UC San Diego
Recent Work

Title
After the Fall: An Ex Post Characterization of Housing Price Declines Across Metropolitan Areas

Permalink
https://escholarship.org/uc/item/7ng9d927

Authors
Carson, Richard T
Dastrup, Samuel R.

Publication Date
2009-12-03

After the Fall
An Ex Post Characterization of Housing Price Declines Across Metropolitan Areas

Richard T. Carson
Samuel R. Dastrup

Department of Economics
University of California, San Diego

December 3, 2009

Abstract

Housing prices have plummeted across the United States. This paper examines differences in the magnitude of housing price decreases across metropolitan areas. A relatively small number of housing market variables observable before the fall are capable of explaining over 70% of the considerable variation in price declines. An additional nonparametric analysis suggests that exceeding particular thresholds for some of the key predictors is associated with much larger price drops. These findings are consistent with historical price patterns and raise questions about the validity of mortgage pricing and risk diversification norms in the US. The analysis points to a set of stylized facts concerning the housing price bubble that need to be explained and suggests fruitful hypotheses for understanding the dramatic housing price declines.

The value of the U.S. housing stock fell $4.4 trillion dollars from 2006 to the first quarter of 2009,[1] leading to considerable turmoil in financial markets across the world. This fall is widely believed to be one of the primary contributing factors behind the current economic crisis. Economic commentators, particularly those focused on the macro economy and the health of financial institutions, tend to treat this fall in housing prices as a nationwide phenomenon.[2] Casual inspection of the magnitude of the change in housing prices across metropolitan areas suggests that the changes are anything but uniform. Across 358 areas the magnitude of the fall varies from essentially zero to over sixty percent. Figure 1 illustrates this variation in a map of the MSAs. In this paper, we investigate whether it is possible to predict differences in percentage change in housing prices across metropolitan areas given a set of predictors available before the fall.

Our analysis is primarily descriptive. While the variables we use are predetermined in the sense of being *ex ante*, the analysis is of course *ex post*. No claims of causality are being made. In this sense, our paper is much less ambitious than several working papers currently in circulation or recently published (e.g.,Glaeser, Gyourko & Saiz 2008, Glaeser, Gyourko 2007, Himmelberg, Mayer & Sinai 2005) that attempt to explain the entire dynamics of housing price changes. Our contribution is to present a set of stylized facts concerning differences in housing price declines across metropolitan areas that need to be explained and to put forward a few initial hypotheses.

Remarkably, over 70% of the variation in housing price declines across metropolitan areas can be accounted for by a relatively small number of variables that were readily available before the fall, including previous appreciation rates, building trends, the prevalence of subprime lending, and changes in median income levels. A nonparametric analysis suggests that exceeding particular thresholds for some of these key predictor variables is associated with much larger price drops. The relationships we find are also consistent with historical episodes. It would be possible to condition lending practices and portfolio risk assessment on these metropolitan level indicators.

Perhaps most importantly, our regression model suggests that lenders drew the wrong implications from increases in home prices between 2000 and 2005. Instead of indicating decreased risk, greater price appreciation is associated with increases in the magnitude of subsequent price declines. We also show that local

---

[1] Federal Reserve Board's Flow of Funds Accounts, September 17, 2009, Table B.100 (Line #3).

[2] Popular press examples of this tone include a March 10, 2009 Wall Street Journal Opinion column "The Fed Didn't Cause the Housing Bubble" by Alan Greenspan. For a set of academic papers that largely adopts this perspective, but which is much more nuanced with respect to the potential importance of local conditions, see the recent B.E. Journal of Economic Analysis and Policy symposium edited by Gabriel, Quigley and Rosenthal (2009).

**Figure 1: Variation in HPI declines for US MSAs**

%DropHPI by MSA
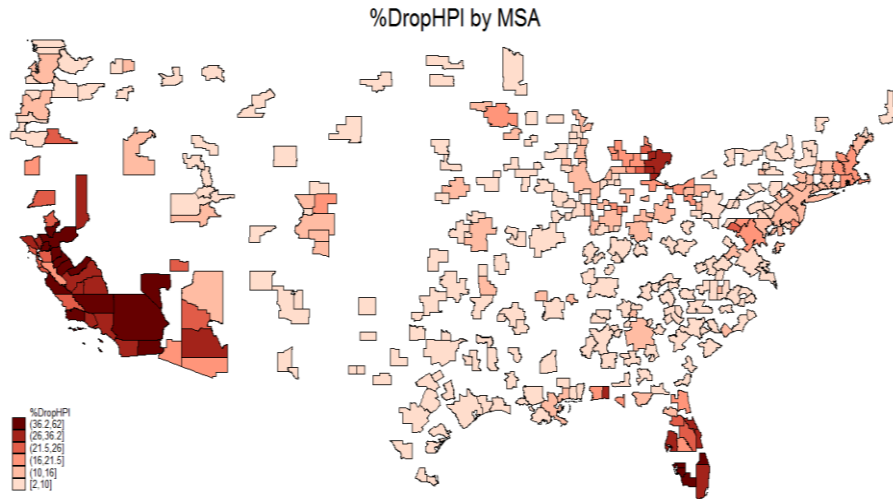
housing market variables are better predictors of declines than regional location, which provides intuition into why common approaches taken to diversify risks in bundling mortgages were unsuccessful.

II. The Null Hypothesis

There is a simple version of the null hypothesis: there is an average drop in housing prices across metropolitan areas but dispersion around that average is unrelated to potentially actionable variables. There are, of course, other potential null hypotheses. One is the possibility of differential regional impacts. This notion lay behind the common diversification strategy in bundling mortgages, with mortgages in the Northeast and Pacific census divisions often thought to be riskier. It should be clear though that this strategy is imperfect at best; as an example, housing prices in Portland Oregon have fallen only 12.6% from their peak while housing prices in Stockton California have fallen 54.7%.[3] Many other pairs of cities within the same U.S. census division show similarly large divergences, illustrating the difficulty with the traditional geographic diversification strategy in assembling home loan portfolios as a way of reducing risk. Figure 2 shows two groups of geographically close metropolitan statistical areas with disparate house price experiences.

---

[3] Price changes based on OFHEO housing price indexes (HPIs) deflated by the non-housing consumer price index, as discussed below.

Figure 2: Examples of HPI variation across MSAs
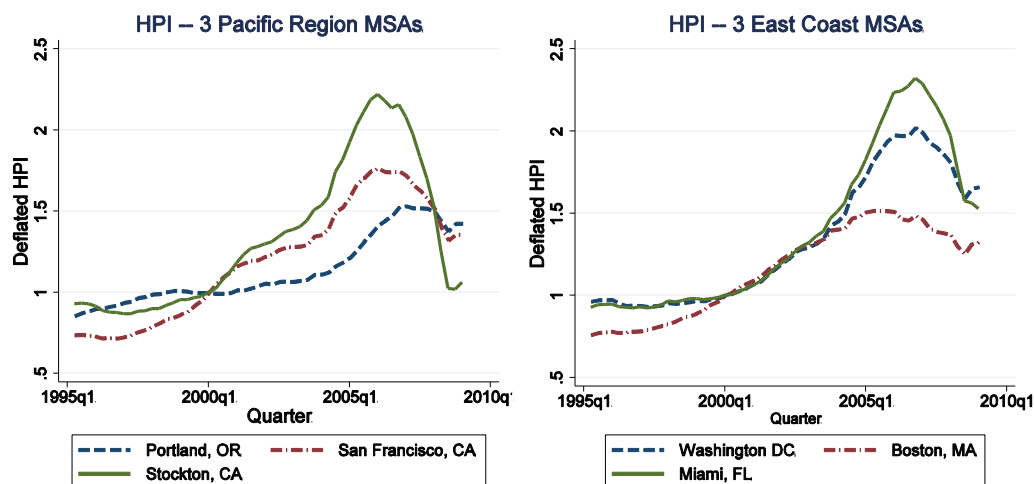
We look at several alternative hypotheses. The first takes the position that the run-up in housing prices post 2000 was in some ways artificial and not driven by economic fundamentals. This suggests that the larger the run-up in housing prices the larger the likely fall. Closely related to this alternative would be variants that revolve around changes in the shape of the distribution of home prices. What we have in mind here are changes in the skewness of the distribution and the level of dispersion. Changing fundamentals of household income and population are also a source of variation across markets that may predict the size of the fall. Another alternative is that overbuilding in a particular metropolitan area relative to population or workforce growth might be a contributing factor that predicts the magnitude of subsequent price declines. Another aspect of home prices that might be important is simply their absolute level as reflected by some summary statistics such as the median or mean, alone or in conjunction with a similar measure of income. One of the most popular potential villains is the fraction of loans that carried substantially higher than "normal" interest rates, reflecting the credit worthiness of their borrowers. The speculative home flipper is a possible villain whose activities might be proxied by the change in the percent of owner occupied housing. Efforts to drive up home ownership have been blamed for the residential real estate bubble so the absolute level of home ownership might be an important predictor. The specific demographics of a metropolitan area, including the poverty rate, ethnic composition, the percent retired, and educational attainment levels, might also have some explanatory power. Finally, in addition to geographic grouping by U.S. Census Bureau divisions, it is possible to look at whether a metropolitan area's size is related to the drop in housing prices.

3

Figure 3: Histogram of *%DropHPI*
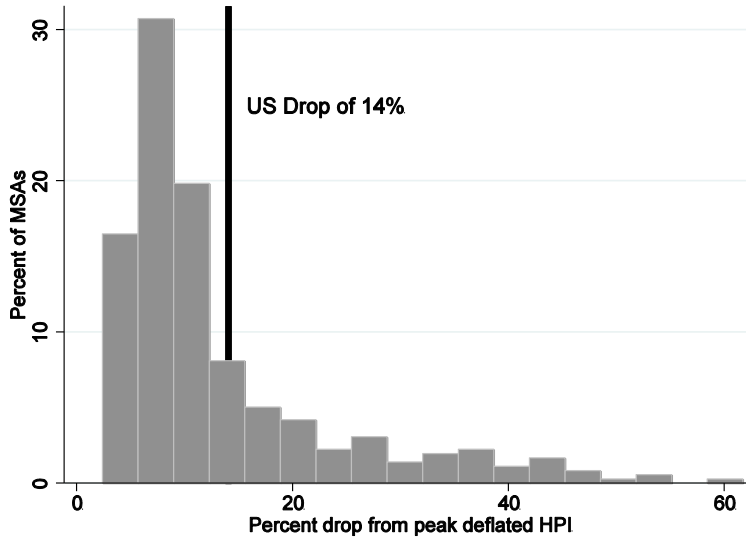
III. Data

We take as our dependent variable, *%DropHPI*, the percentage drop in a market's housing price index (HPI) defined as the percent decline from its highest point between the first quarter 2000 and the third quarter 2008 to its subsequent lowest point. We use the Office of Federal Housing Enterprise Oversight (OFHEO) metropolitan statistical areas and divisions all-transactions index, since this index is available for all Office of Management and Budget metropolitan statistical areas (MSAs). MSAs are defined as groups of counties economically integrated to a core urban area as measured by commuting ties. This definition aligns well with the concept of a market for housing and is the level of aggregation for our analysis. A quarterly index is reported for the 363 metropolitan areas, 358 of which we are able to match to other data sources described below.[4] We use the OFHEO HPI rather than some of its competitors, such as the well-known Case-

---

[4] HPI data is available at www.fhfa.gov; we use the second quarter 2009 revision of the data. OHFEO was subsumed by the Federal Housing Finance Agency in the fall of 2008. To combine the HPI data with other sources, HPI metropolitan divisions are averaged to metropolitan areas for the 11 areas that are subdivided into divisions. While the index is based on conforming loans that are limited by size of the loan, comparisons of histograms of home values used to construct the indexes reported in a 2005 OFHEO bulletin *Inclusion of Expensive Homes in the HPI* (http://www.fhfa.gov/webfiles/1057/Focus2Q05.pdf) to data reported in the Census and American Community Survey indicates that, at least for California coastal cities, the mix of homes in the HPI reflects the underlying distribution of homes.

Figure 4: *%DropHPI* against *%GainHPI*
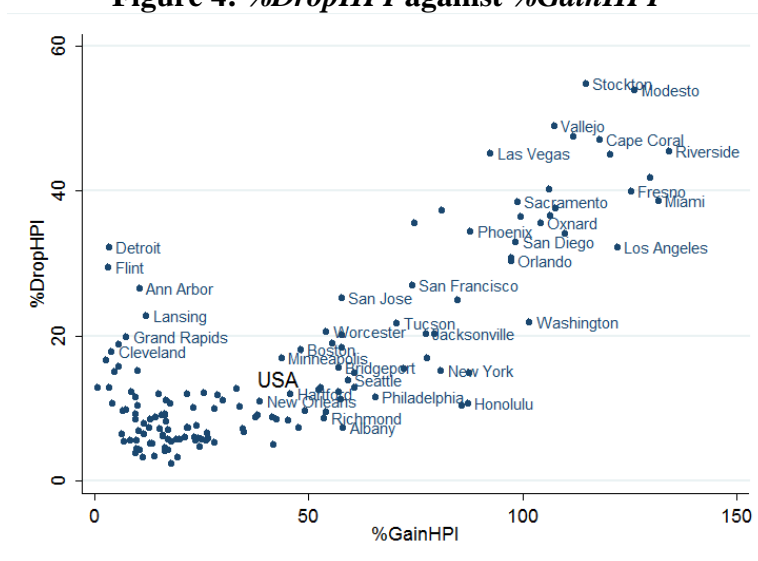
Shiller index, largely because it is available for a much greater number of metropolitan areas.[5] There is considerable variation in *%DropHPI*. It ranges from 2.39 to 61.72 with a mean of 13.35 and a standard deviation of 10.75, implying a coefficient of variation of just under 1 and a distribution with a long right tail. A histogram of this distribution is shown in Figure 3.

   The first predictor, *%GainHPI,* is defined as the relative gain in real HPI between the first quarter of 2000 and fourth quarter of 2006. We use 2006Q4 instead of the MSA specific peak in HPI to avoid the possibility of an artificial correlation with *%DropHPI* due to measurement error in the maximum HPI in each MSA. Further, using a common quarter for all MSAs lessens the importance of hindsight knowledge of the exact date when the peak was reached. While many MSA HPIs peaked in 2006Q4, some had already begun to decline and the national series reached its peak in 2007Q2. Because HPI in most markets is relatively flat between 2006Q2 and 2007Q2, the choice of an exact date is somewhat arbitrary and largely inconsequential in terms of our estimates.[6] *%GainHPI* can be viewed

---

[5] The average correlation between the OFHEO and the Case-Shiller indexes for the cities available for the Case-Shiller series is 0.93 for our sample period starting in 2000 Q1. This correlation would be higher except for marked divergences in the two indices in the post-bust period for Atlanta, Chicago, Dallas, and Denver.

[6] The elasticity of *%DropHPI* with respect to *%GainHPI* defined with respect to 2006Q2 to 2007Q2 differs by at most 0.05. Using the maximum gain in HPI as a regressor has slightly more predictive power than using any specific date; however, qualitative conclusions do not change and quantitative differences are small across definitions. The elasticity of *%DropHPI* with respect to the maximum gain is 0.73.

as representing a shift in a market's home price distribution. Figure 4 shows the predictive relationship of *%GainHPI* and *%DropHPI*.[7] Changes in the shape of market price distributions are captured using Census 2000 estimates of quartiles of reported home values as a baseline and calculating percent changes to the American Community Survey 2005-2007 three year average estimates (ACS).[8] The quartile values are deflated by HPI to capture the changes in the median and interquartile range, holding the mean constant. These variables are *%ΔMedHPI* and *%ΔIQHPI*.[9]

Other variables captured from the 2000 Census for each MSA include median house price level, *MedHPrice($10K)*; median income, *MedInc($1K)*; percent of single family homes that are owner occupied, *%OwnerOcc*; percent of households in poverty, *%Poverty*; racial compostion; average education levels; and population. We impute the percentage of the population that is retired, *%Retired,* as the over 65 population not in the labor force. Percentage change variables are constructed for median income, owner occupancy rates, percent retired, and population using the parallel variables in the ACS. The Census Bureau also collects data on building permits issued by metropolitan area and reports annual population estimates for each MSA. We use these data to create a variable that captures building activity relative to population growth, *%ExcessPermits*. We calculate the difference between the number of units that would have grown the 2000 stock at the same rate as the population and the actual number of permits issued from 2000 to 2005 as a percent of the reported 2000 Census housing stock.[10] Lastly, we calculate the percent of mortgages in each MSA reported to be high priced (have a high interest rate relative to a Treasury benchmark) in the 2005 Housing Mortgage Disclosure Act data MSA level reports as our measure of the prevalence of subprime lending around the peak of housing prices, *%Subprime*.[11]

---

[7] Cities with low gains and also drops of 20% or more seem to be concentrated around Detroit, in areas with economies declining with the big three US auto makers.

[8] Data is available at http://factfinder.census.gov. ACS data is reported by MSA. We aggregate the Census 2000 data from county reports to current MSAs using the geographic relationship files at http://www.census.gov/population/www/metroareas/metroarea.html.

[9] The change in mean price of properties sold may represent a change in the composition of properties sold that is not entirely captured in the OFHEO value weighting methodology. See Calhoun (1996) and Leventis (2008) for an overview of the weighting methodology and a discussion of the differences between the OFHEO and the S&P/Case-Shiller indexes.

[10] That is: $\%ExcessPermits = \left(\frac{Population_{2005}}{Population_{2000}} - 1\right) - \frac{\Sigma_{2000}^{2005}\ Permits}{Housing\ Stock_{2000}}$. Quarterly values for permit data were scaled to reflect the revised annual data. Similar estimates are found using Bureau of Labor Statistics labor force series instead of population.

[11] See Mayer and Pence (2008) for a discussion of data measuring subprime lending. The HMDA reports are available from http://www.ffiec.gov.

In Table 1, we present summary statistics for the 358 MSAs for which all data is available.[12] Substantial variation is evident not only in the magnitude of housing price drops across areas, but in our predictors also. For example, the percent increase in prices from 2000Q1 to 2006Q4, *%GainHPI,* has a mean of 37.65 and ranges from -4.7% to 137.38%, with a standard deviation of 33.32% while building relative to population growth, *%ExcessPermits,* ranges from areas with substantial underbuilding at -10.5% to areas with substantial overbuilding at 14.7%. The prevalence of subprime lending also varies across MSAs, with a 7.8% standard deviation around a mean of 22.5% of loans.

A variety of factors that co-move with house prices are not included in our analysis, primarily due to the difficulty of obtaining relevant data for a full complement of MSAs. A number of studies examine the relationship of user costs to housing prices (see Mayer and Hubbard (2008) for a discussion and review). User cost analyses compare measures of the after tax cost of owning a home to house price to rent ratios or house prices directly. Construction of reliable measures of MSA specific user costs for marginal home buyers is not, however, something we undertake in this paper. The Census and ACS do include a cost of ownership question, which is an incomplete measure of average user cost for homeowners. Another popular measure of housing market changes is an"affordability index," typically a comparison of median prices to median incomes. Figure 5 depicts the high correlation of these various measures of housing price changes, where *%dOwnCost* is the percent change in ownership costs for households with a mortgage from the Census to the ACS, and *%dAffordable* is the percent change from the Census to the ACS of the median price to median income ratio. Similar results as those presented below are obtained when substituting either of these variables for *%GainHPI* in our regressions.[13]

---

[12] Of the 363 OMB defined MSAs, one has an HPI series that does not begin until after Q1 2000 (Hinesville, GA), two lack building permit data in the HUD SOCDS (Lake Havasu City, AZ and Palm Coast, FL), one lacks HMDA data (Sebastian, FL), and one lacks housing statistics in the ACS (Carson City, NV). Also, labor force growth data for the geographically overlapping NECTAs were used for a few New England MSAs for which the BLS reports data by NECTA and not MSA.

[13] As a single predictor, *%dAffordability* has more explanatory power than *%GainHPI*. Results using more comprehensive sets of covariates are reasonably similar. We have chosen to present results using *%GainHPI* and *%ΔMedInc($1k)* as it facilitates interpretation relative to *%dAffordability,* which effectively scales *%GainHPI* by an income measure.

Table 1: Data summary statistics for 358 MSAs in the analysis

	Mean	Std Dev	Min	Max
%DropHPI	13.35	10.75	2.39	61.72
%GainHPI	37.65	33.32	-4.71	137.38
%ΔMedHPI	1.79	7.44	-22.94	30.40
%ΔIQHPI	16.00	13.71	-33.43	63.18
%ExcessPermits	5.03	2.82	-10.48	14.69
MedInc($1K)	40.05	6.71	24.86	74.34
%ΔMedInc($1K)	17.68	7.05	-1.66	41.24
MedHPrice($10K)	10.70	4.42	4.28	42.26
%Subprime	22.53	7.76	3.88	52.64
%OwnerOcc	61.81	5.50	34.38	77.06
%ΔOwnerOcc	-1.79	2.99	-11.09	8.41
%ΔPopulation	6.50	7.09	-15.67	40.31
%Poverty	12.20	4.13	4.38	35.45
%Black	10.06	10.59	0.11	48.52
%Hispanic	9.29	14.30	0.36	94.40
%Asian	2.01	3.28	0.19	45.37
%Retired	11.00	2.86	4.62	28.60
%ΔRetired	0.52	5.06	-13.50	20.06
%HighSchool	58.66	5.89	37.54	69.88
%College	22.53	7.28	10.33	52.38
Pop<250k	0.53			
250k<Pop<750k	0.30			
Pop>750k	0.17			
NewEngland	0.04			
MiddleAtlantic	0.08			
EastNorthCentral	0.17			
WestNorthCentral	0.08			
SouthAtlantic	0.21			
EastSouthCentral	0.08			
WestSouthCentral	0.12			
Mountain	0.09			
Pacific	0.13			

Figure 5: Scatterplots of price increase variables and %DropHPI
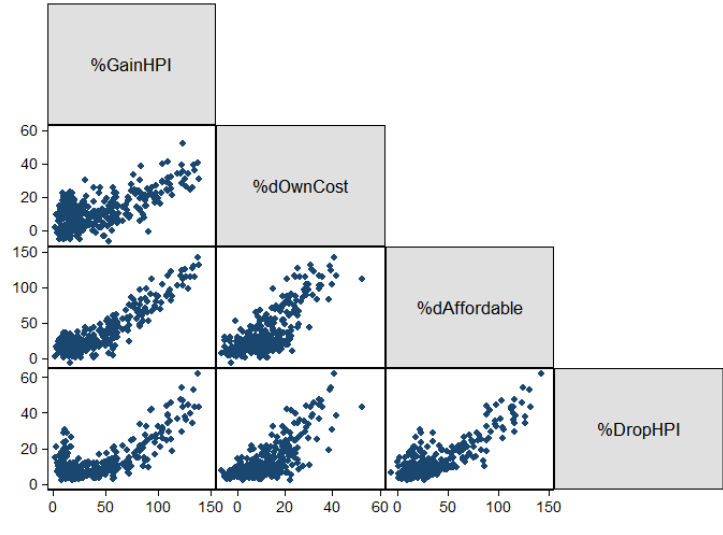
While we include a measure of building permit issuance relative to population growth in our set of predictors, we do not include more detailed characterizations of local regulation of housing supply (Gyourko, Saiz & Summers 2008) or of geographic supply constraints (Glaeser, Gyourko & Saiz 2008, Saiz 2008). Other variables for which MSA level data are not available for a full sample of areas include higher frequency income and employment data, as well as data for lending trends concerning household debt and equity levels and mortgage terms (e.g., prevalence of exotic option ARMs).

IV. Parametric Estimates

To explore the hypothesized relationships between the variables in our dataset, we estimate a series of nested linear regressions. In each case we predict *%DropHPI*, first using *%GainHPI* as a predictor and then progressively including more explanatory variables. We also estimate models based entirely on geography and demographic variation. Estimates are provided in Table 2, which shows the models with housing market variables, and Table 3, which reports those with geographic and demographic controls. All are OLS estimates with White standard errors.[14]

---

[14] Use of STATA's robust regression routine (rreg) based on Tukey's biweight function (Li 1985) on the full model suggests quite similar parameter estimates to the OLS estimates presented here, indicating that a small number of outliers are not driving our results.

The results in Tables 2 and 3 provide several insights. First, one variable, the magnitude *%GainHPI,* explains 55% of the variance in the drop (column 1) and the elasticity [evaluated throughout at the vector of mean covariate values] from the full model (column 8) implies that two thirds of the post first quarter 2000 gain is given up conditional on the other covariates.[15] This finding indicates that rather than large increases in home prices in an area being associated with decreased lender risk, the opposite was true.

It is worth a deeper investigation as to whether this first stylized fact, that greater prior appreciation predicts greater price declines, indicates a causal relationship with respect to *%GainHPI* (or one of its highly correlated cousins). Glaeser, Gyourko, and Saiz document an "enormous mean reversion" from the average 14.6% price appreciation for the 1982-1989 period in their sample of 79 MSAs and report that "for every percentage point of growth in a city's housing prices between 1982 and 1989, prices declined on average 0.33 percentage points between 1989 and [1996]." While we are unable to estimate our full model on prior housing booms due to a lack of historical data for many of our variables and MSAs, we show in section VII that a limited version of the model based on past price increases predicts the magnitude of historical real price declines. Even if this relationship between appreciation rates and subsequent declines is not causal, its persistence in the available data for the two decades prior to this housing bubble suggests the possibility of conditioning lending on it in some fashion; for example, requiring higher down payments in markets with rapid appreciation might simultaneously protect lenders and dampen booms.[16]

Our second stylized fact is that changes in the shape of the housing price distribution are also predictive of *%DropHPI*. The addition of two indicators of the change in shape of the MSA housing price distribution explains an additional 8% of the variance (column 2). In the full model (column 8), the elasticity of making the distribution less right skewed by shifting the median (*%ΔMedHPI*) outward is 0.03, which when coupled with the mean shift of 1.79%, is associated with a small drop in *%DropHPI*. However, *%ΔMedHPI* ranges from -23% to 30% suggesting that *%DropHPI* falls by 1.5 as one goes in the sample from a change in the median price that make the distribution considerably more right skewed to one making the distribution considerably less right skewed. An increase in dispersion, represented by the change in the interquartile range, *%ΔIQHPI*, is associated with a smaller change in *%DropHPI* with an elasticity of -0.04. At the

---

[15] The parameter estimates on *%GainHPI* are remarkably stable as additional variables are added to the model suggesting that this effect is largely orthogonal to the other explanatory variables. This would be expected by construction for the other variables related to the change in the housing price distribution but not for other covariates in the model.

[16] See Kelly (2009) for evidence that high down payments are associated with lower default levels.

Table 2: Housing variables regression results

	1	2	3	4	5	6	Elasticity at Mean
%GainHPI	0.240***	0.252***	0.255***	0.290***	0.287***	0.262***	0.738
	(0.016)	(0.015)	(0.014)	(0.016)	(0.0184)	(0.016)	(0.046)
%ΔMedHPI		0.384***	0.358***	0.442***	0.439***	0.325***	0.043
		(0.048)	(0.047)	(0.052)	(0.053)	(0.049)	(0.006)
%ΔIQHPI		-0.0866***	-0.089***	-0.073**	-0.075**	-0.047*	-0.066
		(0.0287)	(0.027)	(0.028)	(0.031)	(0.028)	(0.034)
%ExcessPermits			0.571***	0.440***	0.432***	0.500***	0.189
			(0.136)	(0.142)	(0.141)	(0.139)	(0.052)
%ΔMedInc($1k)				-0.326***	-0.326***	-0.177***	-0.235
				(0.062)	(0.063)	(0.059)	(0.077)
MedInc($1k)				0.013	-0.001	0.052	0.156
				(0.056)	(0.0793)	(0.098)	(0.294)
MedHPrice($10k)					0.028	0.218*	0.175
					(0.121)	(0.122)	(0.098)
%ΔPopulation					0.030	0.021	0.010
					(0.062)	(0.053)	(0.026)
%Subprime						0.296***	0.500
						(0.051)	(0.083)
%OwnOcc						0.202**	0.936
						(0.079)	(0.370)
%ΔOwnOcc						-0.162	0.022
						(0.139)	(0.019)
Constant	4.306***	4.542***	1.665	5.802**	6.084**	-19.73***	
	(0.614)	(0.718)	(1.014)	(2.577)	(2.774)	(5.418)	
Observations	358	358	358	358	358	358	
R-squared	0.554	0.634	0.656	0.686	0.687	0.734	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Demographic and geographic variables regression results

	7	8	Elasticity at Mean	9	10	11	12
%GainHPI	0.236***	0.238***	0.67				
%ΔMedHPI	0.351***	0.218***	0.03				
%ΔIQHPI	-0.044	-0.037	-0.04				
%ExcessPermits	0.414***	0.243*	0.09				
%ΔMedInc($1k)	-0.170***	-0.089	-0.12				
MedInc($1k)	0.294*	0.246	0.74	0.982***	0.577***		
MedHPrice($10k)	0.147	0.129	0.10				
%ΔPopulation	0.038	0.131**	0.06				
%Subprime	0.298***	0.288***	0.49				
%OwnOcc	-0.015	-0.077	-0.36				
%ΔOwnOcc	-0.389**	-0.448***	0.06				
%Retired	0.113	0.216	0.18	1.356***	1.311***		
%ΔRetired	-0.114	-0.083	0.00				
%Poverty	0.258	0.224	0.20	0.953***	0.520*		
%Black	-0.124***	-0.058	-0.04	0.008	0.065		
%Hispanic	0.002	0.005	0.00	0.275***	0.238**		
%Asian	0.041	0.005	0.00	0.505	0.141		
%HighSchool	0.288***	0.074	0.33	0.464***	0.142		
%College	0.006	-0.068	-0.12	-0.041	-0.113		
250k<Pop<750k		0.845			1.718*	2.974**	
Pop>750k		-0.269			2.167	3.885***	
NewEngland		-0.867			-10.05***	-12.98***	-12.44***
MiddleAtlantic		-1.876			-15.23***	-18.38***	-18.12***
EastNorthCentral		6.505***			-7.568***	-12.52***	-13.01***
WestNorthCentral		1.859			-9.975***	-17.88***	-18.54***
SouthAtlantic		-1.754			-9.115***	-12.87***	-13.02***
EastSouthCentral		-1.531			-12.62***	-19.66***	-20.15***
WestSouthCentral		-3.806**			-17.57***	-20.50***	-20.88***
Mountain		-0.862			-8.431***	-12.47***	-12.97***
Constant	-34.84	-17.24		-82.52***	-30.78	25.29***	27.10***
Observations	358	358		358	358	358	358
R-squared	0.762	0.814		0.368	0.519	0.345	0.321

Robust standard errors suppressed   Omitted Categories: Pacific Division, Pop<250k, <HighSchool
*** p<0.01, ** p<0.05, * p<0.1

average *%ΔIQHPI* value of 16% this is associated with a reduction of 0.6 in *%DropHPI*. Together these variables suggest that making the distribution of home prices more alike in the sense of making the distribution less right skewed and/or reducing the variability of the distribution is associated with larger values for *%DropHPI*. These effects are small for many areas because the shape of the housing price distribution in many areas did not change much other than the shift in the mean, but it is potentially important in some cases.

The OFHEO indexes do not provide information on changes in the shape of the home price distribution, preventing a more detailed investigation of this phenomenon in our data. The S&P/Case Shiller indexes are published for three price tiers, representing the bottom, middle, and top third of market transactions.[17] Figure 6 plots these tiered indexes for each of the 17 cities for which they are available. The pattern evident in this small sample is consistent with our regression results based on our comprehensive dataset. Cities where the bottom third of the price distribution appreciated at a higher rate than the other tiers experience larger subsequent price drops. Little attention has been paid to shifts in the shape of the housing price distribution as a predictor of *%DropHPI* in a bust period. Forces behind the relationship would appear to be a fruitful avenue for future research which would benefit from the publication of tiered indexes for a comprehensive sample of cities.[18]

Overbuilding is often identified as one of the forces behind the drop in housing prices, and our third stylized fact confirms that areas with greater building relative to population growth experience greater price declines. The variable *%ExcessPermits* enters our regression model with highly significant coefficients with the expected positive sign. The parameter estimate on this term falls somewhat as further variables are added, particularly when controlling for census division and MSA size in the full model (column 8) where the coefficient is significance at the .10 level. The elasticity of .09 is small but when coupled with the range of the variable [-10% to 15%] shifts the prediction of *%DropHPI* by around 2 points.[19] Note that the issuance of building permits is unique in being directly controlled by local authorities. A casual comparison of this housing boom and bust with its predecessor of the late 80s/early 90s suggests that over-

---

[17] The indexes are published and described at http://www2.standardandpoors.com/.

[18] The raw data on which the original OFHEO indexes are based could be used to produce these tiered indexes.

[19] This variable is picking up on differences in excess permits and not the influence of a national building boom. One city in a large system of cities has been shown theoretically to have limited influence on housing prices due to migration between cities(Brueckner 1990, Engle, Navarro & Carson 1992). This is not the case if a large number of cities simultaneously increase the size of their housing stock relative to the total population in the system and it may be useful to examine the role of correlated building sprees on housing prices in future work.
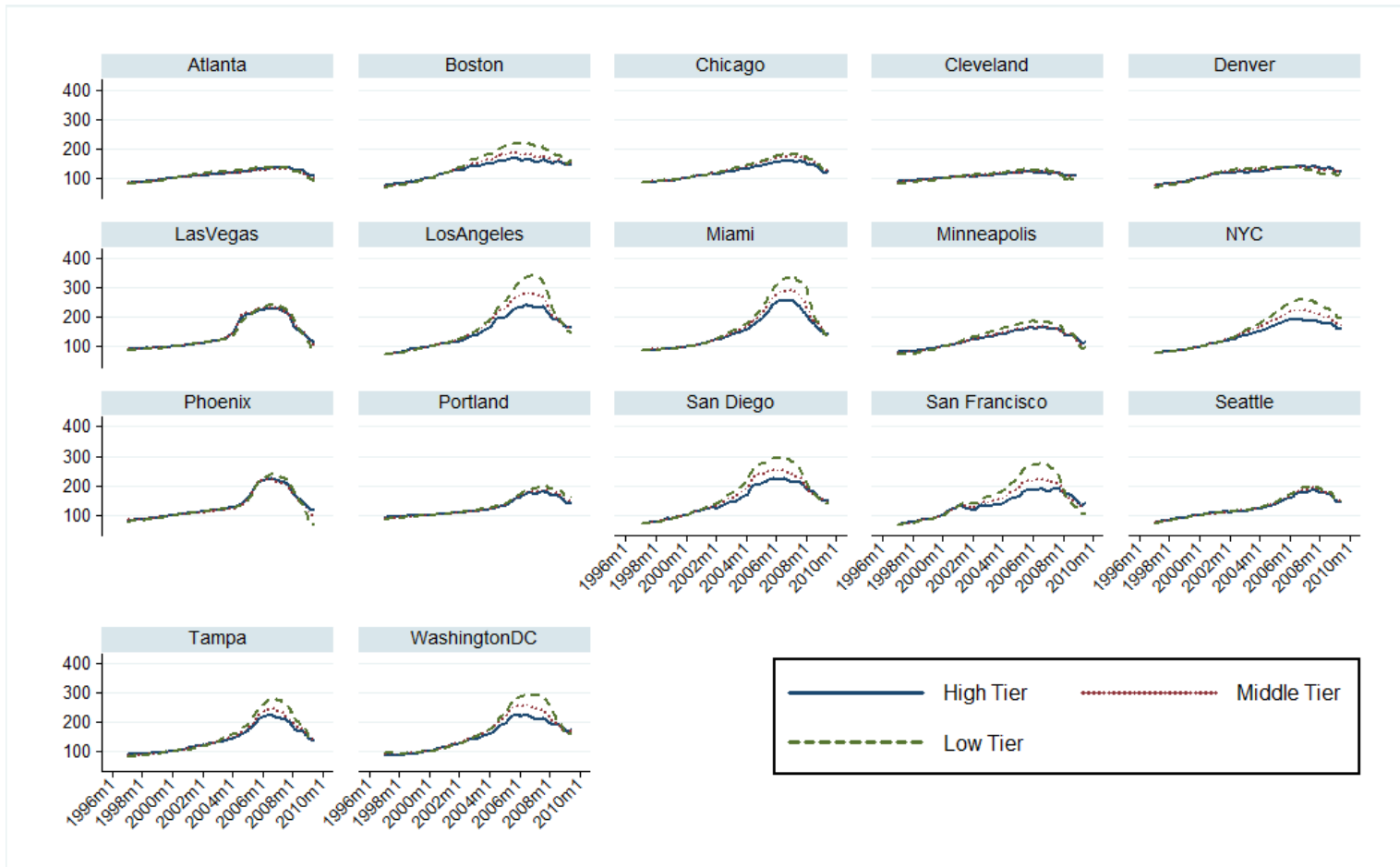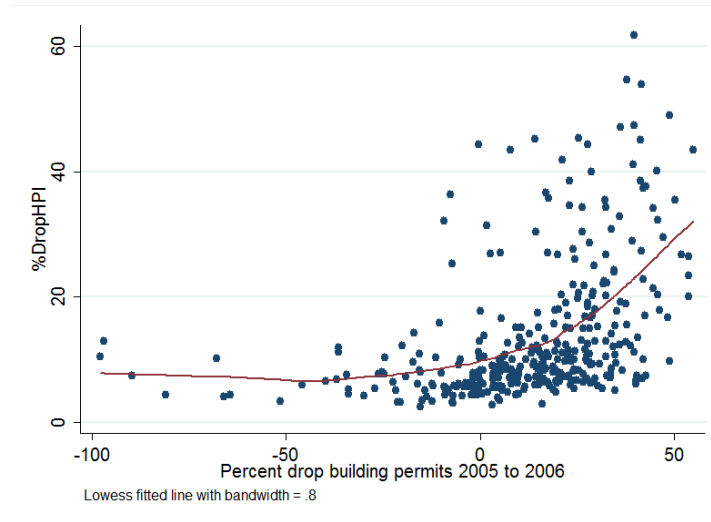
Figure 6: S&P/Case Shiller Tiered HPI by City

Figure 7: Percent decline in building permits 2005-2006 and *%DropHPI*

building played an important role in areas with large price fluctuations, but limitations in the available building permit data in terms of frequency and areas represented prevent a more thorough investigation of a historical relationship. Residential building permits appear to be a leading indicator of HPI declines, as suggested by the scatter plot in Figure 7, which plots the percentage drop in building permits between 2005 and 2006 against *%DropHPI*.[20]

Our next finding is that while the initial level of median income does not consistently predict *%DropHPI,* changes in median income are an important predictor of price declines. *MedInc($lk)* locates the initial level of the income price distribution, and, higher values are associated with larger *%DropHPI* . While the parameter estimates are erratic in size and are not significant across models, the elasticity of .74 is relatively large. In contrast, the parameter estimates for our change in median income variable, *%ΔMedInc($1k)*, are negative and highly significant in models including housing and demographic variables. With the full set of covariates, including census division indicators, the estimate is not significant, but the elasticity of -.12 and the range of the variable [-1.66 to 41.24] suggest that rising income levels have a mitigating effect on price declines. The level of income variable is positive and highly significant in models without housing market variables (columns 9 and 10), suggesting a concentration of factors related to housing price fluctuations in higher income areas. These effects are economically large in many areas, and hence potentially important from a

---

[20] The outliers where permits increased (large negative drops) from 2005 to 2006 are almost all MSAs located near the gulf coast with differences likely due to hurricane Katrina recovery.

policy perspective. As such, attention should be paid to finding ways to more precisely define metropolitan area income measures and the key interaction role (e.g., the *%dAffordability* variable) that income has with measures of changes in housing prices.

Our price level variable, *MedHPrice($1K),* can be taken as an indicator of the original location of the HPI distribution, while our population growth variable, *%ΔPopulation,* captures changes in pressure on the local housing stock. Both enter positively although neither is consistently significant. The elasitity of .10 for median home price in the full model relates price drops to cities with higher initial price levels, which is potentially amplified by the related median income variable. *%ΔPopulation* enters with a positive sign with an elasticity of .06 significant at the 5% level in the full model, suggesting that shrinking cities were not precursors to falling home prices. Similarly, changes in the share of the population that is retired are not predictive of price declines.

The variable associated with bad lending practices is highly significant, while those associated with speculation provide mixed evidence. The first, *%Subprime*, has the expected positive sign. Its associated elasticity is also sizeable at .49. The range spanned by this variable in the sample is quite large [4% to 53%]. This suggests the possibility of a large negative externality of these loans to an area as a whole. If true, there may be major implications for government policy toward loans of this type.[21] The other, *%ΔOwnOcc*, is negative and significant in models controlling for demographic makeup and location (columns 7 and 8). A change in the owner occupancy rate is a measure of speculation, as speculation can decrease the percentage of the housing stock that is owner occupied. The negative sign is consistent with this hypothesis, even though the magnitude of the elasticity is small, .06, and the range of the variable is relatively small [-11% to 8%]. The initial level of owner occupied housing, *%OwnOcc*, is positive and significant in predicting the drop in HPI in the first models without demographic covariates. The coefficient estimate turns negative and insignificant in the model with demographic variables and in the model with population size and census division indicators. The magnitude of the associated elasticity, -.61, is large. Coupled with a fairly wide range that the variable spans [34% to 77%], differences in *%OwnOcc* shift predictions of *%DropHPI* up and down quite a lot.

---

[21] While we used high priced loans reported in HMDA because this data readily available, there are various measures that may also be useful, such as the percent of loans with loose documentation requirements, or that involved low introductory rates. See (Mian & Sufi 2009) for an in depth analysis of the contribution of household borrowing during the price run-ups to the subsequent rise in mortgage default.

An MSA's demographic makeup has predictive power alone (column 9), but is largely insignificant conditional on housing market variables or census division (column 7). Here, *%Black* has a negative and highly significant association with an elasticity of -.04 in the full model with the variable spanning a range of less than 1% to almost 50%. The other demographic variables are insignificant in the full model, although *%HighSchool* is positive and significant in the model without the area size and census division indicators. A model using just levels of the demographic variables explains 37% of the variance and has all of the variables with positive signs except *%College*. All of the variables are significant at the .05 level except *%College*, *%Asian,* and *%Black*.[22]

Our final stylized fact is that census division and MSA size alone are less predictive of price declines than are housing market variables, and add little explanatory power when combined into the full model. We control for location and size with indicator variables for three population size groups and the nine census divisions with the omitted indicators being for small metropolitan areas the Pacific division respectively. In the full model (column 8), neither of the population size indicators are significant at the 5% level, indicating that after controlling for our other covariates, market size is not predictive of *%DropHPI*. Two of the divisions, East North Central [which contains IL, IN, MI, OH, and WI] and West South Central [which contains AR, LA, OK, and TX] are significant, but of opposite signs and roughly similar magnitudes (6.51 and -3.81).

In contrast, a model with just the population size indicators and census divisions (column 11) presents a much different picture. This model explains 35% of the variance (32% with just the divisions). The two population size indicators are highly significant and positive (2.97 and 3.89, for the middle level and large size categories,[23] respectively). The divisional indicators are all highly significant and quite negative relative to the Pacific division with the West South Central being the most negative at -20.50 and the East North Central and Mountain divisions being the closest to the Pacific division at -12.5. While this regression presents a different picture than our full model, suggesting that different areas of the country had different price decline risks, it has less than half of the explanatory power of the models based on market characteristics alone.

---

[22] Race variables were not available for all MSAs in the ACS. Changes in racial composition do not enter significantly in models estimated for the subsample for which they are available with exception of *%Hispanic*, which has a small negative coefficient significant at the 5% level. Changes in other demographics also do not enter significantly in models presented below, except change in *%Highschool*, which has a small negative coefficient that is marginally significant in the full model and a larger and significant coefficient .in models without housing characteristics.

[23] An F-test fails to reject that these coefficients are equal.

Adding geographic and size variables to a model with housing and demographic variables increases the variance explained by 5%. In this full model, MSA level housing market characteristics maintain their significance in describing price decline. In contrast, adding housing and demographic indicators to the division and population size model explains an additional 47% of the variance. We conclude from these results that if one wanted a parsimonious model with reasonably high explanatory power, it would be based on the housing variables alone.

In summary, our stylized facts regarding the magnitude of price declines are: significant positive relationships with prior appreciation, overbuilding, and subprime lending; importance of changes in the shape of local housing price distributions and changes in income levels; and weaker relationships for initial price and income levels, population growth, and speculation prevalence. Additionally, models based on demographic makeup, size, and geographic location alone are less predictive of the size of declines than are those based only on housing market variables. Together, these patterns in the data and the variability of our predictors underscore that to a great extent, the fall in housing prices depended on local market conditions.

There may be other MSA level variables that could be added to our model that would increase its explanatory power, but this would only strengthen the conclusion that there were forces at work that were operating differentially across the metropolitan areas.[24] Our decision to go with the largest possible sample size in terms of metropolitan areas restricted our analysis in term of variable availability, which is higher for larger areas. It did, however, increase the range of variability in the dependent variable and some of the key independent variables and allow us to look at issues related to changes in the shape of the housing price distribution that do not appear to have been previously examined. Certainly, there is a more complete story to be told about the role of construction, building cost, permits, and the relationship between income and implicit rental prices which Glaeser, Gyourko and Saiz have made an excellent start on. This sort of analysis could be extended to a larger set of metropolitan areas and adapted to include the additional contributing factors suggested by the stylized facts above.

---

[24] An examination of the residuals from our model suggests that there may be a couple of localized phenomena associated with large deviations. The first of these is centered on Detroit and several other nearby MSAs where we under predict %DropHPI. This is plausibly a result of the collapse of the Detroit-based automobile sector. The second of these is in the Central Valley of California centered on Merced where there was hope of the rapid expansion of a new University of California campus. The cities that we over-predict have no easily identifiable underlying factor, with the largest over predictions associated with Honolulu followed by McAllen-Edinburg-Mission, TX, Grand Junction, CO and Ocean City, NJ.

VI. A Nonparametric Approach

We now turn to a nonparametric analysis which provides additional insight into differences across regions in housing price declines. We are concerned that our simple linear framework does not consider potentially important nonlinearities and interactions. To address this concern, we utilize the multivariate adaptive regression splines (MARS) methodology, a nonparametric approach to describing the data (Friedman 1991). MARS selects "basis functions" of input variables and uses these functions (splines) as regressors to predict *%DropHPI*, relying on generalized cross-validation to choose among potential models. Relative to most other nonparametric approaches it is optimized to locate threshold effects and its search parameters can be set to locate interaction effects if they exist.

A search over all possible univariate splines using 10 fold cross validation produces an optimal model (MARS 1) predicting *%DropHPI* that utilizes 8 variables in 12 basis functions, with a fit described by a generalized cross validation R-squared of 0.84. [25] The variables used in the selected model are listed in Table 4, along with relative importance scores generated in the model selection and cross validation process. The selected model utilizes primarily housing market variables, although census division is the second most useful predictor, and percent black is also selected as a predictor. The model introduces the variables in a piecewise linear fashion. For three important predictors, *%GainHPI, %Subprime,* and *%ExcessPermits,* the splines highlight nonlinearities over relevant data ranges with a straightforward interpretation. Three of the basis functions, which can be thought of here as piecewise linear regressors, are based on the *%GainHPI* variable,

$$\max(0, 32.97 - \%GainHPI)$$
$$\max(0, \%GainHPI - 66.27)$$
$$\max(0, \%GainHPI - 14.64),$$

with coefficients estimated on these constructed variables of .21, .22, and .21 respectively. This fitted relationship indicates that low levels of price appreciation are not informative for the magnitude of future declines, while higher appreciation rates predict price declines at an increasing rate. MSAs with price growth less than 33% are predicted to have low or no price declines, with a give back rate of .21 predicted for those with growth between 33% and 66.3%, and a highest give back rate of 0.44 for *%GainHPI* larger than 66.27%. For the average seven years of growth from 2000 to the peak of prices, these kink points represent 4.3% and

---

[25] A "penalty" on added variables is introduced in our chosen model selection criterion to encourage parsimony. See Appendix I for a full exposition of the MARS models.

Table 4: Importance of MARS 1 selected predictors

Variable	Score	
%GainHPI	100.00	|||||||||||||||||||||||||||||||||||
Census Division	30.24	||||||||||
%Black	18.62	||||||
%OwnOcc	17.63	||||||
%ExcessPermits	14.94	|||||
%Subprime	13.80	||||||
%ΔPopulation	10.06	|||||
MedHPrice($10k)	5.47	|||

7.8% annual growth rates respectively in real housing prices. The resulting predictive contribution to *%DropHPI* estimated using this spline approach is shown in Figure 8.

Threshold values also for *%ExcessPermits* and *%Subprime* in the MARS model. No increased drop is predicted for MSAs with lower than 1.3% overbuilding, with a .54 higher drop predicted for a one percent increase in *%Excesspermits*. Similarly, no increased drop is predicted through lending where less than 16.5% of 2006 loans were subprime. MSAs where *%Subprime* is above the threshold are predicted to have a .19 higher drop for each percent increase in subprime lending concentration. Figures 9 and 10 depict these fitted relationships. The remaining variables enter the model with some added flexibility, with the relationships described in the housing variables similar to the estimates in our linear model (column 6 in Table 2) and presented fully in Appendix 1.

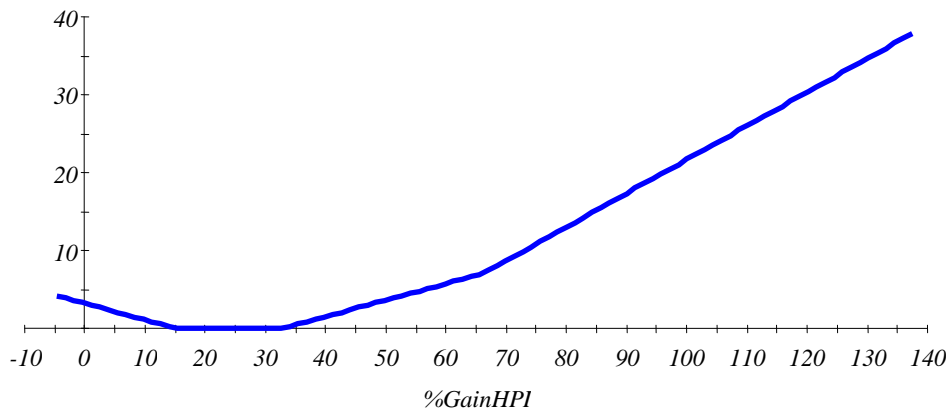**Figure 8: MARS 1 predicted response of *%DropHPI* to *%GainHPI***

20
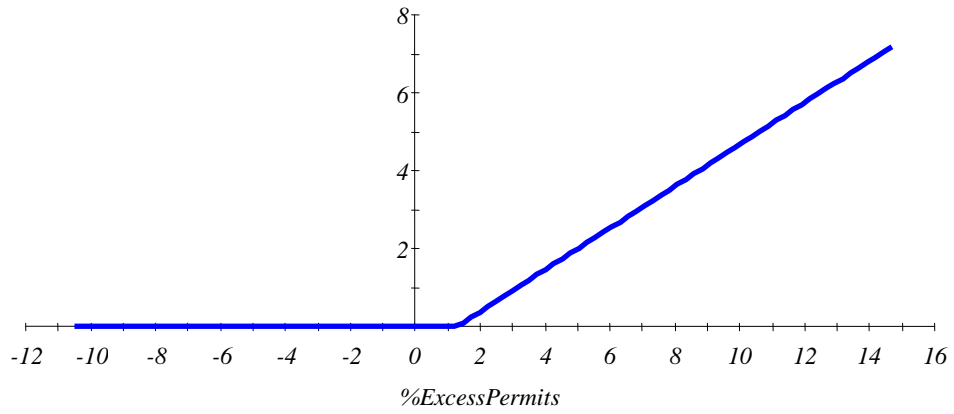
Figure 9: MARS 1 predicted response of *%DropHPI* to *%ExcessPermits*

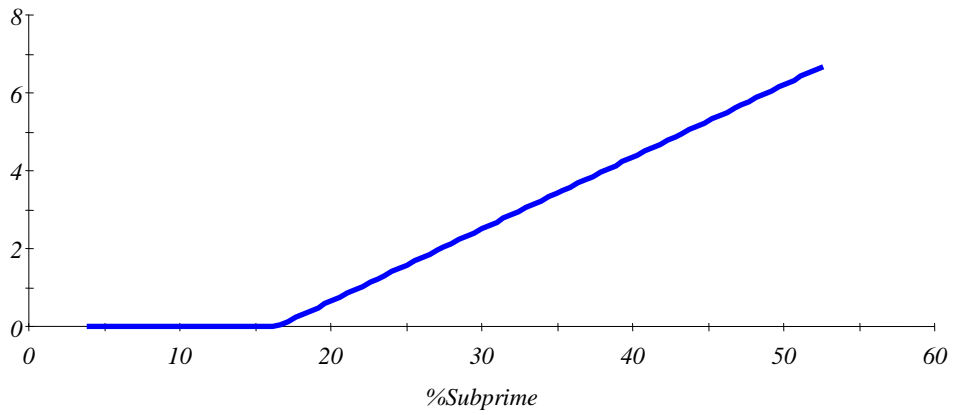**Figure 10: MARS predicted response of %DropHPI to %Subprime**

Table 5: Importance of MARS 2 Selected Predictors

Variable	Score	
%GainHPI	**100.00**	|||||||||||||||||||||||||||||||||||||
%Subprime	**34.19**	||||||||||||
MedHPrice($10k)	**34.09**	||||||||||||
%ΔMedHPI	**28.45**	||||||||||
%ExcessPermits	**23.50**	|||||||||
%HighSchool	**20.40**	|||||||
%ΔPopulation	**16.29**	|||||
%Black	**13.31**	||||
Census Division	**13.24**	||||
%Poverty	**7.88**	||

To explore the importance of predictors based on variable interactions, we allow for two way interactions in the MARS procedure.[26] The resulting model (MARS 2) utilizes 23 basis functions constructed from 10 variables with a generalized cross validated *R*-squared of 0.90. The variables are listed with predictive importance scores in Table 5. Note that a primary effect of allowing interactions is an increase in the predictive importance of the housing market variables relative to the geographic identifier census division.

The first interaction term to enter the model is a hybrid of price appreciation and subprime lending:

$$BF5 = \max(0, \%Subprime - 18.41) * \max(0, \%GainHPI - 32.97).$$

The positive coefficient on this term indicates that MSAs with both *%Subprime* > 18.41 and *%GainHPI* > 32.97 are predicted to have higher price declines as the interaction of the terms increases. The two variables also enter in additional basis functions used to form the model, resulting in the predictive surface shown in Figure 8. The model captures similar interactions in the data between *%DropHPI* and the interaction of *%ExcessPermits* with *%GainHPI*. The resulting predictive surfaces are shown in Figure 11 and Figure 12, which illustrate that simultaneous high levels of risky lending, overbuilding, and rapid price appreciation predict the greatest price declines.

Our nonparametric analysis confirms the importance of *%GainHPI* as the primary predictor of the magnitude of subsequent price decreases, as well as the

---

[26] We do not allow interactions with Census Division dummies, but require the predictive relationship to be estimated for the entire sample.

Figure 11: MARS2 predictive response of *%DropHPI* to *%GainHPI* and *%Subprime* interaction
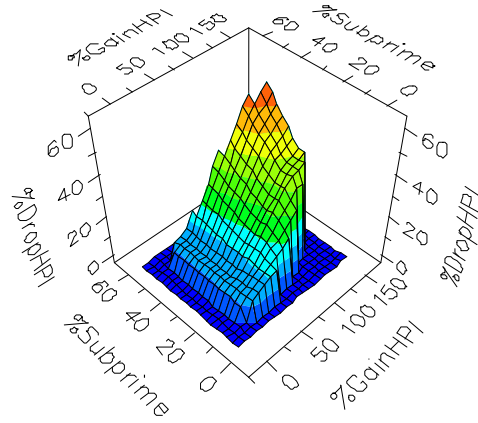
Figure 12: MARS 2 predicted response of *%DropHPI* to *%ExcessPermits* and *%GainHPI* Interaction
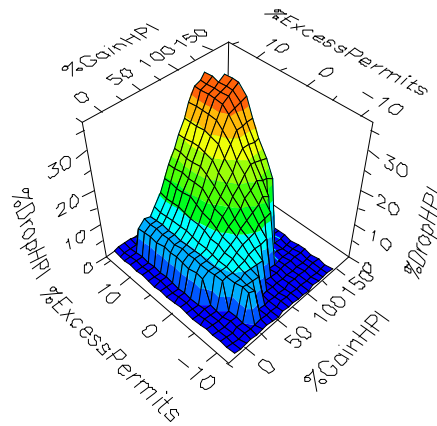
importance in the data of housing market variables in addition to geographic correlation. The smoothing analysis with no interactions uncovers similar covariates as the regression model as being important predictors, while pointing to important nonlinearities in the relationships, particularly of price increases and subprime lending. Allowing interactions in the nonparametric fitting procedure reveals the importance of combinations among excesses in subprime lending, price appreciation, and building for the eventual size of price declines.

VII. Historical Comparisons

One question that arises from our *ex post* analysis of current housing price declines is the extent to which the stylized facts we find are unique to this bust. While historical data for our full sample of variables is not available, we are able to examine correlations of earlier MSA level price declines with previous appreciation rates, demographic variables, and geographic location. The length of the HPI series varies across MSAs, with the longest beginning in 1975Q2 (Los Angeles); there are 130 MSAs with data back to at least 1980Q1, while 327 extend back to at least 1990Q1. Throughout the sample, there are 243 cases where the HPI reaches a high point relative to the previous three and the subsequent one year, which we label as "peaks" in the metropolitan area HPI.[27] For each of these peaks, we calculate the percentage magnitude of the decline from the peak HPI level to the subsequent low value, *%DropHPI*, as well as the percent change in prices from five years prior to the peak as *%GainHPI*. These variables are plotted in Figure 10. We utilize variables from the 1990 Census[28] for levels of housing and demographic data, but are unable to calculate changes in these values coincident to the peaks, such as changes in the housing price distribution shape variables, income levels, and owner occupancy rates. As explained in Appenidx 2, which also presents summary statistics and all regression results for our historic estimation, we are also unable to construct reliable historical estimates of overbuilding and subprime lending. The average magnitude of prior MSA level price declines is 10.18% (slightly lower than the average for the current bust), with a 10.17 standard deviation. Price gains are calculated relative to the index level five years prior to the peak and average 18.65%, not as substantial as those seen from 2000 to the most recent peak, although with a standard deviation of 22.55 there is significant variability historically in the extent of price appreciation prior to peaks.

Historically, the magnitude of MSA level price declines is predicted by appreciation prior to the peak. Figure 13 plots this relationship for the 243 pre 2000 price declines, along with the univariate predictions lines for the historical data (solid), and the prediction (dashed) from the univariate regression for the 2000s data (Model 1 of Table 2). This relationship is robust to the inclusion of demographic variables and indicators for census division and size, evidence that the stylized facts of our model are not unique to the current price declines. We estimate a 30-40% elasticity at the mean for our *%GainHPI* parameter, suggesting

---

[27] These cases also have 5 years of prior HPI data allowing the calculation of a 5 year percent gain, *%GainHPI*.

[28] The HUD SOCDS aggregates many variables from the 1990 Census data for current MSA definitions, allowing us to compare these to our HPI variables calculated for the same geographies.

Figure 13: Pre-2000 HPI drops and prediction line of 2000s model
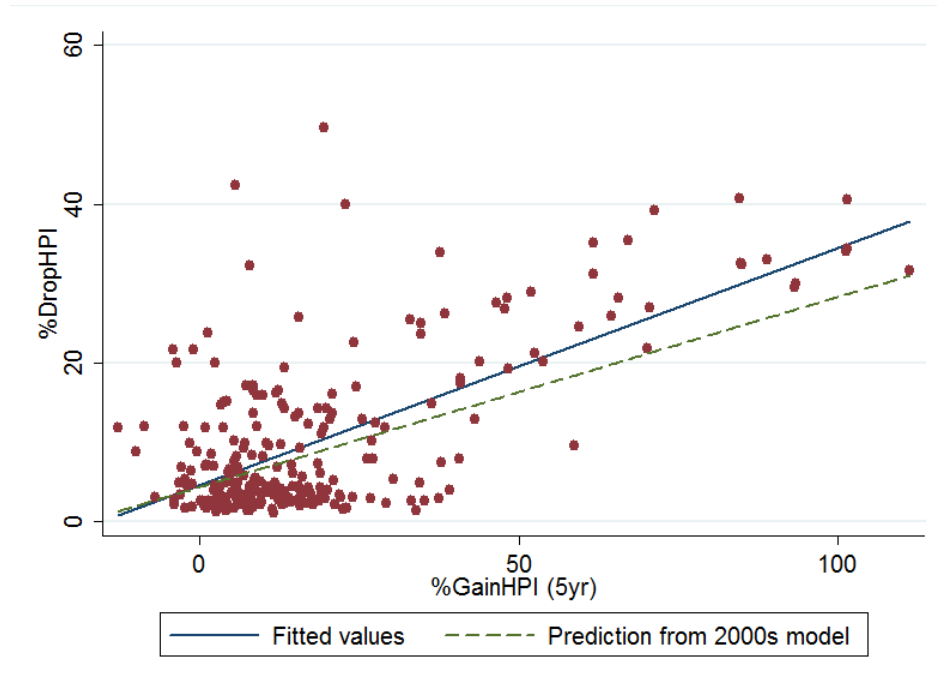
that historically when real prices have declined, 1/3 of the prior 5 years of appreciation has been lost.[29] While this is lower than the 72% we estimate for the current downturn, our finding that local price appreciation prior to a peak gives a better prediction of price decline magnitudes than does geographic location is supported. In the historical data, *%GainHPI* alone captures 44% of the variation in historical price declines which increases to 52% with the inclusion of housing market level variables from the 1990 census. Census division dummies alone explain 37%, and a full model including all available variables explains 60% of variation. In our historical sample, housing market trends are comparable to geographic location and market size in their predictive ability.

While our analysis has focused explicitly on predicting the magnitude of price declines, it is relevant to the discussion of the predictability of the crisis in general to examine historical cases of high house price appreciation outside the context of a peak and subsequent decline. In the quarterly MSA level HPI data, we find 1,630 observatios prior to 2000 where relative to the previous five years,

---

[29] Our estimate is consistent with the 33% give back rate reported by Glaeser, Gyourko, and Saiz for a smaller sample of MSAs mentioned above.

Table 6: Tabulated high appreciation rates and subsequent declines

	No fall of at least 5% within 5 years		Fall of at least 5% within 5 years		Total
5 year annual growth rate less than 4.3%	Count:	11,527	Count:	5,086	16,614
	Row %:	69.4	Row %:	30.6	
5 year annual growth rate above 4.3%	Count:	679	Count:	951	1,630
	Row %:	41.7	Row %:	58.3	
Total		12,207		6,037	18,244
		66.9		33.1	

Pearson $\chi^2(1)$: 515.56

real prices appreciated at a rate of higher than 4.3% a year,[30] the rate our nonparametric model selects as a kink point above which price increases predict the magnitude of subsequent declines. Of these, 951 (58%) saw price declines of at least 5% begin within 5 years. The magnitude of the drop was greater than 10% in 720 (44%) of these cases. Returning to our original post 2000 dataset, we find that 143 (40%) of all MSAs had annual growth rates above 4.3% from 2000 to 2006Q3. Using our historical decline rate of 44%, we would predict price declines of at least 10% in 63 of these MSAs, 103 of which have seen actual declines of 10% or more.[31] In analyzing increases for demand and supply of subprime loans in MSAs with high HPI growth, Goetzmann, Peng, and Yen (2009) develop ARIMA models for nineteen of the Case-Shiller HPIs. They report that 15 of the 19 series fell more than two standard errors below predictions based on a forecast models using data through the end of 2005, which suggests that the large declines experienced were not predictable using standard techniques. They also note, however, that comparable five year forecasts beginning in 2000 under-predict price increases for 14 of 19 cities, with six series more than two standard deviations above forecast. They conclude that "the rate of model failure should have given grounds to doubt the reliability of the confidence bands, at the very least." Our analysis indicates that abnormally high appreciation in metropolitan housing markets was consistently followed by price declines, with greater increases historically followed by greater price drops.

This consistent mean reversion of MSA level price trends after rapid appreciation throughout historical data raises questions about the observed geographic uniformity of mortgage rates and underwriting guidelines. While realized average rates differ across MSAs with compositional differences in borrowers and lenders, an examination of advertized mortgage rates for the leading national lenders suggests that mortgages are not typically priced with

---

[30] Calculated using the median 7 years from 2000 to peak for *%GainHPI.*
[31] An additional 62 MSAs dropped at least 10% but had pre-peak growth rates less than 5.4%.

regard to MSA level risk. As an example, Wells Fargo displays "current interest rates for several common loan types for the purchase of a single family primary residence" on its website, with no indication in the accompanying discussion of risk-based pricing in its "Loan Pricing Disclosure" that rates vary by MSA or state.[32] Similarly, HUD handbooks of mortgage insurance guidelines as well as GSE underwriting guides are silent on local market level risks.[33] One recent exception to this apparent uniform approach to mortgage lending across geography is the ResiLogic model used by Fitch Ratings to "analyze credit risk in securities backed by U.S. residential mortgages." The model used to evaluate the risk in the underlying mortgages was "enhanced" in July 2008 to include state and MSA-level "regional risk multipliers" based on an "analysis of regional risk [which] takes into account individual state's and MSA's economic metrics, such as personal income and distribution, employment growth, housing construction, and other indicators." Consistent with our findings, in the enhanced model, "the largest component of the state and MSA risk multipliers is [the model's] five-year home price forecast"(Sirotic, Somerville & Barberio 2009).

The value to a lender of a mortgage depends on housing price trends through prepayment and default risks (Downing, Stanton & Wallace 2005). The well documented trends of positive short term correlation of housing prices in local markets (Case and Shiller 1989) and the historically consistent mean reversion discussed above both suggest a local market level risk for mortgage values. Our findings of systematic MSA level variation in housing markets suggest that mortgage prices would vary based on local market conditions. The observed uniformity in pricing and underwriting presents a puzzle for future research.

VII. Conclusion

This paper presents a set of stylized facts that need to be explained when analyzing the forces behind this great housing boom and bust. While the associations we have found do not establish causality, given that *%GainHPI*

---

[32] See for example https://www.wellsfargo.com/mortgage/rates/. Similar advertisements were presented by Chase, while listed Citi rates did not differ with a number of zip code inputs.

[33] Both are available from the respective agency and company websites. Section 23.5 in the Freddie Mac Single-Family Seller/Servicer Guide limiting loan to value ratios in neighborhoods deemed by appraisers to be declining in value was deleted effective June 1, 2008. See Stuart (2003) for a history and description of the US mortgage market and a discussion of the "one size fits all" norms in mortgage lending. See Wallison and Ely (2000) for a prescient discussion of the possibility that expansion and standardization of Fannie Mae and Freddie Mac's lending activities would nationalize mortgage risk and lead to a housing market crisis.

explains a sizeable fraction of variability of *%DropHPI*, understanding causality will surely involve being able to predict *%GainHPI*. Further, we find that the key pattern between *%GainHPI* and *%DropHPI* appears to repeat itself One lesson, though, is that while this housing bust has elements of a national phenomenon, it has played out in a local fashion. A small number of variables related to the housing price distribution, building, lending, and area demographics explain a large fraction of differences across metropolitan areas. It is also clear that geographic diversification across census divisions was an inadequate and perhaps misleading approach to understanding the underlying sources of variability in home price fluctuations. Larger increases in housing prices, higher percentages of high priced/low quality loans and building faster than the workforce is growing were all associated with larger drops in housing prices when the bust came. These all make common sense, *ex post*, the question is why they did not *ex ante*.[34] From the vantage point of financial markets, the fundamental question is why there were not differences across MSAs in home loan interest rates, which would have compensated investors for increased risk, and/or down payment requirements, which would have reduced exposure to price declines. From a public policy perspective, the open question is whether increasing interest rates and down payment requirements in markets with high appreciation rates would dampen the boom and bust cycles of housing markets.

---

[34] The National Association of Realtors website posted 10-page "anti bubble reports" for 135 MSAs in the fall of 2005 to "show that the facts simply do not support the possibility of a housing bust." The reports assess housing prices in each market for local member agents and goes to great effort to develop arguments as to why the housing bubble was unlikely to burst. Evidently, most purchasers of these mortgage packages believed similarly.

Appendix I: MARS models

Our MARS 1 model utilizes 8 predictor variables and 12 basis functions, with a generalized cross validated *R*-squared of 0.84. The model predicts *%DropHPI* as the linear regression based on variables defined by the functions listed in Table 7. Subsets of census divisions used as functions are listed in Table 8.

Table 7: MARS 1 basis functions and coefficients

Basis Function	Coefficient
max(0, 32.97 – *%GainHPI*)	0.214
Division is in SubSet1	5.81
max(0, *%GainHPI* – 66.28)	0.221
max(0, *%GainHPI* – 14.64)	0.212
max(0, 53.5404 – *%ΔOwnOcc*)	-0.871
max(0, *%ΔOwnOcc* – 69.12)	1.20
max(0, 2.48 – *%Black*)	-2.35
Division is in SubSet2	3.24
max(0, *%ExcessPermits* – 1.30)	0.537
max(0, *%Subprime* – 16.51)	0.185
max(0, *%ΔPopulation* – 14.00)	0.350
max(0, *MedHPrice($10k)* – 11.99)	0.222
Intercept	-1.065

Table 8: MARS 1 Census Division Subsets

Variable	Divisions
SubSet1	New England, East North Central, Mountain, Pacific
SubSet2	West North Central

Our MARS 2 model utilizes 10 predictor variables and utilizes 23 basis functions, with a generalized cross validated *R*-squared of 0.90. The model predicts *%DropHPI* as the linear regression based on variables defined by the functions listed in Table 9

Table 9: MARS 2 basis functions and coefficients

Basis Function	Coefficient
max(0, *%GainHPI* - 32.97)	1.19
max(0, 32.9685 - *%GainHPI*)	- 0.809
max(0, *%Subprime* - 18.4095)* max(0, *%GainHPI* - 32.97)	0.0119
max(0, 18.4095 - *%Subprime*) * max(0, *%GainHPI* - 32.97)	- 0.0373
max(0, 6.52175 - *%ExcessPermits*)* max(0, *%GainHPI* - 32.97)	- 0.0293
max(0, *%GainHPI* - 13.8038)	- 0.819
Division is East North Central or West North Central	5.18
max(0, *MedHPrice($10K)* - 10.27)* max(0, 13.8038 - *%GainHPI*)	0.603
max(0, *%ΔMedHPI* + 22.9447)* max(0, *MedHPrice($10K)* - 4.28)	0.0505
max(0, 12.4754 - *%ΔPopulation*)* max(0, *%GainHPI* - 32.9685)	- 0.0101
max(0, *%GainHPI* - 103.721)	- 0.279
max(0, 30.5483 - *%Subprime*)* max(0, *MedHPrice($10K)* - 4.28)	- 0.0401
Division is East North Central, West North Central, or Pacific	-2.42
max(0, *%ΔPopulation* - 11.8683)* max(0, 32.9685 - *%GainHPI*)	-0.0383
max(0, *%ΔMedHPI* + 8.38518)* max(0, *MedHPrice($10K)* - 4.28)	-0.0367
max(0, *%ΔMedHPI* - 10.6662)* max(0, *%HighSchool* - 55.8408)	0.123
max(0, *%Black* - 0.111783)* max(0, *%GainHPI* - 32.9685)	-0.0055
max(0, *%GainHPI* - 65.9564)* max(0, *%HighSchool* - 55.8408)	0.0221
max(0, *%GainHPI* - 9.70669)* max(0, *%HighSchool* - 55.8408)	-0.0074
max(0, *%Poverty* - 4.37793)* max(0, 13.8038 - *%GainHPI*)	0.0852
max(0, *%ΔPopulation* - 16.3361)* max(0, 96.2537 - *%GainHPI*)	0.0106
max(0, *%HighSchool* - 57.4469)* max(0, 13.8038 - *%GainHPI*)	0.103
max(0, 23.064 - *%Subprime*)* max(0, *%GainHPI* - 13.8038)	0.0134
Intercept	19.93

Appendix II: Historical Comparisons

Price gains and drops for historical housing cycles and demographic variables from the 1990 Census are used to repeat our nested regression exercise using historical data. While building permit data is reported at annual intervals for many MSAs as far back as 1980, we are unable to calculate a reliable measure of excess permits at the MSA level prior to each boom, as geographic definitions of MSAs are not consistent over time in the permit data and reliable intercensal population or labor force growth estimates are not available prior to 1990. Additionally, the tracking of variability in loan terms only began recently, so we do not know the extent of subprime lending over time by MSA in past decades. Table 10 presents summary statistics for the available variables for the 243 HPI peaks in our historical sample. The average magnitude of prior MSA level price declines is 10.28 (two percentage points lower than the average for the current bust), with a 10.14 standard deviation. Price gains are calculated relative to the index level five

Table 10: Summary Statistics for Pre 2000 Variables

	Mean	Std Dev	Min	Max
%DropHPI	10.18	10.17	1.04	49.59
%GainHPI(5yr)	18.65	22.55	-12.81	111.26
MedInc($1K)	34.55	6.02	17.62	57.99
MedHPrice($10K)	8.33	4.75	3.56	28.62
%OwnerOcc	60.33	5.62	42.27	77.87
%Poverty	13.38	5.07	5.85	41.88
%Black	11.06	10.98	0.10	44.70
%Hispanic	7.98	13.27	0.26	85.24
%HighSchool	53.97	8.61	28.42	75.19
%College	20.02	5.99	10.30	42.84
Pop<250k	0.40			
250k<Pop<750k	0.37			
Pop>750k	0.23			
NewEngland	0.05			
MiddleAtlantic	0.10			
EastNorthCentral	0.13			
WestNorthCentral	0.06			
SouthAtlantic	0.23			
EastSouthCentral	0.07			
WestSouthCentral	0.13			
Mountain	0.08			
Pacific	0.16			

Table 11: Regression Results for Pre 2000 *%DropHPI* -- Housing Variables

	1	4	5	6	Elasticity at Mean
%GainHPI	0.299***	0.253***	0.213***	0.228***	0.419
	(0.0185)	(0.0290)	(0.0347)	(0.0342)	(0.067)
MedInc($1k)		0.273**	0.130	0.207	0.703
		(0.136)	(0.162)	(0.175)	(0.583)
MedHPrice($10k)			0.413*	0.164	0.134
			(0.227)	(0.280)	(0.230)
%OwnOcc				-0.315**	-1.868
				(0.139)	(0.785)
Constant	4.599***	-3.978	-1.726	16.41**	
	(0.588)	(4.152)	(4.500)	(6.973)	
Observations	243	243	243	243	
R-squared	0.440	0.456	0.468	0.494	

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

years prior to the peak, and are not as substantial, on average, as those seen from 2000 to the most recent peak. Average levels of the demographic variables and the proportion of the sample in each size and geographic group compare well to our data for the current markets. Regression estimates for available housing market variables are presented in Table 11 and compare well to our estimates for current declines. Our price increase variable, *%GainHPI*, alone captures 44% of the variation in historical price declines. Estimates for regressions that include demographic, size, and census division controls are reported in Table 12. As before, controlling for census division alone explains less of the variation in decline than controlling for the size of appreciation prior to the peak. Our most complete model has an R-squared of 0.60, and finds that 31% of five year price increases are lost conditional on the variables available in our sample. While this is lower than the 67% we estimate for the current downturn, our finding that local price appreciation gives better prediction of price decline magnitudes than does geographic location is confirmed.

Table 12: Regression Results for Pre 2000 *%DropHPI* -- Demographic and Geographic Variables

	7	8	Elasticity at Mean	9	10	11	12
%GainHPI5yr	0.238***	0.171***	0.314				
MedInc($1k)	0.365	0.760**	2.58	1.125***	0.964***		
MedHPrice($10k)	-0.0743	-0.349	-0.285				
%OwnOcc	-0.286*	-0.154	-0.914				
%Poverty	-0.0985	0.216	0.284	0.167	0.324		
%Black	0.0229	-0.0629	-0.068	-0.0176	-0.0514		
%Hispanic	0.107*	0.00489	0.004	0.176***	0.0236		
%HighSchool	-0.193*	-0.132	-0.699	-0.185	-0.184		
%College	-0.335	-0.376*	-0.740	-0.386*	-0.459**		
250k<Pop<750k		4.001***			3.780***	5.246***	
Pop>750k		4.547**			4.753***	8.714***	
NewEngland		3.396			10.22***	10.13***	12.19***
MiddleAtlantic		-4.866			-3.235	-7.864***	-7.019**
EastNorthCentral		-10.96***			-12.18***	-15.20***	-14.64***
WestNorthCentral		-5.858**			-6.918***	-10.96***	-11.63***
SouthAtlantic		-3.765			-4.712**	-11.23***	-10.82***
EastSouthCentral		-3.804			-4.497*	-12.90***	-11.99***
WestSouthCentral		-1.797			-1.901	-9.023***	-9.737***
Mountain		-1.851			-2.885	-7.302***	-8.669***
Constant	28.36**	6.443		-14.44	-6.747	14.40***	18.21***
Observations	243	243		243	243	243	243
R-squared	0.516	0.602		0.351	0.560	0.485	0.374

Robust standard errors suppressed

Omitted Categories: Pacific Division, Pop<250k, <HighSchool

*** p<0.01, ** p<0.05, * p<0.1

References

Brueckner, J.K. 1990, "Growth Controls and Land Values in an Open City", *Land Economics,* vol. 66, no. 3, Private Markets, Public Decisions: An Assessment of Local Land-Use Controls for the 1990s, pp. 237-248.

Calhoun, C. 1996, "Ofheo House Price Indexes: HPI Technical Description", *Official OFHEO Document.* Office of Federal Housing Enterprise Oversight, Washington D.C.

Case, K.E. & Shiller, R.J. 1989, "The Efficiency of the Market for Single-Family Homes", *The American Economic Review,* vol. 79, no. 1, pp. 125-137.

Dagum, C. 1977, "A New Model of Personal Income Distribution: Specification and Estimation", *Economie Appliquée,* vol. 30, no. 3, pp. 413-436.

Downing, C., Stanton, R. & Wallace, N. 2005, "An Empirical Test of a Two-Factor Mortgage Valuation Model: How Much Do House Prices Matter?", *Real Estate Economics,* vol. 33, no. 4, pp. 681-710.

Engle, R., Navarro, P. & Carson, R. 1992, "On the Theory of Growth Controls", *Journal of Urban Economics,* vol. 32, pp. 269-269.

Friedman, J.H. 1991, "Multivariate Adaptive Regression Splines", *The Annals of Statistics,* vol. 19, no. 1, pp. 1-141.

Gabriel, S.A., Quigley, J.M. & Rosenthal, L.A. 2009, "The Mortgage Meltdown, the Economy, and Public Policy", *The B.E.Journal of Economic Analysis & Policy,* vol. 9, no. 3 (Symposium).

Glaeser, E.L. & Gyourko, J. 2007, "Housing Dynamics", *Harvard Institute of Economic Research,* Discussion Paper Number 2137.

Glaeser, E.L., Gyourko, J. & Saiz, A. 2008, "Housing Supply and Housing Bubbles", *Journal of Urban Economics,* vol. 64, no. 2, pp. 198-217.

Goetzmann, W.N., Peng, L. & Yen, J. 2009, "The Subprime Crisis and House Price Appreciation", National Bureau of Economic Research, Working Paper w15334.

Gyourko, J., Saiz, A. & Summers, A. 2008, "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index", *Urban Studies,* vol. 45, no. 3, pp. 693-729.

Himmelberg, C., Mayer, C. & Sinai, T. 2005, "Assessing High House Prices: Bubbles, Fundamentals and Misperceptions", *Journal of Economic Perspectives,* vol. 19, no. 4, pp. 67-92.

Kelly, A.J. 2008, "'Skin in the Game': Zero Down Payment Mortgage Default", *Journal of Housing Research,* vol.17, no.2, pp. 75-99.

Leventis, A. 2008, "Revisiting the Differences Between the OFHEO and S&P/Case-Shiller House Price Indexes: New Explanations", *Research Paper*, Office of Federal Housing Enterprise Oversight, Washington D.C.

Li, G. 1985, "Robust Regression" in *Exploring Data Tables, Trends, and Shapes*, eds. D.C. Hoaglin, F. Mosteller & J.W. Tukey, Wiley, New York, pp. 281-340.

Mayer, C. & Hubbard, G. 2008, "House Prices, Interest Rates, and Mortgage Market Meltdown", Unpublished Manuscript, Columbia Business School.

Mayer, C.J. & Pence, K. 2008, "Subprime Mortgages: What, Where, and to Whom?" National Bureau of Economic Research, Working Paper w14083.

Mian, A.R. & Sufi, A. 2009, "House Prices, Home Equity-Based Borrowing, and the U.S. Household Leverage Crisis", National Bureau of Economic Research, Working Paper w15283.

Saiz, A. 2008, "On Local Housing Supply Elasticity", *SSRN eLibrary*.

Sirotic, A., Somerville, H. & Barberio, V. 2009, *Resilogic^{TM}:U.S. Residential Mortgage Loss Model Criteria*, www.fitchratings.com.

Stuart, G. 2003, *Discriminating Risk: The US Mortgage Lending Industry in the Twentieth Century,* Cornell University Press, Ithica, NY.

Wallison, P.J. & Ely, B. 2000, *Nationalizing Mortgage Risk*, AEI Press, Washington, D.C.