointly estimate a national total return index of private commercial real estate as well as two parameters that characterize how each metro-level index differs from a national index: the average excess return per quarter of the metro index compared to the national index and the sensitivity of the metro index to the national index. The estimated national index and the two metro-specific parameters allow me to construct the total return index for metro areas. I use a sample of 1,893 apartment properties, 2,597 industrial properties, 1,861 office properties, and 1,034 retail properties from the NCREIF database to construct 34 apartment metro indices, 41 industrial metro indices, 31 office metro indices, and 21 retail metro indices from 1997 to 2014. Note that it is certainly possible to characterize metro indices in more flexible and more complicated ways, but this paper focuses on the simple two-parameter specification to illustrate the effectiveness of the parameter-reduction approach.

I assess the effectiveness of the parameter-reduction approach by evaluating the economic merits of the metro indices. Specifically, do the two parameters sufficiently capture local return dynamics so that the metro indices better explain local property returns than the national index? Are differences in the return indices across metro areas reflecting information on local returns or simply noise? This paper uses three tests to help shed light on these questions and to assess the merits of the metro indices.

The first test is an in-sample test of the incremental explanatory power of metro indices. Specifically, I analyze whether the metro index excess returns (metro index returns net of the national index returns) help explain local properties' excess returns (property returns net of the national index). Significant explanatory power would be consistent with the notion that metro indices capture unique local return dynamics.

To confirm that such explanatory power is not driven by unknown mechanical relationships, I conduct a placebo test as the second test. Specifically, instead of using index excess returns of the metro where each property is located, I use excess returns of a random wrong metro area to redo the first test. If the explanatory power of metro indices is not driven by unknown mechanical relationship, such a placebo test should find no explanatory power of indices of random wrong metros.

Further note that the explanatory power of metro indices in the first test may be partly driven by outliers, as outliers would appear on the left side of the regressions and also affect metro indices, which are on the right side of regressions. I use an out-of-sample test, which is the third test, to mitigate this problem. I split the sample of property returns for each metro area into two mutually exclusive subsamples, say sample A and sample B. I test whether metro excess returns calculated from sample A help explain excess returns of properties in sample B. As samples A and B are mutually exclusive, explanatory power of metro excess returns is not likely due to outliers. The existence of significant explanatory power, therefore, would be strong evidence for the information content of metro indices.

Results from the three tests are that (1) metro-level indices provide incremental explanatory power for returns of local individual properties compared to the national index, (2) placebo indices do not provide any explanatory power, and (3) metro indices have significant out-of-sample explanatory power. Overall, I find strong evidence indicating that the metro-level indices estimated using the parameter-reduction approach have significant economic merits.

This paper makes two original contributions to the literature. First, it is the first to propose the parameter-reduction approach to overcome extreme data scarcity, and the first to use it to estimate metro-level total return indices of private commercial real estate. To my best knowledge, such indices did not exist before this paper. Second, this paper is the first to propose and use the three tests to assess the economic merits of local indices. Results from the three tests help establish that the parameter-reduction approach is capable of providing meaningful and informative local benchmarks for private commercial real estate returns. Note that both the parameter-reduction approach and the tests on economic merits of indices can be easily applied to other non-traded assets.

The rest of this paper is organized as follows. Section II describes the data. Section III discusses the econometric model based on the parameter-reduction approach and reports the national and metro indices. Section IV evaluates the economic merits of the metro indices. Section V concludes.

II. Data

I estimate metro-level total return indices using the 2014:Q4 version of the proprietary database of National Council of Real Estate Investment Fiduciaries (NCREIF). NCREIF is a not-for-profit real estate industry association, which collects, processes, and disseminates information on the operation and transactions of commercial real estate. Its database consists of institutional-grade properties owned or managed by NCREIF members, which are typically large investment companies, pension funds, and life insurance companies, in a fiduciary setting.[2] The database contains information on property attributes, such as property type and location, as well as operational and transactional information, including net operating income (NOI), acquisition cost, appraised values, and capital expenditures, for each property in each quarter during its holding period. All cash flow variables are on an unlevered basis. The 2014:Q4 version consists of 33,338 properties and its sample period is from the third quarter of 1977 to the

[2] Examples of NCREIF members are Blackrock, Citi group, TIAA, New York Life, Invesco, Heitman/JMB, and Cornerstone real estate advisers.

fourth quarter of 2014. The NCREIF property level data have been used in previous research such as Peng (2016).

I calculate the holding period modified IRR (MIRR) for each property in the database whenever possible. I calculate MIRRs instead of IRRs mainly because IRRs are often not well defined for commercial real estate investments as there are often multiple solutions to the present value functions due to the long holding periods and irregular cash flows of real estate investments. To calculate the MIRRs, I first construct the quarterly series of cash flow for each property. In the acquisition quarter of a property, the cash flow is simply the acquisition cost.[3] In each of the subsequent quarters before disposition, the quarterly cash flow is the net operating income (NOI) minus capital expenditures. If there is a partial sale in that quarter, I also add net proceeds from the partial sale to the cash flow. In the disposition quarter, the cash flow is net sale proceeds plus NOI and then minus capital expenditures in that quarter.[4] I then construct a simple total return index for each type of properties and use the index's quarterly returns as both the financing rate and the reinvestment rate to calculate the MIRRs for the same type of properties. When constructing such indices, I first use market values (or appraised values if market values are not available) at the beginning and the end of each quarter and the net cash flow (NOI plus partial sale minus capital expenditures) for each quarter to calculate the quarterly total return for each property. The index's return in that quarter simply equals the equal-weighted average of properties' returns. Finally, I use the quarterly cash flow series and the series of the financing and reinvestment rates to calculate the holding period MIRR for each property.

Properties in the database have three final disposition statuses: true sales (arms' length transaction), other sales (e.g. transfer of ownership to another member, split into multiple properties, consolidation into existing properties, returned to lender, property destroyed,

---

[3] I assume that all acquisitions and dispositions take place at the end of quarters. For a small number of properties, the database shows positive net operating income in the recorded acquisition quarters, possibly because their acquisitions took place in the middle of those quarters. For these properties, I assume the acquisitions took place at the end of the previous quarters.
[4] For a small number of properties, the net operating income in the disposition quarter is 0. I then assume that the dispositions took place at the end of the previous quarters.

etc.), and being held by investors at the end of the sample period (2014:Q4). For 13,398 properties with true sales, I am able to calculate actual MIRRs for 7,103 properties. I am unable to calculate MIRRs for other true sales because (1) there is no recorded sale time, (2) cash flow information is missing for some recorded quarters, or (3) no information is reported for some quarters. For (1) sold properties for which I am unable to calculate MIRRs, (2) properties that were disposed in other ways, and (3) properties that were still held by investors at the end, I calculate their MIRRs in the first five years since acquisition if possible, using their appraised values at the end of year 5. I call them estimated MIRRs. Table 1 counts properties in the NCREIF database according to their final disposition statuses and whether I am able to calculate the true or estimated MIRRs.

As NCREIF members did not report capital expenditures until 1997:Q1, I only use properties acquired on or after 1997:Q1. This reduces the sample to 24,055 properties. Further, I focus on the four main property types with true or estimated MIRRs as well as location information (metro areas,), and thus the sample consists of 7,867 properties, including 2,009 apartment, 2,764 industrial, 1,972 office, and 1,122 retail properties.

I then filter out outliers with extreme MIRRs. To do this, I first exclude properties with their MIRRs being in the bottom and top 2% of the distribution for each property type respectively. I then estimate a quarterly national total return index for each property type using the repeat sales regression, which economists have been using to construct house price indices (see, e.g. Case and Shiller (1987), Clapp and Giaccotto (1992), Goetzmann (1992), Gatzlaff and Haurin (1997), Fisher, Gartzlaff, Geltner and Haurin (2003), Goetzmann and Peng (2006), among many others). I will describe the details of the regression in next section. After estimating the national total return indices, I identify properties with per quarter residuals being three standard deviations away from the average residual, and exclude them from the sample. The cleaned sample consists of 1,893 apartment, 2,597 industrial, 1,861 office, and 1,034 retail properties with holding period MIRRs.

Table 2 reports summary statistics for properties with calculated MIRRs for each property type. There are 1,893 apartment properties located in 125 metro areas in 40 states, which were invested by 54 different NCREIF members; 2,597 industrial properties located in 113 metro areas in 35 states, invested by 57 different members; 1,861 office properties located in 99 metro areas in 40 states, invested by 68 different members; and 1,034 retail properties located in 169 metro areas in 44 states, invested by 50 different investors. This table also reports the quartiles of MIRRs. Figure 1 plots the number of properties with MIRRs in each quarter since 1997 for each property type.

III. Estimating metro indices

III. 1. The econometric model

The main challenge in estimating metro indices is extreme data scarcity for each metro area. I overcome this problem by reducing the number of parameters that I need to estimate. Specifically, instead of estimating index returns for each quarter, I estimate a few parameters that characterize each metro index in how it differs from a national index.

In this paper I simply characterize the differences between a metro index and the national index using two parameters: the average excess return per quarter of the metro index compared to the national index and the sensitivity of the metro index to the national index. It is certainly possible to characterize metro indices in more flexible but more complicated ways to allow more heterogeneity in metro index dynamics. However, this paper focuses on investigating and demonstrating the feasibility and effectiveness of the parameter-reduction approach in estimating metro indices, not to identify the optimal way, if it ever exists, to characterize metro indices, which should be a on-going effort and is saved for future research.[5]

Specifically, I assume that

$$r_{i,m,t} = \alpha_m + \beta_m r_t + \varepsilon_{i,t}, \tag{1}$$

[5] Nonetheless, in unreported analyses, I do allow the sensitivity to be more flexible and let it differ when the national index returns are positive and negative, but I find no statistical evidence for the asymmetric sensitivity and the estimated metro indices are similar.

where $r_{i,m,t}$ is the period t total return of property i (in log), which is located in metro area m; r_t is the total return of the national index in the same period (in log); α_m is the per-period excess return of the metro index; β_m is the sensitivity of the metro index to the national index; and $\varepsilon_{i,t}$ is an error term. Note that I cannot estimate the model in (1) directly as I do not observe the property's total return in each period. So I aggregate both the left and the right sides of (1) from the property's acquisition period buy_i to its disposition period $sell_i$ to have

$$\sum_{t=buy_i+1}^{sell_i} r_{i,m,t} = \alpha_m\left(sell_i - buy_i\right) + \beta_m \sum_{t=buy_i+1}^{sell_i} r_t + \sum_{t=buy_i+1}^{sell_i} \varepsilon_{i,t}\ . \tag{2}$$

Now note that the left side is the total return over the property's holding period, which is a function of the holding period MIRR (quarterly log gross return) and thus observed. The model that can be estimated is the following.

$$\left(sell_i - buy_i\right) \times MIRR_i = \alpha_m\left(sell_i - buy_i\right) + \beta_m \sum_{t=buy_i+1}^{sell_i} r_t + \sum_{t=buy_i+1}^{sell_i} \varepsilon_{i,t} \tag{3}$$

A few things about (3) are worth noting. First, it is a simple extension of the original repeat sales regression by Bailey, Muth and Nourse (1963). In fact, (3) reduces to the original sales regression if we let $\alpha_m = 0$ and $\beta_m = 1$. Second, the metro index return in period t, denoted by $r_{m,t}$, is naturally determined by the national index r_t, the per-period excess return α_m, and the sensitivity β_m as follows.

$$r_{m,t} = \alpha_m + \beta_m r_t \tag{4}$$

The metro index level in dollar amount in period t, denoted with $R_{m,t}$, can be easily calculated as follows.

$$R_{m,t} = \$1 \times \prod_{s=1}^{t} \exp\left(r_{m,s}\right) \tag{5}$$

Third, the model in (3) helps reduce the number of parameters we need to estimate for each metro area to two (α_m and β_m), and thus mitigates the problem of extreme data scarcity. The apparent cost of reducing the number of parameters is that (3) may not capture all heterogeneity in return dynamics across metro areas.

III. 2. Estimation and the national indices

I estimate (3) using the two-step E-M algorithm proposed by Peng (2012) and construct metro indices according to (4) and (5). Specifically, I first estimate the national total return index r_t using holding period MIRRs of properties from all metro areas, treating α_m and β_m for each metro as known (using 0 as the initial value of all α_m and 1 as the initial value of all β_m). I then treat the estimated r_t as known, estimate α_m and β_m as parameters for each metro area separately using that metro's properties. I iterate the above two steps, and each time use the updated estimated values of the national index and α_m and β_m for each metro area until the national cap rate index converges (the sum of squared differences between the current round's and the previous round's estimates is near 0).

It is worth noting that some metro areas have very few properties, which means the estimation of α_m and β_m for them is infeasible or at least inaccurate. To overcome this problem, in the second step discussed above, I estimate α_m and β_m only for metro areas with at least 15 properties. Nonetheless, I still use the entire sample of properties to estimate the national index, using 0 as the value of α_m and 1 as the value of β_m for metro areas with fewer than 15 properties.

Two econometric details are worth mentioning in estimating (3) using the E-M algorithm. First, economists have well recognized possible sample selection problems of the repeat sales regression because sold properties may not be random samples from the population of properties (see, e.g. Clapp, Giaccotto and Tirtiroglu (1991), Clapp and Giaccotto

(1992), Gatzlaff and Haurin (1997), Goetzmann and Peng (2006), among others). I calculate MIRRs for all properties, both sold and unsold, whenever possible and my filtering rules do not depend on the probability of sales. Therefore, the sample selection problem does not seem a main issue. Second, the model in (3) may have heteroskedasticity. As Case and Shiller (1989), Goetzmann (1992), and many others point out, i.i.d error $\varepsilon_{i,t}$ implies that the variance of the error term in (3) increases with the duration of the time between acquisition and disposition. Should that be the case, OLS estimators are not efficient. Case and Shiller (1989) use fitted variance from a regression of squared residuals against duration as the weight. However, empirically I find no significant relationship between the variance of regression residuals and the duration. Therefore I estimate (3) using OLS.

Figure 2 plots the national total return indices for the four property types. The indices have different starting time but the same starting values of $1. It is evident that indices are very volatile. This is partly due to relative small samples I use to estimate the indices. To apply such indices, say to use them as performance benchmarks, one might want to smooth them using techniques such as Kalman smoothing or Kalman filtering. However, the research goal of this paper is to demonstrate the effectiveness of the parameter-reduction approach in estimating metro indices. Therefore, I choose to report and work on the original indices.

Table 3 reports summary statistics of the national indices estimated with the repeat sales regressions (RSR indices). The first quarter for each index is the quarter when there are at least 15 properties. The geometric average quarterly return is 1.58% for apartment, 1.61% for industrial, 1.67% for office, and 2.17% for retail properties. These numbers are slightly lower than the geometric average quarterly returns of the NCREIF total return property-type indices (NPI, http://www.ncreif.org) during the same periods,[6] which are 2.31% for apartment, 2.35% for industrial, 2.29% for office, and 2.55% for retail properties. The arithmetic average quarter return of the RSR national index is 2.40% for

[6] The NCREIF total return indices (NPIs) are constructed using appraised values of properties as well as cash flow information in each quarter of qualified properties in the NCREIF database.

apartment, 3.82% for industrial, 3.31% for office, and 5.30% for retail properties. The arithmetic average returns are significantly higher than the geometric average returns due to temporal volatility of the indices and Jensen's inequality (see, e.g. Goetzmann (1992), Goetzmann and Peng (2002), and Peng (2002)). The autocorrelation of quarterly returns is negative and economically significant: -0.32 for apartment, -0.48 for industrial, and -0.44 for office and retail properties. However, one needs to be cautious not to over-interpret the autocorrelation of repeat sales index returns, as autocorrelation of estimated indices may be biased measures of that of true indices and the direction of the bias is ambiguous (see Case and Shiller (1989) for more details). This table also reports the quartiles of the quarterly returns for each property.

Figure 3 plots the RSR indices against NPIs during the same sample period for each of the four property types. The purpose is to see whether the RSR indices are reasonable. A few things are worth noting. First, the RSR indices and NPIs appear to have similar long-term temporal patterns during the same periods. Second, the RSR indices are more volatile than NPIs. This is partly because the samples that I use to estimate the RSR indices are much smaller than those used to calculate the NPIs so the RSR indices may contain more noise. Another reason is that indices constructed from appraised-values tend to be too smooth (see, e.g. Geltner (1989)). Overall, the RSR indices seem to be reasonable measures of the long-term performance of commercial real estate, while they are very volatile and may not capture short-term dynamics well.

III. 3. Metro indices

I estimate (3) using the E-M algorithm for each property type separately. For each type, while I estimate parameters α_m and β_m for all metro areas with at least 15 properties, I only report the results for the top 10 metro areas with the most properties. The complete results are available upon request from the author. For each of the top 10 metros, I report the number of properties in that metro, the per-quarter excess return α_m and whether it is significant (null hypothesis is that α_m is 0), and the sensitivity of the metro index to the national index β_m and whether it is significant (null hypothesis is that β_m is 1).

Table 4 reports the results for apartment properties. There are 34 metro areas with at least 15 apartment properties, which is about 27% of the 125 metro areas in the apartment sample. 1,555 out of the entire sample of 1,893 properties, or about 82%, are located in these 34 metros. The number of properties located in the top 10 metros ranges between 132 (Atlanta, GA) to 62 (Fort Lauderdale, FL). The per-quarter excess return is statistically significant for 9 out of the 10 metros, and the sensitivity is significant for 4 out of the 10 metros.

Table 5 reports the results for industrial properties. There are 41 metro areas with at least 15 industrial properties, which is about 36% of the 113 metros in the sample of industrial properties. 2,303 out of the entire sample of 2,597 properties, or about 89%, are located in these 41 metros. The number of properties located in the top 10 metros ranges between 257 (Atlanta, GA) to 70 (Baltimore, MD). The per-quarter excess return is statistically significant for 3 out of the 10 metros, and the sensitivity is significant for 6 out of the 10 metros.

Table 6 reports the results for office properties. There are 31 metro areas with at least 15 office properties, which is about 31% of the 99 metros in the office sample. 1,598 out of the entire sample of 1,861 properties, or about 86%, are located in these 31 metros. The number of properties located in the top 10 metros ranges between 191 (Washington-DC) to 60 (Seattle, WA and Santa Ana, CA). The per-quarter excess return is statistically significant for 6 out of the 10 metros, and the sensitivity is significant for 2 out of the 10 metros.

Table 7 reports the results for retail properties. There are 21 metro areas with at least 15 office properties, which is about 12% of the 169 metros in the office sample. 592 out of the entire sample of 1,034 properties, or about 57%, are located in these 21 metros. The number of properties located in the top 10 metros ranges between 66 (Atlanta, GA) to 25 (Orlando, FL). The per-quarter excess return is statistically significant for 2 out of the 10 metros, and the sensitivity is significant for 4 out of the 10 metros.

After estimating the metro parameters, I use the estimated parameters and the national index to estimate total return indices according to (4) and (5) for metros with at least 15 properties. Figure 4 plots all the metro-level indices for each property type for the metros with at least 15 properties. The starting time for each metro is the earliest acquisition quarter of all properties located in that metro, and the ending time is the latest disposition quarter. The starting values of all metro indices are $1.

IV. Economic merits of metro indices

Do the metro indices capture sufficient local return dynamics to the extent that they better explain local property returns than the national indices? Are differences in the return indices across metro areas reflecting information on local returns or simply noise? These are fundamental economic problems this section tries to shed light on.

I use three tests to assess economic merits of the metro indices. First, I analyze whether the metro indices provide additional information that helps explain local property returns compared to the national index. Specifically, using properties located in metro areas with at least 15 properties, I run cross-sectional regressions of the holding period gross return of each property net of the national index against the return of its metro area index over the same period net of the national index as follows,

$$\left(sell_i - buy_i\right) \times MIRR_i - \sum_{t=buy_i+1}^{sell_i} r_t = \alpha + \rho\left(\sum_{t=buy_i+1}^{sell_i} r_{m,t} - \sum_{t=buy_i+1}^{sell_i} r_t\right) + \varepsilon_{i,t}\,. \tag{6}$$

The null hypothesis is that $\rho = 0$, which means metro index returns provide no additional explanatory power for local property's holding period returns compared to the national index.

I estimate (6) for each of the four property types respectively, and report the results in Table 8. The coefficient of the metro index net of the national index is 1.00 for apartment, industrial, and retail, and 1.01 for office properties. All the coefficients are statistically significant at the 1% level. This provides strong evidence that metro-level indices do provide incremental explanatory power for local properties' returns compared to the national index.

Second, to rule out the possibility that the explanatory power of metro indices found in Table 8 is driven by unknown mechanical relationships caused by my parameter-reduction approach, I conduct a placebo test that is identical to regressions in (6) but uses metro indices of random wrong metro areas. Specifically, for each property in the same sample used in Table 8, I exclude the metro area where this property is located, and then randomly select one of other metros and use its index return net the national index return as the explanatory variable on the right side of (6). For each property type, I repeat the placebo test for 1,000 times. I report the mean, the standard deviation, and the quartiles of the estimated ρ for each property type in Table 9. I also plot the histogram of ρ in Figure 5. It is apparent that ρ is not significantly different from 0. This helps rule out the possibility that the explanatory power of metro indices for local property returns is due to unknown mechanical reasons.

The third test is an out-of-sample test, which aims to analyze whether the explanatory power found in Table 8 is indeed driven by information. Note that the explanatory power could be driven by noise. Imagine that a metro area has a few outliers - properties with high MIRRs for idiosyncratic reasons. Such high MIRRs appear on the left side of the (6), and also affect the metro index returns on the right side of (6). This may lead to a significantly positive estimate of ρ even if metro indices do not capture any genuine local return dynamics.

To overcome this problem, I conduct out-of-sample tests as follows. I first randomly split the properties of each metro (with more than 15 properties) into two samples, say A and B. I use properties in sample A and the national index to estimate the two metro-specific parameters α_m and β_m according to (3), and then use the parameters to estimate metro index returns according to (4). After that, I use (6) to test whether the metro-area index estimated from sample A helps explain holding period returns of individual properties in sample B. Since the two samples are mutually exclusive, it is unlikely for them to generate artificially positive ρ. Therefore, a significant and positive ρ would suggest that the metro indices contain information.

I repeat the out-of-sample regression for 1,000 times for each property type, and randomly split properties into two mutually exclusive samples each time. Table 10 reports the mean, the standard deviation, and the quartiles of the 1,000 estimates of ρ for each property type. It is apparent that ρ is significantly positive, which is very strong evidence indicating that metro indices do contain information. On the other hand, note that estimates of ρ are always less than 1. This is partly due to the fact that estimated metro indices contain measurement errors; consequently, estimated ρ is attenuated. I further plot the histogram of ρ in Figure 6 for each property type. It is apparent that ρ is statistically significant and positive, but less than 1. In short, both Table 10 and Figure 6 provide strong evidence that metro indices capture genuine return heterogeneity across metro areas.

V. Conclusion

Benchmarking returns is one of the most important and most challenging research topics in real estate economics. The main challenge is that real estate values and investment returns are observed infrequently. The problem of data scarcity is even more severe for private commercial real estate than for single-family homes, and for estimating local indices than for estimating national indices, to the extent that typical methods, such as the repeat sales regression, cannot be applied directly to construct local indices.

This paper uses a parameter-reduction approach to estimate private commercial real estate total return indices at the metro level. This approach uses a smaller number of parameters to characterize the differences between metro indices and the national index. Consequently, estimates of such parameters and the national index make it possible to construct metro indices. Using a sample of holding period total returns of individual properties from the NCREIF dataset, I construct 34 apartment metro indices, 41 industrial metro indices, 31 office metro indices, and 21 retail metro indices.

The success of the parameter-reduction approach depends on the economic merits of the metro indices it produces. I conduct three tests, a in-sample test, a placebo test, and an out-of-sample test, and find that metro indices provide incremental explanatory power for returns of local individual properties compared to the national index; indices from random wrong metros do not provide any explanatory power; and metro indices have significant out-of-sample explanatory power for local property returns. These results help establish the effectiveness of the parameter-reduction approach in mitigating extreme data scarcity for the construction of indices of non-traded assets. Both the parameter-reduction approach and the three tests on the economic merits of metro indices can be easily applied to other non-traded assets.

References

Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse, 1963, A regression method for real estate price index construction, *Journal of the American Statistical Association* 58, 933-942.

Case, Karl E., and Robert Shiller, 1989, The efficiency of the market for single family homes, *American Economic Review* 79, 125-137.

Case, Karl, and Robert Shiller, 1987, Prices of single family homes since 1970: New indexes for four cities, *New England Economic Review* 46-56.

Clapp, John, 2004, A semiparametric method for estimating local house price indices, *Real Estate Economics* 32, 127-160.

Clapp, John, and Carmelo Giaccotto, 1992, Repeated sales methodology for price trend estimation: An evaluation of sample selectivity, *Journal of Real Estate Finance and Economics* 5, 357–374.

Clapp, John M., Carmelo Giaccotto, and Dogan Tirtiroglu, 1991, Housing price indices based on all transactions compared to repeat subsamples, *Real Estate Economics* 19, 270-285.

Fisher, Jeffrey, Dean Gartzlaff, David Geltner, and Donald Haurin, 2003, Controlling for the impact of variable liquidity in commercial real estate price indices, *Real Estate Economics* 31, 269-303.

Gatzlaff, Dean H, and Donald R. Haurin, 1997, Sample selection bias and repeat-sales index estimates, *Journal of Real Estate Finance and Economics* 14, 33-50.

Geltner, David, 1989, Estimating real estate's systematic risk from aggregate level appraisal-based returns, *AREUEA Journal* 17, 463-481.

Geltner, David, and William Goetzmann, 2000, Two decades of commercial property returns: A ncreif index using independent appraisals, *Journal of Real Estate Finance and Economics* 21, 5-22.

Goetzmann, William N., 1992, The accuracy of real estate indices: Repeat sale estimators, *Journal of Real Estate Finance and Economics* 5, 5-53.

Goetzmann, William N., and Liang Peng, 2002, The bias of the rsr estimator and the accuracy of some alternatives, *Real Estate Economics* 30, 13-39.

Goetzmann, William N., and Liang Peng, 2006, Estimating house price indices in the presence of seller reservation prices, *Review of Economics and Statistics* 88, 100-112.

Peng, Liang, 2002, Gmm repeat sales indices, *Real Estate Economics* 30, 239-261.

Peng, Liang, 2012, Repeat sales regression on heterogeneous properties, *Journal of Real Estate Finance and Economics* 45, 804-827.

Peng, Liang, 2016, The risk and return of commercial real estate: A property level analysis, *Real Estate Economics* 44, 555-583.

Plazzi, Alberto, Walter Torous, and Rossen Valkanov, 2008, The cross-sectional dispersion of commercial real estate returns and rent growth: Time variation and economic flucuations, *Real Estate Economics* 36, 403-439.

Table 1. NCREIF database summary

This table summarizes the numbers of properties that were sold (true sales or other types of sales) and were not sold, and with actual, estimated, and no holding period total return MIRRs.

	True sales	Other sales	Not sold	Total
True MIRRs	7,103	0	0	7,103
Estimated MIRRs	10	973	2,817	3,800
No MIRR	6,285	7,496	8,654	22,435
Total	13,398	8,469	11,471	33,338

Table 2. Summary of properties

This table summarizes properties in the final sample that I use to construct metro total return indices.

	Apartment	Industrial	Office	Retail
All properties	1,893	2,597	1,861	1,034
True MIRR	1,233	1,421	1,109	528
Estimated MIRR	660	1,176	752	506
States	40	35	40	44
Metro areas	125	113	99	169
Investors	54	57	68	50
Annual MIRR: minimum	-26.21%	-34.08%	-28.57%	-29.04%
Annual MIRR: 25%	-6.55%	-5.44%	-8.04%	-6.19%
Annual MIRR: medium	2.77%	3.61%	2.50%	3.22%
Annual MIRR: 75%	9.51%	9.65%	9.82%	11.52%
Annual MIRR: maximum	88.46%	68.49%	89.11%	87.64%

Table 3. National total return indices summary

This table report summary statistics of the national total return indices for apartment, industrial, office, and retail properties.

	Apartment	Industrial	Office	Retail
Starting quarter	1997:Q2	1997:Q1	1997:Q1	1998:Q2
Geometric average quarterly return	1.58%	1.61%	1.67%	2.17%
NPI geometric average	*2.31%*	*2.35%*	*2.29%*	*2.55%*
Arithmetic average quarterly return	2.40%	3.82%	3.31%	5.30%
Standard deviation quarterly returns	13.26%	21.96%	19.03%	27.63%
Quarterly return autocorrelation	-0.32	-0.48	-0.44	-0.44
Minimal quarterly return	-24.10%	-45.49%	-38.09%	-44.63%
25% percentile quarterly return	-6.96%	-9.85%	-10.32%	-12.13%
Median quarterly return	0.19%	1.67%	0.42%	0.71%
75% percentile quarterly return	9.96%	14.43%	16.83%	20.22%
Maximum quarterly return	34.63%	72.24%	65.28%	139.99%

Table 4. Metro index parameter summary: Apartment

This table summarizes the number of apartment properties, estimated per-quarter excess return (null hypothesis of t-tests is that their values are 0), and estimated sensitivity of each metro index to the national index (null hypothesis of t-tests is that their values are 1) for the top 10 metro areas with most properties out of the 34 metro areas (1,555 properties) for which I estimate metro-level indices. ***, **, and * indicate significant levels of 1%, 5%, and 10% respectively.

Metro-area	Properties	Excess return	Sensitivity
GA-Atlanta	132	-0.002	1.38*
TX-Dallas	105	0.007***	0.57***
NY-New York	99	0.005*	1.09
TX-Houston	95	0.003*	0.88
CA-Los Angeles	89	-0.005**	1.26*
TX-Austin	74	0.007***	0.58
AZ-Phoenix	70	-0.012***	1.32**
DC-Washington	68	-0.006**	1.07
WA-Seattle	65	-0.007**	1.10
FL-Fort Lauderdale	62	0.006*	0.91

Table 5. Metro index parameter summary: Industrial

This table summarizes the number of industrial properties, estimated per-quarter excess return (null hypothesis of t-tests is that their values are 0), and estimated sensitivity of each metro index to the national index (null hypothesis of t-tests is that their values are 1) for the top 10 metro areas with most properties out of the 41 metro areas (2,303 properties) for which I estimate metro-level indices. ***, **, and * indicate significant levels of 1%, 5%, and 10% respectively.

Metro-area	Properties	Excess return	Sensitivity
GA-Atlanta	257	0.006***	0.74***
IL-Chicago	192	-0.005***	1.15*
CA-Los Angeles	180	-0.002	1.08
TX-Dallas	158	0.003	0.67***
CA-Riverside	124	-0.000	1.19*
CA-Santa Ana	84	0.002	0.92
CA-Oakland	82	-0.006***	1.00
AZ-Phoenix	76	0.001	1.08
WA-Seattle	71	0.000	0.75*
MD-Baltimore	70	0.000	1.20*

Table 6. Metro index parameter summary: Office

This table summarizes the number of office properties, estimated per-quarter excess return (null hypothesis of t-tests is that their values are 0), and estimated sensitivity of each metro index to the national index (null hypothesis of t-tests is that their values are 1) for the top 10 metro areas with most properties out of the 31 metro areas (1,589 properties) for which I estimate metro-level indices. ***, **, and * indicate significant levels of 1%, 5%, and 10% respectively.

Metro-area	Properties	Excess return	Sensitivity
DC-Washington	191	0.003**	1.15
TX-Dallas	99	-0.003**	0.75**
GA-Atlanta	88	-0.005**	0.80
IL-Chicago	85	-0.000	0.92
CA-Los Angeles	81	-0.007**	1.04
CA-San Diego	80	-0.002	1.17
CA-San Francisco	74	0.005**	0.77
NY-New York	62	0.008***	1.46**
WA-Seattle	60	0.003	1.09
CA-Santa Ana	60	-0.001	1.12

Table 7. Metro index parameter summary: Retail

This table summarizes the number of retail properties, estimated per-quarter excess return (null hypothesis of t-tests is that their values are 0), and estimated sensitivity of each metro index to the national index (null hypothesis of t-tests is that their values are 1) for the top 10 metro areas with most properties out of the 21 metro areas (592 properties) for which I estimate metro-level indices. ***, **, and * indicate significant levels of 1%, 5%, and 10% respectively.

Metro-area	Properties	Excess return	Sensitivity
GA-Atlanta	66	0.000	1.09
IL-Chicago	64	-0.002	0.92
DC-Washington	59	0.003	0.76**
CA-Los Angeles	36	0.002	1.12
MN-Minneapolis	34	0.010***	0.47***
TX-Dallas	31	-0.012***	0.99
AZ-Phoenix	30	-0.004	1.35*
FL-Fort Lauderdale	27	0.001	1.04
TX-Houston	27	0.001	0.74
FL-Orlando	25	-0.002	1.44*

Table 8. Explanatory power of metro indices

This table reports regressions of properties' holding-period gross total returns (log) minus the national total return index during the same periods (log) against properties' respective metro-area total return indices (log) minus the national total return index during the same periods (log) for each property type respectively. Samples consist of properties located in metro areas with at least 15 properties. White (1980) heteroskedasticity-consistent coefficient standard deviations are in parentheses. ***, **, and * indicate significant levels of 1%, 5%, and 10% respectively.

	Apartment	Industrial	Office	Retail
Intercept term	-0.02**	-0.02***	-0.02	-0.00
	(0.01)	(0.01)	(0.01)	(0.02)
Metro - National	1.00***	1.00***	1.01***	1.00***
	(0.08)	(0.06)	(0.08)	(0.09)
Sample size	1,555	2,303	1,589	592
Adjusted R-square	0.09	0.12	0.08	0.16

Table 9. Summary of placebo test coefficients

This table report summary statistics of the slope coefficients from 1,000 rounds of placebo tests on the explanatory power of metro indices for each property type. In each round, I regress properties' holding-period gross total returns (log) minus the national total return index during the same periods (log) against total return indices (log) of random metro areas where properties are not located minus the national total return index during the same periods (log). ***, **, and * indicate significant levels of 1%, 5%, and 10% respectively.

	Apartment	Industrial	Office	Retail
Mean	-0.004	-0.010	-0.017	0.014
(Standard deviation)	(0.076)	(0.054)	(0.082)	(0.089)
Minimum	-0.239	-0.177	-0.305	-0.267
25%	-0.053	-0.048	-0.070	-0.042
Median	-0.08	-0.010	-0.016	0.014
75%	0.043	0.028	0.038	0.068
Maximum	0.225	0.134	0.219	0.276

Table 10. Summary of out-of-sample-test coefficients

This table report summary statistics of the slope coefficients from 1,000 rounds of out-of-sample tests on the explanatory power of metro indices for each property type. In each round, properties in metros with at least 15 properties are randomly split into two samples: A and B. I first use properties in sample A to estimate metro parameters and thus indices. I then regress sample B properties' holding-period gross total returns (log) minus the national total return index during the same periods (log) against metro-area total return indices (log) estimated from sample A minus the national total return index during the same periods (log). ***, **, and * indicate significant levels of 1%, 5%, and 10% respectively.

	Apartment	Industrial	Office	Retail
Mean	0.338***	0.494***	0.321***	0.381***
(Standard deviation)	(0.083)	(0.110)	(0.090)	(0.121)
Minimum	0.048	-0.011	0.074	-0.004
25%	0.281	0.434	0.259	0.293
Median	0.335	0.498	0.318	0.367
75%	0.391	0.568	0.383	0.460
Maximum	0.668	0.848	0.632	0.796

Figure 1. Sample size across time

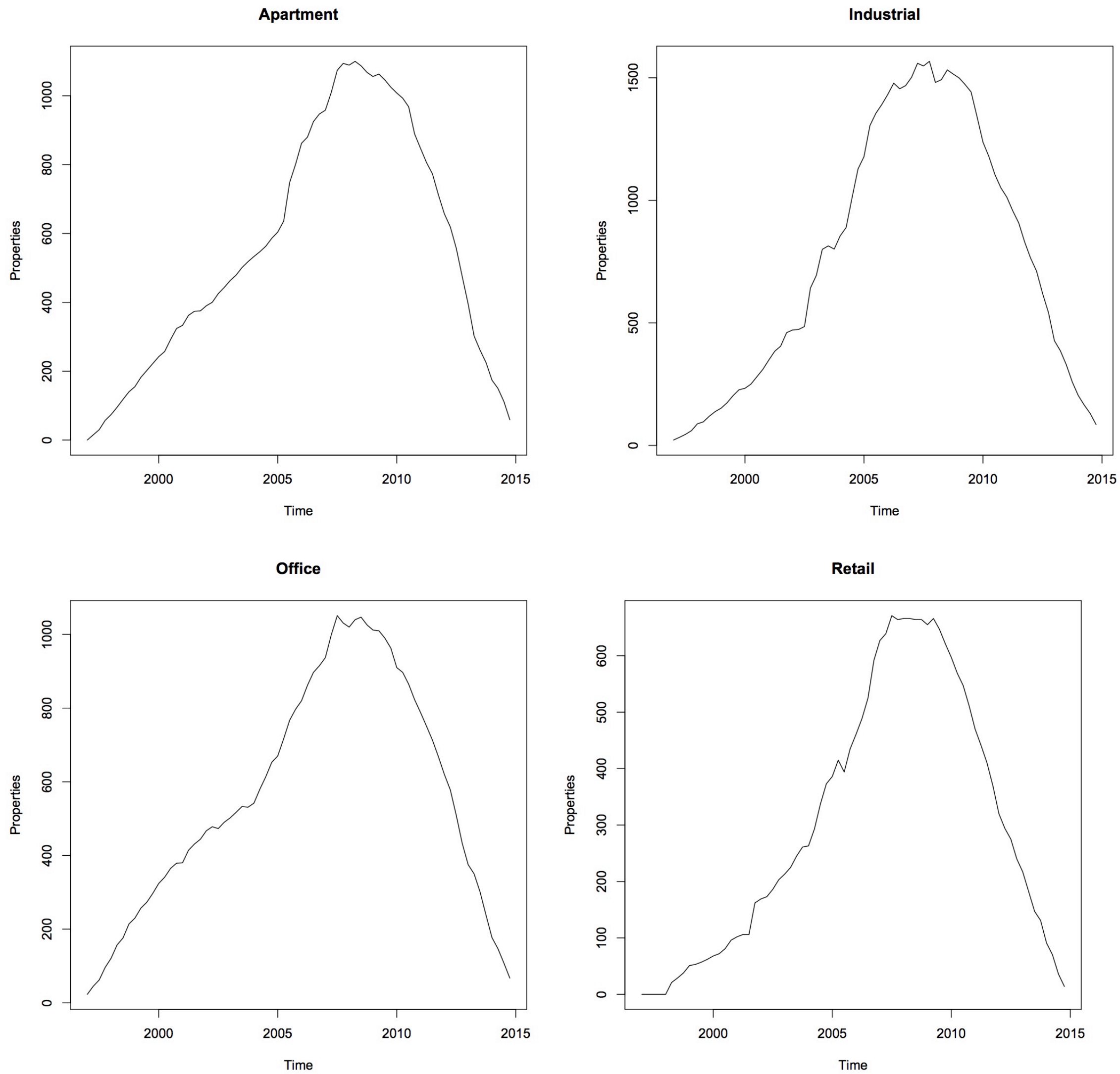

Figure 2. National commercial real estate total return indices

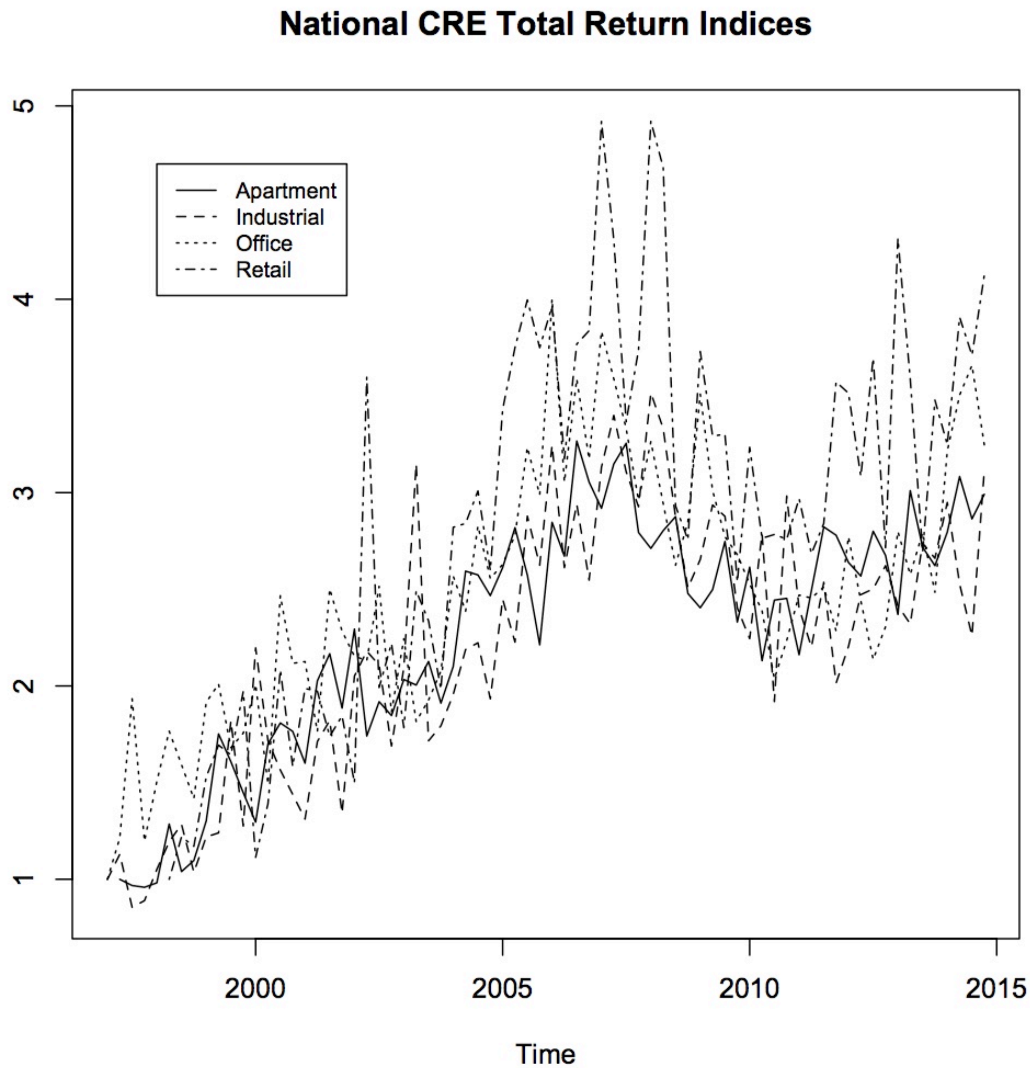

Figure 3. RSR national indices vs. NPIs

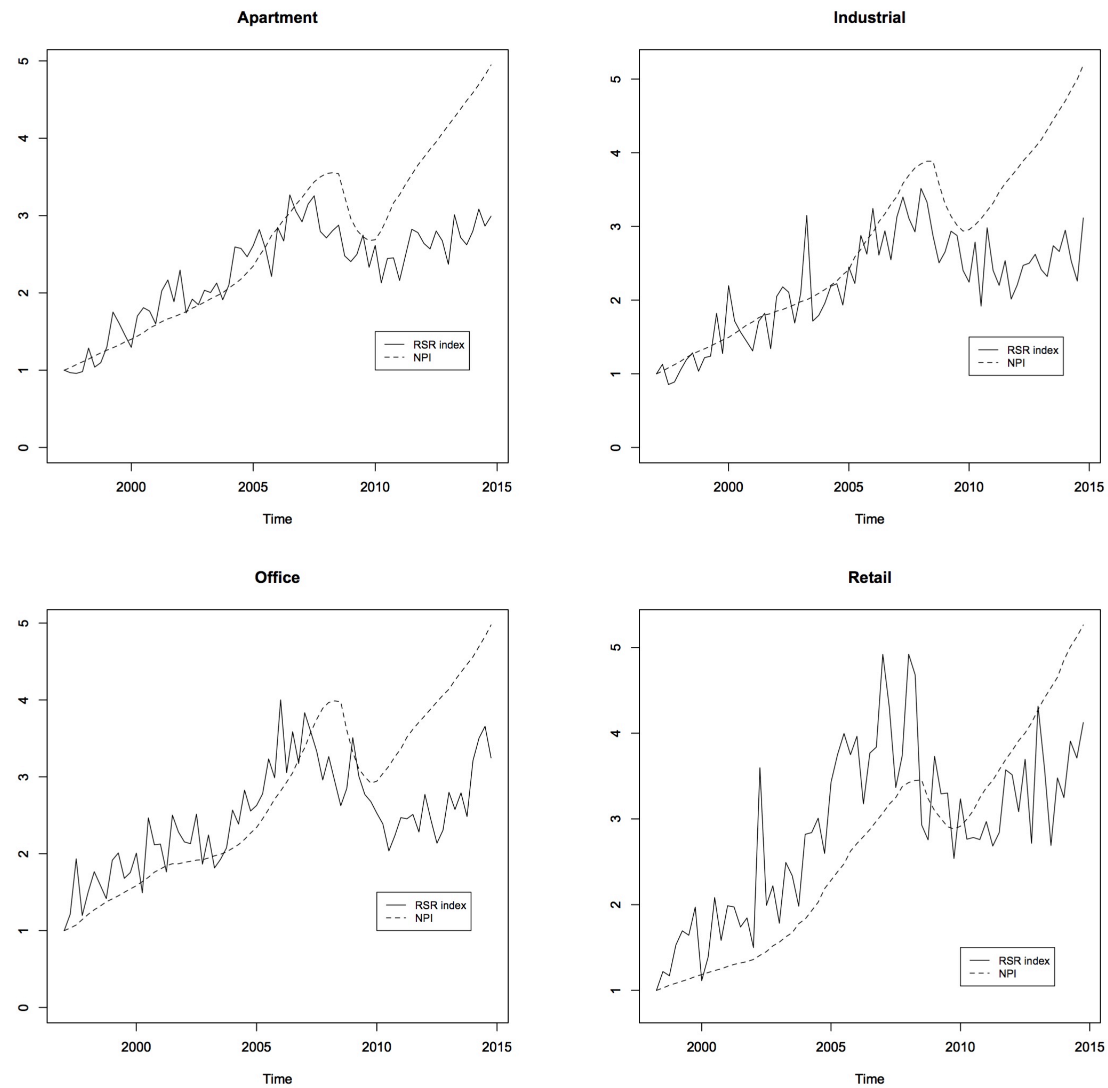

Figure 4. Metro-area total return indices

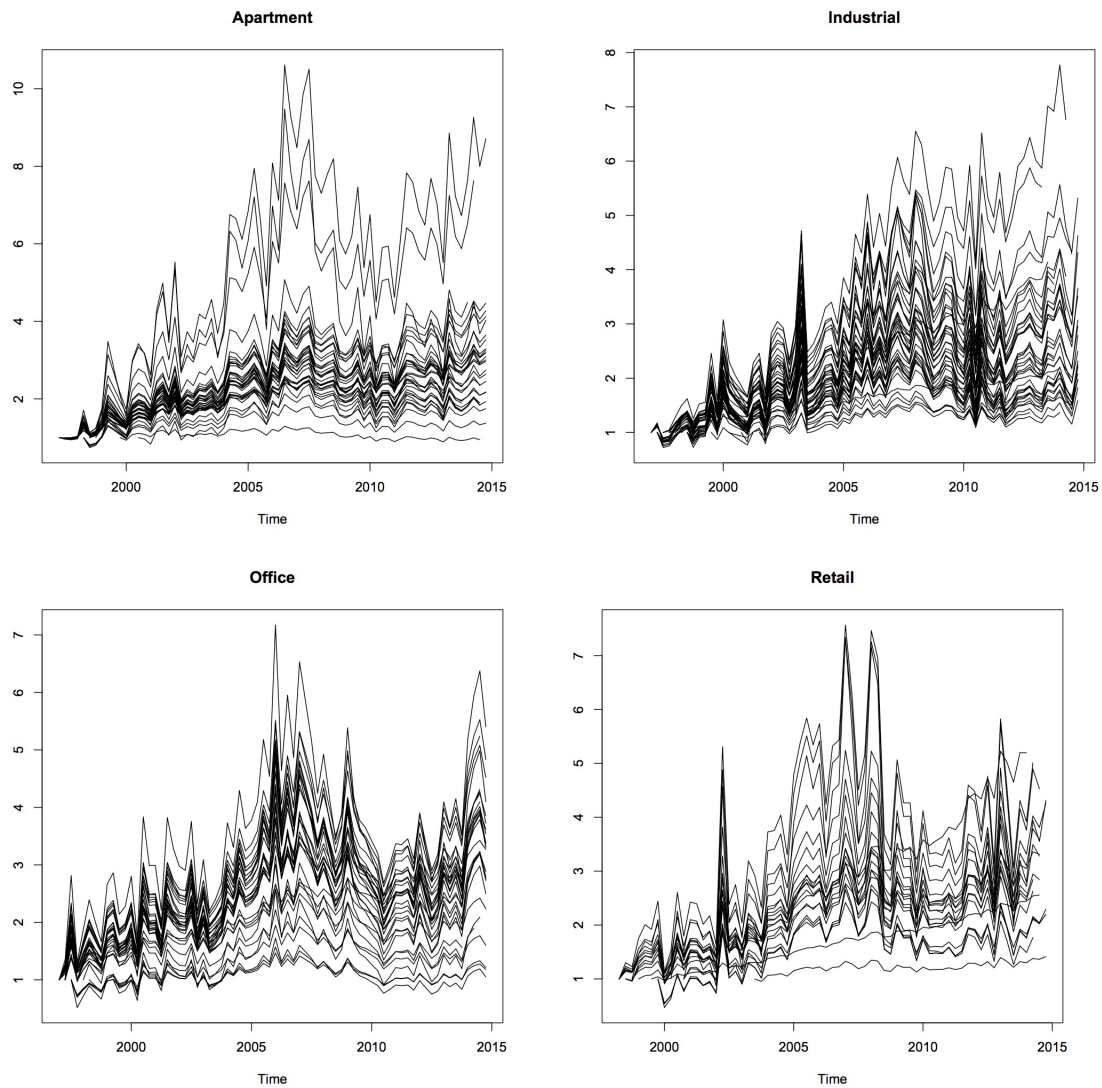

Figure 5. Histograms of placebo-test coefficients

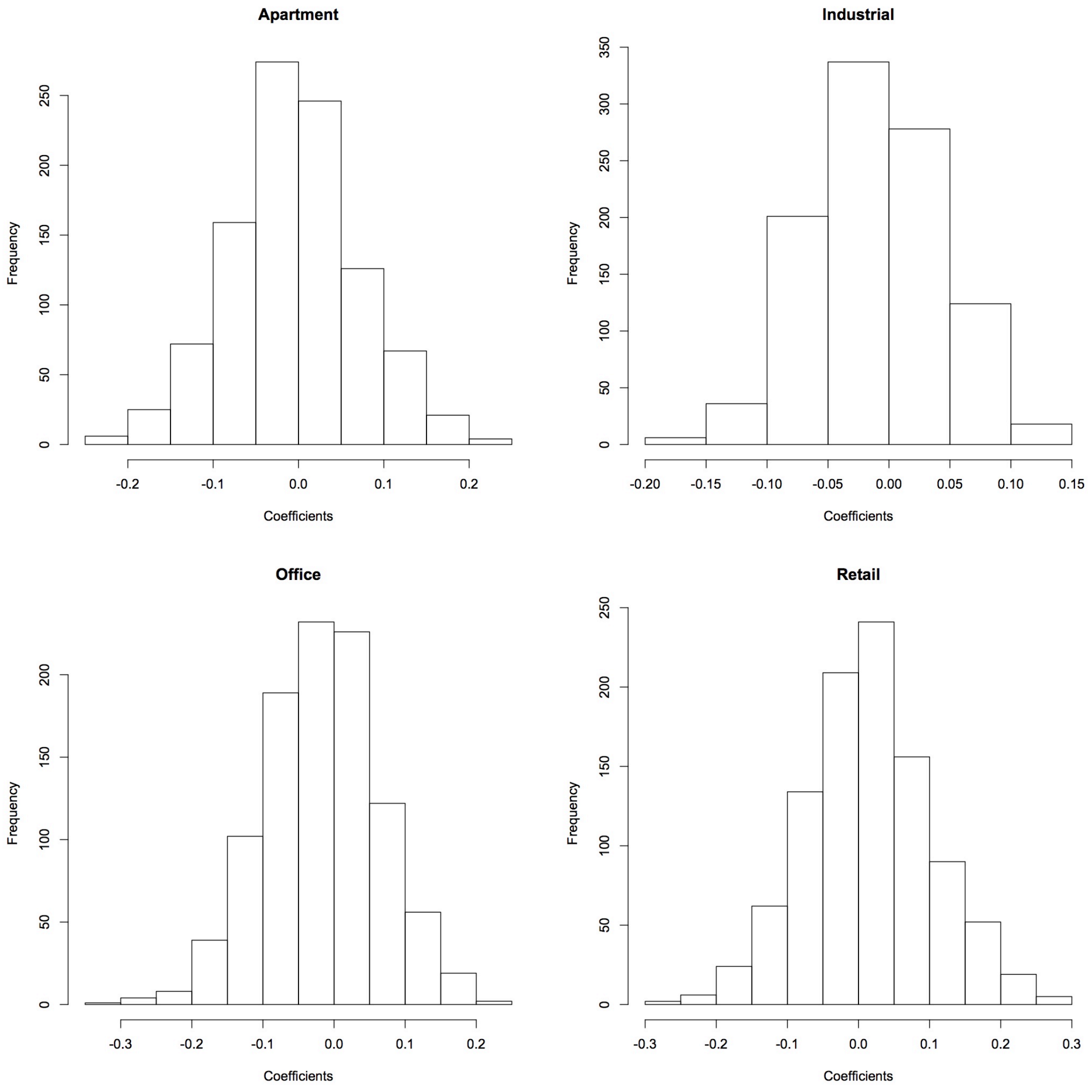

Figure 6. Histograms of out-of-sample-test coefficients