The Predictability of House Prices

Authors Anthony Y. Gu

Abstract

The level and direction of autocorrelation in house price movements differ across areas and change over time. This finding reconciles the conflicting reports in the literature. When quarterly house price indices exhibit negative autocorrelation, autocorrelation shows a positive connection to volatility and a negative connection to rate of return. Autocorrelation between longer time periods is mainly positive; it exhibits a negative relationship with volatility and a positive relationship with rate of return. Volatile house price indices tend to have lower rates of return. It would be possible to obtain excess returns by following a trading strategy based on the estimated autocorrelation.

Introduction

Urban and financial economists generally think that the house market ought to be less efficient than financial markets, because of the unstandardized commodities, implicit rents, transaction costs, carrying costs and taxes. However, it has been difficult for investors to exploit these inefficiencies toward earning excess returns, because house price change has been difficult to estimate. Finding the patterns of house price movements would provide a useful reference for professional real estate investors, and also for mortgage bankers and homeowners.

A few empirical works have tested the efficiency of house markets. The first rigorous testing is published in two papers by Gau (1984, 1985). His examination of commercial real estate in the Vancouver area for the years 1971–1980 concludes that the price changes follow a random walk. Linneman (1986) studies individual homeowners' assessments of their house values in Philadelphia for the years 1975 and 1978. He found that houses tended subsequently to increase in value if they were undervalued relative to a 1975 hedonic regression, but—due to transaction costs—only an insignificant number of units appear to have exhibited profitable arbitrage opportunities. Case and Shiller (1989, 1990) find evidence of positive autocorrelation in real house prices, performing weak and strong form efficiency tests on weighted repeated sales price data for house in Atlanta, Chicago, Dallas and San Francisco during the 1970–1986 period. They point out that the constructed return series is contaminated with errors, and that the estimator will be biased and inconsistent with the noise. That is, if a data set includes only

I_{11}, I_{13}, I_{22} and I_{23}, where I_{it} denote the log price of house i at time t, then the estimated first period return is $(I_{13} - I_{11}) - (I_{23} - I_{22})$, and the second period return is $(I_{23} - I_{22})$. Hence, the first period return is negatively correlated to the second period return. Case and Shiller also show that the contamination can result in positive correlation. They try to solve this problem by proposing a data partition method. Kuo (1996) points out that Case and Shiller's estimator is still not consistent, and that it involves an arbitrary partition of the data set; Kuo tries to solve this problem with a two-step, two-sample method and a Bayesian method. He estimates the autocorrelation and seasonality of house price changes using Case and Shiller's data set, and reports positive autocorrelation as well as some weak seasonality. In comparing the empirical results, he concludes that the estimates are sensitive to the estimation methods used.

Yet the conflicting findings of the authors may not be due to superior or inferior methodologies. The researchers' use of data sets that differ by property type, area and time: may explain why Gau finds randomness, while Case and Shiller and Kuo find positive autocorrelation.

Consider insights provided by research on the stock market, though many previous studies have found significant autocorrelation in stock returns, none has reported that investors could earn excess returns based on estimated autocorrelation. They typically cite transaction costs as the main obstacle. The problem may be, however, that the actual autocorrelation changes in a systematic or random manner over time. If actual autocorrelation were to change over time, then the estimated autocorrelation would not offer adequate information about dependency in price movements. If the changes were random, then the market would still be weak-form efficient, because investors would be unable to forecast the market using the estimated autocorrelation.

This study tests autocorrelation in house price movements across the entire United States, in different time periods, using the largest existing data set. The conflicting results reported by previous researchers might all be valid if the same method reveals different house price behavior in different locations and different time periods. Also examined is the impact that autocorrelation and volatility have on rate of return. A final aspect of this study is an exploration of the possibility of earning abnormal returns based on the estimated autocorrelation.

The data are briefly described in next section, and an explanation of the methodology follows. Next, the estimated autocorrelation and the resulting implications are presented, followed by a description of the regression analysis and an explanation of the results. The final section is the conclusion.

The Data

The data set used in this study is Freddie Mac's Conventional Mortgage Home Price Index (CMHPI). The index is estimated using such a large sample that the problem of spurious autocorrelation, as described in the Case and Shiller (1989,

1990) and Kuo (1996) studies, should not exist; the estimators should be consistent. The index provides a measure of typical price inflation for houses within the U.S. The CMHPI, which was originally jointly developed by Freddie Mac and Fannie Mae, is based on an ever-expanding combined database that currently contains more than twelve million entries. The index is computed based on conventional mortgages that were purchased or securitized by Freddie Mac or Fannie Mae. The included mortgages are for single unit residential houses only. They are "conforming," in that at the time of purchase they met Freddie Mac or Fannie Mae underwriting standards, and did not exceed the allowable loan limits set for the two companies. For example, the 2001 loan limit was $275,000 except for mortgages originated in Alaska, Hawaii, Guam and the U.S. Virgin Islands, where it was $412,500. High priced properties therefore are under-represented, potentially distorting the results in more expensive house markets.[1]

As described by Freddie Mac, the CMHPI uses a statistical method based entirely on "repeat transactions." Any time a home's value—based on either a sale or an appraisal—can be observed twice, the measured change becomes one observation of house price growth over the indicated time period. The index is defined to be the statistically determined set of values that most closely fits many such repeated observations. In other words, the index involves complicated weighted averages of all the house price growth observations. Since the index is limited to properties traded at least twice, it is biased toward price trends for the types of houses that are traded most frequently.[2]

The data set includes quarterly house price indices from the first quarter of 1975 to the first quarter of 1999. There are indices for all fifty states and the District of Columbia, plus separate indices for nine Census divisions and an aggregate index for the nation as a whole. Further details regarding the building of the indices are described by Wang and Zorn (1999).

Methodology

The most commonly used class of time-series model to describe temporary deviations about a trend is:

$$I_t = \psi t + \sum_{j=0}^{\infty} \pi_j \varepsilon_{t-j}. \tag{1}$$

Where I_t represents the natural logarithm of house price index, ψ_t describes the trend and ε_t is a random disturbance. The model is not structural: it should be thought of as a way of capturing the dynamic behavior of I. If $\Sigma \pi_j \varepsilon_{t-j}$ is a stationary stochastic process, then fluctuations in I_t are temporary, and I_t is called "trend stationary." The π_j must approach zero for large j if $\Sigma \pi_j\ \varepsilon_{i-j}$ is to be stationary. Hence, a decline in an index below trend today has no effect on forecasts of the index value $E_t\,(I_{t+j})$ for the distant future; this relationship implies that growth rates for the index must rise above their historical average for several periods until the trend line is reestablished.

Equation (2) represents a simple time-series model that captures permanent fluctuations in a house price index:

$$I_t = \mu + I_{t-1} + \varepsilon_t. \tag{2}$$

The process described by this equation is a random walk with drift.

The size of a random walk component in a house price index can be measured from the variance of its log differences. If the natural logarithm of the house price index I_t is a pure random walk (Equation 2), then the variance of its q-differences grows linearly with the difference q: $\text{var}(I_t - I_{t-q}) = q\sigma^2$. On the other hand, if the natural logarithm of the house price index is stationary about a trend (Equation 1), then the variance of its q-differences approaches a constant, which is twice the unconditional variance of the series: $\text{var}(I_t - I_{t-q}) \rightarrow 2\sigma_I^2$. The variance ratio $(1/q)$*$\text{var}(I_t - I_{t-q})$, as a function of q, should be constant at σ^2 if I_t is a random walk. This ratio should decline toward zero if I_t is trend-stationary and negatively correlated; it should increase with q if I_t is not trend stationary and positively correlated. If fluctuations in an index are partly permanent and partly temporary—a condition that can be modeled as a combination of a stationary series and a random walk—then the variance ratio should settle down to the variance of the shock to the random walk component.

The variance ratio of q-differences can display loosely structured reversion, whereas many other approaches cannot. The fluctuations in a house price index are partly temporary, such that the random walk component is relatively small and a shock today will be partially reversed in the long run. This reversal is likely to be slow, loosely structured and not easily captured in a simple parametric model.

A time series following a random walk should exhibit a unit variance ratio. A variance ratio that is greater than unity indicates that the variances grow more than proportionally with time, and suggests positive autocorrelation; a variance ratio that is smaller than unity indicates that the variances grow less than proportionally with time, and suggests negative autocorrelation. The variance-ratio $VR(q)$ is defined[3] as:

$$VR(q) = \frac{\sigma^2(q)}{\sigma^2(1)}, \tag{3}$$

where $\sigma^2(q)$ is $1/q$ the variance of the q-differences and $\sigma^2(1)$ is the variance of the first differences. Furthermore:

$$\sigma^2(q) = \frac{1}{m} \sum_{t=q}^{nq} (I_t - I_{t-q} - q\hat{\mu})^2, \qquad (4)$$

where:

$$m = q(nq - q + 1)\left(1 - \frac{q}{nq}\right).$$

and

$$\sigma^2(1) = \frac{1}{(nq - 1)} \sum_{t=1}^{nq} (I_t - I_{t-1} - \hat{\mu})^2, \qquad (5)$$

where:

$$\hat{\mu} = \frac{1}{nq}(I_{nq} - I_0).$$

I_0 and I_{nq} are the first and last observations of the time series. The asymptotic standard normal test statistic for the variance-ratio under the hypothesis of homoscedasticity is:

$$z(q) = \frac{VR(q) - 1}{[\phi(q)]^{0.5}} \sim N(0, 1), \qquad (6)$$

where:

$$\phi(q) = \frac{2(2q - 1)(q - 1)}{3q(nq)}.$$

The asymptotic standard normal test statistic for the heteroscedasticity-consistent estimator is:

$$z^*(q) = \frac{VR(q) - 1}{[\phi^*(q)]^{0.5}} \sim N(0, 1), \qquad (7)$$

where:

$$\phi^*(q) = \sum_{j=1}^{q-1} \left[\frac{2(q - j)}{q}\right]^2 \hat{\delta}(j)$$

and

$$\hat{\delta}(j) = \frac{\Sigma(I_t - I_{t-1} - \hat{\mu})^2(I_{t-j} - I_{t-j-1} - \hat{\mu})^2}{[\Sigma(I_t - I_{t-1} - \hat{\mu})^2]^2}.$$

Estimation Results and Their Implications

The results of the heteroscedasticity-robust variance ratio test with the statistic $Z^*(q)$ are reported in Exhibit 1. For $q = 2$, all the indices exhibit negative autocorrelation. This autocorrelation is significant at the 5% level for forty-six of the fifty-one ''areas'' (the fifty states and Washington, D.C.), for six of the nine ''regions,'' and for the U.S. as a whole. This finding indicates that, in most of the states, quarterly house price fluctuations are not random: a price increase in one quarter is likely to be followed by a price decrease in another, and vice versa. A possible economic explanation is the existence of seasonality.[4] It should be noted that Case and Shiller (1987), and Kuo (1996), report higher second quarter returns in the house markets.

As q increases, more indices show positive autocorrelation, and the difference across states becomes more apparent. Positive autocorrelation is exhibited in twenty-four of the fifty-one area indices when $q = 4$, in thirty-four of the area indices when $q = 8$, and in thirty-eight of the area indices when $q = 16$. It is exhibited in eight of the nine regional indices when $q = 4$ and 8, and in all nine regional indices when $q = 16$. The index for the U.S. as a whole shows positive autocorrelation when $q = 4$, 8 and 16.

For $q = 4$, few of the fifty-one areas' variance ratios are significantly different from unity. Only two of the twenty-four greater-than-one ratios, and only five of the twenty-seven smaller-than-one ratios, are significant. This outcome reveals that, for intervals of approximately twelve months, house price changes are not significantly correlated, or are random, in most of the states. For $q = 8$, thirty of the thirty-four greater-than-one ratios are significant and all the smaller-than-one ratios are not significant. This evidence shows that, between intervals of about 24 months, an increase in house price tends to be followed by another increase. For $q = 16$, thirty-five of the thirty-eight greater-than-one ratios are significant and all the smaller-than-one ratios are not significant. This pattern further shows that between long-term intervals, an increase in house price tends to be followed by another increase.

Indices for twenty-four states exhibit positive autocorrelation for all differences of $q > 2$. Among them, California (CA) and Massachusetts (MA) reveal the most positive autocorrelation. Indices for thirteen states exhibit negative autocorrelation for all the differences ($q = 2, 4, 8, 16$); among them, Alabama (AL) and South

Exhibit 1 | The Estimated Variance Ratios and Heteroscedasticity Consistent *Z*-Values

State / Region	*q*(2)*VR*	*Z**(*q*)	*q*(4)*VR*	*Z**(*q*)	*q*(8)*VR*	*Z**(*q*)	*q*(16)*VR*	*Z**(*q*)
Alaska	0.43	−4.65	0.87	−0.60	0.75	−0.76	1.02	0.04
Alabama	0.26	−5.11	0.30	−2.76	0.31	−1.78	0.35	−1.16
Arkansas	0.35	−3.45	0.72	−0.86	1.01	0.02	1.58	0.80
Arizona	0.57	−3.75	1.39	1.98	2.31	4.30	3.64	5.91
California	0.94	−0.42	3.63	10.67	6.80	15.39	10.89	17.93
Colorado	0.61	−2.71	1.93	3.71	3.70	7.04	6.37	9.57
Connecticut	0.73	−2.20	2.47	6.78	4.52	10.65	7.85	14.15
Washington DC	0.34	−4.14	0.86	−0.50	1.42	1.00	2.17	1.89
Delaware	0.32	−4.07	0.84	−0.54	1.10	0.24	1.90	1.39
Florida	0.23	−2.70	0.41	−1.19	0.58	−0.56	0.81	−0.17
Georgia	0.42	−3.11	0.88	−0.36	1.20	0.41	1.90	1.25
Hawaii	0.24	−1.80	0.41	−0.81	0.43	−0.51	0.51	−0.30
Iowa	0.33	−2.18	0.74	−0.49	1.01	0.01	1.29	0.25
Idaho	0.33	−5.05	0.59	−1.79	0.82	−0.51	1.21	0.42
Illinois	0.51	−3.34	1.37	1.48	1.77	2.01	2.24	2.20
Indiana	0.40	−4.33	0.98	−0.09	1.58	1.58	2.11	2.06
Kansas	0.54	−3.41	1.59	2.52	2.88	5.27	4.42	6.53
Kentucky	0.32	−3.54	0.52	−1.44	0.69	−0.60	0.68	−0.43
Louisiana	0.65	−2.76	2.25	5.77	4.24	9.76	7.80	14.03
Massachusetts	0.79	−1.65	2.91	8.49	5.59	13.35	10.57	19.01
Maryland	0.62	−4.14	2.01	6.45	3.53	10.52	5.93	14.04
Maine	0.36	−3.58	0.76	−0.77	0.88	−0.25	1.29	0.42
Michigan	0.59	−1.74	1.78	1.94	2.86	3.00	3.77	3.07
Minnesota	0.42	−3.76	1.07	0.26	1.77	1.90	2.40	2.36
Missouri	0.29	−3.21	0.25	−1.95	0.42	−0.98	0.52	−0.56
Mississippi	0.34	−3.79	0.29	−2.33	0.28	−1.56	0.31	−1.01
Montana	0.38	−3.54	0.85	−0.49	0.95	−0.12	1.35	0.53
North Carolina	0.38	−2.93	0.88	−0.32	1.31	0.55	1.88	1.08
North Dakota	0.34	−3.20	0.39	−1.72	0.23	−1.42	0.20	−1.01
Nebraska	0.26	−2.15	0.45	−0.94	0.57	−0.47	0.80	−0.15
New Hampshire	0.30	−4.94	0.83	−0.69	1.21	0.56	2.10	2.02
New Jersey	0.59	−3.01	1.93	3.94	3.36	6.54	5.73	8.96
New Mexico	0.35	−4.64	0.89	−0.44	1.24	0.66	1.85	1.56
Nevada	0.31	−2.71	0.78	−0.50	1.31	0.46	1.99	1.01
New York	0.50	−4.24	1.39	1.89	2.25	4.01	4.22	7.04
Ohio	0.52	−3.27	1.62	2.43	2.64	4.21	3.35	4.13
Oklahoma	0.68	−2.44	2.30	5.70	4.36	9.65	8.02	13.75
Oregon	0.52	−2.70	1.58	1.89	2.78	3.79	4.23	4.69

Exhibit 1 | (continued)

The Estimated Variance Ratios and Heteroscedasticity Consistent Z-Values

State / Region	q(2)VR	Z*(q)	q(4)VR	Z*(q)	q(8)VR	Z*(q)	q(16)VR	Z*(q)
Pennsylvania	0.50	−3.64	1.12	0.49	1.88	2.42	3.02	3.78
Rhode Island	0.43	−3.94	1.16	0.63	2.05	2.77	3.39	4.28
South Carolina	0.30	−4.62	0.44	−2.13	0.61	−0.97	0.97	−0.06
South Dakota	0.27	−4.18	0.31	−2.28	0.24	−1.64	0.26	−1.10
Tennessee	0.35	−3.73	0.51	−1.62	0.66	−0.75	0.90	−0.16
Texas	0.61	−2.83	1.98	4.15	3.59	7.17	6.65	10.69
Utah	0.54	−4.11	1.64	3.28	2.97	6.64	5.08	9.38
Virginia	0.55	−3.05	1.66	2.59	2.98	5.08	4.80	6.65
Vermont	0.29	−3.34	0.51	−1.31	0.60	−0.70	0.76	−0.29
Washington	0.79	−1.34	2.64	6.06	4.70	8.95	6.67	9.38
Wisconsin	0.54	−7.22	1.03	0.27	1.09	0.53	1.08	0.32
West Virginia	0.22	−3.03	0.29	−1.59	0.17	−1.23	0.18	−0.82
Wyoming	0.42	−4.25	1.10	0.44	1.82	2.25	3.12	4.01
New England	0.78	−1.87	2.78	8.65	5.38	13.90	10.15	19.82
Middle Atlantic	0.60	−3.41	1.92	4.47	3.55	8.12	6.30	11.53
South Atlantic	0.41	−2.87	1.20	0.56	2.06	1.93	3.35	2.93
East South Central	0.41	−3.08	0.69	−0.95	0.94	−0.11	1.19	0.26
West South Central	0.70	−2.13	2.43	5.94	4.56	9.68	8.43	13.80
West North Central	0.58	−2.96	1.53	2.17	2.63	4.35	3.43	4.42
East North Central	0.74	−1.60	2.45	5.09	4.25	7.47	5.78	7.51
Mountain	0.64	−2.80	2.07	4.74	3.92	8.51	6.72	11.38
Pacific	0.93	−0.49	3.57	10.16	6.71	14.79	10.46	16.75
United States	0.73	−1.99	2.56	6.74	4.88	10.96	7.89	13.31
No. Positive[a]								
50 States & DC	0	0	24	2	34	30	38	35
Nine Regions	0	0	8	7	8	7	9	8
United States	0	0	1	1	1	1	1	1
No. Negative[a]								
50 States & DC	51	46	27	5	17	0	13	0
Nine Regions	9	6	1	0	1	0	0	0
United States	1	1	0	0	0	0	0	0

Notes:

A value greater than 1.645 indicates statistical significance at the 0.1 level.

A value greater than 1.96 indicates statistical significance at the 0.05 level.

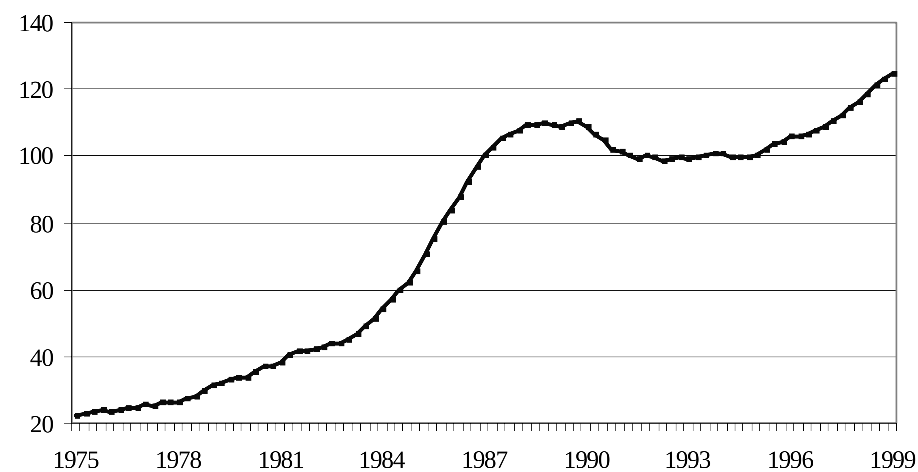
A value greater than 2.575 indicates statistical significance at the 0.01 level.

[a] Values are significant.

Dakota (SD) reveal the most significant negative autocorrelation. As shown in Exhibit 2, which plots quarterly index values for these six representative states, the plots of the CA and MA indices exhibit relatively smooth and long trends, while those for AL and SD exhibit more zigzags, particularly from 1975 to 1985.[5] It is worth noting that Alabama and South Dakota are less populous sates; their relatively fewer house transactions may account for the jagged noisiness during

Exhibit 2 | Indices of the Representative States

A. Massachusetts, positive autocorrelation for all lags, smoother trends.



B. California, positive autocorrelation for all lags, smoother trends.

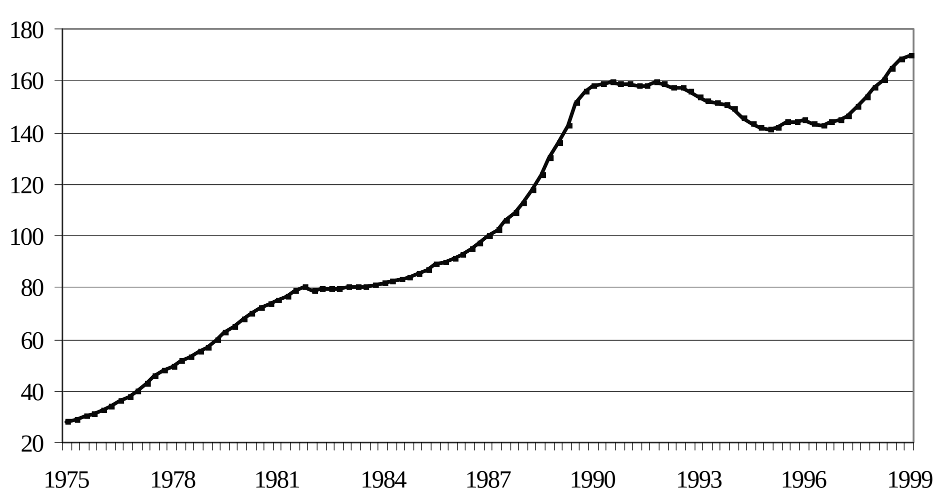
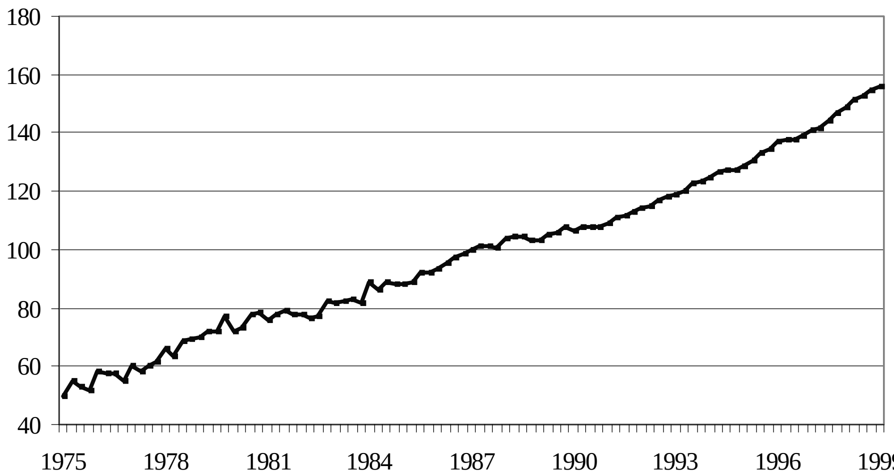


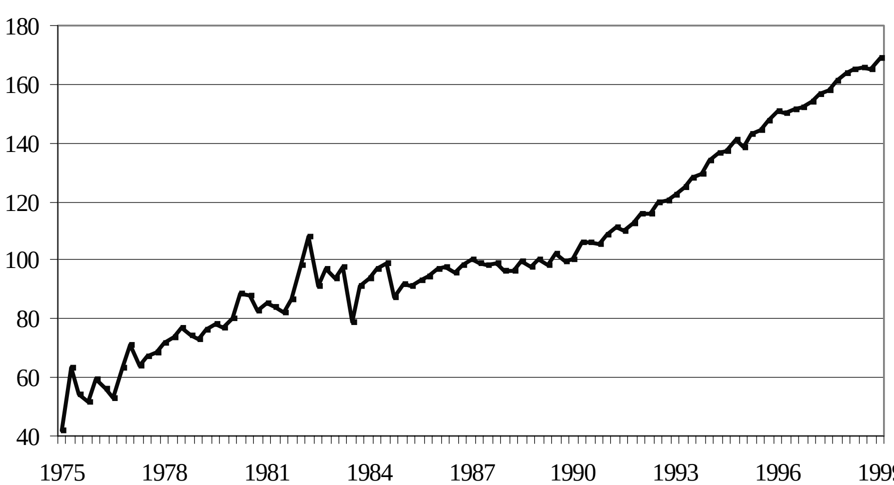
Exhibit 2 | Continued

Indices of the Representative States

C. Alabama, negative autocorrelation for all lags, more zigzags.



D. South Dakota, negative autocorrelation for all lags, more zigzags.

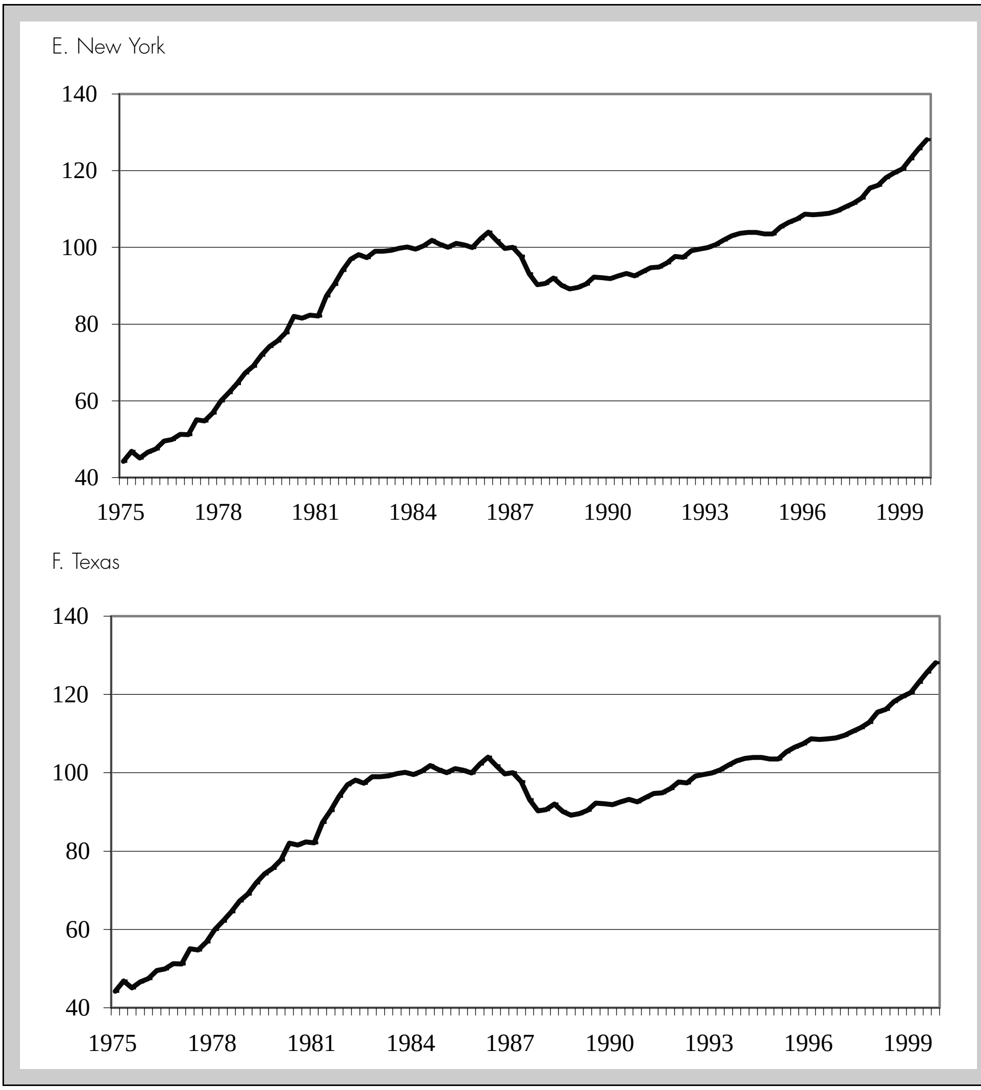


the early period. By contrast, the far greater number of transactions in Massachusetts and, especially, California may result in smooth graphs despite the relatively small time intervals.[6]

One way to test whether autocorrelation changes over time is to divide the indices into two equal time periods, and then to estimate autocorrelation separately for the two periods. One subset is from the first quarter of 1975 through the last

Exhibit 2 | Continued

Indices of the Representative States



quarter of 1986; the other is from the first quarter of 1987 through the last quarter of 1998.

As hypothesized, autocorrelation changes over time; there is a marked difference in measured autocorrelation from the first sub-period to the second. Exhibit 3 presents the heteroscedasticity-robust Z-values "$Z^*(q)$" of the estimated variance ratios for the two sub-periods. Notice that a positive (negative) *Z*-value indicates positive (negative) autocorrelation. Column 1 in the table lists the fifty-one areas and the regions. In row 1, ‘‘First’’ represents the first twelve years (1975–1986), and ‘‘Second’’ represents the second twelve years (1987–1998).

As shown in Exhibit 3, most of the indices exhibit significant changes. Among the fifty-one area indices, for $q = 2$, autocorrelation in twenty-eight indices changed direction (from negative to positive) from the first period to the second, and seven of these have significant coefficients for both directions.

Autocorrelation in nine indices changed from significant to insignificant, or the opposite, under similar conditions. For $q = 4$, twenty-six changed directions; five of them have significant coefficients for both directions, and ten changed from significant to insignificant (or the opposite). For $q = 8$, twenty-three changed directions, and eight changed from significant to insignificant (or the opposite). For $q = 16$, twenty changed directions, and thirteen changed from significant to insignificant (or the opposite).

Fourteen area indices exhibit positive autocorrelation for all the differences ($q =$ 2, 4, 8 and 16) during both periods (but for $q > 2$, only four areas showed positive autocorrelation for the whole period). Only two indices exhibit negative autocorrelation for all the differences during both periods (while thirteen are negative for the whole period). Similar phenomena occur among the nine regions: six show positive autocorrelation for all the differences during both periods. The index for the U.S. as a whole exhibits consistent and significant positive autocorrelation for all the differences during both periods; this result may reflect the generally upward trend in house price during the period. The evidence of positive autocorrelation supports what Case and Shiller (1989, 1990) and Kuo (1996) reported, and what Young and Graff (1996) found in the case of commercial real estate.

However, there is a concern that the estimated autocorrelation may not be accurate if the distribution of the time series is not normal, because the variance ratio method is sensitive to the family of distributions under consideration.[7] Young and Graff (1996) have reported that the probability distributions of returns on investments in houses are not normal. Normality tests on the natural logarithms of the indices provide no evidence that the probability distributions are not normal for 63.93% of the series.

Volatility, Return and Autocorrelation

This section examines the relationship between autocorrelation and volatility, and that between autocorrelation and rate of return. Equation (8) states the regression model:

$$AC_{iq} = \alpha + \beta_1\sigma_i + \beta_2\sigma_i^2 + \beta_3R_i, \tag{8}$$

Exhibit 3 | The Estimated Heteroscedasticity Consistent *Z*-Values for the Two Sub Periods

State / Region	First $Z^*(q = 2)$	Second $Z^*(q = 2)$	First $Z^*(q = 4)$	Second $Z^*(q = 4)$	First $Z^*(q = 8)$	Second $Z^*(q = 8)$	First $Z^*(q = 16)$	Second $Z^*(q = 16)$
Alaska	−1.31	−0.13	−0.96	−0.52	−0.86	−0.88	−0.82	−0.78
Alabama	−3.20	0.20	−2.72	−0.26	−1.73	0.09	−1.06	0.52
Arkansas	−1.76	1.08	−1.18	0.32	−0.36	1.41	0.47	2.57
Arizona	0.87	1.86	0.84	3.30	2.86	6.22	4.21	8.66
California	4.27	4.72	7.25	7.83	11.95	11.16	15.95	11.48
Colorado	0.61	3.39	1.98	4.63	5.07	6.88	6.86	9.19
Connecticut	0.50	1.99	0.43	2.25	1.24	1.97	1.45	2.88
Washington DC	−2.48	1.51	−1.27	2.82	0.14	3.27	1.33	3.60
Delaware	−2.62	3.73	−1.46	5.30	−1.25	6.77	−0.93	6.99
Florida	−1.87	0.87	−1.22	0.99	−0.54	−0.43	−0.15	−0.18
Georgia	−1.14	2.26	−0.96	3.76	−0.56	5.00	−0.32	6.45
Hawaii	−1.27	4.10	−0.90	7.24	−0.64	11.82	−0.46	13.42
Iowa	−1.04	0.38	−0.45	−0.15	0.16	0.12	0.59	0.70
Idaho	−2.63	0.50	−1.97	0.07	−0.78	1.50	0.20	2.84
Illinois	0.04	1.83	1.08	2.58	1.78	3.48	2.99	2.60
Indiana	−0.47	−2.25	0.68	−1.57	2.93	−1.19	4.69	−0.76
Kansas	0.41	0.99	2.08	2.56	5.56	5.18	8.08	8.23
Kentucky	−1.75	−1.19	−1.40	−0.44	−0.43	0.51	−0.12	1.03
Louisiana	1.77	2.83	3.92	4.13	7.79	4.82	11.17	6.49
Massachusetts	2.33	3.58	4.96	5.04	6.83	6.75	3.63	8.87
Maryland	−1.04	3.57	0.57	5.30	2.91	6.61	5.46	6.77
Maine	−1.68	2.53	−1.21	2.94	−1.02	2.65	−0.77	3.17
Michigan	0.83	2.14	1.90	2.70	3.42	5.59	4.68	7.04
Minnesota	−1.05	1.33	0.16	1.07	2.14	2.40	3.13	4.12

Exhibit 3 | (continued)

The Estimated Heteroscedasticity Consistent Z-Values for the Two Sub Periods

State / Region	First $Z^*(q = 2)$	Second $Z^*(q = 2)$	First $Z^*(q = 4)$	Second $Z^*(q = 4)$	First $Z^*(q = 8)$	Second $Z^*(q = 8)$	First $Z^*(q = 16)$	Second $Z^*(q = 16)$
Missouri	−1.84	1.81	−2.02	2.63	−1.05	5.30	−0.57	8.51
Mississippi	−1.73	−1.12	−2.36	0.17	−1.62	0.86	−1.11	2.28
Montana	−1.37	0.91	−0.59	−0.12	−0.26	1.09	0.30	2.62
North Carolina	−1.40	2.58	−0.85	3.81	−0.17	6.31	0.06	9.65
North Dakota	−1.32	−3.57	−1.62	−1.72	−1.32	−0.70	−0.92	0.41
Nebraska	−1.38	1.79	−0.96	0.40	−0.47	1.62	−0.11	3.03
New Hampshire	−3.37	3.27	−1.82	4.78	−1.29	6.56	−0.69	8.66
New Jersey	−0.92	2.04	−0.08	2.40	0.24	2.00	−0.34	1.98
New Mexico	−2.57	3.05	−1.11	4.34	−0.10	6.67	0.99	8.61
Nevada	−1.47	1.57	−0.60	1.44	0.48	2.64	1.33	3.27
New York	−1.30	2.33	−0.66	2.49	0.19	1.49	0.88	1.89
Ohio	0.44	0.05	2.65	−0.70	5.34	−0.93	7.10	−0.57
Oklahoma	1.79	1.66	3.75	1.27	7.70	0.70	10.05	1.27
Oregon	0.21	1.64	1.76	1.67	4.20	2.10	5.83	2.01
Pennsylvania	−1.38	2.38	−2.09	2.65	−1.14	2.37	−0.49	2.75
Rhode Island	−2.29	2.30	−1.93	2.12	−0.92	1.72	−0.43	2.35
South Carolina	−2.76	0.64	−2.29	0.15	−1.18	0.52	−0.33	0.99
South Dakota	−2.47	−2.10	−2.19	−1.16	−1.54	−0.32	−1.00	0.88
Tennessee	−1.74	1.56	−1.86	3.68	−1.14	6.92	−0.50	10.47
Texas	0.23	1.69	1.68	0.64	4.39	0.38	6.92	0.46
Utah	−0.71	3.29	0.80	4.87	3.54	7.24	5.94	10.12
Virginia	−0.95	2.97	−0.45	4.20	0.85	4.56	1.63	3.86
Vermont	−2.10	2.02	−1.57	3.35	−1.11	3.74	−0.81	2.63
Washington	3.35	2.10	5.52	3.11	10.03	4.13	13.77	3.68

Exhibit 3 | (continued)

The Estimated Heteroscedasticity Consistent *Z*-Values for the Two Sub Periods

State / Region	First $Z^*(q = 2)$	Second $Z^*(q = 2)$	First $Z^*(q = 4)$	Second $Z^*(q = 4)$	First $Z^*(q = 8)$	Second $Z^*(q = 8)$	First $Z^*(q = 16)$	Second $Z^*(q = 16)$
Wisconsin	1.48	1.22	0.61	0.32	1.52	0.72	2.01	0.86
West Virginia	−2.08	−2.48	−1.54	−2.32	−1.17	−1.72	−0.73	−0.90
Wyoming	−0.83	−0.98	0.12	−0.68	2.13	0.02	4.06	0.94
New England	2.02	3.38	3.69	4.39	5.92	5.18	5.62	6.81
Middle Atlantic	−1.11	2.05	−0.59	1.98	0.85	1.36	0.68	1.61
South Atlantic	−1.34	3.12	−0.32	4.49	0.77	4.04	1.42	3.87
East South Central	−0.92	1.63	−1.09	2.84	−0.22	4.84	0.32	7.55
West South Central	1.79	1.83	3.99	0.95	8.00	0.73	11.15	0.99
West North Central	1.03	2.88	1.86	3.21	4.59	5.45	5.77	8.52
East North Central	2.97	1.32	5.23	0.75	8.79	1.56	11.79	1.36
Mountain	1.27	2.98	3.20	3.71	7.16	5.76	10.53	8.04
Pacific	4.36	4.35	7.45	7.30	12.52	10.52	16.77	10.28
United States	2.59	3.64	5.14	4.91	10.07	5.33	14.19	5.07
No. Changed Directions[a]								
50 States & DC	27	7	24	5	23	0	20	0
Nine Regions	3	0	3	0	1	0	0	0
No. Changed Significance[b]								
50 States & DC		9		10		8		13
Nine Regions		3		2		3		4

Notes:
A value greater than 1.645 indicates statistical significance at the 0.1 level.
A value greater than 1.96 indicates statistical significance at the 0.05 level.
A value greater than 2.575 indicates statistical significance at the 0.01 level.
[a]Values are significant.
[b]Values are significant for the same direction.

where AC_q represents the level of autocorrelation in index i for q differences, which is calculated as the absolute value of the estimated variance ratio minus 1. The absolute value can measure the extent of deviation from randomness with the same scale. Using the absolute value of the dependency measurement can reveal both the effects and the magnitudes of the effects. In addition, it demonstrates the connection between the independent variables and the level of autocorrelation, which is the purpose of the regression analysis. σ_i and σ_i^2 are the standard deviation and variance of the natural logarithm of the quarterly indices; including both variables in the model can capture any nonlinearity that may exist in the relationship between volatility and autocorrelation. R_i is the average quarterly rate of return. The quarterly rate of return is calculated as:

$$R_t = \ln P_t - \ln P_{t-1} = I_t - I_{t-1}, \quad (9)$$

where:

R_t = Rate of return during interval t;
P_t = Value of the index at the end of period t; and
P_{t-1} = Value of the index at the end of period $t - 1$.

Values of all the variables are for the entire 1975–1999 period. Exhibit 4 presents the results of the regression analysis. For $q > 2$, autocorrelation is significant and negatively related to volatility, a result indicating that volatile indices tend to reveal lower autocorrelation. This finding is consistent with what LeBaron (1992) found in stock markets. Exhibit 5 provides the return volatility of the fifty states house price indices.[8] Autocorrelation is significant and positively related to the rate of return, a result implying that higher rates of house price appreciation tend to occur in streams of price increases.

The result is the opposite, however, for $q = 2$: autocorrelation has a significant positive relationship with volatility, and a significant negative relationship with rate of return. This outcome may be due to the fact that all the indices exhibit negative autocorrelations for $q = 2$, and most of these are significant. This relationship is also seen in the index values and the estimated autocorrelations shown in Exhibit 1. For example, indices with the most significant negative autocorrelations for $q = 2$, such as those for Alabama (AL) and South Dakota (SD), experienced obvious quarterly fluctuations. At the same time, indices with non-significant autocorrelation for $q = 2$, such as those for California (CA) and Massachusetts (MA), experienced little quarterly fluctuation. Negative autocorrelation indicates that price fluctuates more frequently than random; hence, the level of autocorrelation and volatility can move in the same direction. Because negative autocorrelation is negatively related to rate of return, volatile indices tend to be indicators of lower rates of return.

The connections between volatility and autocorrelation, and between return and autocorrelation, are opposite for all q differences. The implication is that, in

Exhibit 4 | Estimated Relations between Volatility, Return and Autocorrelation

		Independent Variables			
Absolute value of (VR − 1)	Intercept	Std. Dev.	Variance[a]	Return	Adj. R^2
$q(2)$	0.61	11.14	−0.72	−23.29	0.38
	(4.31)	(2.98)***	(−2.03)**	(−2.99)***	
$q(4)$	0.09	−23.27	2.07	71.44	0.11
	(0.18)	(−1.65)*	(1.56)	(2.43)**	
$q(8)$	0.39	−86.48	6.78	199.00	0.25
	(0.34)	(−2.65)***	(2.20)**	(2.93)***	
$q(16)$	0.89	−155.92	11.24	360.74	0.26
	(0.40)	(−2.47)**	(1.89)*	(2.75)***	

Notes: The dependent variables are the absolute values of the estimated variance ratios minus one, for $q = 2, 4, 8$ and 16, respectively. The independent variables include the standard deviation and variance of the quarterly returns of the indices, and the rate of quarterly return of the indices. *t*-values are in parenthesis.

*Statistically significant at the 0.1 level.

**Statistically significant at the 0.05 level.

***Statistically significant at the 0.01 level.

[a]The estimated coefficients are divided by 100.

general, more volatile or more risky indices tend to be associated with lower rates of return. Examining the data generally reveals the negative connection between volatility and rate of return. For example, returns of the CA and MA indices are much higher than those of AL and SD for both the whole period (CA 599%, MA 559%, AL 310%, SD 394%) and the first twelve-year period (CA 348%, MA 439%, AL 198%, SD 234%). The AL and SD indices are much more stable for the second twelve-year period, and they exhibit higher returns (156% and 168%, respectively) relative to those of CA and MA (172% and 126%, respectively). This result may be because, as Turner (2000) points out, house price volatility discourages homeownership. This phenomenon is similar to what has been found in the stock market (French, Schwert and Stambaugh, 1987; Campbell and Hentschel, 1992; and Haugen, 1999).

The significantly positive or negative autocorrelation in the indices' returns suggests that there might be a pattern of house price movements and, therefore, that investors would be able to develop some trading strategies to exploit the pattern if it were possible to trade the indices.[9] Consider a simple trading strategy based on the estimated autocorrelations of the indices: starting with a buy, and selling at the end of the period if something was bought but there was no chance to sell it. Short selling is not allowed. Transaction costs, tax effects and implicit rents are ignored, because relevant data are not available. For a series with q

Exhibit 5 | Volatility of the 50 States House Price Indices

State	Volatility	State	Volatility	State	Volatility	State	Volatility	State	Volatility
Alaska	0.62	Hawaii	0.95	Maine	0.43	New Jersey	0.24	South Dakota	0.68
Alabama	0.29	Iowa	0.27	Michigan	0.19	New Mexico	0.30	Tennessee	0.21
Arkansas	0.25	Idaho	0.29	Minnesota	0.19	Nevada	0.27	Texas	0.20
Arizona	0.21	Illinois	0.16	Missouri	0.36	New York	0.27	Utah	0.22
California	0.21	Indiana	0.13	Mississippi	0.43	Ohio	0.12	Virginia	0.14
Colorado	0.19	Kansas	0.15	Montana	0.41	Oklahoma	0.20	Vermont	0.41
Connecticut	0.25	Kentucky	0.23	North Carolina	0.14	Oregon	0.27	Washington	0.20
Delaware	0.27	Louisiana	0.18	North Dakota	0.79	Pennsylvania	0.19	Wisconsin	0.19
Florida	0.35	Massachusetts	0.22	Nebraska	0.31	Rhode Island	0.36	West Virginia	0.91
Georgia	0.14	Maryland	0.14	New Hampshire	0.49	South Carolina	0.19	Wyoming	0.41

Note: Reported as ten times the standard deviation of index return.

differences exhibiting negative autocorrelation, the rule would be to buy when a price decrease is observed and to sell when a price increase is observed. For a series with q differences exhibiting positive autocorrelation, the rule would be to buy when there is a price increase after a sequence of price decreases, and to sell when there is a price decrease following a sequence of price increases.

This simple strategy can be applied to the indices exhibiting the most significant autocorrelation. For example, there are eighteen repeat transactions on the Alabama index for $q = 2$. Following the simple strategy yields a total nominal return of 89.93% for the whole period, much less than the 214.18% return from a buy and hold strategy. On the other hand, there are only four repeat transactions on the California index for $q = 4$. The result is a total return of 538.29% for the whole period, somewhat higher than the 507.76% return from buying and holding. Applications of this simple strategy on other q differences, and on other indices, all result in returns lower than would result from buying and holding. Sophisticated programs might result in better returns.

Conclusion

House price changes in the U.S. exhibit some patterns, but the patterns differ across geographic areas. At the extremes, the indices for California and Massachusetts reveal long smooth waves, while that of Alabama is more closely configured around a line with more zigzags; most of the other indices show results somewhere in between.

For the entire 1975 to 1999 period, quarterly house price movements are negatively correlated for all the areas, and most of the negative correlations are significant, a result indicating that a quarter with high price increase may be followed by a quarter with price decrease, or vice versa. The implication is that a potential buyer may want to wait for a couple of months if there have been significant price increases, or buy as soon as possible if sharp price decreases have been observed for a few months. A seller would adopt the converse strategy.

The size and direction of autocorrelation changes over time. Most of the indices exhibit significant autocorrelation changes, in direction or level, in moving from the first twelve-year sub-period to the second twelve-year sub-period. Using the real price level for the tests, or adjusting the house price indices with nationwide CPI values, does not contradict that evidence.

When a simple trading strategy is applied to the indices exhibiting the most significant autocorrelation, the only excess return obtained is for the index of California for $q = 4$. While this kind of trading strategy is statistically feasible, it is far from certain whether abnormal returns could really be earned since trading on the index does not actually exist. However, it is reasonable to expect that one or more house indices and their derivatives might be traded in the future—as are the S&P 100, 500 and the Dow 30—as suggested by Case, Shiller and Weiss (1996).

When the indices exhibit more numerous directional changes than randomness would predict, and when the indices are volatile, the negative autocorrelation tends to be positively related to volatility and negatively related to the rate of return. When the indices move up or down in streams, the positive autocorrelation is negatively related to volatility and positively related to the rate of return. These results imply that more volatile indices tend to be associated with lower rates of return, consistent with results seen in studies of the stock market.

The evidence that house prices, in different areas and over different time periods, reveal different autocorrelations implies that the disparate results of previous researchers may all be relevant. It would be interesting to test Case and Shiller's data using Gau's method, and to test Gau's data using the Case and Shiller and Kuo methodology, to see whether randomness would be found in Gau's Vancouver data and positive autocorrelation would be found in Case and Shiller's data for the four metropolitan areas. If all the existing methods lead to findings of similar autocorrelation using the same data set—such as the Freddie Mac indices used in this study—the next stage of research would be to find methodologies that approach accurate autocorrelation estimates and that identify patterns of changes in autocorrelation.

Endnotes

[1] This compelling observation was offered by an anonymous referee, whom the author thanks.

[2] One problem with using appraised values is that, if seasonality exists, a resulting house price index may exhibit upward or downward bias depending on the time of appraisal. As noted by Graff (1998), this bias can be substantial, and it can not be eliminated simply through the use of a larger sample.

[3] The variance ratio method is created by Lo and MacKinlay (1988), and modified by Liu and He (1991).

[4] The author thanks an anonymous referee for offering this insight.

[5] Plots of two large states, New York and Texas indices are also presented as a referee suggested.

[6] The author thanks an anonymous referee for suggesting this possible explanation.

[7] Again, the author is grateful for the explanation suggested by an anonymous referee.

[8] Thank the referee for the suggesting this table.

[9] Case, Shiller and Weiss (1996) suggest the establishment of futures or options markets for residential real estate prices.

References

Campbell, J. Y. and L. Hentschel, No News is Good News, *Journal of Financial Economics*, 1992, 31, 281–318.

Case, K. E. and R. J. Shiller, The Efficiency of the Market for Single-Family Homes, *American Economic Review*, 1989, 79, 125–37.

——., Forecasting Prices and Excess Returns in the House Market, *Journal of the American Real Estate and Urban Economics Association*, 1990, 3, 253–73.

Case, K. E., R. J. Shiller and A. N. Weiss, Mortgage Default Risk and Real Estate Prices: The Use of Index-based Futures and Options in Real Estate, *Journal of House Research*, 1996, 7:2, 243–58.

Gau, G., Weak Form Test of the Efficiency of Real Estate Investment Markets, *Financial Review*, 1984, 10, 301–20.

——., Public Information and Abnormal Returns in Real Estate Investment, *Journal of the American Real Estate and Urban Economics Association*, 1985, 15–31.

Graff, R, A., The Impact of Seasonality on Investment Statistics Derived From Quarterly Returns, *Journal of Real Estate Portfolio Management*, 1998, 4, 1–16.

French, K. R., G. W. Schwert, and R. F. Stambaugh, Expected Stock Returns and Volatility, *Journal of Financial Economics*, 1987, 19, 3–29.

Haugen, R. A., *The Inefficient Stock Market: What Pays Off and Why*, New Jersey, Prentice Hall, 1999.

Kuo, C. L., Autocorrelation and Seasonality in the Real Estate Market, *Journal of Real Estate Finance and Economics*, 1996, 12: 139–62.

LeBaron, B., Some Relations between Volatility and Autocorrelation in Stock Market Returns, *Journal of Business*, 1992, 55, 199–219.

Linneman, P., An Empirical Test of the Efficiency of the House Market, *Journal of Urban Economics*, 1986, 20, 140–154.

Liu, C. Y. and J. He, A Variance-Ratio Test of Random Walks in Foreign Exchange Rates, *Journal of Finance*, 1991, 46, 773–85.

Lo, A. W. and A C. MacKinlay, Stock Market prices Do Not Follow Random Walks: Evidence From a Simple Specification Test, *Review of Financial Studies*. 1988, 1, 41–66.

Turner, T. M., Does Investment Risk Affect the House Decision of Families?, Unpublished manuscript, University of California–Davis, 2000.

Wang, F. T. and P. M. Zorn, Estimating House Price Growth with Repeat Sales Data: What's the Aim of the Game? *Journal of House Economics*, 1999, 6, 93–118.

Young, S. M. and R. A. Graff, Systematic Behavior in Real Estate Investment Risk: Performance Persistence in NCREIF Returns, *Journal of Real Estate Research*, 1996, 12, 369–81.

The author thanks the anonymous reviewers, along with Chionglong Kuo and Joseph Trefzger, for very helpful comments and suggestions.

Anthony Y. Gu, State University of New York, Geneseo, NY 14454 or gu@geneseo.edu.