

## MIT Open Access Articles

*Estimating Real Estate Price Movements for High Frequency Tradable Indexes in a Scarce Data Environment*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Bokhari, Sheharyar and David Geltner. "Estimating Real Estate Price Movements for High Frequency Tradable Indexes in a Scarce Data Environment." The Journal of Real Estate Finance and Economics (2010) : 1-22.

**As Published:** <http://www.springerlink.com/content/uxl087168u781086/>

**Publisher:** Springer Science + Business Media B.V.

**Persistent URL:** <http://hdl.handle.net/1721.1/64714>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike 3.0



# Estimating Real Estate Price Movements for High Frequency Tradable Indexes in a Scarce Data Environment

By  
Sheharyar Bokhari & David Geltner\*

*This version June 2010*

## **Abstract**

Indexes of commercial property prices face much scarcer transactions data than housing indexes, yet the advent of tradable derivatives on commercial property places a premium on both high frequency and accuracy of such indexes. The dilemma is that with scarce data a low-frequency return index (such as annual) is necessary to accumulate enough sales data in each period. This paper presents an approach to address this problem using a two-stage frequency conversion procedure, by first estimating lower-frequency indexes staggered in time, and then applying a generalized inverse estimator to convert from lower to higher frequency return series. The two-stage procedure can improve the accuracy of high-frequency indexes in scarce data environments. In this paper the method is demonstrated and analyzed by application to empirical commercial property repeat-sales data.

## **Key Words:**

Real estate price indexes; Frequency-conversion; Transactions-based-index estimation; Derivatives; Noise filter.

---

\* The authors are both with the MIT Center for Real Estate, Commercial Real Estate Data Laboratory, 77 Massachusetts Ave, Cambridge MA 02139. Geltner is the contact author, at [dgeltner@mit.edu](mailto:dgeltner@mit.edu) (phone 617-253-8308, fax 617-258-6991).

# Estimating Real Estate Price Movements for High Frequency Tradable Indexes in a Scarce Data Environment

Sheharyar Bokhari & David Geltner

## 1. Introduction & Background

In the world of transaction price indexes used to track market movements in real estate, it is a fundamental fact of statistics that there is an inherent trade-off between the frequency of a price-change index and the amount of “noise” or statistical “error” in the individual periodic price-change or “capital return” estimates.<sup>1</sup> Geltner & Ling (2006) discussed the trade-off that arises, as higher-frequency indexes are more useful, but *ceteris paribus* are more noisy and noise makes indexes less useful. More generally, the fundamental problem is transaction data scarcity for index estimation, and this is a particular problem with commercial property price indexes, because commercial transactions are much scarcer than housing transactions.<sup>2</sup> However, the greater utility of higher frequency indexes has recently come to the fore with the advent of tradable derivatives based on real estate price indexes.<sup>3</sup> Tradability increases the value of frequent, up-to-date information about market movements, because the lower transactions and management costs of synthetic investment via index derivatives compared to direct cash investment in physical property allows profit to be made at higher frequency based on the market movements tracked by the index. Higher-frequency indexes also allow more

---

<sup>1</sup> The terms “noise” and “error” are used more or less interchangeably in this paper.

<sup>2</sup> There are over 100 million single-family homes in the U.S., but less than 2 million commercial properties.

<sup>3</sup> Over-the-counter trading of the IPD Index of commercial property in the UK took off in 2004 and after growing rapidly through 2007 the market remained active through the financial crisis of 2008-09. Trading on the appraisal-based NCREIF Property Index (NPI) of commercial property in the US commenced in the summer of 2007. The Moody's/REAL Commercial Property Price Index, launched in September 2007 based on Real Capital Analytics Inc (RCA) data, is also designed to be a tradable index and is, like the Case-Shiller house price index, a repeat-sales transaction price-based index.

frequent “marking” of the value of derivatives contracts, which in turn allows smaller margin requirements, which increases the utility of the derivatives.<sup>4</sup>

In the present paper we propose a two-stage frequency-conversion estimation procedure in which, after a first-stage regression is run to construct a *low frequency* index, a second-stage operation is performed to convert a staggered series of such low-frequency indexes to a *higher-frequency index*. The first-stage regression can be performed using any desirable index-estimation technique and based on either hedonic or repeat sales data.<sup>5</sup> The proposed frequency conversion procedure is optimal in the sense that it minimizes noise at each stage or frequency. We find that while the resulting high-frequency index does not have as high a signal/noise ratio (SNR) as the underlying low-frequency indexes, it adds no noise in an absolute sense to what is in the low-frequency indexes, and it generally has less noise than direct high-frequency estimation. The 2-stage procedure thus preserves essentially all of the advantage of the low-frequency estimation while providing the additional advantage of a higher-frequency index.

The rest of the paper is organized into the following sections. In section 2, we briefly review the existing literature on estimating real estate price indexes and introduce the frequency-conversion technique, which we label as the “Generalized Inverse Estimator” (GIE) based on the mathematical procedure it employs. In sections 3 and 4 respectively, we discuss the hypothesized merits of the proposed procedure and provide an empirical comparison between the proposed technique and other popular methods of

---

<sup>4</sup> For example, margin requirements in a swap contract are dictated by the likely net magnitude of the next payment owed, which is essentially a function of the periodic volatility of the index, and volatility (per period) is a decreasing function of index frequency (simply because there is less time for market price change deviations around prior expectations to accumulate between index return reports that cover shorter time spans). Lower margin requirements allow greater use of synthetic leverage which facilitates greater liquidity in the derivatives market.

<sup>5</sup> It could even be based on appraisal data if the reappraisals occur staggered throughout the year.

high frequency price index estimation. We conclude in section 5 that the frequency conversion procedure tends to be more accurate at the higher frequency than direct high-frequency estimation in a data scarce environment.

## **2. Prior Work and the Proposed Frequency Conversion Procedure**

Goetzmann (1992) introduced into the real estate literature what is perhaps the major approach to date for addressing small-sample problems in price indexes, namely, the use of biased ridge or Stein-like estimators in a Bayesian framework. Other approaches that have been explored in recent years include various types of parsimonious regression specifications that effectively parameterize the historical time dimension (see e.g. Schwann (1998), McMillen et al (2001), and Francke (2009)), as well as procedures that make use of temporal and spatial correlation in real estate markets (see for instance, Clapp (2004) and Pace et al (2004)). Some such techniques show promise, but are perhaps more appropriate in the housing market than in commercial property markets. Spatial correlation is more straightforward in housing markets, and the need for transparency in a tradable index can make it problematical to estimate the index on sales outside of the subject market segment. Another concern that is of particular importance in indexes supporting derivatives trading is that the index estimation procedure should minimize the constraints placed on the temporal structure and dynamics of the estimated returns series, allowing each consecutive periodic return estimate to be as independent as possible, in particular so as to avoid lag bias and to capture turning points in the market

as they occur even if these are inconsistent with prior temporal patterns in the index.<sup>6</sup>

Most of the previously noted recent techniques are unable to fully address these issues.

Bayesian procedures such as that introduced by Goetzmann (1992) can have the desirable feature of not inducing a lag bias and not hampering the contemporaneous representation in the index of turning points in the market. Such a technique, applied if necessary at the underlying low-frequency first stage, can therefore complement the frequency-conversion procedure we propose, and we explore such synergy in the present paper, finding that the frequency-conversion technique can further enhance indexes that are already optimized by such Bayesian methods.

We now introduce the frequency conversion procedure.<sup>7</sup> For illustrative purposes, we derive a quarterly-frequency index from four underlying staggered annual-frequency indexes. Other frequency conversions are equally possible in principle (e.g., from quarterly to monthly, or semi-annual to quarterly). Also for illustrative purposes and because they represent a scarce-data environment, we use a repeat sales database in this paper to assess the frequency-conversion technique. However, as mentioned earlier, the application of the proposed procedure is in principle not limited to any particular type of dataset, sample-size or choice of first-stage estimation methodology.

---

<sup>6</sup> This is particularly important to allow the derivatives to hedge the type of risk that traders on the short side of the derivatives market are typically trying to manage. For example, developers or investment managers seek to hedge against exposure to unexpected and unpredictable downturns in the commercial property market.

<sup>7</sup> It should be noted that the procedure introduced here is similar to what has been recently suggested in the regional economics literature, where Pavia et al (2008) provide a method for estimating quarterly accounts of regions from the national quarterly and annual regional accounts. The two methods are similar in that both use the generalized inverse, in the regional economics case to construct a quarterly regional series with movements that closely track the underlying annual figures.

As noted, commercial property transaction price data in particular is scarce (e.g., compared to housing data). To the extent the market wants to trade specific segments, such as, say, San Francisco office buildings, the transaction sample becomes so small that we may need to accumulate a full year's worth of data before we have enough to produce a good transactions-based estimate of market price movement. This is the type of context in which we propose the following two-stage procedure to produce a quarterly index.

## **2.1 The Proposed Methodology:**

We begin by estimating annual indexes in four versions with quarterly staggered starting dates, beginning in January, April, July, and October. Label these four annual indexes: “CY”, “FYM”, “FYJ”, and “FYS”, to refer to “calendar years” and “fiscal years” identified by their ending months. Each index is a true annual index, not a rolling or moving average within itself, but consisting of independent consecutive annual returns.<sup>8</sup> It is important to use time-weighted dummy variables in the low-frequency stage in order to eliminate temporal aggregation. For example, for the calendar year (CY) index beginning January 1<sup>st</sup>, a repeat-sale observation of a property that is bought September 30 2004 and sold September 30 2007 has time-dummy values of zero prior to CY2004 and subsequent to CY2007, and dummy-variable values of 0.25 for CY2004, 1.0 for CY2005 and CY2006, and 0.75 for CY2007.<sup>9</sup> The result will look something like what is pictured in Exhibit 1 for an example index based on the Real Capital Analytics repeat-sales database for San Francisco Bay area office property. If properly specified, these annual

---

<sup>8</sup> That is, independent *within* each index. Obviously, there is temporal overlap *across* the indexes.

<sup>9</sup> This specification, attributable to Bryon & Colwell (1982), eliminates the averaging of the values within the years, and effectively pegs the returns to end-of-year points in time. See Geltner & Pollakowski (2007) for more details.

indexes generally have no lag bias and essentially represent end-of-year to end-of-year price changes.<sup>10</sup> Each of these indexes also has as little noise as is possible given the amount of data that can be accumulated over the annual spans of time. It is of course important for the low-frequency indexes to minimize noise and, while not the focus of the present paper, the annual-frequency indexes depicted here employ the previously noted Goetzmann Bayesian ridge regression method, which as noted does not introduce a general lag bias.<sup>11</sup>

-----  
Insert Exhibit 1 about here.  
-----

Next, a frequency-conversion is applied to this suite of annual-frequency indexes to obtain a quarterly-frequency price index implied by the four staggered annual indexes. We want to perform this frequency conversion in the most accurate way possible, with as little additional noise and bias as possible. How can we use those staggered annual indexes to derive an up-to-date quarterly-frequency index? Looking at the staggered annual-frequency index levels pictured in Exhibit 1, one is tempted to try to construct a quarterly-frequency index by simply averaging across the levels of the four indexes at each point in time. (Try to fit a curve “between” the four index levels.) But such a process would entail a delay of three quarters in computing the most recent quarterly return (while we accumulate all four annual indexes spanning that quarter), which for derivatives trading purposes would defeat the purpose of the higher-frequency index.

---

<sup>10</sup> However, it should be noted that in the early stages of a sharp downturn in the market, loss aversion behavior on the part of property owners can cause a data imbalance that can make it difficult for an annual-frequency index to fully register the downturn at first. This consideration will be discussed shortly.

<sup>11</sup> Note that according to the Goetzmann Bayesian approach, the ridge is not necessary when the resulting indexes are sufficiently smooth without the ridge. This turns out to be the case for the eight regional indexes examined later in the present paper, but not for the MSA-level indexes such as the San Francisco Office index depicted in Exhibit 1.



Such a levels-averaging procedure would also considerably smooth the true quarterly returns (it would effectively be a time-centered rolling average of the annual returns).

The approach we propose to the frequency-conversion procedure is a matrix operation which can be conceived of as a second-stage “repeat-sales” regression at the quarterly frequency using the four staggered annual indexes as the input repeated-measures data.<sup>12</sup> Each annual return on each of the four staggered indexes is treated like a “repeat-sale” observation in this “second-stage regression”. If we have  $T$  years of history, we will have  $4T-3$  such repeated-measures observations (the row dimension of the second-stage regression data matrix), and we will have  $4T$  quarters for which we have time-dummies (the columns dimension in the regression data matrix, the quarters of history for which we want to estimate returns). We are missing “1<sup>st</sup>-sales” observations for the first three quarters of the history, the quarters that precede the starting dates for all of the annual indexes other than the one that starts earliest in time (the CY index in our present example), as the staggered annual indexes each must start one quarter after the previous. Obviously, with fewer rows than columns in the estimation data matrix, our regression is “under-identified”, that is, the system has fewer equations than unknowns.<sup>13</sup> Basic linear algebra tells us that such a system has an infinite number of exact solutions (that is, quarterly index return estimates that will cause the predicted quarterly values to exactly match the repeated-measures observations on the left-hand-side of the second-stage regression, i.e., a regression  $R^2$  of 100%, a perfect fit to the data which is the low-

---

<sup>12</sup> As noted, the input annual indexes from the first stage do not need to be repeat-sales indexes, in principle. They could be any good transactions-price based type of index, such as a hedonic index. They provide “repeated measures” in the sense that they provide repeated observations of index levels or price changes over time.

<sup>13</sup> We cannot simply drop out the first three quarters from the second-stage index, as that will then impute the first three annual returns entirely to the first quarter (only) of the index history and thereby bias the estimation of all of the quarterly returns.

frequency index returns). However, of all of those infinite solutions, there is a particular solution that minimizes the variance of the estimated parameters, i.e., that minimizes the additional noise in the quarterly returns, noise added by the frequency-conversion procedure. This solution is obtained using what is called the “Moore-Penrose pseudoinverse” matrix of the data (see original papers by Penrose (1955, 1956) and its applications in Albert (1972)). We shall refer to this frequency-conversion method as the “Generalized Inverse Estimator”, or GIE for short. This estimator is the “Best Linear Minimum Bias Estimator” (BLMBE) (see (Chipman, 1964) for proof).

How good is the GIE as a frequency-conversion method? It adds effectively no noise and very little bias to the underlying annual-frequency returns. Appendix B shows a way to see that the bias resulting from such an estimation decreases as the number of index periods to be estimated ( $T$ ) becomes large, approaching zero as the history approaches an infinite number of periods. From the simulation analysis in Exhibit 2, it can be seen that with even small values of  $T$ , the amount of bias is small and economically insignificant. In the simulation in Exhibit 2 the history consists of less than seven years, converting to 27 quarters. Exhibit 2 depicts a typical randomly-generated history of true quarterly market values (the thick black line, which in the real world would be unobservable), the corresponding staggered annual index levels (thin, dashed lines, here without any noise, to reveal any noise added purely by the frequency-conversion second stage), and the resulting second-stage GIE-estimated quarterly index levels (thin red line with triangles, labeled “ATQ” for “Annual-to-Quarterly”).<sup>14</sup> Clearly,

---

<sup>14</sup> In Exhibit 2 the first (CY) annual index starts arbitrarily at a value of 1.0, and the subsequent three staggered annual indexes are pegged to start at the interpolated level of the just-prior annual index at the time of the subsequent index’s start date. This is merely a convention and does not impact the quarterly

the derived GIE quarterly index almost exactly matches the true quarterly market value levels. The slight deviation reflects the bias. Numerous simulations of random histories and varying market patterns over time give similar results to those depicted in Exhibit 2. The GIE-based frequency-conversion adds only minimal and economically insignificant bias to the staggered annual indexes, while increasing the index frequency to quarterly. Unlike other techniques that require a procedure for choosing an optimal value for a parameter, the GIE is already optimal in the class of linear estimators, and it is relatively simple to compute (see Appendix A).

-----  
Insert Exhibit 2 about here.  
-----

## 2.2 An Illustration of the Setup for the 2<sup>nd</sup> Stage Regression

To clarify and summarize the proposed procedure, consider this illustration. Suppose the following staggered annual returns were estimated from the 1<sup>st</sup> stage regression:

Annual Period	CY Returns	Annual Period	FYM Return	Annual Period	FYJ Return	Annual Period	FYS Return
1Q00-4Q00	a	2Q00-1Q01	b	3Q00-2Q01	c	4Q00-3Q01	d
1Q01-4Q01	e						

Then, as shown in the next table below, the left-hand side variable for the 2<sup>nd</sup> stage regression will be the stack of annual returns (left-most column in the data table) and the right-hand side variables would be time-dummies that are set equal to 1 for the four

---

return estimates, as all indexes are only indicators of relative price movements across time, not absolute price levels.

quarters that make up a particular annual return observation (the other columns in the table below).

Stacked-Returns	1Q 2000	2Q 2000	3Q 2000	4Q 2000	1Q 2001	2Q 2001	3Q 2001	4Q 2001
a	1	1	1	1	0	0	0	0
b	0	1	1	1	1	0	0	0
c	0	0	1	1	1	1	0	0
d	0	0	0	1	1	1	1	0
e	0	0	0	0	1	1	1	1

As seen above, there are more quarterly returns to estimate than there are staggered annual return observations. Specifically, for the ATQ frequency conversion, there are always 3 extra parameters to estimate. Appendix A outlines the method used for this estimation and Appendix B shows a way to see that the bias resulting from such an estimation decreases as  $T$  becomes large. Indeed, intuitively, the reader can convince oneself that as  $T$  becomes large, the percentage difference between  $T$  and  $T + 3$  decreases. Hence, the system gets closer to being effectively exactly identified and thus the bias goes down over time. From the simulation analysis in Exhibit 2, it can be seen that with even small values of  $T$ , the amount of the bias is small. It should also be noted that since the 2<sup>nd</sup>-stage regression fits the stacked returns observations exactly, any noise in the estimation of the staggered annual returns gets carried over to the estimated quarterly returns, but no new noise is added. Thus, there is a direct relationship between statistically reliable 1<sup>st</sup> and 2<sup>nd</sup> stage estimations. For this reason the frequency conversion procedure should be viewed as a complement or “add-on” to good lower-frequency index estimation, not a substitute for such current best practice.

### 2.3 General Characteristics of the Resulting Derived Quarterly Index:

Based on the foregoing argument, the 2-stage derived quarterly index (which we shall refer to here as the “GIE/ATQ”, for “annual-to-quarterly”) offers the prospect of being more precise than a directly-estimated single-stage quarterly index, as it is based fundamentally on a year’s worth of data instead of just a quarter’s worth. However, though better than direct quarterly estimation in terms of precision or noise minimization, in one sense the 2-stage index cannot be as “good” as the corresponding underlying annual-frequency indexes, if we define the index quality by the signal/noise ratio. But the GIE/ATQ will provide information more frequently than the annual indexes, and this may make a lower signal/noise ratio worthwhile. To see this, consider the following.

In the signal/noise ratio (SNR) the numerator is defined theoretically as the periodic return volatility (longitudinal standard deviation) of the (true) market price changes, and the denominator is defined as the standard deviation of the error in the estimated periodic returns.<sup>15</sup> The GIE/ATQ frequency-conversion procedure gives a SNR denominator for the estimated quarterly index which is no larger than that of the underlying annual-frequency indexes (due to the exact matching of the underlying annual returns noted in the previous section). That is, the standard deviation of the error in the

---

<sup>15</sup> The theoretical SNR cannot be observed or quantified in the real world, where the true market returns cannot be observed, and hence the true market volatility (SNR numerator) cannot be observed. Empirical estimates of the theoretical SNR are confounded by the fact that the volatility of any empirically estimated index will itself be “contaminated” by the noise in the estimated index (the denominator in the SNR). Furthermore, the denominator of the theoretical SNR should equal the theoretical cross-sectional standard deviation in the return estimates, which is not exactly what is measured by the regression’s standard errors of its coefficients. To see this, consider conceptually a “perfect” index whose return estimates always exactly equal the unobservable true market returns each period. The regression producing such an index would have zero in the denominator of its theoretical SNR and yet would still have positive standard errors for its coefficients for any empirical estimation sample, as there is noise in the estimation database (cross-sectional dispersion in the price changes), causing the regression to have non-zero residuals in the data. In spite of these practical limitations, the theoretical SNR is a useful construct for conceptual analysis purposes (and also in simulation analysis, where “true” returns can be simulated and observed).

second-stage quarterly return estimates is no larger than that in the first-stage annual return estimates, as evident in the simulation depicted in Exhibit 2 by the fact that the ATQ adds essentially no error. But the numerator of the SNR is governed by the fundamental dynamics of the (true) real estate market. These dynamics dictate that the periodic return volatility will be smaller for higher frequency returns. For example, if the market follows a random walk (serially uncorrelated returns), the quarterly volatility will be  $1/\text{SQRT}(4) = 1/2$  the annual volatility. This means that, even though the theoretical SNR denominator does not increase at all (no additional estimation error), the SNR in the ATQ would still be one-half that in the underlying annual indexes. If the market has some sluggishness or inertia (positive autocorrelation in the quarterly returns, as is likely in real estate markets) then the SNR will be even more reduced in the ATQ below that in the annual indexes.

Importantly, the SNR of the GIE/ATQ can still be greater than that of a directly-estimated (single-stage) quarterly index. To see this, suppose price observations occur uniformly over time. Then there will be four times as much data for estimating the typical annual return in the annual-frequency indexes compared to the typical quarterly return in the directly-estimated quarterly-frequency index. By the basic “Square Root of N Rule” of statistics, this implies that the directly-estimated quarterly index will tend to have  $\text{SQRT}(4) = 2$  times greater standard deviation of error in its (quarterly) return estimates than the annual indexes have in their (annual) return estimates. Thus, the SNR for the direct quarterly index will have a denominator twice that of the annual indexes and therefore twice that of the GIE/ATQ 2-stage quarterly index. Of course, either way of producing a quarterly index will still be subject to the same numerator in the theoretical

SNR, which is purely a function of the true market volatility. Thus, the GIE/ATQ will have a lower SNR than the underlying annual-frequency indexes, but it will have a higher theoretical SNR than a directly-estimated quarterly index. In data-scarce situations, this can make an important difference.<sup>16</sup>

To return to the essential point of the contribution of this technique, while the GIE/ATQ does not have as good a SNR as the annual indexes, it does provide more frequent returns than the annual indexes (quarterly instead of annual), and thereby does provide additional information.<sup>17</sup> Thus, there is a useful trade-off between the staggered annual indexes and the derived quarterly index. The GIE/ATQ gives up some SNR information usefulness in the accuracy of its return estimates, but in return provides higher frequency return information.

#### **2.4 An Illustrative Example of Annual-to-Quarterly Derivation in Data-Scarce Markets:**

To gain a more concrete feeling for the above-described methodology and application, consider one of the smaller (and therefore more data-scarce) markets among the 29 Moody's/REAL Commercial Property Price Indexes that are based on the RCA repeat-sales database: San Francisco Bay metro area office properties.<sup>18</sup>

---

<sup>16</sup> The fact that the GIE/ATQ theoretical SNR is greater than that of direct quarterly estimation does not mean that in every empirical instance it would necessarily be greater. It should be noted that formal definition and computation of a "standard error" for the GIE is not straightforward. As noted, the regression is under-identified, which means there are no "residuals" in the second stage.

<sup>17</sup> Among the four staggered annual indexes we do get new information every quarter, but that information is only for the entire previous 4-quarter span, which is not as useful as information about the most recent quarter itself, which is what is provided by the ATQ. For example, a turning point in the most recent quarter will not necessarily show up in the most recent annual index, as the latter is still influenced by market movement earlier in the 4-quarter span it covers.

<sup>18</sup> The Moody's/REAL Commercial Property Price Index is produced by Moody's Investor Services under license from Real Estate Analytics LLC (REAL). During the 2006-09 period the San Francisco Office index averaged 12 repeat-sales transaction price observations (second-sales) per quarter, and in the most recent quarter (3Q09) there were only 2 observations.

-----  
Insert Exhibit 3a about here.  
-----

First consider direct quarterly repeat-sales estimation of a San Francisco office market price index. Exhibit 3a depicts a standard Case-Shiller version of such an index based on the 3-stage WLS procedure first proposed by Case and Shiller (1987). This index is labeled “CS” in the chart and is indicated by the thin blue line with solid diamonds. The scarcity of the data gives the CS index so much noise that the resulting spikiness practically obscures the signal of the fall, rise, and fall again in that office property market subsequent to the dot-com bust, the following recovery, and the 2008-09 financial crisis and recession. The thicker green line marked by open diamonds labeled CSG adds the Goetzmann Bayesian ridge noise filter to the basic CS approach. The CSG index allows most of the market price trajectory signal to come through. But this index still may be excessively noisy for supporting derivatives trading, where index noise equates to “basis risk” that undercuts the value of hedging and adds spurious volatility that will turn off synthetic investors.

Now observe how the 2-stage procedure works in the San Francisco office market example. Exhibit 3b depicts the CSG-based direct-quarterly index which we just described together with the GIE/ATQ-based 2-stage quarterly index (indicated by the red line marked by open triangles). Exhibit 3b also shows the staggered annual-frequency indexes that underlie the ATQ (as thin fainter solid lines without markers). These indexes are themselves CSG-based repeat-sales indexes of the same methodology as the direct-quarterly index, only estimated at the annual rather than quarterly frequency. Thus, the annual indexes that underlie the ATQ index already employ the state-of-the-art noise



filtering of the Goetzmann procedure. In Exhibit 3b, note how the ATQ index is generally consistent with the annual returns that span each quarter.<sup>19</sup> However, the quarterly index picks up and quantifies the changes implied by *changes* in the staggered annual indexes. For example, while the CY annual index ending at the end of 4Q2007 was positive (up 14.7%), it was less positive than the immediately preceding FYS annual index ending in 3Q2007 (up 26.7%). The resulting derived ATQ quarterly index indicated a downturn in 4Q2007 (-2.1%). Meanwhile, the direct-quarterly CSG index picked up a sharp downturn in 3Q2007, a quarter sooner than the ATQ, but the CSG then indicated a positive rebound of +1.6% in 4Q2007, which is probably noise. The smoother pattern in the ATQ index suggests less noise and therefore less spurious quarterly return estimates.

-----  
Insert Exhibit 3b about here.  
-----

At first it may seem odd that the derived quarterly index can be negative when all of the staggered annual indexes that underlie it are positive. The intuition behind a result such as the above example is that an annual index could still be increasing as a result of rises during the earlier quarters of its 4-quarter time-span, with a drop in the last quarter that does not wipe out all of the previous three quarters' gains. When the most recent annual index is rising at a lower rate than the next-most-recent annual index, it can (although does not necessarily) indicate that the most recent quarter was negative. The derived quarterly return (ATQ) methodology is designed to discover and quantify such

---

<sup>19</sup> In fact, as noted previously, the ATQ returns are *exactly* equivalent to the corresponding underlying annual-frequency returns over each 4-quarter span of time covered by each of the staggered annual indices periodic returns. The depicted ATQ index level in the exhibit does not exactly touch each annual index periodic end-point value only because of the arbitrary starting value for the annual indexes. Note that the ATQ and CY indexes exactly match at the end of each calendar year, as both these indexes have the same starting value of 1.0 at the same time (2000Q4). The same would be true of the other three annual indexes if we set their starting values equal to the level of the ATQ at their starting points in time during 2001.

situations in an optimal (i.e., “BLMBE”) manner. As noted, simple curve-fitting of the annual indexes introduces excessive smoothing, and will not be able to pick up in a timely manner the kind of turning point just described.

The San Francisco office market depicted in Exhibit 3b offers an excellent example of both the strengths and weaknesses of the GIE/ATQ method versus direct quarterly estimation using state-of-the-art methods such as the CSG index depicted in the chart. Even though it uses Bayesian noise filtering, the CSG index is relatively noisy, as indicated by its spikiness during much of the history depicted (even when the transaction data was most plentiful). The CSG index differs importantly from the staggered annual indexes estimated from the same repeat-sales data. Arguably, the direct quarterly index does not as well represent what was going on in the San Francisco Bay office market during a number of individual quarters of the 2001-2008 period. For example, down movements of -9.1% in 2Q2005, -1.7% in 1Q2006, and -4.3% in 3Q2006 seem out of step with the strong bull market of that period, while up movements of +6.3% in 2Q2001 in the midst of the Bay Area’s tech bust, and +1.6% in 4Q2007 and +1.2% in 3Q2009 seem out of step with the big downturn of 2007-09. The staggered annual indexes and the ATQ seem to better represent the tech-bust-related fall in the Bay Area office market during 2001-03 and the strong bull market of the 2004-07 period, and indeed in this particular case the ATQ appears visually to be about as good as the annual indexes (by the smoothness of the index lines’ appearance), in addition to being more frequent. The directly-estimated quarterly index has considerably greater quarterly volatility than the ATQ, a likely indication of greater noise in the former index.

On the other hand, in spite of the anomalous uptick in 4Q2007, the directly-estimated quarterly index shows some sign of slightly temporally leading the ATQ and the annual-frequency indexes. This is most notable in the direct quarterly index's beginning to turn down in 3Q2007, one quarter ahead of the ATQ in the 2007-09 market crash. Thus, there is some suggestion in our San Francisco office example that the GIE/ATQ method may not be quite as "quick" as direct quarterly estimation at picking up a sharp market downturn, although it appears to rapidly catch up.

### **3. Hypothesized Strengths & Weaknesses of the Frequency Conversion Approach**

The preceding section presented a concrete example of both the strengths and weaknesses of the 2-stage/frequency-conversion procedure for providing higher-frequency market information in small markets. The suggestion is that the advantage for the 2-stage approach over direct (single-stage) high-frequency estimation would lie in the GIE's greater precision (less noise). However, even though the 2-stage procedure is more accurate in theory than direct high-frequency estimation, either procedure may be more accurate in a given specific empirical instance, particularly if the effective increase in sample size is small, which would be the case if the change in frequency is not great. In the empirical analysis in this paper the increase in going from quarterly to annual estimation is a fourfold increase in frequency (doubling of the "square root of N"), and we shall see what sort of results obtain.

While precision is a potential strength of the ATQ, there may be a weakness as well. The preceding examination of the San Francisco office index during the 2007-09 market downturn suggested that perhaps direct single-stage quarterly estimation is better

at capturing the early stages of a sharp downturn in the market. Recall that the directly-estimated index turned down one quarter sooner than the ATQ in the San Francisco office market in 2007. In other words, the hypothesis would be that direct quarterly estimation might show a slight temporal lead ahead of annual estimation (and the resulting ATQ) in such market circumstances. This could result from the effect of loss aversion behavior on the part of property owners during the early stages of a sharp market downturn. Property owners react conservatively, not revising their reservation prices downward (perhaps even ratcheting them upwards, effectively pulling out of the asset market). Unless and until property owners are under pressure to sell in a down market, the result is a sharp drop-off in trading volume.

This has two impacts relevant for transaction price index estimation. First, the relatively few transactions that do clear during the early stages of the downturn reflect relatively positive or eager buyers. This dampens the price reduction actually realized in the market (as reflected in the prices observable in closed transactions). But it does not prevent a directly-estimated high-frequency index from reflecting that market price reduction (such as it is), as best such an index can do so (given the data scarcity, which increases the noise in the index), in the sense that the index does not have a lag bias.

The second effect of the fall-off in sales volume, however, poses a particular issue for annual-frequency indexes as compared to higher-frequency directly-estimated indexes. An annual index reflects an entire 4-quarter span of time in each periodic return, and in the downturn/loss-aversion circumstance just described the most recent part of that 4-quarter time span has markedly fewer transaction observations than the earlier part of the span. Thus, the data used to estimate the annual index's most recent annual return is

dominated by the earlier, pre-downturn sales transactions. Even though the annual index uses Bryon-Colwell-type time-weighted dummy variables (as described previously), the sparser data in the more recent part of the time span may make it difficult for the annual index to fully reflect the recent market movement. Such a difficulty in the annual indexes would then carry over into the quarterly GIE/ATQ indexes derived from them.

#### **4. An Empirical Comparison of Frequency Conversion versus Direct Estimation in Data-Scarce Markets**

The RCA repeat-sales database and the Moody's/REAL Commercial Property Price Indexes based on that data present an opportunity to begin an empirical comparison of the two approaches. As noted, computation of index estimated returns standard errors is not straightforward for the GIE/ATQ, and “apples-to-apples” comparisons of estimated standard errors across the two procedures is not attempted in the present paper.<sup>20</sup>

However, there are two statistical characteristics of an estimated real estate asset market price index that can provide practical, objective information about the quality of the index. These two characteristics are the volatility and the first-order autocorrelation of the index's estimated returns series. Based on statistical considerations, we know that noise or random error in the index return estimates will tend to increase the observed volatility in the index returns and to drive the index returns' first-order autocorrelation down toward negative 50%.<sup>21</sup>

---

<sup>20</sup> For one thing, consider that the second-stage regression itself has no residuals, as it makes a perfect fit to the staggered lower-frequency indexes that are its dependent variable. Furthermore, as noted, the objective of a price index regression is not the minimization of transaction price residuals *per se*, but rather the minimization of error in the coefficient estimates (the index's periodic returns). While bootstrapping or simulation could be employed, the present paper focuses on the empirical analysis to follow.

<sup>21</sup> These are basic characteristics of the statistics of indexes. (See, e.g., Geltner & Miller *et al* (2007), Chapter 25.)

Considering this, it would seem reasonable to compare the two index estimation methodologies based on the volatilities and first-order autocorrelations of the resulting estimated historical indexes. Lower volatility, and higher first-order autocorrelation, would be indicative of an index that is likely to have less noise or error in its individual periodic returns. For example, in the San Francisco office index that we considered previously in Exhibit 3b, the GIE/ATQ index has 4.4% quarterly volatility, versus 6.8% in the directly-estimated CSG index that seemed more noisy.

Among the Moody's/REAL Commercial Property Price Indexes there are 16 indexes (including the San Francisco office index we have previously examined) that are currently published at only the annual frequency (with four staggered versions, as described above), because the available transaction price data is deemed to be insufficient to support quarterly estimation. An examination of the relative values of the quarterly volatilities and first-order autocorrelations resulting from estimation of quarterly indexes by the two alternative procedures across these 16 market segments can provide an interesting empirical comparison of the two procedures in a realistic setting.

The 16 annual-frequency Moody's/REAL indexes include eight at the MSA level and eight at the multi-state regional level. The eight MSA-level indexes are: four different property sectors (apartment, industrial, office, retail) for Southern California (Los Angeles and San Diego combined), three other MSA-level office indexes (New York, Washington DC, and San Francisco), and one other apartment index (for Southern Florida, which combines Miami, Ft Lauderdale, West Palm Beach, Tampa Bay, and

Orlando). The eight multi-state regional indexes include the four property sectors each within each of two NCREIF-defined regions: the East and the South.<sup>22</sup>

-----  
Insert Exhibit 4 about here.  
-----

Exhibit 4 summarizes the comparison of the precision of the two approaches based on a volatility and autocorrelation comparison of the two quarterly index procedures (labeled “FC” and “DirQ” in the exhibit, for “Frequency-Conversion” and “Direct-Quarterly”). The volatility test is defined as the ratio of the FC quarterly volatility divided by the DirQ quarterly volatility. The autocorrelation test is defined by the arithmetic difference between the FC first-order autocorrelation minus that of the DirQ. Both tests are applied separately to the entire 33-quarter available history 2001-2009Q3 and to the more recent 19-quarter period 2005-09Q3. The RCA repeat-sales database “matured” to a considerable degree by 2005, with many more repeat-sales observations available since that time (until the recent crash and liquidity crisis). The comparison is made for each of the 16 indexes and also averaged across the eight MSA-level and eight regional-level indexes.

This comparison indicates that the 2-stage GIE/ATQ frequency conversion approach provided lower volatility and higher 1<sup>st</sup>-order autocorrelation in almost all cases, suggesting that this approach is more precise (less noisy). Of the 64 individual index comparisons (16 indexes X 2 time frames X 2 tests), the GIE/ATQ performed

---

<sup>22</sup> The East Region includes all the 15 states north and east of Georgia, Tennessee, and Ohio. The South Region includes the 9 states encompassed inclusively between and within Florida, Georgia, Tennessee, Arkansas, Oklahoma, and Texas. There is thus some geographical overlap between the MSA-level and regional-level indexes, in the sense that three of the eight MSA-level indexes are also within two of the regional-level indexes. The New York and Washington DC office indexes are within the East Office regional index, and the South-Florida Apartments index is within the South Apartment regional index.

better than the DirQ in all but one case (the AC test for the Southern California Industrial Index during 2005-09).

However, while the frequency conversion procedure seems clearly to be less noisy at the quarterly frequency on the basis of the Exhibit 4 comparison, recall that we raised a possible weak point about the 2-stage procedure in its ability to quickly and fully reflect the early stages of a sudden and sharp market downturn, such as occurred during 2007-09 in the U.S. commercial property markets. We suggested that during such times property-owner loss-aversion behavior could cause the underlying annual-frequency indexes to experience difficulty fully reflecting a late-period price drop in the market.

-----  
Insert Exhibit 5 about here.  
-----

Exhibit 5 presents some empirical evidence relevant to this point from the same 16 Moody's/REAL market-segment indexes examined in Exhibit 4. The exhibit shows the percentage price change from the 2007 peak (within each index) through the most recently available 3Q2009 data as tracked in each market by the frequency-conversion index and the directly-estimated quarterly index. The exhibit also presents two measures of the lead-lag relationship. In the left-hand columns are the calendar quarters of the peak for each index, and in the right-most column is the lead-minus-lag cross-correlation between the two indexes. We see that the direct quarterly index turned down first in six out of the 16 indexes (but only with a one quarter difference in five out of the six cases), while the frequency conversion index beat the direct quarterly in two cases (Washington DC & New York Metro office), with the two methods indicating the same peak quarter in eight cases. In the last column, if the lead-minus-lag cross-correlation is negative, it



indicates that the correlation of the direct-quarterly index with the 2-stage index one quarter later is greater than the correlation of the converse, suggesting that the direct-quarterly index leads the frequency-conversion index. This is the case in 13 out of the 16 indexes, which suggests that the direct-quarterly index does show some tendency to slightly lead the frequency-conversion index in time.

## **5. Conclusion**

This paper has described a methodology for estimating higher frequency (e.g., quarterly) price indexes from staggered lower-frequency (e.g., annual) indexes. The application examined here is to provide higher-frequency information about market movements in data-scarce environments that otherwise require low-frequency indexes. The proposed frequency-conversion approach takes advantage of the lower frequency to, in effect, accumulate more data over the longer-interval time periods which can be used to estimate returns with less error. Then it applies the Moore-Penrose pseudoinverse matrix in a second-stage operation in which the staggered low-frequency indexes are converted into a higher-frequency index. Linear algebra theory establishes that this frequency conversion procedure exactly matches the lower-frequency index returns and is optimal in the sense that it minimizes any variance or bias added in the second stage. Numerical simulation and empirical comparisons described here confirm that the two-stage frequency-conversion technique results in less noise than direct high-frequency estimation in realistic annual-to-quarterly indexes for practical U.S. commercial property price indexes such as the Moody's/REAL CPPI annual indexes (e.g., situations with second-sales observational frequency averaging in the mid-20s or less per quarter). The

result is higher-frequency indexes that, while they have signal/noise ratios lower than the underlying low-frequency indexes, nevertheless add higher frequency information that may be useful in the marketplace, especially in the context of tradable derivatives. The only major drawback is that the frequency-conversion procedure may tend to slightly lag behind direct quarterly estimation, particularly during the early stage of a sharp market downturn. The lag appears to generally be no more than one quarter.

Finally, we would propose two strands of possibly productive directions for future research. First, throughout this paper no consideration was given to the covariance structure among the observations. Thus, more efficient estimators may exist if reasonable distribution assumptions were made and accounted for in the estimation of the high frequency series. Second, exploration of approaches that employ multiple imputation techniques in a Bayesian framework or a Markov Chain Monte Carlo context might lead to a better way of estimating high frequency indexes in a data scarce environment and in quantifying the noise that remains in the resulting indexes. With this in mind the current paper presents only a first step which may be improved upon by subsequent researchers, but which in itself appears to already have some practical value.

Appendix C presents charts of the GIE/ATQ and direct-quarterly indexes for all 16 annual-frequency Moody's/REAL Indexes.

## Appendix A: The Moore-Penrose Pseudoinverse or the Generalized Inverse

The Moore-Penrose pseudoinverse is a general way of solving the following system of linear equations:

$$y = \mathbf{X}b, \quad y \in \mathbf{R}^n; b \in \mathbf{R}^k; \mathbf{X} \in \mathbf{R}^{n \times k} \quad (1)$$

It can be shown that there is a general solution to these equations of the form:

$$b = \mathbf{X}^\dagger y \quad (2)$$

The  $\mathbf{X}^\dagger$  matrix is the unique Moore-Penrose pseudoinverse of  $\mathbf{X}$  that satisfies the following properties:

1.  $\mathbf{X} \mathbf{X}^\dagger \mathbf{X} = \mathbf{X}$  ( $\mathbf{X} \mathbf{X}^\dagger$  is not necessarily the identity matrix)
2.  $\mathbf{X}^\dagger \mathbf{X} \mathbf{X}^\dagger = \mathbf{X}^\dagger$
3.  $(\mathbf{X} \mathbf{X}^\dagger)^T = \mathbf{X} \mathbf{X}^\dagger$  ( $\mathbf{X} \mathbf{X}^\dagger$  is Hermitian)
4.  $(\mathbf{X}^\dagger \mathbf{X})^T = \mathbf{X}^\dagger \mathbf{X}$  ( $\mathbf{X}^\dagger \mathbf{X}$  is also Hermitian)

The solution given by equation (2) is a minimum norm least squares solution. When  $\mathbf{X}$  is of full rank (i.e., rank is at most  $\min(n, k)$ ), the generalized inverse can be calculated as follows:

- Case 1: When  $n = k$  (same number of equations as unknowns) :  $\mathbf{X}^\dagger = \mathbf{X}^{-1}$   
Case 2: When  $n < k$  (fewer equations than unknowns) :  $\mathbf{X}^\dagger = \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1}$   
Case 3: When  $n > k$  (more equations than unknowns) :  $\mathbf{X}^\dagger = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$

In the application for deriving higher frequency indexes from staggered lower frequency indexes, Case 2 provides the relevant calculation. Furthermore, it should be noted that when the rank of  $\mathbf{X}$  is less than  $k$ , no unbiased linear estimator,  $b$ , exists. However, for such a case, the generalized inverse provides a minimum bias estimation.<sup>23</sup> For the basic references on the Moore-Penrose pseudoinverse see the references by Penrose (1955, 1956), Chipman (1964), and Albert (1972) in the bibliography.

---

<sup>23</sup> Properties of the generalized inverse can be found in Penrose (1954) and equation (2) first appeared in Penrose (1956). Proofs of Cases 1 – 3 can be found in Albert (1972) and a proof of minimum biasedness is given in Chipman (1964).

## Appendix B:

### A Note on the Bias in the Generalized Inverse Estimator (GIE)

Here we consider the case relevant to our present purposes, i.e. where  $\mathbf{X}^\dagger = \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1}$ . Therefore, in our application, the solution (or estimation) of the second-stage regression (equation (2) of Appendix A) can be re-written as:

$$\mathbf{b} = \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{y}$$

Considering that the true value of the predicted variable ( $\mathbf{y}$ ) is by definition:  $\mathbf{X}\mathbf{b}_{\text{True}}$ , therefore the expected value of  $\mathbf{b}$  is:

$$E[\mathbf{b} | \mathbf{X}] = \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{b}_{\text{True}}$$

Let  $\mathbf{R} = \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X}$  be the “resolution” matrix, which would have otherwise been the  $k$  by  $k$  identity ( $\mathbf{I}$ ) matrix if  $\mathbf{X}$  had been of full column rank. In our case, the resolution matrix is instead a symmetric matrix describing how the generalized inverse solution “smears” out the  $\mathbf{b}_{\text{True}}$  into a recovered vector  $\mathbf{b}$ . The bias in the generalized inverse solution is

$$\begin{aligned} E[\mathbf{b} | \mathbf{X}] - \mathbf{b}_{\text{True}} &= \mathbf{R} \mathbf{b}_{\text{True}} - \mathbf{b}_{\text{True}} \\ &= (\mathbf{R} - \mathbf{I}) \mathbf{b}_{\text{True}} \end{aligned}$$

We can formulate a bound on the norm of the bias:

$$\| E[\mathbf{b} | \mathbf{X}] - \mathbf{b}_{\text{True}} \| \leq \| \mathbf{R} - \mathbf{I} \| \| \mathbf{b}_{\text{True}} \|$$

Computing  $\| \mathbf{R} - \mathbf{I} \|$  can give us an idea of how much bias has been introduced by the generalized inverse solution. However, the bound is not very useful since we typically have no knowledge of  $\| \mathbf{b}_{\text{True}} \|$ .

In practice, we can use the resolution matrix,  $\mathbf{R}$ , for two purposes. First, we can examine the diagonal elements of  $\mathbf{R}$ . Diagonal elements that are close to one correspond to coefficients for which we can expect good resolution. Conversely, if any of the diagonal elements are small, then the corresponding coefficients will be poorly resolved.

For the particular data matrix used in this study, i.e  $\mathbf{X}$  is Toeplitz, the diagonal elements of  $\mathbf{R}$  approach one very fast. For instance, for the annual to quarterly conversion, a 24 by 27 matrix (24 observations, 27 quarterly return estimates), the diagonal elements of  $\mathbf{R}$  have a value of 0.89. For a 50 by 53 matrix, the diagonal elements have a value of 0.94. By induction, as the number of periods to be estimated ( $T$ ) go to infinity, and the percentage difference between  $T$  and  $T-3$  becomes negligible, the diagonal elements of  $\mathbf{R}$  approach a value of 1. Hence, the bias goes to zero as the system gets closer to being effectively identified.

## **Acknowledgments**

The authors thank Henry Pollakowski and Jungsoo Park for advice and assistance, and Real Estate Analytics LLC (REAL) and Real Capital Analytics Inc (RCA) for both data and financial support. Note: Commercial use of the index construction methodology described in this paper has been licensed by MIT exclusively to REAL.

## Bibliography

Albert, A. (1972). *Regression and the Moore-Penrose Pseudoinverse*. Academic Press, New York.

Aster, R.C, Borchers, B. and Thurber, C. H. (2005). *Parameter Estimation and Inverse Problems*. Elsevier Academic Press

T.Bryan & P.Colwell, "Housing Price Indices", in C.F.Sirmans (ed.), *Research in Real Estate*, vol.2. Greenwich, CT: JAI Press, 1982.

Chipman, J. (1964). On Least Squares with Insufficient Observations. *Journal of the American Statistical Association* 59, No. 208, 1078-1111

K.Case & R.Shiller "Prices of Single Family Homes Since 1970: New Indexes for Four Cities." *New England Economic Review*: 45–56, Sept./Oct. 1987.

J.Clapp, "A Semiparametric Method for Estimating Local House Price Indices", *Real Estate Economics* 32(1): 127-160, 2004.

M.Francke, "Repeat Sales Index for Thin Markets: A Structural Time Series Approach", *Journal of Real Estate Finance & Economics*: forthcoming (Maastricht Special Issue), 2009 (or 2010).

D.Geltner, "Bias & Precision of Estimates of Housing Investment Risk Based on Repeat-Sales Indexes: A Simulation Analysis", *Journal of Real Estate Finance & Economics*, 14(1/2): 155-172, January/March 1997.

D.Geltner & D.Ling, "Considerations in the Design & Construction of Investment Real Estate Research Indices", *Journal of Real Estate Research*, 28(4):411-444, 2006.

D.Geltner, N.Miller, J.Clayton, & P.Eichholtz, *Commercial Real Estate Analysis & Investments* 2<sup>nd</sup> Edition, South-Western College Publishing Co/Cengage Learning, Cincinnati, 2007.

W.Goetzmann, "The Accuracy of Real Estate Indices: Repeat-Sale Estimators", *Journal of Real Estate Finance & Economics* 5(1): 5-54, 1992

D.McMillen & J.Dombrow, "A Flexible Fourier Approach to Repeat Sales Price Indexes", *Real Estate Economics* 29(2): 207-225, 2001.

Pace, RK & JP LeSage, "Spatial Statistics and Real Estate", *Journal of Real Estate Finance and Economics*, 29(2):147-148, 2004.

Pavia, JM & Bernardi Cabrer (2008), "On distributing quarterly national growth among regions", *Environment and Planning A*, 1 - 16

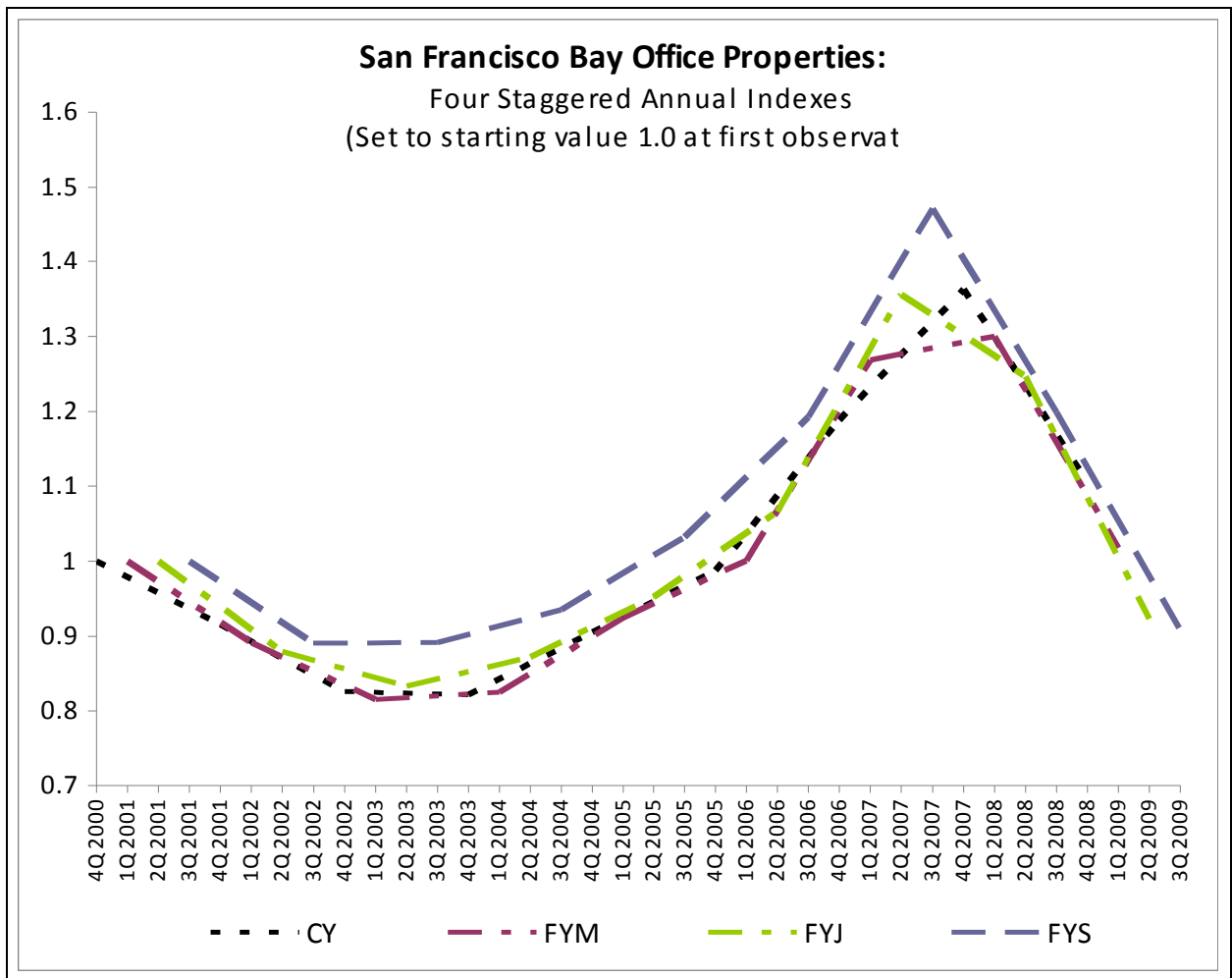
Penrose, R. (1955), "A Generalized Inverse for Matrices", *Proceedings of the Cambridge Philosophical Society* 51, 406-413

Penrose, R. (1956), "On best approximate solutions of linear matrix equations", *Proceedings of the Cambridge Philosophical Society* 51, 17-19

Schwann, G. (1998), "A Real Estate Price Index for Thin Markets", *Journal of Real Estate Finance & Economics* 16(3): 269-287.

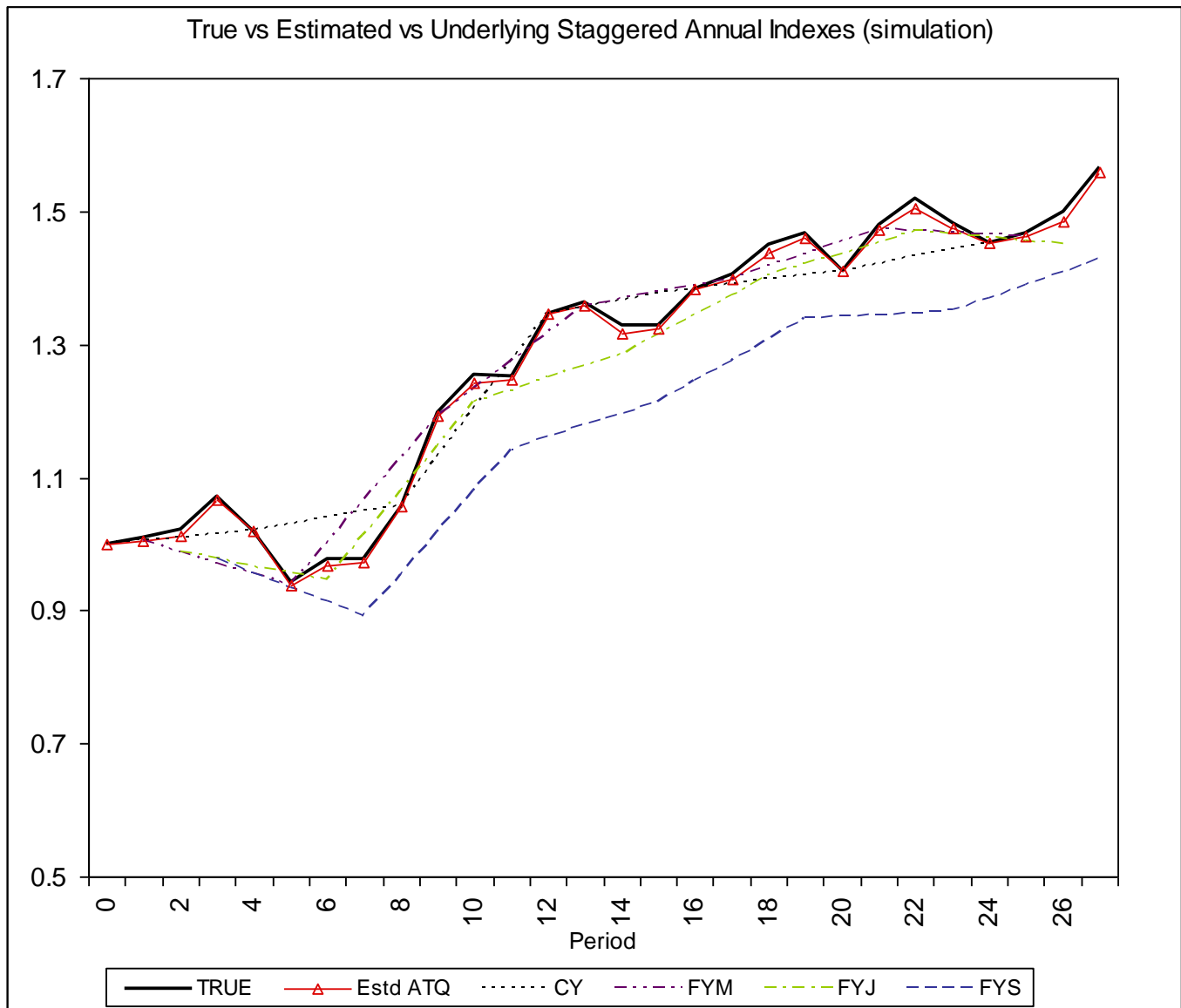
## Exhibits

**Exhibit 1:** (CY is the calendar year index ending December 31 each year, FYM is the index ending March 31 each year, FYJ is the index ending June 30 each year, and FYS is the index ending September 30 each year.) Indexes based on Case-Shiller/Goetzmann (CSG) method (3-stage WLS repeat-sales with Byron-Colwell time-weighted dummy variables and Goetzmann Bayesian ridge procedure).

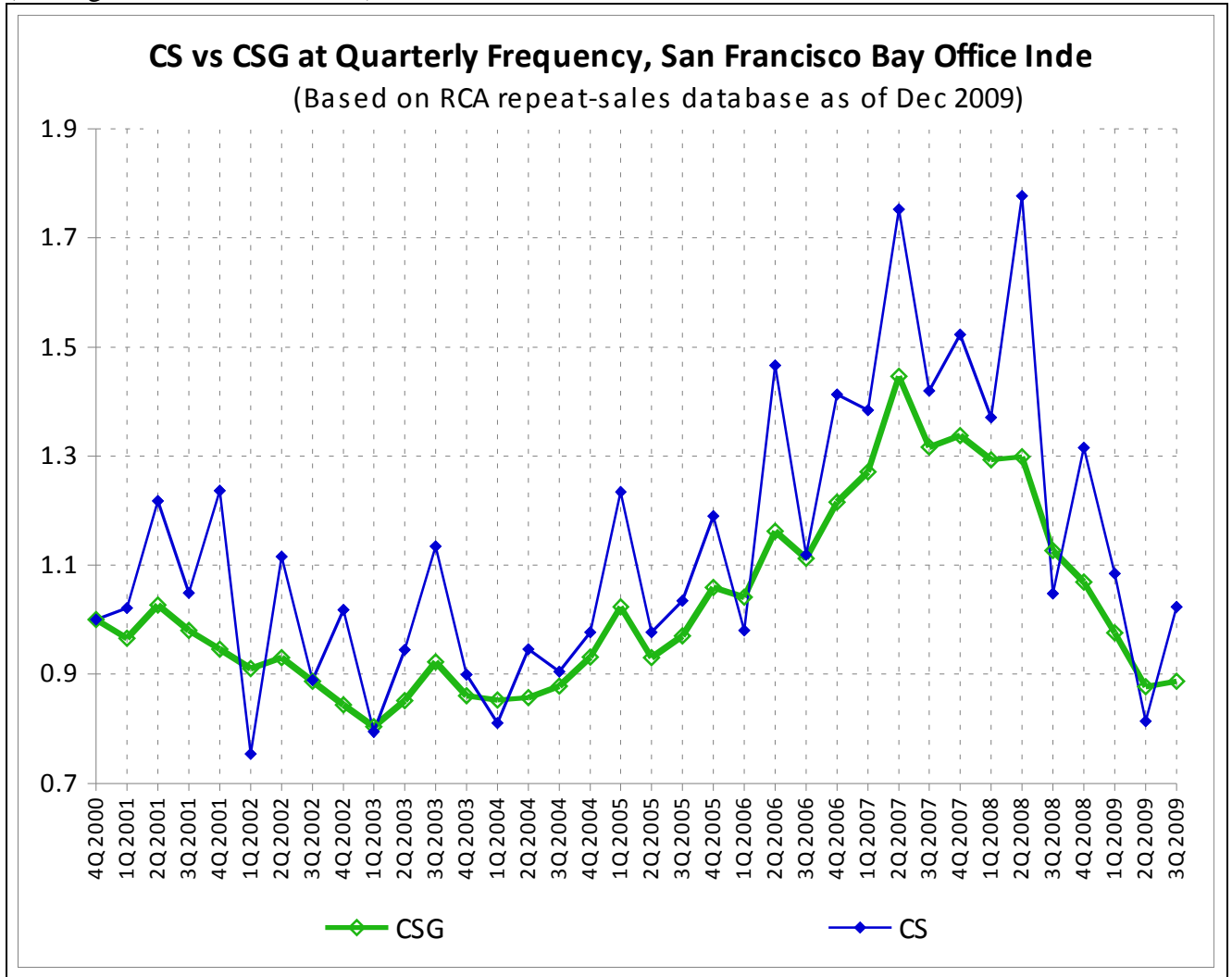




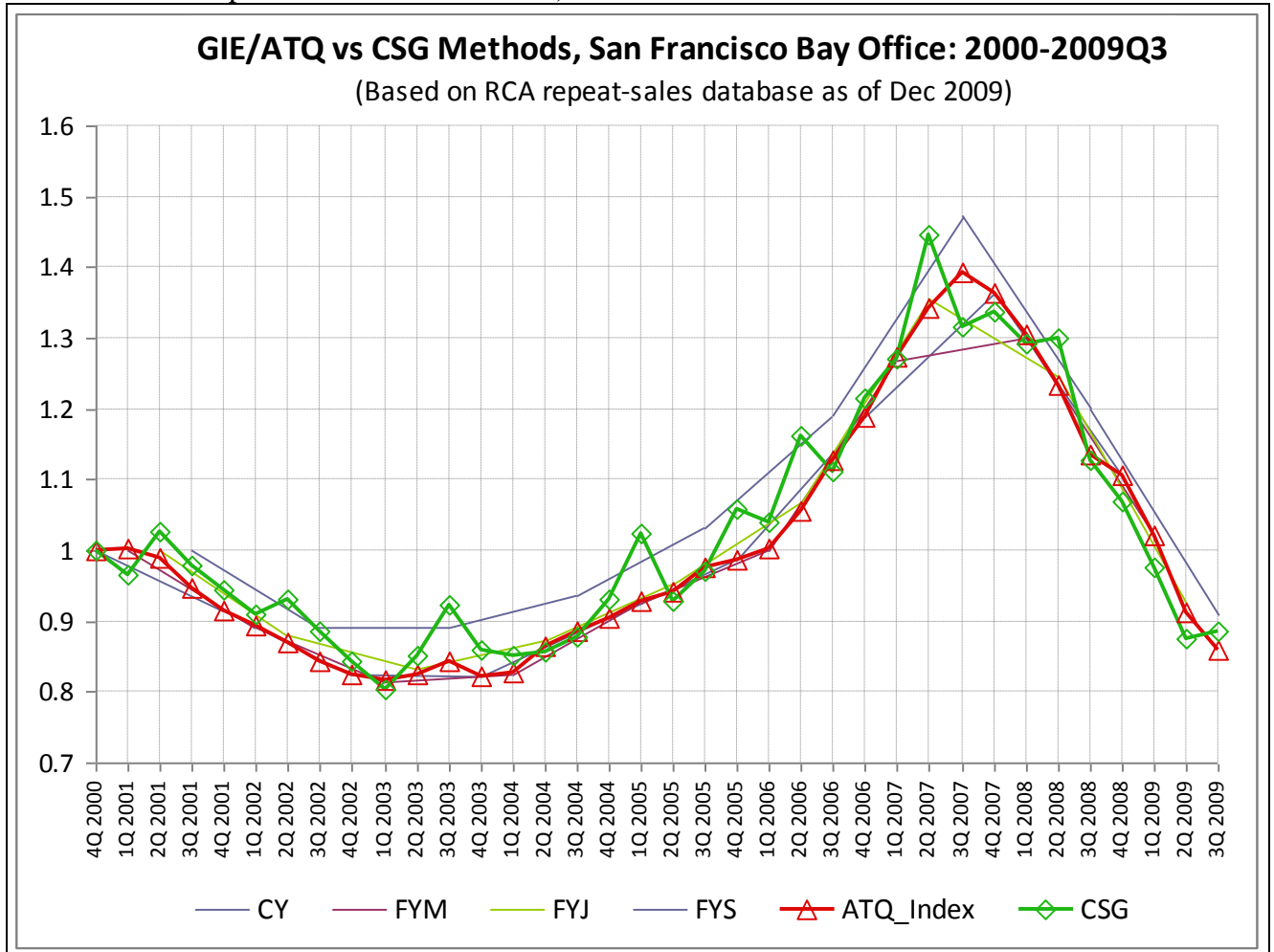
**Exhibit 2:** Simulation of True vs Estimated Quarterly Price Indexes: Annual to quarterly (ATQ) GIE-based frequency conversion.



**Exhibit 3a:** San Francisco office property price indexes based on RCA repeat-sales database, direct quarterly estimation using: (i) Case-Shiller 3-stage WLS estimation (CS, blue solid diamonds), versus Case-Shiller enhanced with Goetzmann Bayesian procedure (CSG, green hollow diamonds)



**Exhibit 3b:** San Francisco office property price index based on RCA repeat-sales database, CSG direct quarterly estimation (green diamonds) and derived GIE/ATQ estimation (red triangles), together with the staggered annual indexes in the background: 2001Q1-2009Q3. (Note: Staggered annual indexes underlying the GIE/ATQ are based on the CSG methodology at the annual frequency, and all are set to arbitrary starting value of 1.0 at their inception dates as in Exhibit 1.)



**Exhibit 4:** Comparison based on Volatility and Autocorrelation Tests of 16 quarterly indexes in data-scarce markets: Frequency Conversion (GIE/ATQ) versus Direct Quarterly (CSG) estimation.

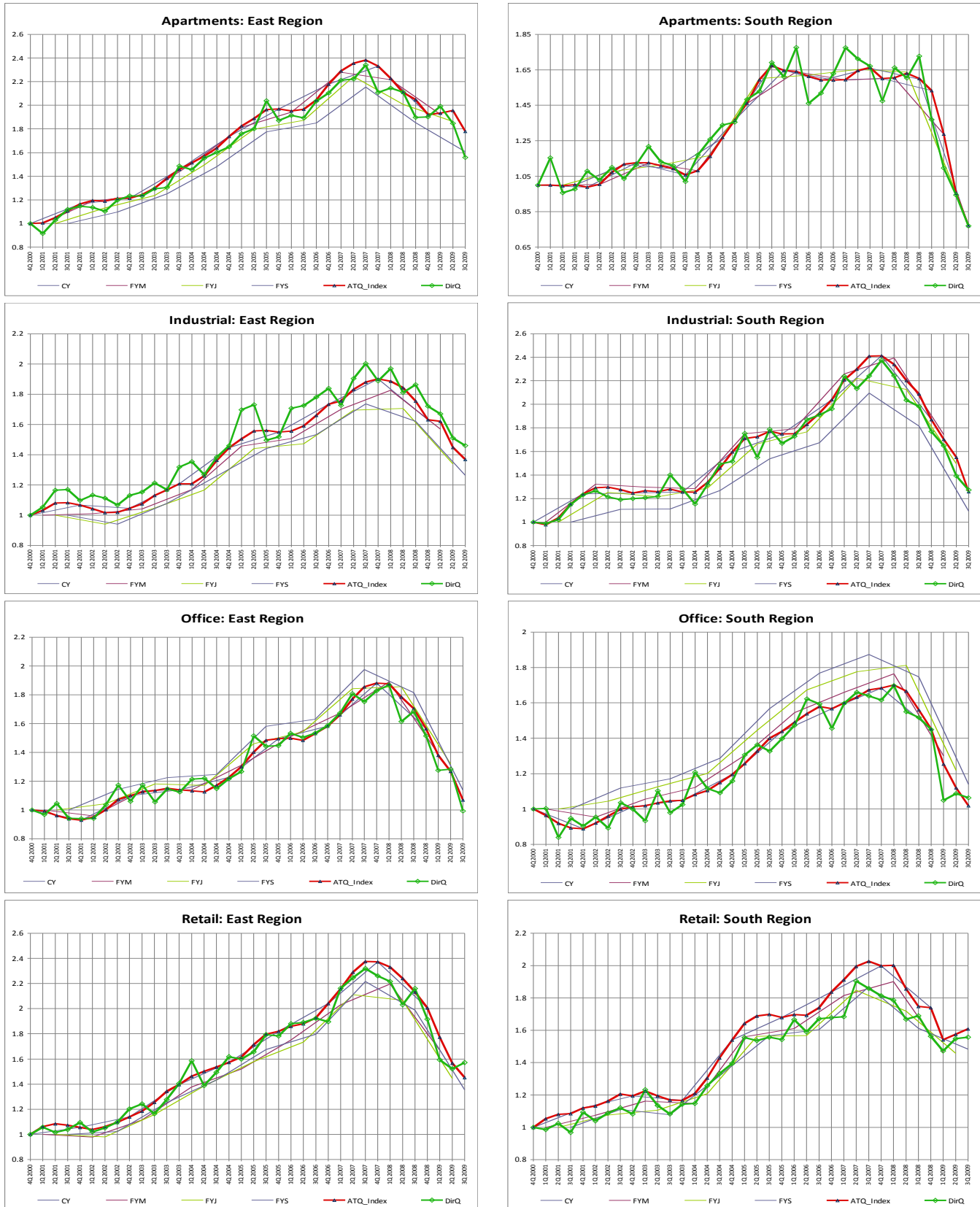
Index Comparison: Frequency-Conversion (FC) vs Direct-Quarterly (DirQ)					
Estimation: Volatility & Autocorrelation Evidence on Index Precision...					
	Data*:	Vol Test**:		AC(1) Test***:	
MSA-level indexes:					
Index:		2005-09	2001-09	2005-09	2001-09
NY Office	20 (4)	0.55	0.47	80%	98%
DC Office	14 (5)	0.69	0.57	71%	90%
SF Office	12 (2)	0.67	0.64	91%	82%
SC Office	27 (7)	0.63	0.58	77%	92%
SC Industrial	30 (15)	0.59	0.40	-48%	77%
SC Retail	16 (5)	0.61	0.55	60%	85%
SC Apts	43 (16)	0.64	0.50	88%	109%
FL Apts	18 (7)	0.62	0.62	76%	77%
Average:	23 (8)	0.63	0.54	62%	89%
Regional-level indexes:					
Index:		2005-09	2001-09	2005-09	2001-09
E Office	55 (19)	0.69	0.60	87%	103%
S Office	39 (13)	0.65	0.49	86%	113%
E Industrial	37 (14)	0.54	0.56	67%	69%
S Industrial	30 (8)	0.75	0.73	91%	77%
E Retail	29 (14)	0.81	0.67	67%	83%
S Retail	38 (9)	0.78	0.69	50%	78%
E Apts	56 (21)	0.59	0.56	62%	74%
S Apts	58 (23)	0.75	0.68	71%	83%
Average:	43 (15)	0.70	0.62	73%	85%
*Avg number of 2nd-sales obs/qtr 2006-09. Database was “immature” with considerably fewer 2 <sup>nd</sup> -sales observations prior to 2005. (In parentheses number of obs in most recent 3Q09 qtr.)					
** Ratio of FC volatility/DirQ volatility: <1 ==> <b>FC better</b> ; >1 ==> <i>DirQ better</i> .					
*** Difference: AC(1)FC - AC(1)DirQ. >0 ==> <b>FC better</b> ; <0 ==> <i>DirQ better</i> .					

**Exhibit 5:** Comparison of 16 quarterly indexes in data-scarce markets: Frequency Conversion (GIE/ATQ) versus Direct Quarterly (CSG) estimation, based on 2007-09 Downturn Magnitude & Lead Minus Lag Cross-Correlation.

<b>Index Comparison: Frequency-Conversion (FC) vs Direct-Quarterly (DirQ) Estimation: Capturing 2007-09 Downturn, &amp; Lead-Lag Evidence.</b>					
	2007 Peak*:		Downturn Magnitude: Peak-3Q09:		Lead – Lag Correl**:
Index:	FC	DirQ	FC	DirQ	
NY Office	<b>3Q07</b>	<b>4Q07</b>	-43.5%	-53.7%	<b>8%</b>
DC Office	<b>4Q07</b>	<b>1Q08</b>	-24.4%	-26.3%	<b>6%</b>
SF Office	<i>3Q07</i>	<i>2Q07</i>	-38.3%	-38.7%	<i>-7%</i>
SC Office	<i>4Q07</i>	<i>3Q07</i>	-39.5%	-40.9%	<i>-22%</i>
SC Industrial	4Q07	4Q07	-31.3%	-35.9%	<i>-32%</i>
SC Retail	<i>4Q07</i>	<i>3Q07</i>	-33.7%	-34.9%	<i>-32%</i>
SC Apts	3Q07	3Q07	-26.0%	-27.8%	<i>-17%</i>
FL Apts	3Q07	3Q07	-53.4%	-56.2%	<i>-55%</i>
Average:			-36.3%	-39.3%	<i>-19%</i>
Index:	FC	DirQ	FC	DirQ	Lead-Lag
E Office	1Q08	1Q08	-43.0%	-46.9%	<b>0%</b>
S Office	1Q08	1Q08	-40.1%	-37.3%	<i>-31%</i>
E Industrial	<i>4Q07</i>	<i>3Q07</i>	-28.0%	-27.2%	<i>-26%</i>
S Industrial	3Q07	3Q07	-47.9%	-43.2%	<i>-8%</i>
E Retail	3Q07	3Q07	-38.9%	-32.2%	<i>-21%</i>
S Retail	<i>3Q07</i>	<i>2Q07</i>	-20.6%	-18.2%	<i>-47%</i>
E Apts	3Q07	3Q07	-25.3%	-33.4%	<i>-41%</i>
S Apts	<i>3Q07</i>	<i>1Q07</i>	-53.6%	-56.6%	<i>-22%</i>
Average:			-37.2%	-36.9%	<i>-24%</i>
*Peak qtr before 2007 downturn. Sooner is presumably better, so: <b>FC earlier → FC leads; DirQ earlier → DirQ leads.</b>					
** Difference: Correl(FC(t),DirQ(t+1)) – Correl(DirQ(t),FC(t+1)): <b>Positive ==&gt; FC leads; Negative ==&gt; DirQ leads.</b>					

## Appendix C;

Charts of all 16 Moody's/REAL Annual Index Markets Showing Frequency Conversion (GIE/ATQ, red triangles) Index & single-stage CSG (here labeled "DirQ", green diamonds) Index, and underlying staggered annual-frequency CSG indexes



## Appendix C (cont.)

