



CCRSI

CoStar Commercial
Repeat-Sale Indices

METHODOLOGY

TABLE OF CONTENTS

Executive Summary	3
Background	4
Data	5
Chart 1: Pair Count by Holding Period	5
Chart 2: Pair Count by Average Annual Price Change	6
Chart 3: Pair Counts by Property Types & Dollar Volume	7
Chart 4: Transaction Value Buckets – All Types	7
Chart 5: Transaction Value Buckets – By Property Type	7
Methodology	10
Table 1: Scenarios	11
Chart 6: Possible Ways to Extract High-Value Segment Based on Price	12
Chart 7: Share of Pair Count & Share of Transaction Value	13
Chart 8: CoStar Commercial Repeat-Sale National Indices	14
References	15
Appendix	16
Geometric Repeat-Sale Regression	16
Appendix Chart 1: Two Cases of Ridge Results	18
Arithmetic Repeat-Sale Regression	19

EXECUTIVE SUMMARY

This paper presents the methodology behind the new CoStar commercial real estate (CRE) repeat-sale price indices. The indices are based on CoStar's extensive CRE database that has 20 years of historical sales comparables across all price ranges and covers all types of CRE transactions across the United States.

The CRE market is characterized by heterogeneity and is highly segmented. This has proven to be a challenge to the research community, which has primarily focused on investment grade or high-value transactions—a very small fraction of total CRE transactions. Because of this narrowly focused view, the valuation of the CRE market tends to be a generalization based on the performance of a small segment.

Our goal was to capture the multifaceted and diverse picture of the commercial real estate market. In the course of the development of our indices, we explored all available repeat-sale indexing methodologies to determine the best approach. We found that the traditional approach to determining market tiers based on transaction prices resulted in a biased index. In response to this finding, we developed a unique methodology segregating high-value and low-value transactions based on the physical characteristics of properties rather than prices.

1. BACKGROUND

The CoStar Commercial Repeat-Sale Index was developed by using the repeat-sale regression technique, which has been increasingly accepted as the most clear-cut and sufficiently rigorous method to meet investors' requirements. The repeat-sale analysis, based on properties that have sold more than once without any significant change in building characteristics between sales, is fundamentally comparable to stock and bond indices, which are based on stock (or bond) price changes from one period to the next. In real estate, the most well-known repeat-sale index is the S&P/Case-Shiller home price index. Not only has it become the barometer of the health of the nation's housing market; it has also been used by the Chicago Mercantile Exchange to support and facilitate derivatives trading against housing. OFHEO's home price index is another example of a repeat-sale based index that has been widely used and cited in the residential housing market.

In the CRE market place, the most commonly used price index to date is the one produced by the National Council of Real Estate Investments Fiduciaries (NCREIF). The NCREIF index is appraisal-based rather than transaction-based and is derived from a small data sample that consists solely of large and prime properties owned by pension fund investors. The Moody's/Real CPPI was the first CRE repeat-sale index, developed in 2007. However, the use of this index is limited by the lack of comprehensive data coverage. CPPI is based on transaction data from Real Capital Analytics (RCA), which has approximately 10 years of history and focuses only on high-value transactions. RCA data initially covered transactions of \$5 million-up and extended to \$2.5 million-up in 2005.

As we mentioned in the executive summary, CoStar began collecting CRE transaction data 20 years ago and covers CRE transactions across the United States in all price ranges. This large database with a long history contains a large number of repeat transactions from which we were able to develop consistent and comprehensive price indices.

2. DATA

Accurate identification of repeat-sale pairs is critical to building an accurate index that reflects market conditions excluding all other factors. We therefore applied three stages of filtering to the CoStar transaction database to obtain the final sales pairs. In the first stage, we extracted a dataset containing properties that were sold more than once. We compared multiple building characteristics to ensure that two transaction pairs were indeed the same asset. In the second stage, we set up 32 exclusionary criteria to filter out non-representative transactions, such as portfolio sales, non-arm’s-length transactions, and build-to-suit transactions.

The third stage of data filtering mainly targeted “flipped” properties — those that sold more than once within a short period of time — and “outliers,” properties with abnormal price increases. Both were identified empirically. As **Chart 1** shows, most transactions occurred after a 12-month holding period, which is consistent with generally accepted practice because of US tax considerations. Therefore, transactions occurring in less than a 12-month period are filtered out as “flippers.”

Chart 1

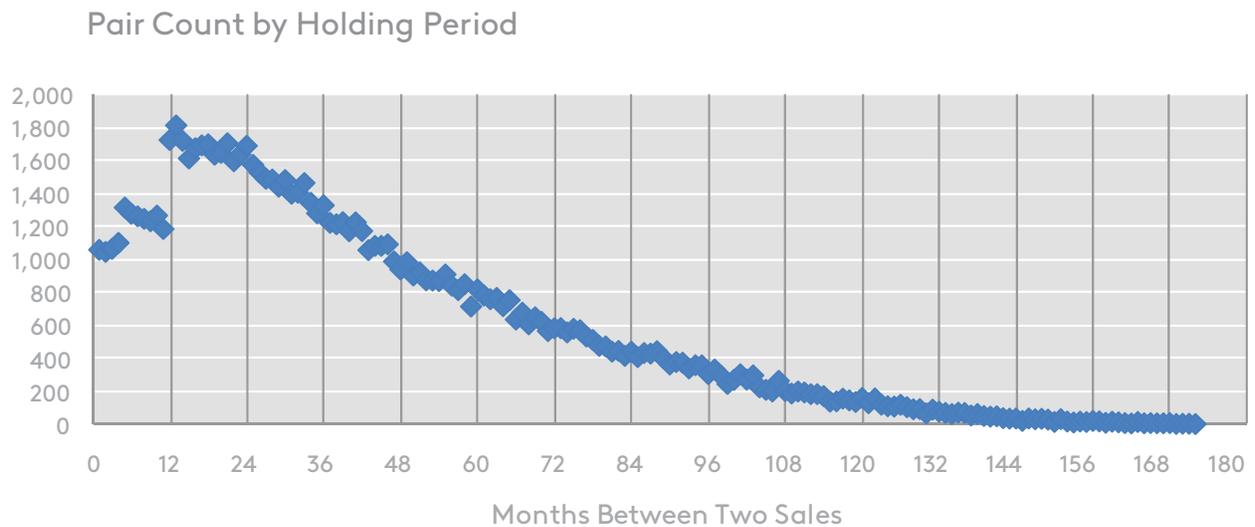
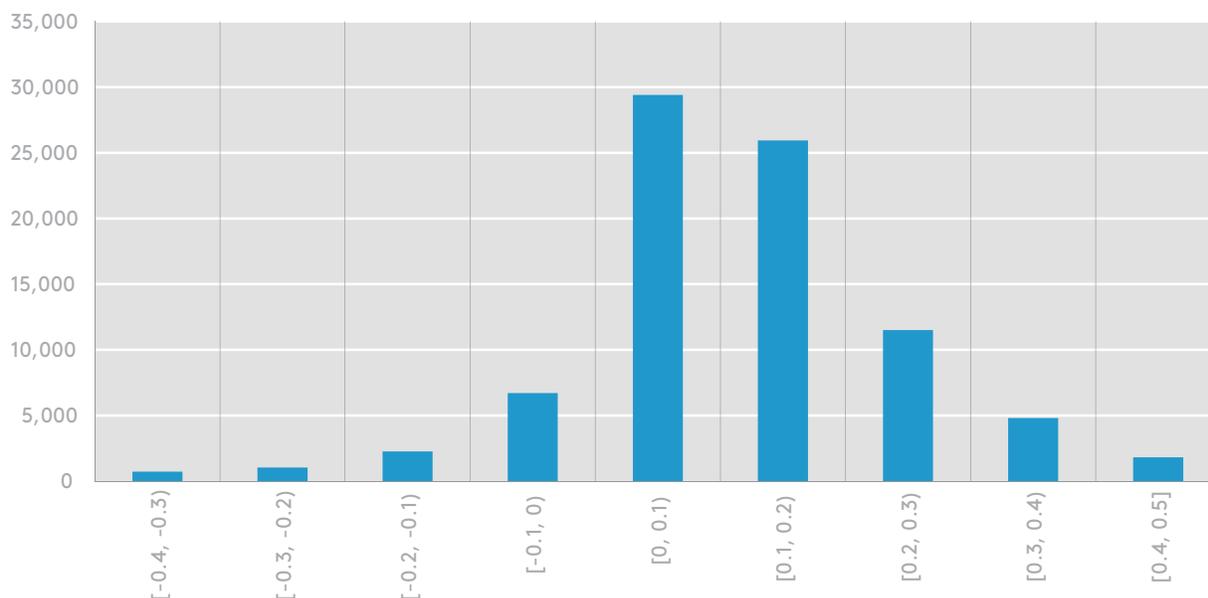


Chart 2 shows pair distribution by the average annual price change. Most of the pairs cluster around the 10% to 20% range. The pair count then decreases quickly in both directions. Roughly 98% of the total pairs fall into the range between -40% and 50%. We thus excluded the pairs at the extreme ends of the spectrum as “outliers” for all property type except for land. The range for land is -50% to 60% because it has a much wider and flatter distribution.

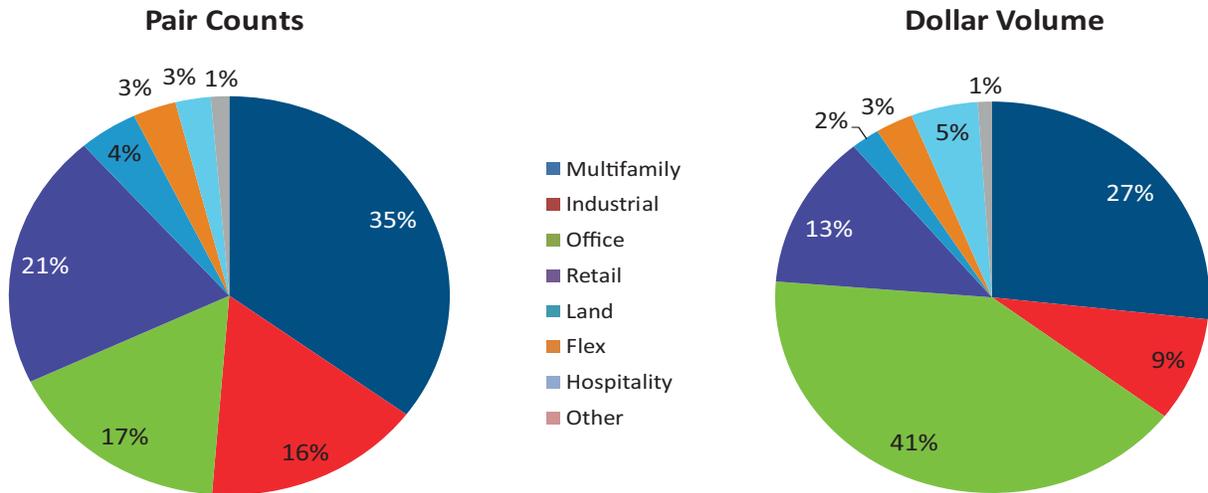
Chart 2

Pair Count by Average Annual Price Change



After the filtering process, our final dataset had a total of 85,428 repeat-sale observations covering the period 1996–2010. **Chart 3** shows the distribution of repeat-sale pair counts by property type and the corresponding share of transaction value of each property type. Multifamily ranks highest in the number of repeat-sale transactions, while office leads in the total value of transactions. This can be explained by the fact that while office transactions occur less frequently than multifamily sales, they tend to sell at higher prices.

Chart 3



The relationship between the number of transaction activities and transaction value can be further demonstrated by the distribution of repeat-sale pair counts over price brackets and the value of each bracket. As shown in **Chart 4**, most of the transaction activities are concentrated in the low price bracket. As prices increase, the number of transactions diminishes rapidly. Transaction value, on the other hand is concentrated in the high price bracket. In particular, those transactions greater than \$5 million account for the major share of total transaction value, even though the number of these high priced transactions represents a small fraction of the total number of transactions.

Chart 4

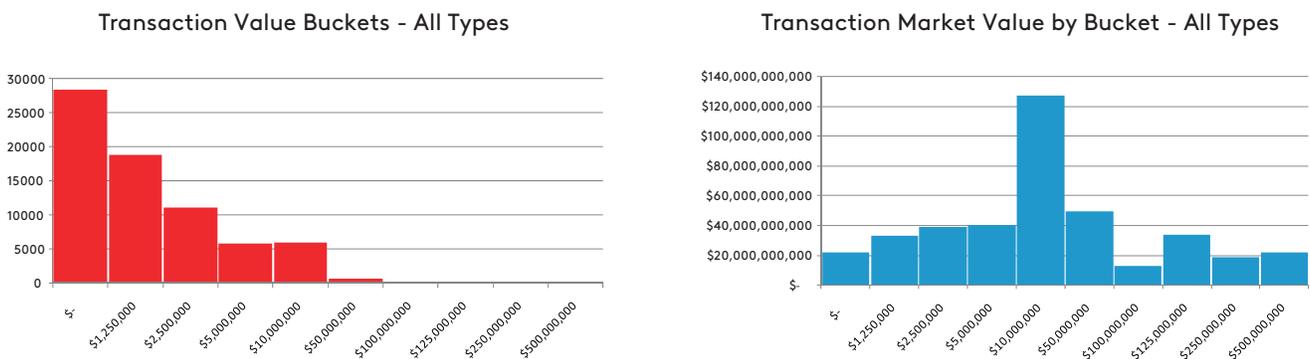


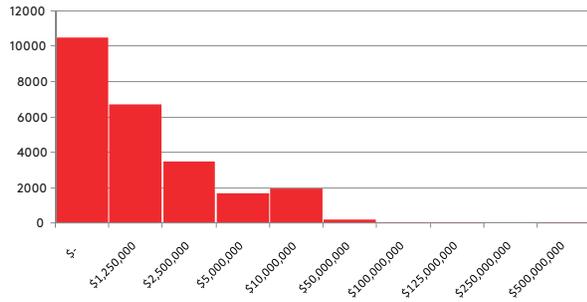
Chart 5 shows the distribution pattern for major property types. The divergence of the concentrations of transaction activity number and value is significant for multi-family. But it is most pronounced for the office property type. The industrial property type, on the other hand, does not show such an extreme divergence between transaction activity and value, because it generally sells within a narrow price range. We expect that retail would show a divergence similar to office. However, the repeat-sale pairs of high-end retail are underrepresented. This is due to the facts that retail transactions are generally included in multi-asset portfolio sales and the physical characteristics of retail properties change significantly from one sale to the next.

Clearly, the heterogeneity of the CRE market is multi-dimensional. There are a variety of property types, and there is also a wide range of difference of quality and price within each property type group. The lower priced segments and the higher priced segments of the CRE market attract different investors with different investment goals, resulting in widely varying performances. Therefore, the price movements in the CRE market are complicated and difficult to capture in a single index.

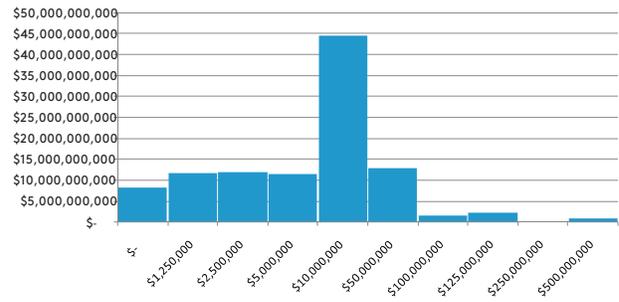
The CoStar Repeat-Sale Indices were therefore developed to capture the diversity of property types and the unique behavior of the relationship between transaction activity and transaction values among low-end and high-end properties.

Chart 5

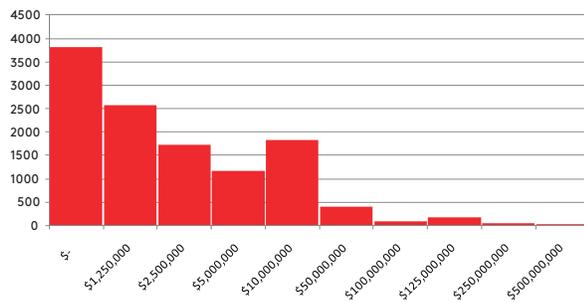
Transaction Value Buckets - Multi-Family



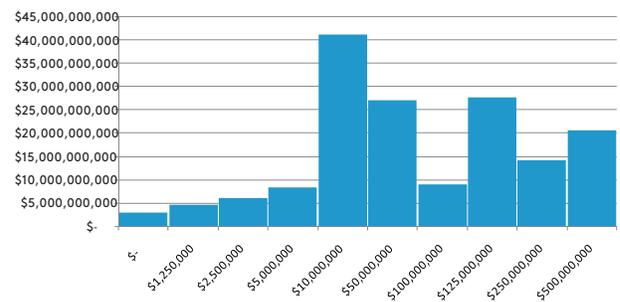
Transaction Market Value by Bucket - Multi-Family



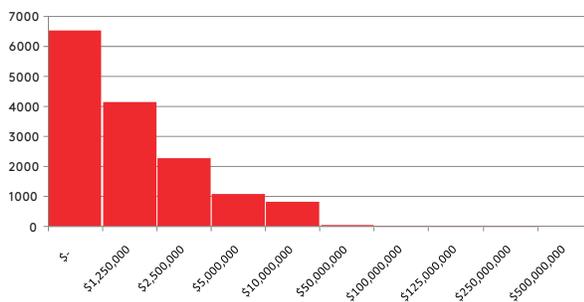
Transaction Value Buckets - Office



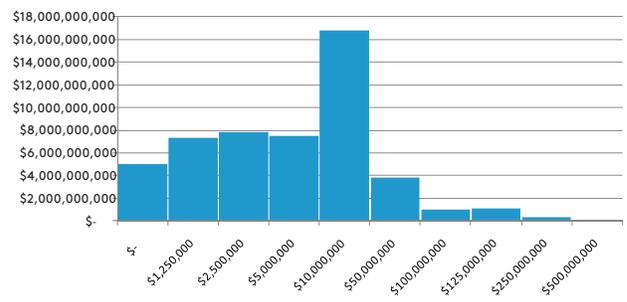
Transaction Market Value by Bucket - Office



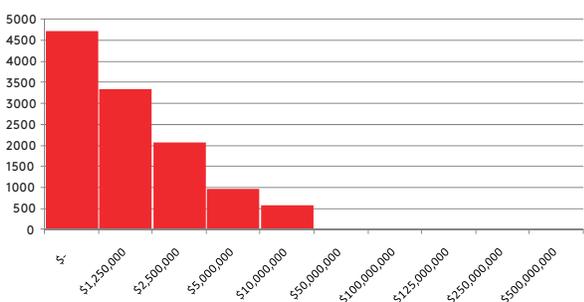
Transaction Value Buckets - Retail



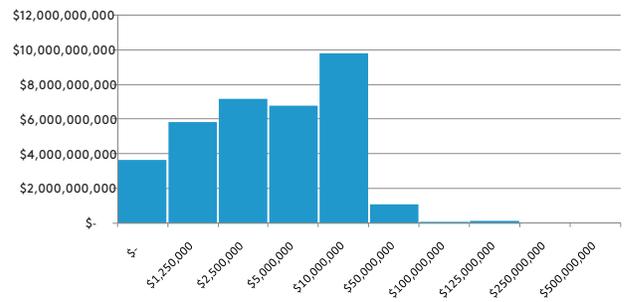
Transaction Market Value by Bucket - Retail



Transaction Value Buckets - Industrial



Transaction Market Value by Bucket - Industrial



METHODOLOGY

Since repeat-sale regression was first introduced by Bailey, Muth and Nourse (1963), significant advances have been made by Case-Shiller (1987), Case-Shiller (1989), Shiller (1991), Goetzmann (1992), Clapp-Giacotto (1992), Gatzlaff-Haurin (1997), and others. To date, there are two types of repeat-sale regression methodologies. The first most commonly used method treats every transaction equally, regardless of the value of the transaction. The resulting index is an equal-weighted, geometric mean index. The second method, known as the arithmetic mean repeat-sale regression, introduced by Shiller in 1991, weighs price change by the value of each transaction. The S&P/Case-Shiller repeat-sale housing index, for example, is an arithmetic mean index. We explored both approaches in developing CoStar index. See **Appendix 1** for a mathematic presentation of the methodologies we used.

Generally speaking, the geometric, equal-weighted methodology is more relevant for measuring the performance of individual properties, while the arithmetic, value-weighted index is a better measure of overall market performance. For capturing CRE price movement each method has its pros and cons. Due to the divergence in concentration of transaction activity and value in CRE, a single equal-weighted index is inadequate. In an equal-weighted index every transaction has the same impact on the results regardless of transaction price, which results in a bias towards low-value deals where transaction activity is highest.

The value-weighted index, on the other hand, captures the impact of high-value transactions, which can be relevant to CRE research due to the heavy influence of the high-end properties on overall market value. However, the value-weighted index has the potential to be overly “noisy” since a few very expensive sales will have a disproportionate impact. Therefore, a value-weighted index has limited applicability unless there are sufficient data to mitigate the noise.

Table 1 illustrates the differences in the two approaches. All three scenarios under Up Market on the left-hand side have the same appreciation, but the sales prices are different. The equal-weighted results show the same overall price change for all three scenarios, while the value-weighted results are tied closely to the appreciation of higher prices. The same holds true in the Down Market scenarios. The value-weighted method is sensitive to price variations in the high end and the equal-weighted method focuses only on the appreciation.

Table 1

UP MARKET

Scenario 1			
	1st Sale Price	Appreciation	2nd Sale Price
obs#1	\$ 10,000,000	5%	\$ 10,500,000
obs#2	10000000	0.1	11000000

Equal Weighted = 7.47%

Value Weighted = 7.50%

Scenario 2			
	1st Sale Price	Appreciation	2nd Sale Price
obs#1	\$ 10,000,000	5%	\$ 10,500,000
obs#2	100000000	0.1	110000000

Equal Weighted = 7.47%

Value Weighted = 9.55%

Scenario 3			
	1st Sale Price	Appreciation	2nd Sale Price
obs#1	\$100,000,000	5%	\$ 105,000,000
obs#2	10000000	0.1	11000000

Equal Weighted = 7.47%

Value Weighted = 5.45%

DOWN MARKET

Scenario 1			
	1st Sale Price	Appreciation	2nd Sale Price
obs#1	\$ 10,000,000	-5%	\$ 9,500,000
obs#2	10000000	-0.1	9000000

Equal Weighted = -7.53%

Value Weighted = -7.50%

Scenario 2			
	1st Sale Price	Appreciation	2nd Sale Price
obs#1	\$ 10,000,000	-0.05	\$ 9,500,000
obs#2	100000000	-0.1	90000000

Equal Weighted = -7.53%

Value Weighted = -9.55%

Scenario 3			
	1st Sale Price	Appreciation	2nd Sale Price
obs#1	\$ 100,000,000	-5%	\$ 95,000,000
obs#2	10000000	-0.1	9000000

Equal Weighted = -7.53%

Value Weighted = -5.45%

In the CRE marketplace high-end transactions are often extremely volatile. This inherent volatility may distort a value-weighted index. Since an equal-weighted index is skewed toward the low end, and the value-weighted index is skewed to the high end and potentially volatile, we decided that an equal-weighted index for separate market segments (low end/high end) would be a more comprehensive approach to capturing the diversity of the market.

However, our efforts to divide transactions into market segments illuminated some issues in the current practice of segmenting properties by price breakpoints. In the CRE marketplace, \$5 million and \$2.5 million are the traditional cut-off points for determining what is high-end and what is low-end. In the course of our research we discovered that segmenting the market data by price breakpoints has great potential to bias the resulting index.

Chart 6 presents the results of our simulations demonstrating how using fixed pricing cut-offs produces bias. If, for example, the first sale price is the cut-off point, the resulting index will capture falling prices, but not rising prices, resulting in downward bias. Conversely, using the second-sale price as the breakpoint, results in an upward bias. If both first and second sale prices are considered, and are greater than a certain price breakpoint, then price movements are pushed below the breakpoint into a lower priced market segment, resulting in an understatement of volatility. Using either the first sale or the second sale price as breakpoints, on the other hand, draws in price movements from the market segment below the breakpoints, resulting in an overstatement of volatility.

In summary, the simulation results of four possible cuts show that using the first sale price generates the lowest average annual growth, 2.9%, whereas using the second sales price results in the highest annual growth rate, 7.7%. While the either/or approach generates the highest volatility, 12.2%, using both sales prices results in the lowest volatility, 9.6%.

Chart 6

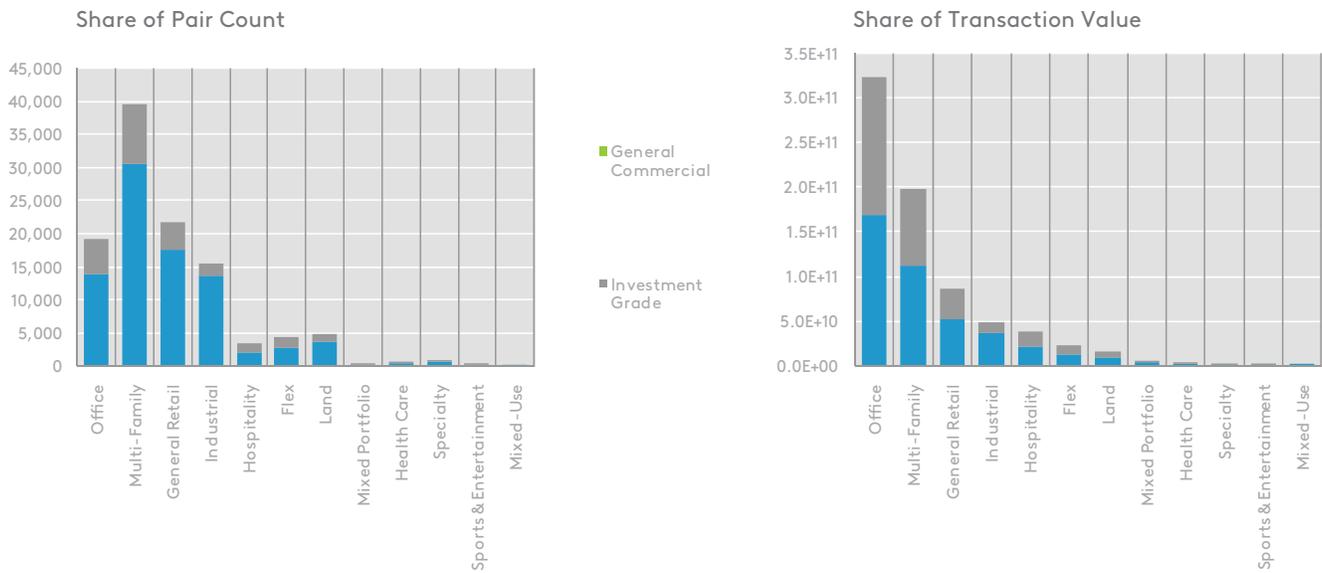
Possible Ways to Extract High-Value Segment Based on Price

<p>First Sale Price > 2.5 Million</p> <p>DOWNWARD</p> <p>Peak to Trough: 45% Average Annual Growth: 2.9% Volatility: 11.6%</p>	<p>Second Sale Price > 2.5 Million</p> <p>UPWARD</p> <p>Peak to Trough: 31% Average Annual Growth: 7.7% Volatility: 9.8%</p>
<p>Both First Sale Price and Second Sale Price > 2.5 Million</p> <p>UNDER STATE VOLATILITY</p> <p>Peak to Trough: 34% Average Annual Growth: 4.6% Volatility: 9.6%</p>	<p>Either First Sale Price or Second Sale Price > 2.5 Million</p> <p>OVER STATE VOLATILITY</p> <p>Peak to Trough: 43% Average Annual Growth: 6.0% Volatility: 12.2%</p>

Using price as a differentiator in developing a price index creates a circular reference. Given this issue, and the varying results from using price cut-offs to define market segments described above, we decided to try a different approach to differentiate market segments at the national level. Rather than use pricing breakpoints, we divided the market into two segments —investment grade and general commercial — defined by building characteristics.

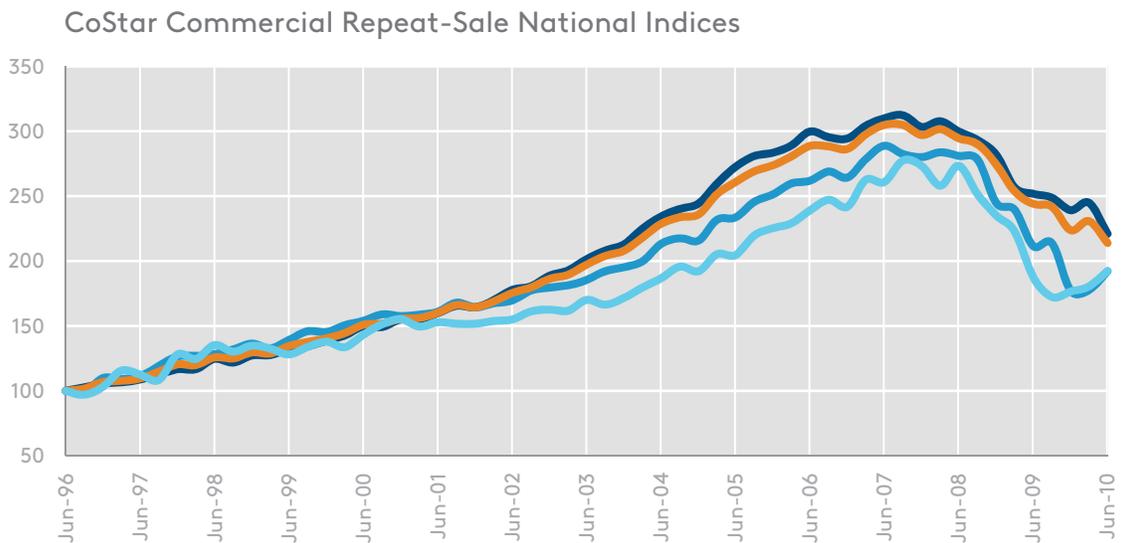
The investment grade properties comprise roughly 30% of the total transactions and capture 77% of the transaction value. The remaining general commercial properties tend to be smaller and fall into low price brackets. **Chart 7** shows the pair count share and the value share of investment grade properties by property type. Multifamily still leads in the number of repeat-sale transactions, while office leads in the total value of transactions. This reflects the same total composition of the dataset as shown in **Chart 3**.

Chart 7



The CoStar equal-weighted index for the investment grade segment does closely resemble a composite value-weighted index since both indices capture the price movement of high-value properties **(Chart 8)**. We found, however, that using the equal-weighted method to track two mutually exclusive market segments, gives a clearer picture of the market, is less complicated, and is the more consistent approach.

Chart 8



In conclusion, the CoStar indices have been developed to capture the multifaceted and diverse nature of CRE. CoStar's rich dataset made it possible for us to approach the challenge from multiple angles. The resulting indices will provide accurate benchmarks to all levels of CRE investment activities.

REFERENCES:

1. Bailey, M.J., R.F. Muth, and H.O. Nourse. 1963. "A Regression Method for Real Estate Price Index Construction." *Journal of the American Statistical Association*, 58: 933-942.
2. Case, K.E., and R.J. Shiller. 1987. "Price of Single-Family Homes since 1970: New Indexes for Four Cities." *New England Economic Review*, 1987: 45-56.
3. Case, K.E., and R.J. Shiller. 1989. "The Efficiency of the Market for Single-Family Homes." *American Economic Review*, 79: 125-137.
4. Shiller, R.J. 1991. "Arithmetic Repeat-sales Price Estimators." *Journal of Housing Economics* (1): 110-126.
5. Goetzmann, W.H. 1992. "The Accuracy of Real Estate Indices: Repeated Sale Estimators." *Journal of Real Estate Finance and Economics*, 5(1): 5-53.
6. Clapp, J.M., and C. Giacotto. 1992. "Estimating Price Trends for Residential Property: A Comparison of Repeated Sales and Assessed Value Methods." *Journal of Real Estate Finance and Economics*, 5(4):357-374.
7. Gatzlaff, D., and D. Haurin. 1997. "Sample Selection Bias and Repeated-Sales Index Estimates." *Journal of Real Estate Finance and Economics*, 14(1): 23-40.
8. Calhoun, C.A. 1996. "OFHEO House Price Indexes: HPI Technical Description," http://www.fhfa.gov/webfiles/896/hpi_tech.pdf.
9. Standard & Poor's. 2006. "S&P/Case-Shiller® metro Area Home Price Indices," http://www2.standardandpoors.com/spf/pdf/index/SPCS_MetroArea_HomePrices_Methodology.pdf.
10. Geltner, D., and Pollakowski, H. 2007. "A Set of Indexes for Trading Commercial Real Estate Based on the Real Capital Analytics Transaction Prices Database," MIT Center for Real Estate.
11. Meese, R.A., and N.E. Wallace. 1997. "The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches." *Journal of Real Estate Finance and Economics*, 14: 51-73.

APPENDIX

A. Geometric Repeat-Sale Regression

With geometric repeat-sale regression (GRS), the logarithm of cumulative price appreciation for a property between two sales is expressed as:

$$Y = \ln\left(\frac{P(t_s)}{P(t_f)}\right) = \sum_t \ln[1 + R(t)]X(t) = \sum_t \beta(t)X(t) \quad (1)$$

$$X(t) = \begin{cases} 1 & t_f < t \leq t_s \\ 0 & \text{otherwise} \end{cases}$$

where $P(t_f)$ is the 1st sale price, $P(t_s)$ is the 2nd sale price, $R(t)$ is the appreciation rate at period t , $X(t)$ is a dummy variable, and $\beta(t)$ is the regression coefficient. Because the appreciation rate R is calculated based on the ratio of 1st and 2nd sale prices, all transactions are equally weighted independent of the values of the transactions. With estimated appreciation rate R , the price index is

$$Index(t) = \prod_{t=0}^t [1 + R(t)]$$

In matrix format, the representation of Equation (1) can be expressed as:

$$Y = X \cdot \beta + \varepsilon \quad (2)$$

Assuming the error term follows the Gaussian diffusion process, β can be estimated by the ordinary least square (OLS):

$$\beta_{OLS} = (X'X)^{-1}X'Y \quad (3)$$

For repeated sales occurring between long time durations, the appreciation rate $R(t)$ estimated from Equations (1) and (3) is the average over the period. This average may not capture the appreciation rate at a particular given time. To correct for the heteroskedasticity problem (the longer the time span between two transactions, the more idiosyncratic the errors), a second regression is estimated to find the variance dependence on the time span between two sales:

$$V = \langle \varepsilon^2 \rangle = A + B \times t + C \times t^2 \quad (4)$$

With the expected variance for each pair, a third OLS regression is estimated to obtain intervalweighted coefficients:

$$\frac{Y}{\sqrt{V}} = \sum \frac{\beta(t)X(t)}{\sqrt{V}} \quad (5)$$

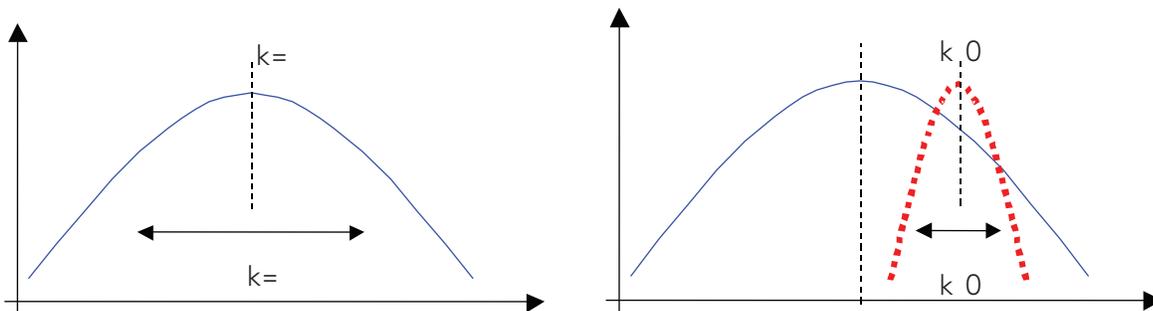
Equations (2), (4) and (5) complete the interval-weighted geometric repeat-sale regression (I-GRS).

The ordinary least square estimator β_{OLS} , **Equation (3)**, can be unreliable when collinearity is present in its parameter matrix. Goetzmann (1992) first suggested the ridge estimator as an additional step to repeat-sale regression if observations are limited. The method was also used in Geltner and Pollakowsky (2007). By introducing a bias in the estimation,

$$\beta_k = (X'X + kI)^{-1} X'Y, \quad 0 \leq k \leq \frac{2\sigma^2}{\beta_{OLS}'\beta_{OLS}} \quad (6)$$

where I is a unit matrix and σ is the variance of β_{OLS} , the mean square error of ridge estimator β_k can be smaller than that of β_{OLS} . The effect of ridge estimation is graphically illustrated in **Figure (1)**. Adding a ridge component reduces the variance of the estimation, but the mean values of the coefficients are shifted from OLS estimates. As long as the bias, together with the reduced variance, falls within the variance of the OLS estimates, the results of the ridge method can be considered better alternatives.

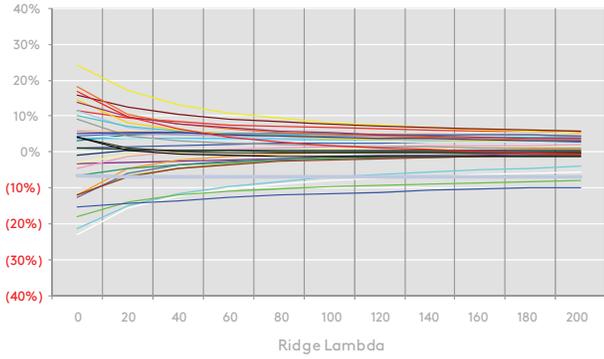
Figure 1



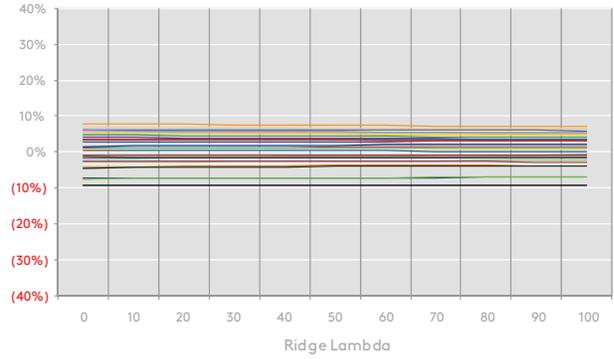
To use or not use the ridge estimator is a trade-off between efficiency and bias, as the selection of k value is not necessarily deterministic. We used ridge correction in the development of local market indices. For aggregated sub-indices, such as the property type index, although there exists some $k > 0$ for which β_k is better than β_{OLS} , the improvements are barely noticeable. **Appendix Chart 1** shows two cases of ridge results. Apparently, ridge estimator improves regional retail index but makes no difference at national retail index.

Appendix Chart 1

Mid-West Retail Regression Coefficients by Ridge Lambda



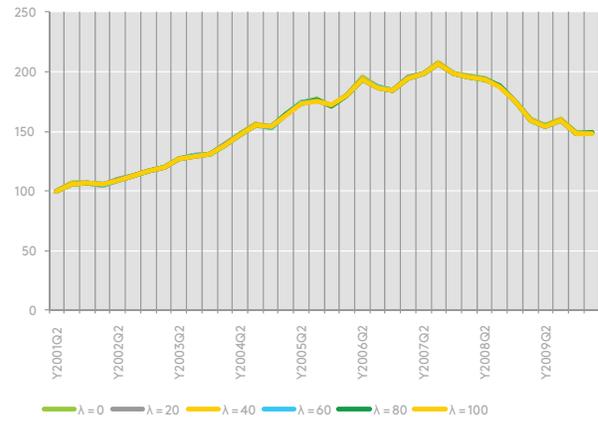
National Retail Regression Coefficients by Ridge Lambda



Mid-West Retail Indices by Ridge Lambda



Mid-West Retail Indices by Ridge Lambda



B. Arithmetic Repeat-Sale Regression

Arithmetic repeat-sale regression (ARS) is analogous to GRS, but with some modified assumptions. A primary objective of ARS is to provide a value-weighted repeat-sale price index similar to market-cap weighted stock indices.

Following Shiller (1991), we define:

$$\beta_t = \frac{P_0}{P_t}. \quad (7)$$

Here β_t is the inverse of accumulated price appreciation between time t and a base period ($t = 0$). For the base period, $\beta_0 \equiv 1$. The price index is given by

$$Index(t) = \frac{1}{\beta_t}.$$

The price difference of two repeat-sales, after adjusting by the inverse index value β , should be zero with a random distribution error ε ,

$$\beta_k P_{nf} - \beta_t P_{ns} = \varepsilon_n \quad (8)$$

where n denotes property n , and f, s denote time periods of the repeat-sales. To estimate β , we define a dummy variable

$$\Omega_{nt} = \begin{cases} -1 & \text{for first repeat sale of property } n \text{ in time period } t \\ 1 & \text{for second repeat sale of property } n \text{ in time period } t \\ 0 & \text{otherwise} \end{cases}$$

and a matrix Z with elements $Z_{nt} = \Omega_{nt}$ ($t=1, 2, \dots$). **Equation (8)** can be expressed as

$$Y = X \cdot \beta + \varepsilon$$

where the elements of matrix X and Y are

$$X_{nt} = P_{nt} \cdot \Omega_{nt} \quad (t=1, 2, \dots),$$

$$Y_{nt} = -P_{nt} \cdot \Omega_{nt} \quad (t=0).$$

The ARS method calculates the value of β from

$$\hat{\beta} = (Z' \cdot X)^{-1} Z' \cdot Y \quad (9)$$

The resulting $\hat{\beta}$ is an arithmetic value-weighted average of the inverse of price appreciation. For repeated sales occurring between long time durations, the appreciation rate estimated from **Equations (7)-(9)** is again the average over the period. Interval adjustment for ARS is similar to that of GRS as given in **Equations (4) and (5)**.

Finally, ARS often requires a robust weighting procedure to mitigate influence of sale pairs with extreme price change. Instead of correcting absolute errors, we correct normalized error to preserve the essence of the arithmetic-mean. Our robust specification is as follows:

The price deviation of Sample i due to the arithmetic-mean of the reciprocal growth rate $\hat{\beta}$ is

$$\varepsilon_{p_i} = \hat{\beta}_{t2} P_{2i} - \hat{\beta}_{t1} P_{1i}$$

Define
$$\varepsilon_{\hat{\beta}_i} = \frac{\varepsilon_{p_i}}{P_{2i}} = \hat{\beta}_{t2} - \hat{\beta}_{t1} \frac{P_{1i}}{P_{2i}} \quad \text{and} \quad \sigma_{\beta} = \sqrt{\frac{\sum \varepsilon_{\hat{\beta}_i}^2}{N-1}}$$

the normalized error of $\hat{\beta}$ is
$$\delta_i = \frac{\varepsilon_{\hat{\beta}_i}}{\sigma_{\beta}}$$

If $\varepsilon_i \gg 1$, the sample is likely to be an outlier that is not representative of the market trend. A second regression is done with transaction prices weighted to reduce the error

$$\text{weight} = \frac{1}{1 + e^{\frac{\hat{\delta}_i - \hat{\delta}_0}{r}}} \quad \text{where } \hat{\delta}_0 = 4, \text{ and } r = 0.5 .$$