Research Support
School of Business Administration

Nov. 1994

EXPECTATIONS, EFFICIENCY, AND EUPHORIA IN THE HOUSING MARKET

Working Paper #9405-37

Dennis R. Capozza
Paul J. Seguin

University of Michigan

NOT FOR QUOTATION
COMMENTS INVITED

Expectations, Efficiency, and Euphoria in the Housing Market

Dennis R. Capozza*
Stephen M. Ross Professor of Real Estate and Professor of Finance
School of Business Administration
University of Michigan
Ann Arbor, MI 48109-1234
313 764 1269
Dennis_Capozza@ccmail.bus.umich.edu

Paul J. Seguin
Associate Professor of Finance
School of Business Administration
University of Michigan
Ann Arbor, MI 48109-1234
313 764 9286
Paul_Seguin@ccmail.bus.umich.edu

August 1, 1994

* Capozza gratefully acknowledges the financial support of Capital Holding Inc. and KCBG Financial Sciences, while Seguin gratefully acknowledges summer research support from the Michigan Business School. Both authors thank Timothy R. Burch, Roger Kormendi and Nejat Seyhun for their helpful comments and Phil Meguire and Alastair Mcfarlane for their expert research assistance with this research. The usual disclaimer applies.

Expectations, Efficiency, and Euphoria in the Housing Market

Abstract

This paper studies expectations of capital appreciation in the housing market. We show that expectations impounded in rent-to-price ratio at the beginning of the decade successfully predict appreciation rates, but only if we first control for fluctuations within a transactions and information cost band. We also provide evidence of informational inefficiency: participants in housing markets appear to overreact to income growth and under-react to population growth in the prior decade.

During the last three decades real appreciation rates of owner occupied housing prices in metropolitan areas of the US have varied widely from minus thirty percent to plus 70 percent. This variation in appreciation rates is of direct concern for homeowners, builders, and realtors whose wealth or livelihood depends on home values. The mortgage lending industry including financial institutions, investors, mortgage bankers and insurers, and investment bankers is also interested because appreciation rates play a crucial role in defaults and prepayments.

Further, the pricing of most mortgage instruments – whole loans, mortgage backed securities, mortgage derivatives (CMOs, REMICS, etc.) – depends on the level and path of home values. Long term price swings also affect numerous policy issues like housing affordability and tenure choice, and they are extremely important to the reverse mortgage industry. The annuity that is paid to a homeowner with a reverse mortgage accumulates a mortgage balance that can eventually exceed the value of the house securing the loan if the borrower lives long enough. Thus the pricing of reverse mortgages is closely tied to expected appreciation rates of the underlying asset over a 10-15 year horizon.

Given their importance in the pricing and trading of numerous financial assets, changes and predictability of housing prices have attracted considerable academic attention. For example, one line of research has attempted to explain the cross-sectional variation in home values. Some studies estimate a reduced form equation using a list of variables that might affect supply or demand (Ozanne and Thibodeau (1983), Hendershott and Thibodeau (1990), and Rose (1989)).

Not surprisingly results vary widely among the studies. Inconsistent results have been found for income, population, population growth, topographical and land use restrictions, and amenities (e.g. climate, crime rates, etc.).

An alternative approach to explaining house values has been to base estimates on the asset pricing approach to valuing real estate developed in Capozza and Helsley (1990) and Capozza and Sick (1991). Under this approach, housing must compete with other financial assets and therefore must satisfy capital market equilibrium conditions. Growth of population and income as well as both systematic and unsystematic risk are important in this approach while amenities are not. Examples of this approach include Capozza and Schwann (1989, 1990) and Abraham and Hendershott (1992).

Interest among academics in long run trends in housing prices was also spurred by a controversial recent study by Mankiw and Weil (1989), who conclude by predicting that real house prices will decline by 50% in the 1990s. Mankiw and Weil note that the surge in demand for housing caused by the postwar baby boom could be easily predicted 20 years in advance since the eventual population in high demand cohorts (over 20 years old) can be estimated from the cohort data of earlier years. Mankiw and Weil hypothesize that if expectations are rational then home buyers should anticipate the effects of the population bulge of the postwar baby boom on the price of housing and bid prices up long in advance. Their empirical results reject efficiency, i.e. home prices appear to rise contemporaneously rather than in advance of predictable cohort shifts, suggesting that the housing market is informationally inefficient, at least with respect to this class of information.

In this study, we also address the issue of informational efficiency in the housing market. However, unlike Mankiw and Weil, we use census data disaggregated by metropolitan areas to analyze decadal appreciation rates. Exploiting disaggregated data on cross-sectional variations in appreciation rates is beneficial for a number of reasons. First, as with labor markets, supply and demand factors in real estate markets vary from locale to locale. By using aggregated data, many of these market specific factors are canceled out (diversified away) in the aggregation process. For example, while high household formation rates are easily predicted in the aggregate, at the metro area level most of the variation in decadal population growth is due to intercity migration rather than the baby boom and varies from -9% to +130%. Thus, metro area rates on which local expectations are likely to be based are highly volatile and swamp the variation of the postwar baby boom.

Second, the use of cross-sectional disaggregated data provides at least two econometric benefits. By using disaggregated data, the number of usable observations, and hence the statistical power of our tests, are increased. Also, by using cross-sectional data, we circumvent a number of potentially troublesome time-series problems encountered in the stock dividend literature[1].

The primary objective of this study is to examine the efficiency of real estate markets, in general, and the rationality of expectations, in particular. The specific proposition we examine is that expected risk–adjusted total returns on housing should be equal across metro areas. To derive empirical predictions from this proposition, we borrow from the stock dividend literature and show that if this proposition is true, then rent/price ratios should predict future appreciation rates. It is this empirical prediction that is the basis for our tests.

However, due to high transactions and information costs as well as asymmetric information, appreciation rates calculated from observed transaction prices will contain both equilibrium components, which we posit are predicted by rent/price ratios, and fluctuations within a transactions and information cost boundary around a normal level. Thus, our dependent variable is measured with considerable error. Further, if, as we believe, these fluctuations are correlated with the rent/price ratio, then results from standard econometric techniques will be biased.

[1] A number of studies in the finance literature have tested the predictive power of the dividend yield to some aggregate stock index. Though some studies have reported success, including Fama and French (1988), a number of recent studies have argued that inference is difficult given the existence of only one time series and the relative variability of prices compared to dividends. See Hodrick (1992) and Goetzmann and Jorion (1993) for examples.

To evade this bias, we propose and adopt a two stage procedure. We first regress observed price levels on a number of variables previously identified as being related to property value. The residuals from this cross-sectional regression are estimates of the observation error (deviation of observed value from their true value absent all market frictions), and so are used as an independent variable in our main specification. When we regress appreciation rates on these first stage residuals and the rent/price ratio, we find that both the rent/price ratio and the residuals from the first stage are highly significant and of the correct sign. However, both vary from their hypothesized value, indicating some form of inefficiency even after purging the effects of market frictions. To explore the source of these inefficiencies, we add lagged income and population growth to the equation. We find that both lagged growth rates enter the equation significantly suggesting that market participants do not fully incorporate available information.

In the following section, we review dividend theory and explicitly state our testable hypotheses. The third section describes the data while the fourth presents the results. Some brief conclusions and ongoing work and presented in the final section.

A Brief Review of Dividend Theory

A fundamental tenet in corporate finance is the dividend irrelevance proposition of Miller and Modigliani (1961). They point out that in the absence of taxes and transactions costs a corporation's dividend policy does not affect the value of its shares. Further, even in the presence of taxes and transaction costs dividend policy does not affect share value. For example, suppose that, for tax reasons, investors prefer capital gains to dividends. Then firms would:

> *adjust their dividend policies to take advantage of the negative effects of dividends by adjusting their dividend policies to supply the levels of yield that are most in demand. As a result the supply of shares at each level of yield will come to match the demand for shares at that level of yield, and investors as a group will be happy with the available range of yields. After equilibrium is reached, no corporation will be able to affect its share price by changing its dividend policy.* (Black and Scholes, 1974)

Empirically, the finance literature has generally shown that equity prices are efficient with respect to this form of information, and it is difficult to detect any difference in risk adjusted returns between high and low dividend securities (see Black and Scholes (1974) and Miller and Scholes (1984)).

The focus of this study is not to test the dividend irrelevance proposition in the housing market. However we do make use of the concept to test the role of expectations. Recall that the dimension of efficient markets we consider here is whether total risk-adjusted expected returns are equal across urban areas. Briefly, if there are differences in expected total returns across urban areas, then capital should flow to those areas with higher expected returns, increasing current price levels, and decreasing future expected total returns. Since expected total returns are the sum of the dividend or rent yield and an expected appreciation rate, urban areas where rent/price ratios are high should have lower expected appreciation. More formally, since:

$$E\{TR_{it}\} = \frac{R_{it}}{P_{it}} + \frac{E\{\Delta P_{it}\}}{P_{it}}$$

where TR_{it} is the total return to housing in area i over time period t, R is the level of rent in area i over time period t, P_{it} is the price of housing, and E is the expectation operator, then:

$$\frac{E\{\Delta P_{it}\}}{P_{it}} = E\{TR_{it}\} - \frac{R_{it}}{P_{it}}.$$

This identity, which is the basis for our empirical tests, indicates that expected capital gains should be negatively related to the rent/price ratio. Therefore, if information about existing rent ratios have been efficiently impounded into housing prices, then the rent/price ratio should have significant predictive power for future capital gains. Further, the relation is exact: investors should expect an area with a one-percent larger rent ratio to experience a one percent smaller appreciation rate per period.

Since expected appreciation rates are not observable[2], we need to use realized, rather than expected, capital gains. Thus, we examine whether expectations are rational by testing whether an area with a one-percent larger rent ratio should, on average, experience a one percent smaller appreciation rate per period, or a ten percent smaller appreciation rate per decade.

Several problems arise from using observed changes. First, the observed rents are not on the same houses as the observed prices. Since owned and rental units tend to be of different quality, there is measurement error in the data on the rent/price ratio. We do attempt to control for average quality between urban areas using housing characteristics available in the census data, but this does not fully correct for quality differences within each area. As a result we expect the coefficient on rent/price to be biased downward.

Second, though arbitrage is normally assumed to keep prices close to long run equilibrium, all assets trade within a band determined by transaction and information costs. In real estate markets these costs are large relative to securities markets, so real estate trades occur at prices within a wide price band. Aside from adding noise to the observations, with a concomitant reduction in statistical power, the existence of wide bands can induce severe bias. To illustrate, assume that prices are near the upper boundary of the transactions cost band. Observed prices will be high and the observed rent/price ratio will be low. In the next period prices will, on average, fall in the middle of the band, decreasing the estimate of capital appreciation.

Unfortunately, this error in observed capital appreciation is highly positively correlated with the observation error inherent in the rent/price ratio at the beginning of the period, which we use as the predictive variable. Therefore, regressing the observed appreciation

[2]An alternative approach to expectations in real estate markets has been to survey owners and renters directly (Case and Shiller (1988), Collins, Lipman, and Groeneman, (1992)). This approach is helpful for assessing the average expectations of owners and renters and for understanding the differences among cohorts. However markets reflect only the expectations of the marginal buyer and seller. Since we cannot identify these marginal traders from a survey, we cannot infer from surveys how the expectations that influence prices are formed.

rate on the observed beginning of period rent/price ratio will result in positively biased coefficient estimates[3].

Data

Our unique sample is a cross sectional time series of 64 Standardized Metropolitan Areas (SMAs) in the US from 1960 to 1980. The data were collected primarily from the decennial census but supplemented with series from other sources. A complete description of the data and its sources appear in the Data Appendix.

The key variables for our analyses are the decadal percentage change in real house values, $\Delta\ln(\text{VALUE})$, calculated as differences in the logs of reported prices deflated by the level of the CPI, and the rent/price ratio (R/P). Since our data is decadal, subscripts relative to "t" refer to decades, so "t-1" indicates an observation from the previous decade.

Results

To illustrate the magnitude of the biases outlined above, we initially estimated a simple one-step specification by regressing the real percentage change in housing values over the decade, $\Delta\ln(\text{VALUE})_t$ against the gross rental rate measured as of the start of the decade, R/P_{t-1}, and separate intercepts for each of the two final years over which the real decadal appreciation was calculated, Y_{70} and Y_{80}. Separate decadal intercepts are included to mitigate the effect of macro-wide factors in the specification including changes in aggregate demographics, especially changes in age cohorts (Mankiw and Weil (1989)). Data from

[3]To illustrate, assume that we want to regress y on x, but can only observe X = x + u and Y = y + v. In the standard error-in-variable case, where E{x'u} = E{y'v} = E{u'v} = 0, then OLS estimates of the slope parameter are biased towards zero by a factor proportional to $\frac{E\{x'x\}}{E\{x'x\} + E\{v'v\}}$. However, in this case, E{u'v} > 0, so OLS slope estimates will suffer from both a positive bias, and a bias towards zero. For example, in the case of a single regressor (with zero mean, for ease of exposition only), it is easy to show that the expected value of the OLS slope estimate is $(\beta + \frac{E\{v'u\}}{\sigma_x^2})\ (1 + \frac{\sigma_v^2}{\sigma_x^2})^{-1}$. The first term captures the positive bias while the second captures the bias towards zero.

two decades of price appreciation for each of the 64 SMAs was used, yielding (with t-statistics in parentheses):

$$\Delta\ln(\text{VALUE}) = \underset{(0.18)}{0.017\, Y_{70}} + \underset{(2.20)}{0.244\, Y_{80}} + \underset{(0.47)}{0.654\, R/P_{t-1}}$$

The intercept associated with appreciation over the 1960s (Y_{70}) is small and insignificant indicating little average real housing price appreciation over this decade. In contrast, real housing appreciated by more than 24% over the 1970s, and this increase is significant. Of primary importance, however, is the coefficient associated with the rent ratio as of the beginning of the decade. Not only is the coefficient insignificantly different from zero, but it is also of the wrong sign. Though insignificance could be due to a bias towards zero attributable to measurement error, we believe the positive sign is also attributable to a positive bias in the coefficient estimate due to correlated measurement errors, as detailed in footnote 2.

To mitigate the measurement error problems, we utilize a new two-step estimation technique that is analogous to an Instrumental Variables technique. A second benefit of our technique is that it allows us to incorporate information about the transactions cost band into the regression. In the first step, we regress the (logs of) housing price *levels* on the set of explanatory variables used by Capozza and Helsley (1990) and Capozza and Sick (1991) in their investigation of urban growth with uncertainty. These variables include income, population, population growth rate, property tax rate, utility cost rate, age, number of rooms, number of baths, developable area in the city, and construction cost.

Results of the cross-sectional regression on housing price levels on these instruments with intercepts that vary by decade appear in table 1. The high level of the R^2 (>.8) suggests that the model is successful in capturing much of the cross-sectional dispersion in housing values. Further, the model seems well specified since all significant variables are of the correct sign: higher house values are associated with higher real income, population, population growth, number of baths and replacement construction costs, and lower levels of taxes and land availability. Numerous alternative specifications were estimated, but conclusions remain unaltered.

The primary motivation for the first stage is not to provide a model of housing values, *per se*, but to generate residuals which are the difference between reported housing values and their expected value conditional on many factors. Thus, these residuals can be thought of as estimates of the transitory component on house value attributable to transactions costs.

As a simple test of this intuition, we regress the change in real housing value, $\Delta\ln(\text{VALUE})_t$, over a decade against the estimated residual as of the start of the decade, $\hat{e}_{t-1}$, and separate intercepts for each of the two final years over which the real decadal appreciation was calculated, Y_{70} and Y_{80}. Data from two decades of price appreciation for each of the 64 SMAs was again used, yielding (with t-statistics in parentheses):

$$\Delta\ln(\text{VALUE}) = \underset{(3.15)}{0.062\, Y_{70}} + \underset{(15.11)}{0.295\, Y_{80}} - \underset{(-3.64)}{0.053\, \hat{e}_{t-1}}$$

We argue that when the residuals from the first stage regression are positive, house values are above their expected value, so they are close to the upper bound defined by transactions costs. In the next decade, prices will appear, on average, lower (towards the center of the transactions band). Thus a higher residual at the beginning of the decade should be related to a lower level of observed capital appreciation over the decade. Consistent with our intuition, the coefficient associated with the first-stage residuals is negative and statistically significant.

We note, however, that the coefficient associated with the first-stage residuals is much closer to zero than its hypothesized value of -1.[4] There are two intuitively different yet possibly econometrically equivalent explanations for this result. It is possible that the first-

[4]Using notation developed above, assume that you want to regress $(y_t - y_{t-1})$ on x_{t-1}, but you can only observe $Y_t = y_t + v_t$ and $Y_{t-1} = y_{t-1} + v_{t-1}$. If (i) you can measure v_{t-1} exactly, (ii) v_{t-1} is orthogonal to both v_t and x_{t-1}, and (iii) you regress $(Y_t - Y_{t-1})$ on both x_{t-1} and v_{t-1}, then the estimated coefficient associated with must be exactly equal to -1. To see this, rewrite the regression as $(y_t + v_t - y_{t-1} - 1v_{t-1}) = \beta x + \delta v_{t-1} + e$. The estimate of δ equals the sum of the estimates from regressing each of the LHS variables on v_{t-1}. Due to the orthogonality assumptions, these will equal 0, 0, 0 and -1 respectively.

stage residuals are less than perfect in identifying the transitory difference between reported value and true underlying value. Since this implies that the first stage residual measures the desired quantity with error, its estimated coefficient will be biased towards zero. Alternatively, it may be possible that residuals for the same SMA but adjoining decades are highly autocorrelated. In this case the coefficient estimate associated with one residual will equal -1 plus the first order autocorrelation of residuals for the same SMA.

The two explanations are econometrically equivalent in that both are the result of an omitted variable in the first-step regression. Because some variable has been omitted, the residual measures the position of the reported price within the transaction band with error. Further, if the value of this omitted variable is highly correlated across observations for the same SMA, residuals for the same SMA will absorb this high correlation. Unfortunately, this omitted variable may be inherently unobservable, and its identification is beyond the scope of this paper.

We next present our main specification that adds the first stage residuals to the model linking capital appreciation over the decade to the rent/price ratio as of the beginning of the decade. Again using the same 128 observations, estimation via OLS yields:

$$\begin{array}{lcccc} \Delta\ln(\text{VALUE}) = & 0.276\,Y_{70} & +\ 0.546\,Y_{80} & -\ 0.073\,\hat{e}_{t-1} & -\ 3.148\,R/P_{t-1} \\ & (2.52) & (4.29) & (-4.16) & (-1.99) \end{array}$$

As above, the estimated coefficient associated with the first-step residual is again negative and significant, consistent with our transactions cost interpretation. Unlike the original specification, however, intercepts associated with appreciation over the 1960s (Y_{70}) and the 1970s (Y_{80}) are both larger positive numbers reflecting appreciation rates of about 2.5% and 4.5% per year for the two decades respectively, and are both significant. The movement in these estimates away from zero is consistent with a mitigation of the errors-in-variables problem[5].

[5]Since estimates of an intercept and estimates of a slope coefficient are negatively correlated, a negative movement away from zero for a slope will tend to occur with a positive movement in the intercept.

Of primary importance, however, is the fact that the rent/price ratio as of the beginning of the decade is now negative and significant as predicted by the theory. Thus, current rent/price ratios appear to have some power to predict subsequent capital appreciation. However, since we are using annual rental rates and decadal capital appreciation rates, the estimated coefficient on the rent/price coefficient should be approximately -10 or roughly three times the estimated magnitude.

The fact that the estimated coefficient is only one third the expected size could arise from a number of sources that are not necessarily mutually exclusive. First, as we mention above, rent/price ratios are very difficult to measure for a variety of reasons, so this variable is still subject to measurement errors inducing a downward bias. Second, if homeowners have a strong preference for capital gains, perhaps due to asymmetric tax treatment, then a 1% decrease in rent yield will result in a less than one percent increase in required capital gains. However, implicit rents are never taxed while capital gains may be taxed on realization. Thus, we believe that housing prices and rents will be determined by marginal investors for whom rents are preferred to capital gains and therefore view this second explanation as being unlikely

A third possibility is that our results are indicative of a rejection of one of the fundamental assumptions unpinning our investigation, namely, that expectations are rational and home buyers are processing publicly available information correctly. Specifically, if expected total returns are not constant across SMAs, but systematically vary in the cross-section, and this variation is positively correlated with the beginning of the decade rent-to-value ratio, then the coefficient estimate on the ratio would suffer from a positive bias. In the next section, we present some supplemental tests to investigate this possibility.

Euphoria

We use decadal data on the growth of income and population to test for cross-sectional variation in expected total required returns. Both income growth and population growth are highly autocorrelated (.78 and .52 respectively), so there is much predictability in these time series and much dispersion in cross-sectional forecasted values. In table 2 we add lagged population growth and lagged income growth to our specification.

As above, the estimated coefficient associated with the first-step residual is again negative and significant, consistent with our transactions cost interpretation. Though the rent/price ratio as of the beginning of the decade remains below its theoretic value of -10, it remains negative and significant and is now almost half its theorized magnitude indicating a modest improvement in the fit of the model.

Of utmost relevance, however, is the result that both of the lagged growth variables enter the model significantly. This is strong evidence against the hypothesis that required total returns to residential real estate are constant and that observed returns vary randomly around an aggregate average. Instead, total returns vary systematically with the two factors introduced.

There are a number of possible conclusions that can be drawn from this result. First, it might be argued that systematic cross-sectional variations in expected total returns is rational, in that it reflects cross-sectional differences in the risk characteristics of the underlying real estate assets. However, as capital market theory points out, expected returns are awarded only for differences in exposure to systematic risk factors. Thus, this first argument is valid only under the unlikely scenario that cross-sectional differences in the two lagged growth rates are somehow correlated with cross-sectional exposures to contemporaneous systematic risk factors and thus, risk premia. A second argument may be that the growth rates may affect the location of observed prices within the transaction band. Thus, the inclusion of these variables merely reduces the errors-in-variables problem. We agree that this would be a valid conclusion if contemporaneous growth rates were employed, but lagged rates are used. Further, the impact of these factors on the location of beginning-of-period prices was accommodated in the first-step regression.

Thus, we conclude that the significance of these factors indicates that agents have not correctly processed the information content of the lagged values of these variables into the rent-to-value ratios. More specifically, the signs of these lagged variables indicate that owners undereact to population growth and overreact to income changes. This last finding is consistent with Abraham and Hendershott (1992) who use annual data on repeat sales. They find a lag structure on income that produces cyclical movements in house prices. Our results suggest that the cycle from annual data is also reflected in decadal data.

Conclusion

In this study, we analyze single family house price appreciation for 64 metro areas over three census periods. Data on house prices, rents, household income, population, housing characteristics (rooms, baths, age), property taxes and utility costs were collected from the decennial censuses for 1960 to 1980. We use these data to test whether equilibrium total returns (the service flow or rent from the property plus price appreciation) are approximately equal across markets. Specifically, we test whether areas with low rent/price ratios have higher expected appreciation rates.

However since real estate markets are subject to very large transactions and information costs, prices fluctuate in a wide band around long run equilibrium. Unlike other studies of the housing market we control for fluctuations in the transactions cost band by using a two stage procedure in which we first estimate an SMA's position within the transactions cost band and then use this information in estimating the relation between rent/price ratios and subsequent rates of appreciation.

Our results indicate that both the component of reported values attributable to transitory movements within the transactions cost band and the rent/price ratio are both valuable predictors of subsequent house price movements. Each percent increase in the rent/price ratio reduces the subsequent decadal appreciation by about 4.3%. Since rent/price ranges from 3% to 10%, cities with the lowest rent/price ratios can be expected to appreciate roughly 30% more per decade than those with the highest rent/price ratios.

We next test whether expectations incorporate all available information by adding lagged population and lagged income growth to the specification. Lagged income growth enters significantly with a negative sign, and lagged population growth enters with a positive sign. While investors and homeowners appear to process past experience into an expected appreciation rate, there appear to be systematic bias to expectations. The effect of population growth is not fully impounded in rent/price ratios. Further, the negative sign associated with lagged income growth suggests a certain degree of unsubstantiated euphoria: when income growth has been high, owners set rent/price ratios as if they systematically overestimate subsequent appreciation rates.

Like most studies of the housing market our results are not fully consistent with asset market efficiency. However, since information and transactions costs are unusually large for this asset class, this conclusion may not be surprising. Further given capital constraints of residential real estate owners and short sale restrictions, the forces of arbitrage usually in place to eliminate inefficiencies are highly impaired.

References

Abraham, J. and P. Hendershott, "Patterns and Determinants of Metropolitan House Prices, 1977-91," mimeo, Ohio State University, 1992.

Black, F. and M. Scholes, "The Effect of Dividend Yield and Dividend Policy on Common Stock Prices and Returns", *Journal of Financial Economics*, 1974, 1, 1-22.

Capozza, D.R. and R. Helsley, "The Stochastic City", *Journal of Urban Economics*, 1990, 28, 187-203.

Capozza D. R. and G. Schwann, "The Asset Approach to Pricing Urban Land: Empirical Evidence", *Journal of the American Real Estate and Urban Economics Association*, 1989, 17, 2, 161-174.

Capozza D. R. and G. Schwann, "The Value of Risk in Real Estate Markets", *Journal of Real Estate Finance and Economics*, 1990, 3, 117-140.

Capozza, D. R. and G. Sick, "The Risk Structure of Land Markets", University of Michigan, 1991.

Case, K. and R. Shiller, "The Behavior of Home Buyers in Boom and Post-Boom Markets," *New England Economic Review*, 1988, 29-46.

Collins, S., B. Lipman, and S. Groeneman, "Experience and Expectation: The Home Price Predictions of Owners and Renters," National association of Realtors, 1992.

Fama, E. F. and K. R. French, "Dividend Yields and Expected Stock Returns", 1988, 22, 3-25.

Goetzmann, W. N. and P. Jorion, 1993, "Testing the Predictive Power of Dividend Yields", Journal of Finance, 1993, 48, 663-679.

Hendershott, P. and T. Thibodeau, "The Relationship between Median and Constant Quality House Prices," *J. of the American Real Estate and Urban Economics Association,* 1990, 18, 323-334.

Hodrick, R. J., "Dividend Yields and Expected Stock Retruns: Alternative Procedures for Inference and Measurement", 1992, 5, 357-386.

Mankiw G. and A. Weil, "The Baby Boom, the Baby Bust, and the Housing Market", *Regional Science and Urban Economics,* 1989, 19, 235-258.

Miller, M. and F. Modigliani, "Dividend Policy, Growth, and the Valuation of Shares", *Journal of Business*, 1961, 34, 411-433.

Miller, M. and M. Scholes, "Dividend and Taxes: Some Empirical Evidence", *Journal of Political Economy*, 1982, 90, 1118-1141.

Ozanne, L. and T. Thibodeau, "Explaining Metropolitan Housing Price Differences," *Journal of Urban Economics*, 1983, 13, 51-66.

Rose, L. "Topographical Constraints and Urban Land Supply Indexes," *J. of Urban Economics,* 1989, 26, 335-347.

Table 1
First Step Regression to Determine Location in Transactions Cost Band

For each of the 64 SMA's and each of three usable census data points, we regress the (logs of) housing price *levels* on the set of explanatory variables used by Capozza and Helsley (1990) and Capozza and Sick (1991) in their investigation of urban growth with uncertainty. These variables include income, population, population growth rate, property tax rate, utility cost rate, age, number of rooms, number of baths, developable area in the city, and construction cost. Detailed descriptions of the source and/or computation of these variables appear in the data appendix, while summary statistics for these variables appear in Table 1. The notation (x100) means the estimated coefficient has been scaled up by a factor of 100). There are 192 usable observations. The R^2= .802 and the regression F-statistic is 60.51 which is significant at any standard significance level.

Explanitory Variable	Coefficient Estimate	T-statistic
Y_{60}: *Intercept for 1960 Census Observations*	0.790	0.77
Y_{70}: *Intercept for 1970 Census Observations*	0.550	0.53
Y_{80}: *Intercept for 1980 Census Observations*	0.846	0.81
ln(Real Income)	0.971	8.56***
ln(Population)	0.028	1.87*
Δln(Population)	0.216	3.31***
Tax Rate: *Median value of property taxes (Monthly)*	-6.867	-2.70***
Utility Rate: *Median heating, water, gas, electricity*	1.624	0.38
Age: *Median age of owner occupied buildings* (x100)	-0.146	-0.65
Median number of rooms (x100)	-0.684	-0.23
Median number of baths	0.096	3.56***
Conditional Land Supply Index	-0.442	-5.52***
ln(Local construction cost index / CPI level)	0.603	2.97***

Table 2
Second Step Regression with Added Explanitory Variables

For each of the 64 usable SMA's and each of two observation intervals (1960 to 1970 and 1970 to 1980), we regress the appreciation in value over the interval (measured as the first difference in the log of the value of the median house price) on a set of explanatory variables including lagged income growth, lagged population growth, separate intercepts for each of the two final years over which the real decadal appreciation was calculated, and the beginning-of-period rent-to-value ratio. Detailed descriptions of the source and/or computation of these variables appear in the data appendix, while summary statistics for these variables appear in Table 1. Also included in the specification is the estimated resdiual from the specification reported in Table 2. This residual proxies for that portion of the reported beginning-of-period value that is transitory and due to fluctuations of reported transactions prices within a band defined by transactions costs. There are 128 usable observations. The R^2= .485 and the regression F-statistic is 18.82 which is significant at any standard significance level.

Explanitory Variable	Coefficient Estimate	T-statistic
Y_{70}: *Intercept for 1960-1970 Observations*	0.558	-3.71***
Y_{80}: *Intercept for 1970-1980 Observations*	0.769	5.18***
Rent-to-price Ratio_{t-1}	-4.317	-2.70***
$\Delta\ln(\text{Population})_{t-1}$	0.176	2.74***
$\Delta\ln(\text{Income / CPI deflator})_{t-1}$	-0.611	-3.30***
$\hat{e}_{t-1}$: Residual from First Step Regression (Table 2)	-0.080	-4.72***

Appendix A
DESCRIPTION OF DATA AND COMPUTATIONS

CITIES STUDIED

The following SMSAs are included in the sample.

1 Akron
2 Albany NY
3 Albuquerque
4 Anaheim-Santa Ana
5 Atlanta
6 Baltimore
7 Birmingham
8 Boston (includes Brockton MA)
9 Buffalo-Niagara Falls
10 Charleston SC
11 Charlotte
12 Chattanooga
13 Chicago (not including Gary-Hammond IN or Kenosha WI)
14 Cincinnati
15 Cleveland
16 Columbus OH
17 Dallas-Fort Worth @
18 Denver
19 Des Moines
20 Detroit
21 El Paso
22 Fort Lauderdale
23 Grand Rapids
24 Hartford-New Britain @
22 Honolulu
26 Houston
27 Indianapolis
28 Jacksonville FL
29 Kansas City
30 Knoxville
31 Lansing MI
32 Las Vegas
33 Los Angeles
34 Louisville
35 Memphis
36 Miami
37 Milwaukee
38 Minneapolis-St. Paul
39 Nashville
40 New Orleans
41 New York-Northern New Jersey
42 Oklahoma City
43 Omaha
44 Orlando
45 Philadelphia
46 Phoenix
47 Pittsburgh
48 Portland OR
49 Providence RI
50 Riverside-San Bernardino
51 Rochester NY
52 Saint Louis
53 Salt Lake City-Ogden @
54 San Antonio
55 San Diego
56 San Francisco-Oakland
57 San Jose
58 Seattle-Tacoma @
59 Syracuse
60 Tampa-St. Petersburg
61 Toledo
62 Tulsa
63 Washington DC
64 West Palm Beach FL

@ = SMSAs formed in 1980 from 2 prior distinct SMSAs. Pre-1980 data are a weighted average of data for each component SMSA.

New York-Northern New Jersey includes the following SMSAs: New York (which includes Westchester and Rockland counties), Nassau-Suffolk, Jersey City, Newark, and Paterson-Passaic. In some cases, the value for 41 is a weighted average of reported values for each of component SMSAs.

SOURCES

The following abbreviations are used:

CCDB	City and County Data Book, a Supplement to the Statistical Abstract.
CHMHC	Census of Housing, Vol. II, Metropolitan Housing Characteristics.
CHUS	Census of Housing, Vol. II, General Characteristics. Part I, US Summary.
ERP	Economic Report of the President to the Congress.
PUMS	Public Use Microdata Sample. Computer file of household micro data, consisting of a 0.1% stratified random sample from the 1970 and 1980 Censuses of Housing and Population, and a 1% stratified random sample from the 1940 and 1950 Censuses of Population and Housing. The 1960 PUMS contains no SMSA identifier and so is unusable.
REIS	Machine readable (DOS) files included in the Regional Economic Information System, prepared and distributed by the Bureau of Economic Analysis, US Department of Commerce.
SA	Statistical Abstract of the United States.

DATA

Data on the following variables are used in this study. Unless otherwise indicated, all data are for metro areas and are taken from the decennial Census of Housing and Population. Owner-occupied dwellings are defined by the Census Bureau as being sited on 10 acres or less of land, and consist largely of 1 unit structures, as condominiums were rare before the 1980 Census. All dollar amounts are not adjusted for inflation unless otherwise indicated.

Baths	Median number of baths in owner-occupied dwellings. Available only on the 1980 PUMS.
CLSI	Conditional Land Supply Index. Ranges from .5 (for a city that completely occupies an island) to 1.0 (for a city on a featureless plain). Source: for 38 cities, values are given in Table 2 of Rose [1989]. For remaining 26 cities, values are informal subjective estimates by Capozza and Meguire.
CPI	Consumer Price Index, all urban residents, average of monthly values. Source: 1940-60: Ibbotson Associates, Stocks, Bonds, Bills and Inflation, Table B-10, last column. 1960-90: 1992 ERP
Gross_Rent	Median of the sum of monthly contract rent on rented unfurnished dwellings, plus estimated monthly cost of utilities. Data for rented 1 unit detached houses were reported only for the 1970 Census. Data for other years is for all rented dwellings. 1940: CHUS, Table 104, last column. 1950: 1956 CCDB, Table 2 (Anaheim, Fort Lauderdale, Las Vegas, Orlando and West Palm Beach) or Table 3 (all other), col. 15. 1960: 1967 CCDB, Table 3. 1960-80: CHMHC, Table A-2. 1990: CD-ROM put out by Census Bureau.
HCCI	Historical Construction Cost Index, all US, 1/1/75 = 1.0. Measured as of July 1 of each year. Source: R.S. Means & Co., Building Construction Cost Data, 1983, 1992 editions.
NCON	Local construction cost, expressed in terms of US average=100. Source: 1980-90: R. S. Means & Co., Building Construction Cost Data, quarterly publication.
Pers Income	1970-90: = GNP+Transfer Payments+Interest on Government Debt-Retained Earnings of Corporations. In per capita terms. Source: REIS, File CA-25, series 030. 1940-50: Same as 1970-90, except computed for principal state surrounding SMSA. Source: REIS, Table SA-52.
Population	Total population by SMSA. 1970-90: REIS, File CA5, series 020. 1960: SA, 1983-4, Table 20, col. 1.
Rooms	Median number of rooms in owner-occupied dwellings. Source: CHMHC, Table A-6.
Tax&Insur	Median value of the sum of property taxes and property/casualty insurance owed on owner-occupied dwellings, at monthly rates. Only 1980 data are available. Available only on the 1980 PUMS.
Utility_Cost	Median of the sum of the monthly cost of heating, water, gas, and + electricity. For owner-occupied dwellings only. Source: computed from estimated monthly cost of each of component appearing on the 1980 PUMS.
Value	Value of owner-occupied dwellings, estimated by owner-occupant. Source: 1960: 1967 CCDB, Table 3, col. nn. 1970: 1972 CCDB, Table n, col. nn.

	1980: 1987 CCDB Table n, col. nn.
	1990: CD-ROM put out by Census Bureau.
Yearbuilt	Median coded value for year in which owner-occupied dwellings were built. Available only from the 1970 and 1980 PUMS. Coding is as follows: 1=1979-80, 2=1975-78, 3=1970-74, 4=1960-69, 5=1950-59, 6=1940-49, 7=1939 or earlier

COMPUTED VARIABLES

The following variables are computed from the above data as follows:

Age:	if Yearbuilt=1, =1; if 1<Yearbuilt<7, = Census year - midpoint of YearBuilt; if Yearbuilt=7, = Census year - 1925.
Rent	= 12*Gross_Rent/Value
ln(Value)	= LOG(VALUE/CPI)
Δln(Value)	= ln(Value)(t) - ln(Value)(t-10)
ln(Income)	= LOG(FAMERN/CPI)
Δln(Income)	= ln(Income)(t) - ln(Income)(t-10)
Δln(pop)	= LOG(Population(t)/Population(t-10))
Taxrate	= (Taxes & Insurance)/Value
Utilityrate	= Utility_Cost/Value

Pre-1980 values of TAXRATE, UTILITYRATE and NCON are set to their 1980 values.

Annual population for the periods 1951-9 and 1961-8 is estimated as follows. For the period 1970-90, regress the growth rate of annual SMSA population on the growth rates of the populations of those states any part of which are included in the SMSA. Include a ML correction for AR(1) errors. Using state population growth rates for the period 1949-50, forecast the 1949-50 SMSA population growth rate. Conditioning on this forecast, forecast the growth rate of population for 1951-69, accounting for estimated AR(1) error process. Prorate the difference between the cumulated annual forecasted growth rates for 1951-60 and the actual Growth rate over 1950-60 over the forecast growth rates. Repeat for 1961-9. For 1950 and 1960, let Forecast_Population = Population. Then forecast population using

$$\text{Forecasted\_Population}(t) = \text{Forecasted\_Population}(t-1)*e^{[\text{forecast\_population\_growth}(t)]}$$