The 1998–2005 Housing "Bubble" and the Current "Correction": What's Different This Time?

Authors William C. Wheaton and Gleb Nechayev

Abstract This paper examines the inflation in housing prices between 1998 and 2005 and investigates whether this run-up in prices can be "explained" by increases in demand fundamentals such as population, income growth, and the decline in interest rates over this period. Time series models are estimated for 59 MSA markets and price changes from 1998 to 2005 are dynamically forecast using actual economic fundamentals to drive the models. In all 59 markets, the growth in fundamentals from 1998 to 2005 forecasts price growth that is far below that which actually occurred. An examination of the 2005 forecast errors reveals they are greater in larger MSAs, in MSAs where second home and speculative buying was prevalent, and in MSAs where indicators suggest the sub-prime mortgage market was most active. These latter factors are unique to the recent housing market and hence make it difficult to asses if and how far housing prices will "correct" after 2005.

Recently there has been renewed interest in the behavior of the housing market in the United States. The most common discussion is over whether prices have risen "too" high relative to some benchmark. This paper avoids a normative discussion and focuses instead on the "reasons" for the rapid rise in prices between 1998 and 2006. The degree to which economic fundamentals contributed to the rise is assessed by using time series analysis for each of 59 markets. It becomes clear that prices rose far faster than would be called for by changes in fundamentals. The findings reveal that the degree of "excess" price increase is strongly related to two relatively new factors influencing the housing market: the widespread availability of risk-priced mortgage credit, and the unusually strong purchase of houses as second homes and investments. While these associations are quite strong statistically, inferring causality is difficult. It also makes it difficult to judge the duration of depth of the current housing correction.

The argument that there is a "bubble" in the recent housing market has been most strongly put by Shiller (2006).[1] He argues that the housing rent-to-price (R/P) ratio has very gradually trended downward since 1913, but dropped more steeply since 1995. He compares this with real interest rates and stock dividends and concludes that R/P ratios are too low and pricing is ‘‘excessive.’’ A clever paper by Smith and Smith (2006), however, reinforces a major measurement issue with the R/P approach—that median or average rent levels are just not comparable to price levels. The typical apartment has no land and around 700 square feet, while the median house has both a back yard and more than twice that square footage. Furthermore, these characteristics have changed over time. In a carefully constructed matched sample of single-family housing sales and rental units, they find that current R/P levels seem in general to be overly high—10% or more!

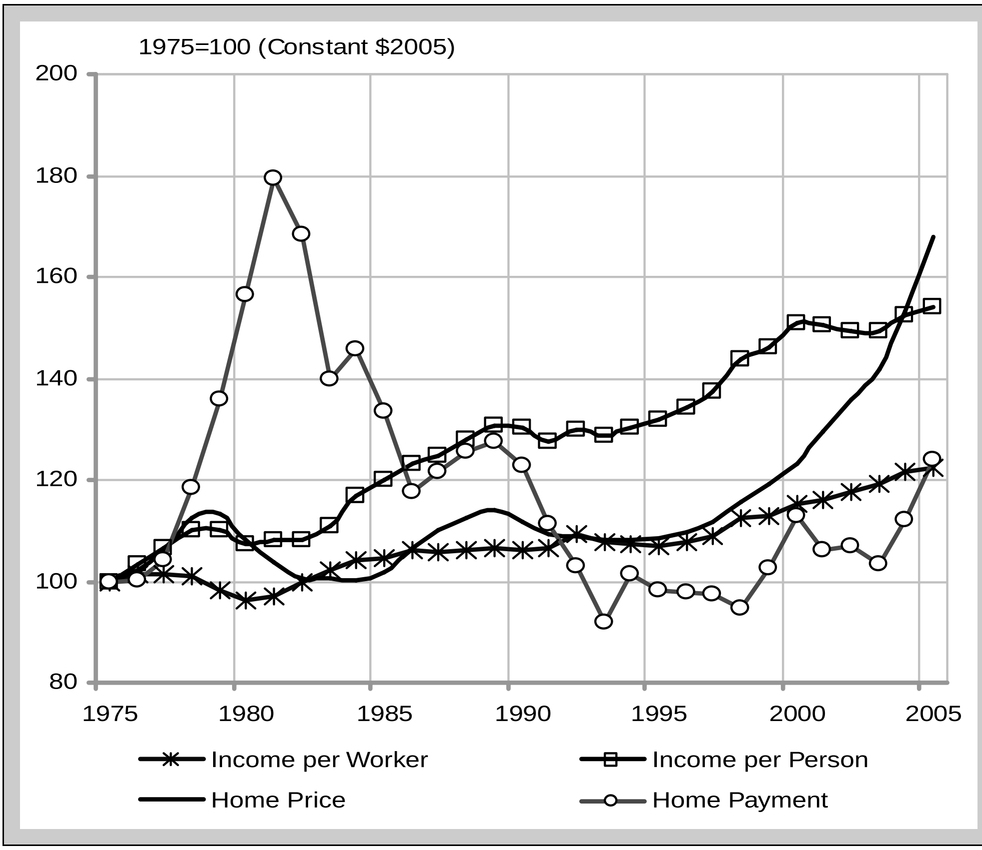
On the other side, Himmelberg, Mayer, and Sinai (HMS, 2005) argue that the movements in real interest rates and incomes since 1976 have almost perfectly matched the movement in prices (i.e., prices times a real interest rate have grown almost the same as income). To HMS, prices are just perfectly following demand fundamentals in the long run, although any discussion inferring market balance has to factor in a demand elasticity. More importantly, yearly payment variations are quite erratic and not closely related to fundamentals.

The current paper is organized as follows. In the next section, there is a brief review of the data most often used to compare house prices and fundamentals and like HMS conclude that the story is very different depending on in what MSA the comparisons are being made. The modeling strategy is discussed next, along the strategy for uncovering the true historic relationships between fundamentals and prices—unique to each MSA. The estimation results follow including an examination of the forecast errors made by the models between 1998 and 2005. Two factors that are unique to the housing market of the last eight years are examined that have not been part of previous housing market booms or corrections. The first is the emergence of a risk-priced sub-prime mortgage market that has helped millions of households become home owners. The second is an unusual growth in the demand for second homes or investment houses. These two factors are highly correlated with the forecast errors in ways that would suggest causality. The paper closes with cautions about inferring causality, and with uncertainty about the future of the housing market in the U.S.

How Much Have House Prices Risen Relative to Economic Fundamentals?

Average house prices in the U.S. rose only 18% when adjusted for inflation over the 23 years from 1975 until 1998 (Exhibit 1).[2] During this same period, average (not median) income per worker increased 12% and income per capita almost 40%. Household income (which is measured with less precision) grew between these two. Thus with certainty, average housing prices over this earlier period grew in line with income. Since 1998, however, prices have risen almost 50% while incomes increased only between 5% and 11%. These last eight years have indeed been remarkable.

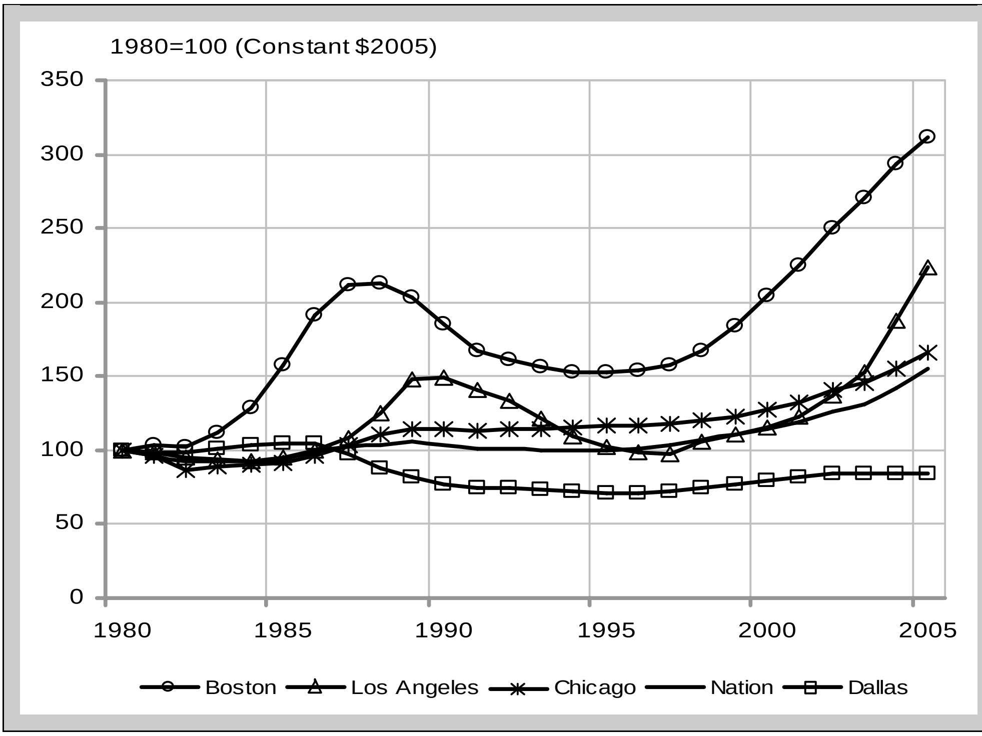
Exhibit 1 | Prices, Income, Payments



As HMS point out, this does not mean that housing has become less ''affordable.'' The annual cost of owning a dollar's worth of housing incorporates interest rates, taxes, and the likely net appreciation of the asset. Without going into the full calculation, which invariably must make some assumptions about owner estimation of future appreciation, the average housing price is simply multiplied by the mortgage interest rate each year (the ''home payment'' series in Exhibit 1) to arrive at the annual cash cost of owning a home. From 1975 through 1998, this ''annual cash cost'' of went through great fluctuation, but on net did not change. In 1998, housing was more ''affordable'' relative to income than in 1975. Since 1998, the annual payment cost has risen about 25%—outstripping income gains, but only slightly as falling mortgage rates almost offset rising prices.

These kinds of comparisons vary radically by geography. Inflation-adjusted house prices have shown dramatically different growth around the country (Exhibit 2). From 1979 though 1998, real home prices rose 74% in Boston, 10% in Los

Exhibit 2 | Prices in Markets



Angeles, 11% in Chicago and fell 21% in Dallas. Since 1998, the increases have been 83% in Boston, 123% in Los Angeles, 42% in Chicago, and only 12% in Dallas. Exhibit 3 shows that across these four (and other) areas, there was no where near as much variation in real income growth. Real income per capita or per worker grew gradually and quite similarly across regions. Once the change in mortgage rates is factored in, the issue of affordability exists mostly in California and the East Coast states.

Modeling House Prices and Economic Fundamentals

An initial test of why house prices are rising is to see if all the changes in the economic variables known to impact housing demand (fundamentals) can ''forecast'' the recent rise in prices over the last eight years. The best way to do this is to estimate an econometric forecasting model using data only through some date (here 1998). Then prices are forecast forward with this model using the *actual* 1998–2005 changes in that economic data. The question is whether the model picks up the rapid rise in the last eight years.[3]

There is a long literature on house price forecasting models. As summarized by Capozza, Hendershott, and Mack (CHM, 2004), there is wide consensus that as

Exhibit 3 | Growth Rates

	1980:Q1–1998:Q4 (76 quarters)			1999:Q1–2005:Q4 (28 quarters)		
Market	*W*	*Y*	*P*	*W*	*Y*	*P*
Atlanta	32%	49%	9%	9%	0%	21%
Boston	42%	60%	74%	14%	11%	83%
Chicago	23%	34%	11%	7%	1%	42%
Dallas	26%	36%	−26%	14%	3%	12%
Denver	22%	37%	6%	16%	9%	33%
Houston	27%	25%	−38%	15%	7%	19%
Los Angeles	31%	13%	10%	12%	6%	123%
New York	49%	48%	56%	7%	6%	89%
Philadelphia	27%	44%	21%	7%	8%	60%
San Diego	13%	31%	1%	10%	17%	128%
San Francisco	46%	51%	43%	16%	9%	93%
Washington	26%	39%	6%	9%	15%	107%

Notes: The sources are the Bureau of the Census, Bureau of Labor Statistics, and OFHEO.
W = Cumulative % change in real income / employment.
Y = Cumulative % change in real income / population.
P = Cumulative % change in real home price index.

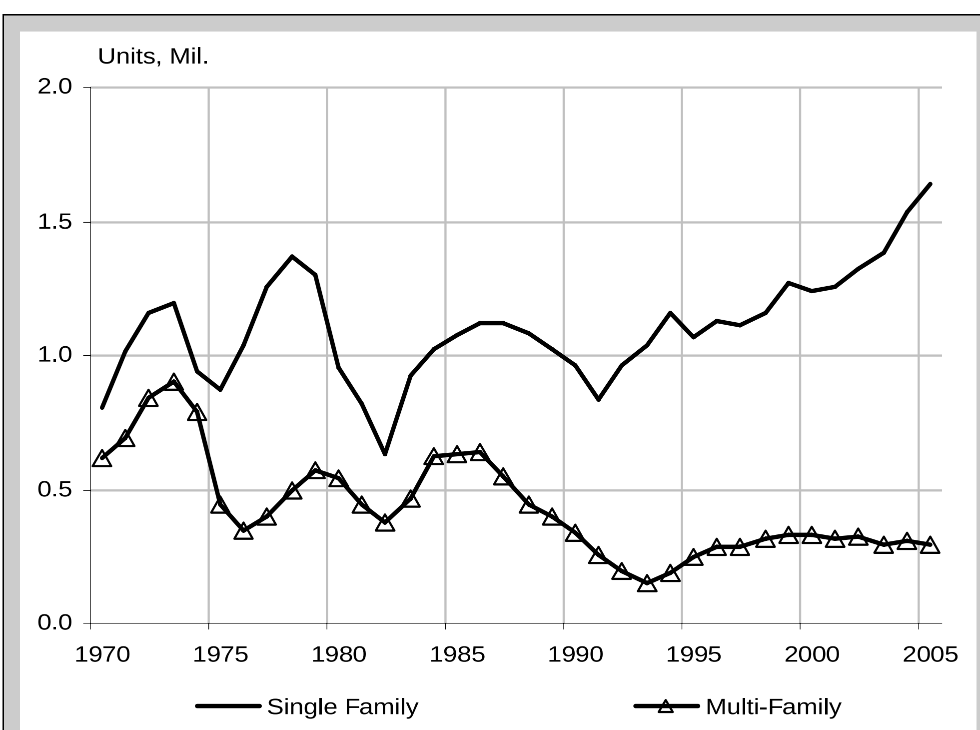
markets grow in employment and population, rents and prices should also increase. There is likewise consensus that they should rise with income and move inversely to some measure of the cost of capital. Like many authors in the past, CHM apply an Error Correction Model (ECM) approach that uses these variables. The two-stage ECM approach however has been recently questioned by Gallin (2006). Instead, the current study favors a single-stage univariate model that simply assumes that housing prices react to demand shocks with a response pattern that is unique to a market. In such a case, prices are being directly predicted by economic variables and possibly lagged prices—all either in levels or differences.

A possible criticism of this approach is that the supply side needs to be incorporated more explicitly by jointly modeling prices and the housing stock using a two-variable VAR. Early work by DiPasquale and Wheaton (1994) used a structured VAR, while more recent papers by Evenson (2004) and Harter-Dreiman (2004) use unstructured VAR models. These models are able to estimate an implied supply elasticity for each market. Glaeser, Gyrouko, and Saks (2005) argue that a declining supply elasticity has caused prices to rise ''excessively'' of late—due largely to increased local development regulations. Identifying changes in supply elasticities is possible; however, only if a unique instrument (e.g., "regulation") can be found for the estimated VAR supply equation. Such an instrument just does not exist for a wide enough range of markets and without it, a univariate model is likely to be just as accurate in forecasting house prices, and a lot more simple.[4]

It is also important to point out several important facts about the supply side of the current housing market. First, the number of constructed single-family houses in 2005 is seen to be at an *all time record* (Exhibit 4). When added to multi-family units, total construction in 2005 was near two other previous peaks (1973, 1978). The "robustness" of recent supply will become even more apparent when later compared to household formation. Thus there is little evidence of at least a national supply "shortage" that might this time be the instigating factor behind recent price inflation.

The first univariate model is in price levels as shown in Equation (1). Current prices are regressed on lagged prices and a set of current economic variables (X_t). If the model works as theory would suggest, multiple price lags may need to be allowed to correct both for the well-known autocorrelation of prices, as well as to implicitly pick up the role of supply in the reaction to demand changes (Granger

Exhibit 4 | U.S. Housing Construction by Type



and Newbold, 1974). Hence in Equation (1) the α_j might be at first positive and then turn negative as new supply helps to quell the impact of demand shocks. For this to produce a stable "impulse response," the sum of the α_j must be less than one. In any given market, the elasticity (or inelasticity) of supply is implicitly picked up in the price response pattern. With perfectly elastic supply, eventually prices return to their original level after a positive increase in an X_t variable. With inelastic supply, they rise and generally remain high. Without lagged prices, the model is effectively a moving equilibrium equation, in which autocorrelation will need to be corrected with AR1 estimation.

An alternative version of the model is in Equation (2) where each variable is differenced (in logs). This is a standard way of correcting for autocorrelation without applying AR1 estimation. Lagged price differences are also added (as suggested by Granger and Newbold, 1974) to possibly control for second-order autocorrelation.

$$LogP_t = \sum_{j=1}^{n} \alpha_j LogP_{t-j} + \beta' LogX_t. \qquad (1)$$

$$Log(P_t/P_{t-1}) = \sum_{j=1}^{n} \alpha_j Log(P_{t-j}/P_{t-j-1}) + \beta' Log(X_t/X_{t-1}). \qquad (2)$$

Results

The results of estimating Equations (1) and (2) are extremely dependent on the combination of model and estimation technique employed.

When Model 1 is estimated in levels with no lagged prices, AR1 estimation produces autocorrelation Rho values that are extremely close to 1.0, but the correction (Hildreth-Lu) generates coefficient estimates for the X variables that are significant and correctly signed. Without the AR1 correction, the signs of the coefficients on economic fundamentals are generally wrong and hence the model (Model 1a) would be useless for assessing their role in determining prices.

When lagged prices are included into the levels equation, the R^2 values jump enormously and AR1 estimation still produces a very significant Rho. The AR1 correction still is necessary in order to yield significant and correctly signed X variables. With OLS estimation, the X variables are incorrectly signed and only lagged prices are significant. This model is called Model 1b.

In Model 2a, all variables are measured in log changes rather than levels and with no lagged price changes. Both OLS and AR1 estimation produce similar results as there is little second-order autocorrelation. The results yield significant and correctly signed *X* variables with similar point estimates to those obtained in Model 1a when estimated with AR1.

When the variables are measured in log changes but with the addition of lagged price changes (Model 2b), either estimation technique still yields little autocorrelation, and the economic variables are generally significant and correctly signed. Lagged price changes are most often not significant, consistent with finding no autocorrelation in Model 2a.

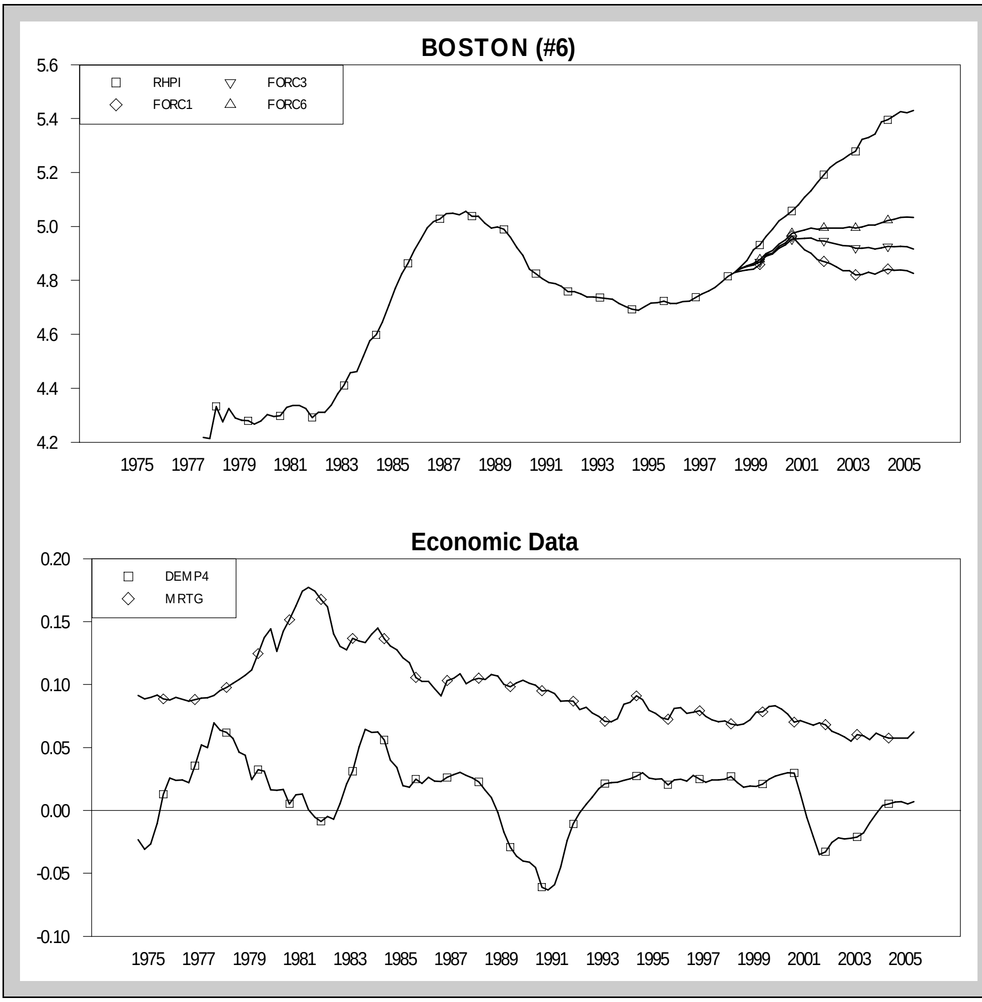
The lesson learned here is that price changes are stationary, and have little autocorrelation. Hence the estimation method chosen matters less and lagged dependent variables increase R^2 values, but have little impact on the *X* variable coefficients. Price levels, on the other hand, are extremely auto-correlated. If the error specification is corrected with AR1 then the models produce results comparable to the equations in differences. If the error is ''corrected'' using lagged dependent variables with OLS (as offered by Granger and Newbold, 1974) the model fails most tests of reasonableness. AR1 estimation hence appears to be essential to any model based on price levels.

These models are estimated individually for each of 59 MSA markets (a table is found in Appendix 1). The OFHEO repeat sales indices are used for prices, and MSA total employment (*LEMP*) is used for economic variables, total personal income divided by employment (*LRINCEMP*), and the 30-year fixed mortgage rate (*LMRTG*). Both real and nominal rates were examined, but nominal work better in this instance. To illustrate the models and their application in more detail, four estimated equations (described above) are presented for the Boston, Chicago, and Phoenix markets in Appendix 2. Exhibits 5–7 show the data and forecast results for each of these areas. In the exhibits, the upward triangles represent Model 1b, the diamonds Model 2a, and the downward triangles Model 2b. Model 1a is virtually identical to 2a. The line with squares is the actual movement in log (price level). The forecast with Model 2a or 2b is actually made in differences, but then has been converted to levels for comparability with the levels forecasts. The forecast errors for all 59 metropolitan markets (converted to a percent of 2006 prices) are found in Appendix 1, using a broader set of six equations that the included the four models described here (Models 1a, 1b, 2a, 2b).

In the case of Boston, the first frame in Figure 4 shows that the model forecast completely misses the recent price growth if the forecast is started in 1998.[5] The recovery in prices from the trough in 1994 is forecast to continue for three more years (adding 20% to real prices) and then a correction sets in after 2001. In fact, prices have risen a full 50% (in constant dollars) without a correction since 2001.

The model's forecast is the case of Boston is completely explainable, by turning to the second frame in Exhibit 5. Like many regions, starting in mid 2001 there is a downturn in the local economy (employment, square symbols) almost as severe as the downturn in 1989–1992 and more severe than that in 1980–1983.

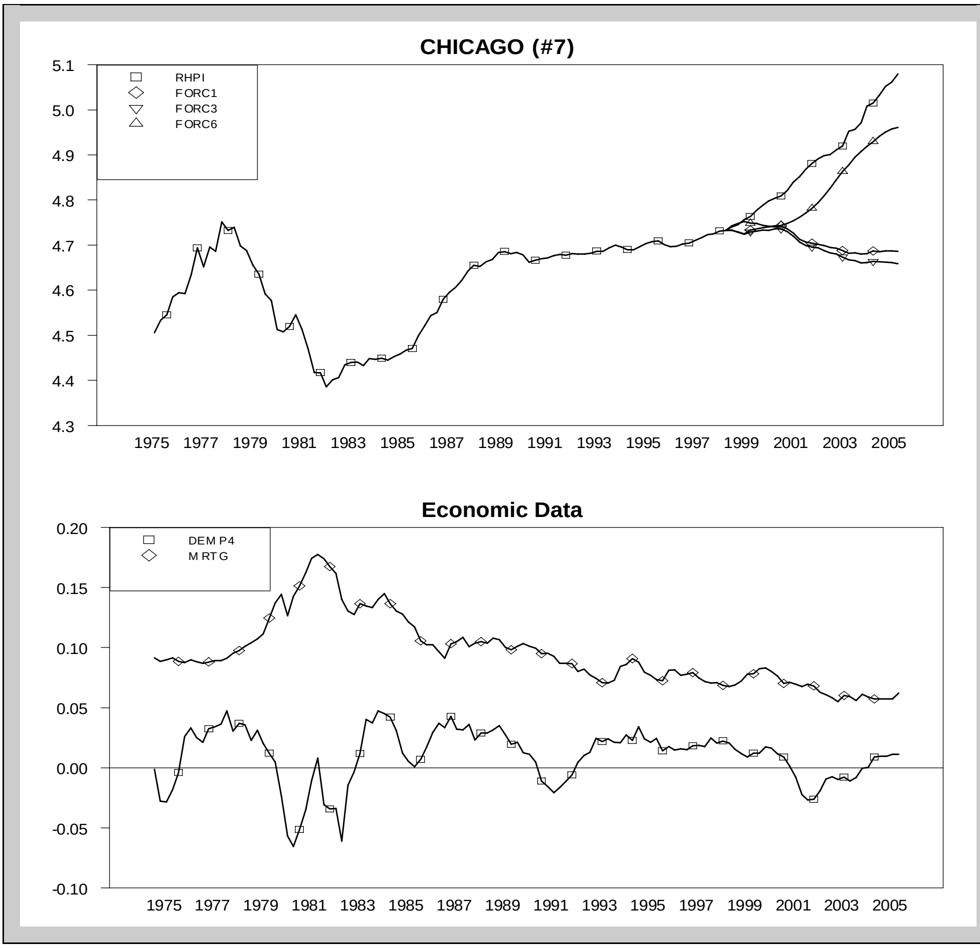
Exhibit 5 | Boston Model Results



Furthermore, the recovery from 2003 to 2005 is actually less robust than that after the previous two episodes. Finally, the fall in interest rates (diamond symbols) from 2001 to 2005 appears no greater than that which occurred in the four years following after the previous two recessions. Given these movements in fundamentals, it is easy to understand how the model calls for a correction after 2001. The forecast error for Model 1b (42%) is less than that for Model 1a or 2a (62%), but is high enough so as to completely miss the recent housing boom.

To varying degrees, this story is repeated in all of the 59 MSAs. In Chicago, prices were essentially flat in real terms through much of the 1980s and early 1990s and then picked up recently. Model 1b is able to capture some of this recent

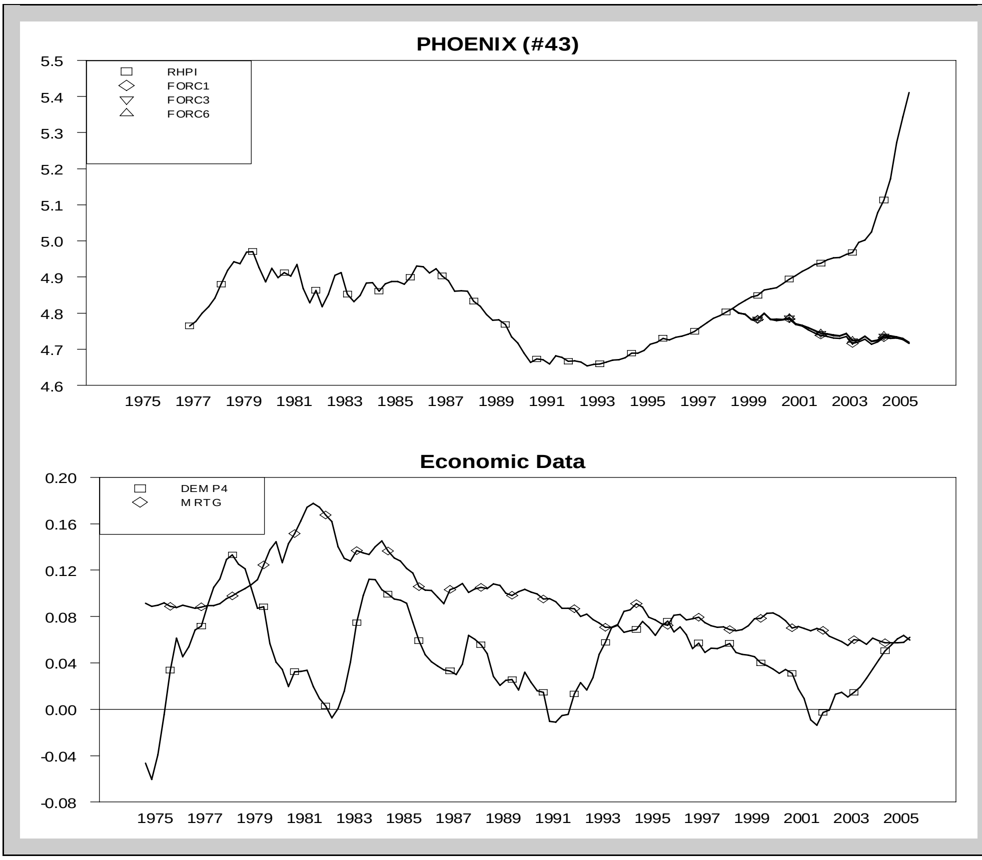
Exhibit 6 | Chicago Model Results



price growth, but still has a large 12% forecast error. The error with the other models is far greater (40%).

Phoenix is typical of many southern and western markets (e.g., Florida, San Diego, Las Vegas). Prices have stagnated or fallen in real dollars continuously since the late 1970s, only to soar in recent years. The result across the models for these types of areas is a very large forecast error. Again Model 1b generally has the least error, but in the case of Phoenix and many such markets, the error is greater than 70%. In most such markets the ‘‘explanation’’ for the model's under-forecast is also the same: the growth in economic fundamentals during 1998–2005 has been generally weak and not as strong as earlier historical periods during which house prices were not surging.

Exhibit 7 | Phoenix Model Results

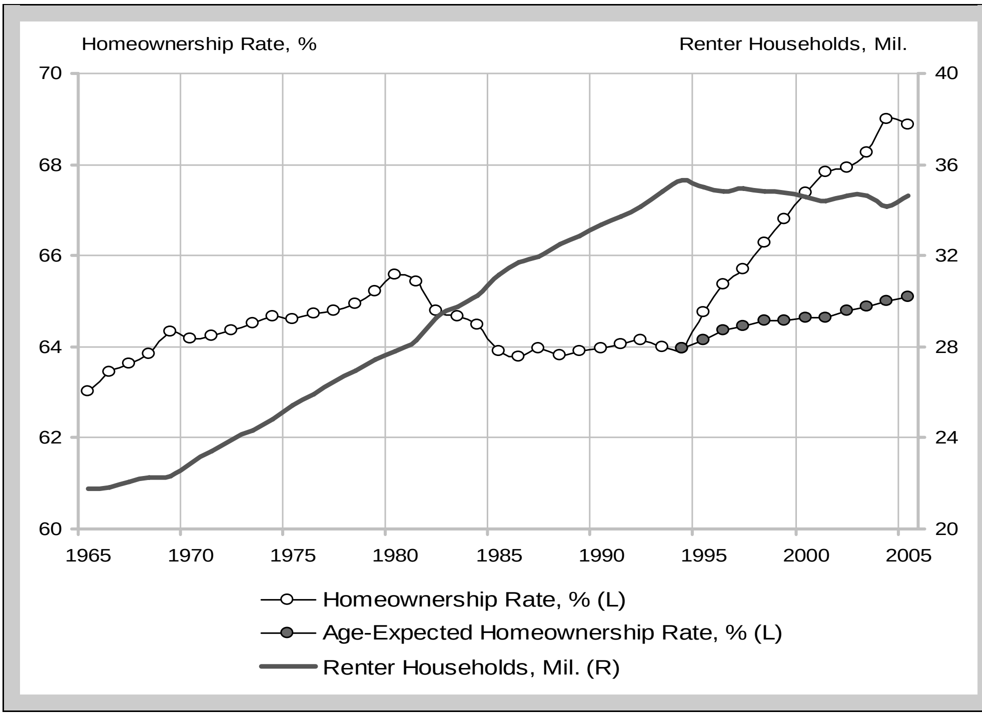


Appendix 1 contains the forecast errors for each market as a percentage of current prices, using a number of different models in addition to those described. These forecast errors are so systematically significant (in the same direction) as to suggest that something completely new and different is characterizing the housing market in the last decade. Two hypotheses are offered about what these new factors might be.

Credit-Propelled Homeownership: 1998–2005

One factor that has changed in the last ten years that would easily explain an unusual growth in prices is the soaring national rate of homeownership. Between 1965 and 1995, the homeownership rate fluctuated between 62% and 64% with little discernable trend. Since 1995 it has jumped 5 percentage points and at the end of 2005 stood at 69% (Exhibit 8). This movement was so pronounced that

Exhibit 8 | U.S. Homeownership and Renter Households



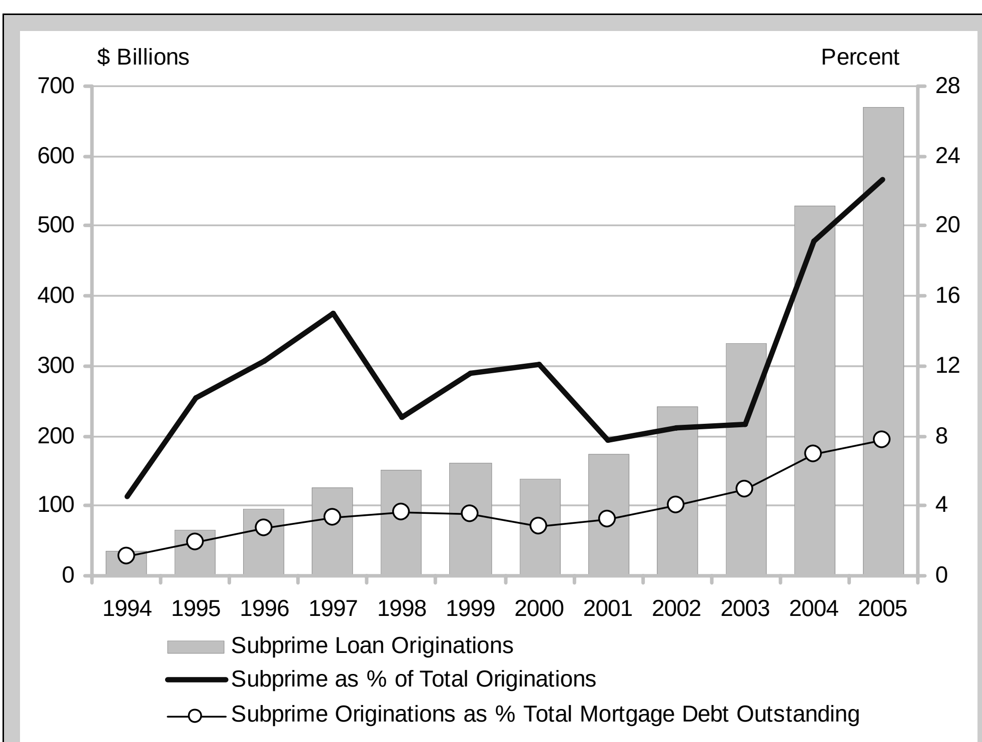
over the last decade the total number of renters in the U.S. actually declined for the first time since WWII. What this has meant is that each year almost 500,000 renter households have switched to owning. A movement in demand this large would certainly soften rents and put great pressure on prices, as Shiller (2006) and others have documented. The question then is what has changed in the last decade to explain homeownership growth?

The evolving demographic makeup of the country can account for only a small part of the overall homeownership increase. If age-specific ownership rates are weighed up by the changing age distribution (Exhibit 8, age-expected ownership rate), there is only a very gentle 1% increase in projected ownership from 1995 to 2005. Per capita income growth might also be an explanation, but incomes grew more in the 1960s and 1970s than in the last decade, so it too would be hard to use as an explanation. Since it is also difficult to disentangle factors in an aggregate time series, the recent changes in homeownership are examined across states from 2000 to 2005.[6] The findings reveal that there is absolutely no statistical relationship between the growth in ownership and that of per capita income. Something new and different seems to be driving homeownership in the last decade.

Most probably the growth in U.S. homeownership and hence prices has been driven by an explosive growth in credit availability, particularly the recent emergence of the so-called ''sub-prime'' lending market. The emergence of this market in the mid 1990s is perfectly timed with the beginning of the sharp rise in U.S. homeownership (Exhibit 9). Prior to this time, most households with poor credit ratings, or households seeking very aggressive underwriting were simply rationed out of the mortgage market. Since that time, ''risk-based pricing'' has provided ample credit in these situations—albeit at significantly higher rates—and often with subordinated loans. There seems to be no end to investors' appetite for securitized pools of such ''risky'' loans. By 2005, almost a quarter of all loans originated each year were sub-prime and the stock of sub-prime loans had reached 8% of total U.S. mortgage debt (Exhibit 9).

A number of studies have begun to explore the emergence of the sub-prime credit market. Some have investigated whether pricing adequately reflects the market's higher default rates, others the question of whether lenders are somehow predatory in attracting borrowers. In 2004, a special issue of the *Journal of Real Estate Finance and Economics* was devoted to this relatively new market. Several more

Exhibit 9 | Sub-prime Mortgage Originations



recent studies have tried to identify geographic differences in the penetration of this type of lending (e.g., Ho and Pennington-Cross, 2006) and borrower self-selection between credit markets (Ben-Shahar, 2006).

This study would ideally like to have some pure measure of sub-prime credit ''availability'' or lending ''supply'' *across* metropolitan markets to see if it can explain some of the differences in home price growth that cannot be explained by local level economic variables. Any effort in this direction, however, will immediately open up the difficult problem of causation. If there is a statistical relationship between the strength of sub-prime lending and residual house price growth, is it the case that an exogenous change in lending is generating the surge in demand and prices? Or, do the positive residuals reflect an unmeasured demand shock, which then logically leads to the growth in lending outlets and originations?

Home Mortgage Disclosure Act (HMDA) data, which contains basic information on all mortgages originated yearly in the U.S., is used to examine this issue. Since 2004, the HMDA data contains specific information about rate spreads and ''high-priced'' mortgages, as well as the originators who specialize therein. Using the 2004–2005 HMDA databases, and restricting the analysis to home purchase only loans, the variables listed in Exhibit 10 were developed for each of the 59 markets.[7]

In evaluating this data, one must be careful to realize that most variables do not at all represent the pure ''supply'' of credit; rather they represent some equilibrium outcome of supply against demand. Consider, for example, the fraction of loans that are ''high-priced'' (defined as carrying at least a 300 or 500 bps premium

Exhibit 10 | HMDA Data

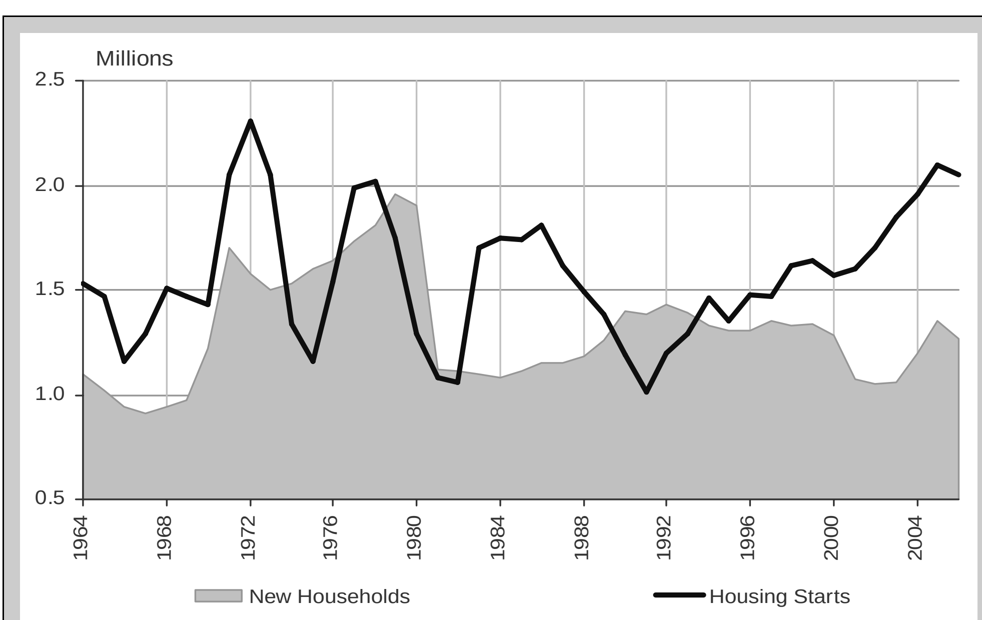
Variable	Definition
Variable l1	High-priced loans as % of all originations.
Variable l2	% loans originated by lenders defined as "sub-prime."
Variable l4	High-priced first lien loans as % of all high-priced originations.
Variable l5	High-priced first lien loans as % of all first lien originations.
Variable l6	High-priced first lien loans as % of all originations.
Variable l7	Loans secured by subordinate lien as % of all originations.
Variable l8	High-priced subordinate loans as % of all high-priced originations.
Variable l9	High-priced subordinate loans as % of all subordinate originations.
Variable l10	High-priced subordinate loans as % of all originations.
Variable lti	Average loan-to-income ratio.
Variable lti_sp	Average loan-to-income ratio for "sub-prime" loans.

above the appropriate conforming loan rate). If a particular market did experience a sudden influx of sub-prime credit supply, the average loan rate in that market could conceivably fall. An increase in sub-prime credit supply would generate both a ''mix effect'' to higher rate loans, but with more total credit available, the actual rate for any given type of loan could be lower. The overall outcome is hence in doubt.

Second and Investment Home Buying: 1998–2005

The current housing market has also seen a record number of housing sales to investors and second home buyers. This is a statistic that is extremely difficult to determine and generally is not available consistently over time. Such activity is estimated with two approaches. First, household formation can be compared with unit production. Exhibit 11 shows that in the last few years, total housing production has outstripped household formation by record amounts. Now to be sure, over the long run production should exceed household formation to allow for demolitions and replacements. In the last few years, however, production has been 30%–60% in excess of new households. With rising prices, purchases of continued excess production must reflect some form of ex-post housing demand from non-owner occupiers.

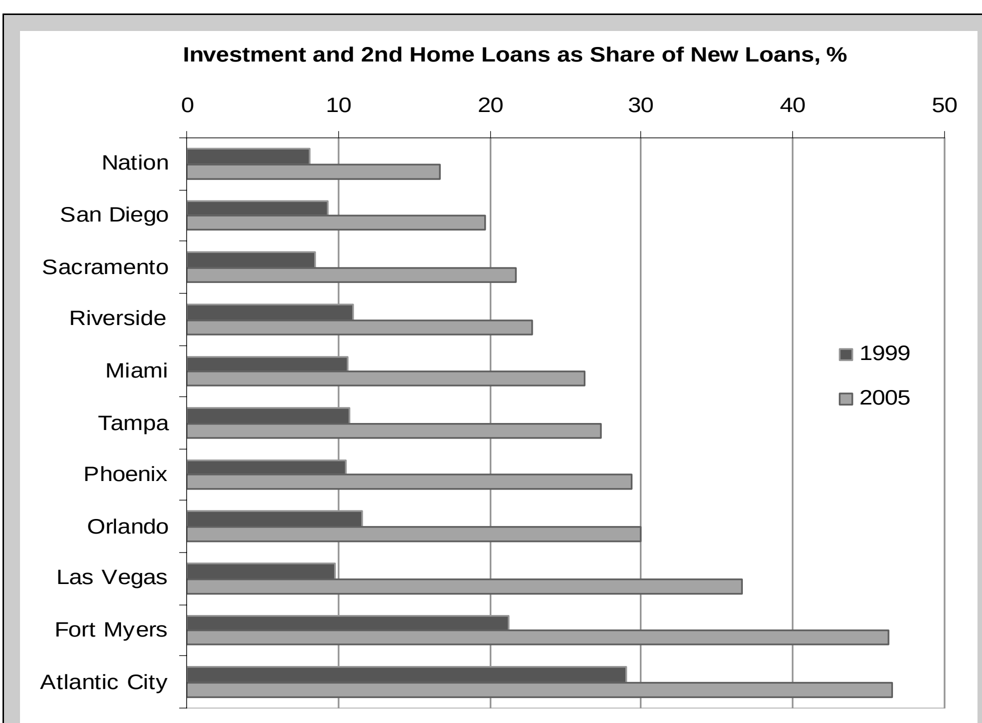
Exhibit 11 | Total Housing Construction and New Households



A more direct way to examine such buying is to look at loan origination records (again for home purchases), wherein the borrower must declare (by law) whether the financing is for purchase of a primary home, second home, or investment property. This data is available from Loan Performance Inc. and goes back to the late 1990s. The sum of ''investor'' and ''second home'' originations as a share of all originations has increased sharply since 1999 (Exhibit 12). It is important to note that these shares are orders of magnitude higher than the Census-reported share of ''seasonal'' housing units in the stock (nationally around 3.5%). The investor and second home share also has doubled in the last five years, despite the fact that the reported data is just for 1–4 family units. These shares and their growth would likely be greater had condominiums sales been included. In addition, they also miss second home purchases financed with expanded primary home loans or with all-equity.

What is important about second homes and investment properties is that such buying can more significantly affect a market's ''net supply'' or vacancy. Most primary home purchasers make ''churn'' moves from one house to another; hence, a transaction has little impact on market vacancy. A purchase/sale by a second home owner can subtract/add more directly to vacancy, and many studies indicate

Exhibit 12 | Investor and Second Home Buying



that prices are sensitive to small movements in housing vacancy and trading volume (Leung, Lau, and Leong, 2002).

It is interesting to speculate on the causes of this recent buying trend. Some suggest that it has resulted from baby-boomers ''pre-buying'' for retirement. Other's offer up that housing has become viewed as a ''safe investment'' relative to stocks and bonds, particularly after the 2000 decline in the stock market. With only aggregate time series, true explanations will be impossible to identify. In this paper, however, the findings reveal that second home buying is at least strongly related statistically to a particular market's price growth—beyond the influence of local economic fundamentals. Causal inference, however, is again a bit tricky. If residual home prices are rising for some other reason, that could generate investment buying in addition to the converse. There clearly can be joint causality between these variables.

A Cross-Sectional Analysis of 1998–2005 Forecast Errors

Appendix 3 presents two cross-section equations. In each the dependent variable is the 2005 forecast model error (percentage difference). On the RHS, the market size (2000 housing units, variable HU00) is always significant and the larger metropolitan areas have more residual error or ''excess'' appreciation. Similarly in virtually any equation that was specified, the share of loans in 2005 that are for second or investment homes (variable *INV2*) is always very significant. If the increase in this share from 1999–2005 is used, there are similar results. If the 1999 share is used, the coefficient is still highly significant but smaller in magnitude. In other words, no matter how measured, second home buying in a market is strongly associated with ''excess'' price appreciation.

Assessing the role of credit availability is more complicated. HUD has identified certain organizations as sub-prime credit originators, and when their average origination share (2004–2005) is used in the equation (the variable *L2*), the anticipated signs are obtained. Likewise if the share of loans in a market that are subordinated is used (variable *L7*), this also works as well. This is to be expected since many subordinated loans are sub-prime. What does not work well (in the sense of having an unanticipated negative signed coefficient) is any measure of the share of loans that carry a ''higher'' contract interest rate (such as variables *L1*, *L4–L6*, *L8*, or *L9*). In these cases, the interest rate may be endogenous. The final variable that works extremely well is the underwriting of loans; in particular, subordinated ones. While the HUD data does not have information on loan-to-value, it does record loan balance to income (LTI). When the ratio of loan balance to income is high, borrowers are being extended ever greater credit and this variable works well whether overall ratios (LTI) are examined or just that for subordinated loans (*LTI_SP*). What is encouraging is that these three measures together explain almost two-thirds of the residual forecast error. The statistical relationships are strong.

Discussion of Causality

Clearly, the fact that all 59 MSA have significant positive residuals on price forecasts made from 1998, suggests that something new and different has been driving the U.S. housing market in the last eight or more years. Local level changes in income, employment, and national interest rates just do not explain the movement in prices. The magnitude of the forecast errors in 2005 has also been shown to be *associated with* greater buying in a market for second home and investment use, and with the greater use of the sub-prime mortgage market and looser loan underwriting. Does this constitute at least some evidence of a ‘‘bubble’’?

If *joint* causation between the price residuals and these factors is assumed, then a bubble explanation is at least consistent with the statistics. As an example of joint causation, suppose a positive shock to credit supply in a market sets off home buying, which then in turn encourages further home buying (and the need for more credit outlets) because ownership is perceived as a good investment. The expansion of credit itself is just an overlooked or unmeasured change in fundamentals, but the subsequent investment buying is a potential bubble. Similarly, if baby-boomers are pre-buying second homes for true use in retirement, then there is another unmeasured positive change to demand fundamentals. However, if they are buying largely because housing has been perceived as a good investment vehicle in these markets, there is again the potential for a bubble.

The only way to rule out some form of a bubble would be to positively identify that the statistical relationships identified in this study are causal. Is it a surge in credit supply and some change in second home preferences that is responsible for greater home buying and hence ‘‘excess’’ price appreciation? If the converse is also true, it would imply that excess appreciation is driving demand—the traditional meaning of a bubble. Finding instruments or counter-factual evidence to rule out investment buying in this case, however, seems hopelessly difficult.

If not a bubble, the new factors driving the housing market make any prognosis for the future extremely difficult. If the market experiences some negative shock going forward—for example, from rising interest rates—this time there are two wild cards in the deck. Any economic slowdown could generate a much larger increase in foreclosures than happened in the past as sub-prime borrowers are more likely to default. In addition, if the market starts to turn down, investors could bailout from housing far more easily than primary home owners. With these two added factors, a ‘‘correction’’ could turn into a deeper housing slump than in the past.

Appendix 1

Actual to Forecasted, % Difference

Market	EQ1	EQ2	EQ3	EQ4	EQ5	EQ6
ALBUQU	28.0	27.8	26.4	18.3	19.0	10.1
ATLANT	33.7	34.0	33.2	22.0	24.8	32.0
AUSTIN	67.3	74.0	55.5	42.7	45.4	30.5
BALTIM	75.5	75.5	75.6	64.7	65.6	64.7
BIRMIN	17.7	18.2	17.5	7.4	1.2	5.2
BOSTON	82.3	68.1	82.4	82.8	82.3	76.6
CHICAG	52.2	51.9	49.1	37.5	5.4	14.7
CHRLTE	13.3	13.6	12.5	4.4	5.0	4.1
CINCIN	23.8	24.0	23.7	16.4	−2.0	−3.4
CLEVEL	20.3	19.9	19.7	11.7	−16.2	4.5
COLUMB	18.7	19.4	18.5	6.0	9.3	13.8
DALLAS	40.3	41.0	37.9	39.1	37.6	57.4
DAYTON	18.4	17.7	18.3	7.2	−18.4	−15.0
DENVER	47.8	42.6	47.7	31.1	28.2	107.1
DETROI	19.0	21.4	20.0	29.7	−16.7	−17.9
ELPASO	15.7	14.3	15.7	5.1	8.5	0.3
FORTLA	144.5	134.6	143.6	141.9	140.6	140.8
FORTWO	21.9	27.0	25.0	23.1	24.8	62.6
GREENS	22.1	21.3	19.7	4.8	8.1	1.9
GREENV	11.1	10.3	10.7	7.4	7.5	3.6
HONOLU	39.2	40.5	41.5	10.7	12.5	11.0
HOUSTO	39.5	40.9	38.0	29.9	95.3	85.7
INDIAN	10.8	11.7	10.8	5.0	−3.8	3.4
JACKSO	86.6	89.6	87.5	66.1	71.0	79.1
KANSAS	39.7	38.7	38.5	29.6	−45.7	−10.7
LANGEL	76.1	102.3	99.1	94.6	107.8	133.0
LISLAN	77.5	93.0	87.9	85.8	72.3	57.5
LOUISV	19.8	19.9	20.0	10.7	−23.7	−20.0
LVEGAS	91.2	94.1	93.2	100.1	95.7	97.8
MEMPHI	13.9	14.6	12.6	10.8	−20.7	0.1
MIAMI	109.7	111.4	110.9	117.3	112.0	107.5
MINNEA	59.9	61.5	60.8	58.3	24.8	22.2
NASHVI	20.9	20.4	20.5	1.6	3.3	−1.0

Appendix 1 (continued)
Actual to Forecasted, % Difference

Market	EQ1	EQ2	EQ3	EQ4	EQ5	EQ6
NEWYRK	67.0	71.0	67.7	67.2	25.3	25.6
NORFOL	65.0	64.9	65.3	55.2	65.2	34.3
OAKLAN	86.2	100.9	102.3	104.4	100.5	89.8
OKLAHO	27.1	32.2	27.1	21.6	19.6	154.6
ORANGE	84.6	115.6	117.4	127.7	94.5	145.4
ORLAND	88.7	90.1	88.2	85.9	86.3	62.8
PHILAD	60.9	59.0	61.2	61.6	61.1	30.9
PHOENI	100.5	100.6	100.6	99.9	100.2	77.7
PITTSB	18.2	19.7	18.2	7.5	21.3	20.4
PORTLA	35.6	35.8	35.3	29.1	26.2	−9.8
RALEIG	20.8	20.6	19.4	2.5	−0.7	−8.6
RICHMO	49.4	50.6	50.1	34.8	27.4	31.8
RIVERS	114.6	103.6	111.8	113.1	163.9	192.0
SACRAM	112.6	126.9	125.0	119.9	124.7	82.5
SALTLA	17.2	17.5	14.4	−1.8	−0.7	−10.9
SANTON	47.1	48.4	48.8	26.3	53.7	63.7
SDIEGO	149.6	153.4	148.3	146.9	147.9	77.7
SEATTL	65.8	71.7	67.1	63.1	62.6	52.7
SFRANC	56.1	54.5	68.2	91.1	69.0	73.7
SJOSE	48.5	57.9	86.9	91.3	83.4	−20.9
SLOUIS	41.5	46.8	38.5	44.8	14.6	31.8
TAMPA	97.4	96.2	97.8	94.1	79.1	81.9
TUCSON	82.7	83.5	78.0	69.2	71.6	68.8
TULSA	24.5	27.4	24.4	19.8	33.4	48.3
WASHIN	93.2	91.5	92.9	92.7	92.7	88.6
WBEACH	163.1	163.7	160.5	139.2	139.0	108.0

Appendix 2
Model 1

		Boston	Chicago	Phoenix
Panel A: Models 1A				
	Coefficient	−52.987	−4.799	−141.864
Constant	Std. Error	31.998	2.437	37.227
	T-Statistic	−1.656	−1.970	−3.811
	Coefficient	2.307	1.085	1.067
DLEMP	Std. Error	0.340	0.242	0.220
	T-Statistic	6.783	4.490	4.842
	Coefficient	0.622	0.101	0.708
LRINCEMP	Std. Error	0.272	0.209	0.212
	T-Statistic	2.287	0.482	3.336
	Coefficient	−13.127	−3.499	−13.891
LMRTG	Std. Error	4.600	3.751	3.860
	T-Statistic	−2.854	−0.933	−3.599
	Coefficient	1.000	0.974	1.000
RHO	Std. Error	0.020	0.030	0.006
	T-Statistic	51.238	33.013	181.851
Centered R^2		0.9925	0.9685	0.9609
Durbin-Watson		1.8802	2.2756	1.9520
Panel B: Model 1B				
	Coefficient	−4.737	0.345	−133.418
Constant	Std. Error	2.568	0.151	43.197
	T-Statistic	−1.845	2.283	−3.089
	Coefficient	0.278	1.825	0.046
RHPI (1)	Std. Error	0.120	0.101	0.110
	T-Statistic	2.324	18.017	0.420
	Coefficient	0.396	−1.466	−0.114
RHPI (2)	Std. Error	0.103	0.200	0.102
	T-Statistic	3.853	−7.327	−1.116
	Coefficient	0.081	0.891	0.005
RHPI (3)	Std Error	0.085	0.196	0.103
	T-Statistic	0.958	4.556	0.046
	Coefficient	−0.080	−0.301	0.127
RHPI (4)	Std. Error	0.084	0.092	0.100
	T-Statistic	−0.946	−3.270	1.263

Appendix 2 (continued)
Model 1

		Boston	Chicago	Phoenix
LEMP	Coefficient	0.588	−0.045	0.970
	Std. Error	0.316	0.043	0.289
	T-Statistic	1.862	−1.056	3.358
LRINCEMP	Coefficient	0.341	0.044	0.699
	Std. Error	0.196	0.067	0.232
	T-Statistic	1.739	0.655	3.016
LMRTG	Coefficient	−7.274	−3.290	−14.959
	Std. Error	3.331	0.955	4.135
	T-Statistic	−2.183	−3.447	−3.618
Centered R^2		0.9963	0.9798	0.9631
Durbin-Watson		1.7712	2.0270	2.0329

Appendix 2
Model 2

		Boston	Chicago	Phoenix
Panel A: Model 2A				
Constant	Coefficient	−0.004	−0.002	−0.014
	Std. Error	0.003	0.002	0.004
	T-Statistic	−1.227	−0.976	−3.835
DLEMP	Coefficient	2.308	1.192	1.067
	Std. Error	0.338	0.266	0.219
	T-Statistic	6.831	4.487	4.875
DLRINCE	Coefficient	0.622	0.170	0.708
	Std. Error	0.270	0.217	0.211
	T-Statistic	2.304	0.784	3.359
DLMRTG	Coefficient	−13.132	−4.070	−13.905
	Std. Error	4.570	3.761	3.836
	T-Statistic	−2.874	−1.082	−3.625
Centered R^2		0.3921	0.1964	0.3009
Durbin-Watson		1.8807	2.3678	1.9523

Appendix 2 (continued)
Model 2

		Boston	Chicago	Phoenix
Panel B: Model 2B				
	Coefficient	−0.002	−0.003	−0.013
Constant	Std. Error	0.002	0.002	0.004
	T-Statistic	−0.866	−1.309	−3.099
	Coefficient	0.660	1.055	0.971
DLEMP	Std. Error	0.322	0.327	0.287
	T-Statistic	2.050	3.232	3.383
	Coefficient	0.385	0.145	0.700
DLRINCE	Std. Error	0.198	0.203	0.230
	T-Statistic	1.942	0.715	3.038
	Coefficient	−7.696	−5.017	−14.970
DLMRTG	Std. Error	3.327	3.478	4.107
	T-Statistic	−2.314	−1.443	−3.645
	Coefficient	0.269	0.007	0.046
DRHPI (1)	Std. Error	0.119	0.114	0.109
	T-Statistic	2.264	0.065	0.423
	Coefficient	0.390	0.197	−0.114
DRHPI (2)	Std. Error	0.102	0.101	0.102
	T-Statistic	3.833	1.948	−1.124
	Coefficient	0.080	−0.266	0.005
DRHPI (3)	Std. Error	0.084	0.100	0.103
	T-Statistic	0.961	−2.671	0.045
	Coefficient	−0.078	0.245	0.127
DRHPI (4)	Std. Error	0.083	0.100	0.100
	T-Statistic	−0.944	2.461	1.271
Centered R^2		0.6328	0.3624	0.3181
Durbin-Watson		1.7848	2.0444	2.0331

Appendix 3

		Equation 1	Equation 2
	Coefficient	−136.612	−166.498
Constant	Std. Error	26.405	30.006
	T-Statistic	−5.174	−5.549
	Coefficient	1.04E-05	8.80E-06
HU00	Std. Error	5.40E-06	5.30E-06
	T-Statistic	1.922	1.649
	Coefficient	3.137	3.114
INV2	Std. Error	0.561	0.514
	T-Statistic	5.596	6.062
	Coefficient	57.451	68.920
LTI_SP	Std. Error	13.670	13.997
	T-Statistic	4.203	4.924
	Coefficient	0.961	—
SP2	Std Error	0.716	—
	T-Statistic	1.342	—
	Coefficient	—	1.641
SP7	Std Error	—	0.721
	T-Statistic	—	2.278
Centered R^2		0.6162	0.6382

Notes:
LTI_SP: Loan-to-income ratio for subprime loans as identified by HUD.
SP2: Subprime loan share identified by HUD based on originators.
SP7: Subprime loan share identified by share that are subordinate loans.
INV2: Share of 2005 home purchase loans made for 2nd or investment use.
HU00: 2000 number of housing units in MSA.

Endnotes

1 Prices rising faster than what fundamentals dictate could be due to two reasons. First, there could simply be overlooked fundamentals. Second, speculative demand could be resulting from the rise in prices. In the classic definition, the first does not constitute a bubble, while the second does.

2 All prices used in this paper are based upon the OFHEO price index for the U.S. or selected MSA. The OFHEO index uses repeated sales of identical properties. This approach is known to have some biases, as discussed by Goetzmann (19920 and most recently Harding, Rosenthal, and Sirmans (2007). These biases are likely to be less than those inherent in the alternative—a median sales price index produced by the National Association of Realtors.

[3] The models could be estimated with data through 2005, but this would create some bias for the experiment. The forecasts of such a model from 1998 to 2005 would contain the influence of the hypothesized recent ''unknown'' factors in so far as they are partially correlated with economic fundamentals and hence alter the parameters of the latter.

[4] Malpezzi (1996) has produced such an instrument for measuring regulation, but it is available for only half of the markets in this study and its construction has been the subject of debate.

[5] In the top frame of Exhibit 6, prices are in natural logs of an index that starts at 100 in 1978. The bottom frame depicts the growth rate of employment and the mortgage interest rate.

[6] Homeownership at the MSA level is measured only with the decade census. Using state level data for the change from 2000 to 2005: $\Delta HO = 1.01 - .0003\Delta Y$, $R^2 = .002$. (ΔHO and ΔY are measured as cumulative five-year percentage changes.)

[7] The HMDA database since the early 1990s has maintained information on mortgages originated by entities specializing in ''sub-prime'' loans. Since this list is quite consistent year to year, it could be used to identify a ''sub-prime share'' of loans in each market and year. Since 2004, the HMDA database also contains information on loan pricing. This study uses the list of originators and ''higher priced'' loans in only the 2004 and 2005 HMDA databases, averaging them for the cross-section comparison. In the construction of all variables, only home purchase loans were included. See Avery and Canner (2005).

References

Avery, R. and G. Canner. New Information Reported under HMDA and its Application in Fair Lending Enforcement. *Federal Reserve Bulletin*, 2005, Summer, 344–94.

Ben-Shahar, D. Screening Mortgage Default Risk: A Unified Theoretical Framework. *Journal of Real Estate Research*, 2006, 28:3, 215–41.

Capozza, D., P. Hendershott, and C. Mack. An Anatomy of Price Dynamics in Illiquid Markets: Analysis and Evidence from Local Housing Markets. *Real Estate Economics*, 2004, 32:1, 1–21.

DiPasquale, D. and W. Wheaton. Housing Market Dynamics and the Future of House Prices. *Journal of Urban Economics*, 1994, 35:2, 1–27.

Evenson, B. House Price Variation across Local Markets: A Supply Side View. University of Nevada at Las Vegas Working Paper, 2004.

Gallin, J. The Long Run Relationship between House Prices and Income. *Real Estate Economics*, 2006, 34:3, 417–39.

Glaeser, E., J. Gyrouko, and Raven E. Saks. Why Have Housing Prices Gone Up? NBER Working Paper 11129, February 2005.

Goetzmann, W.N. The Accuracy of Real Estate Indices: Repeat Sales Estimators. *Journal of Real Estate Finance and Economics*, 1992, 5, 5–54.

Granger, C.W.J. and P. Newbold. Spurious Regressions in Econometrics. *Journal of Econometrics*, 1974, 26, 1045–66.

Harding, J., S. Rosenthal, and C.F. Sirmans. Depreciation of Housing Capital, Maintenance, and House Price Inflation. *Journal of Urban Economics*, 2007, 61:2, 567–87.

Harter-Dreiman, M., Drawing Inferences about Housing Supply Elasticity from House Price Responses to Income Shocks. *Journal of Urban Economics*, 2004, 55:2, 316–37.

Hildreth, C. and J.Y. Lu. Demand Relations with Autocorrelated Disturbances. Technical Bulletin 276, Dept. of Agricultural Economics, Michigan State University, 1960.

Himmelberg, C., C. Mayer, and T. Sinai. Assessing High House Prices: Bubbles, Fundamentals and Misperceptions. Federal Reserve Bank of New York Bulletin, No. 218, 2005, September.

Ho, G. and A. Pennington-Cross. The Impact of Local Predatory Lending Laws on the Flow of Sub-prime Credit. *Journal of Urban Economics*, 2006, 60:2, 210–29.

Leung, C., G. Lau, and Y. Leong. Testing Alternative Theories of the Property Price-Trading Volume Correlation. *Journal of Real Estate Research*, 2003, 23:3, 253–65

Malpezzi, S. Housing Prices, Externalities, and Regulation in U.S. Metropolitan Areas. *Journal of Housing Research*, 1996, 7:2, 209–41.

Shiller, R. Long-Term Perspective on the Current Boom in Home Prices. Berkeley Electronic Press, 2006, March.

Smith, M. and G. Smith. Bubble, Bubble, Where's the Housing Bubble? Brookings Panel on Economic Activity, 2006, March.

The authors are indebted to the MIT Center for Real Estate, Loan Performance Inc. and Torto Wheaton Research. They remain responsible for all results and conclusions derived there from.

William C. Wheaton, Massachusetts Institute of Technology, Cambridge, MA 02139 or wheaton@mit.edu.

Gleb Nechayev, Torto Wheaton Research, Boston, MA 02110-3036 or gnechayev@twr.com.