Title
Mortgage default and student outcomes, the solar home price premium, and the magnitude of housing price declines

Author
Dastrup, Samuel R.

Publication Date
2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Mortgage Default and Student Outcomes, the Solar Home Price Premium, and the Magnitude of Housing Price Declines

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy

in

Economics

by

Samuel Reed Dastrup

Committee in charge:

Professor Richard T. Carson, Chair
Professor Julian R. Betts
Professor Gordon B. Dahl
Professor Steven P. Erie
Professor Joshua Graff Zivin
Professor Joel Watson

2011

The dissertation of Samuel Reed Dastrup is approved, and is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2011

DEDICATION

To Emily. To C, D, & J. To my Parents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am grateful for the direction, support, and guidance of Richard Carson, the chair of my committee, and coauthor of Chapter 3, included with his permission. I also thank Josh Graff Zivin for helpful guidance, suggestions, and generosity. It is with his, Matthew Kahn, and Dora Costa's permission as coauthors that I include Chapter 2. Thanks to Julian Betts for opening doors that made Chapter 1 possible. It is included with his permission as a coauthor. Andrew Zau has been a reliable and patient resource for the research in Chapter 1.

I also thank generous mentors Gordon Dahl and Joel Watson, and the many other UC San Diego Economics Department faculty who have invested in my development. Many thanks go out to other classmates and colleagues who provided helpful discussion and feedback.

I especially thank Graton Gathright and Christopher Wignall, whose advice and friendship were invaluable in writing this dissertation. I am lastly grateful to Emily Dastrup for many discussions, encouragement, and editing help.

VITA

2005	Bachelor of Arts, Economics, Brigham Young University
2011	Doctor of Philosophy, Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Mortgage Default and Student Outcomes, the Solar Home Price Premium, and the Magnitude of Housing Price Declines

by

Samuel Reed Dastrup

Doctor of Philosophy in Economics

University of California, San Diego, 2011

Professor Richard T. Carson, Chair

This dissertation presents three studies related to housing. The first quantifies the effect of mortgage default on elementary student's academic outcomes. We first document socioeconomic differences between students experiencing default and those that did not in San Diego Unified School District in the aftermath of the housing market crash that began in 2006. We identify the impact of default using student-level fixed effects to compare the same student before and after default. Students

experiencing default have lower math scores and greater absenteeism in the year of the default. We also find that classroom and neighborhood mortgage default rates are negatively related to education outcomes.

The second study uses a large sample of homes in the San Diego and Sacramento, California areas to provide some of the first capitalization estimates of the sales value of homes with solar panels relative to comparable homes without solar panels. Using both hedonics and a repeat sales index approach we find that solar panels are capitalized at roughly a 3.5% premium. This premium is larger in communities with a greater share of college graduates and of registered Prius hybrid vehicles.

The final study examines differences in the magnitude of recent 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, which raises questions about the validity of mortgage pricing policy 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.

Chapter 1

Student and Peer Education Outcomes and Stress at Home; Evidence from Mortgage Default in San Diego

1.1 Introduction

Owner's equity in household real estate in the United States decreased by $6.8 trillion from 2006 to the beginning of 2009. This housing market collapse triggered financial market paralysis and contributed to the onset and severity of the Great Recession, which saw unemployment more than double from under 5 percent in the spring of 2008 to 10.1 percent in the fall of 2009. Together, the housing market run-up, collapse, and the declining economy left a historically large percentage of homeowners unable or unwilling to make scheduled mortgage payments. Over 14 percent of the nation's mortgages were in arrears or had begun the foreclosure process as of the end of 2010.[1]

In this paper, we investigate an impact of household financial stress during the housing market collapse and subsequent recession on family wellbeing. Specifically, we measure the change in educational outcomes for elementary students residing in

---

[1] Homeowner equity change based on the Federal Reserve's March 10, 2011 Flow of Funds Accounts of the United States. Unemployment rates as reported on the Bureau of Labor Statics website in June, 2011. Delinquency rate based on November 18th, 2010 Mortgage Bankers Association press release.

homes which begin the foreclosure process. We employ longitudinal data of individual academic outcomes linked to property records in San Diego Unified School District, one of the nation's largest and most diverse school districts in a city that experienced substantial housing market turmoil. Controlling for observable time-varying influences on test scores, students living in single family, owner-occupied homes that receive a Notice of Default during the school year score on average 0.06 standard deviations lower in math and 0.04 standard deviations lower in reading on the yearly California Standards Test in the years following default relative to their own average scores over all years. We similarly estimate that these students are similarly from school 0.22 percentage points more frequently in the year of the default relative to their own absentee rate. We find that the decline in math scores is larger when the student's teacher has two or fewer years experience, but that there is no difference in impact of default for areas with lower census tract group median income.

We also find evidence of spillover effects of mortgage default onto neighborhood and classroom peers. After controlling for other influences on outcomes, including a student's own default exposure, the proportion of a student's neighbors experiencing a single family owner occupied default correlates to a small but statistically significant decline in math score and increase in absence rate. Peer influences operate not only through the neighborhood, but also through the classroom: an increase in the proportion of classmates experiencing a single family default correlates with a small but statistically significant decline in both math and reading scores and increase in absence rates. These results on peer outcomes suggest negative external social effects of mortgage default in addition to the consequences to delinquent families.

We also compare the demographic, neighborhood, and school characteristics of students experiencing default to those who do not. Students experiencing default, whether living in single family owner occupied or rental units, are more often Hispanic or black, disproportionately English learners, live in smaller homes (measured before default), and less frequently have parents with college or post-graduate degrees. They live in neighborhoods with lower median incomes, higher average default levels, higher historical levels of unemployment, and lower education levels. While these students attend schools where their peers are more likely to qualify for reduced price lunch and be English learners, their teachers are more likely to have advanced degrees and have no less experience on average than those of other students. In making these comparisons, we note that our data set spans from the 2001-2002 school year through 2008-2009. Price declines and rampant default began first among relatively less expensive homes and in the subprime market which was disproportionately composed of minority borrowers. More recently, and only near the end of our sample period, default has increased among more expensive homes and formerly prime borrowers.[2]

We note that we are not able to differentiate financial distress due to adverse life events such as job loss, divorce, or unexpected major medical expenses that result

[2] We anticipate analyzing this potential shift in the composition of defaulters as well as exploring long-term impacts as data becomes available for the 2009-2010 school year and beyond.

in a mortgage default from financial stresses due to housing price declines and changes in availability of credit alone. The role of rising unemployment is noteworthy in this regard, as a dramatic increase in persistent unemployment is an important characteristic of the recent recession. Our dataset does not allow us to identify separately the potential for parent unemployment to alter children's education outcomes independently of housing-related financial stress. Rather, we view our estimates as measuring the effect of an increase in household financial stress caused by any number factors and resulting in prolonged mortgage delinquency. These factors are closely related -- where during a housing boom, increasing housing values provide homeowners with a buffer against job loss or unexpected major expenses, declining prices and the accompanying credit freeze leave property owners more financially vulnerable.

Our paper also raises a cautionary flag to outcome-based education reform policy. An example of the movement in education policy towards data-driven outcome evaluation is found in the U.S. Department of Education's application for States to receive "Race to the Top" funding. The Attorney General in applying states must sign the declaration:

> I certify that the State does not have any legal, statutory, or regulatory barriers at the State level to linking data on student achievement [test scores]… to teachers and principals for the purpose of teacher and principal evaluation.

There is a growing literature that evaluates the appropriateness of the use of student outcome data in teacher and school evaluation.[3] Our results demonstrate that local economic fluctuations may change unobservable student potential in any given year. Measured outcomes-based policies that do not correctly adjust for influences on outcomes that schools do not control or affect -- local housing market and economic conditions, for example -- risk both punishing teachers and schools that perform well in the face of external challenges and providing rewards for improvements caused by factors unrelated to principal and teacher performance.

1.2 Mortgage default, family stress, and education outcomes

The academic autopsy of the ongoing housing market decline is well underway, and policy makers are redefining the regulatory structure and policy approach to the housing market.[4] The role of mortgage default and subsequent foreclosure is central to understanding and shaping a policy response to the housing market collapse. An exchange between Gerardi, Ross, and Willen (2011) and Been, Chan, Ellen, and Madar (2011) in the Journal of Policy Analysis and Management is an example of the continuing academic discussion of the causes, housing market

---

[3] For example, Koedel and Betts (2010) use San Diego Unified School District data to demonstrate that the problem for such assessment of non-random assignment of students to teachers can be mitigated using detailed longitudinal data.

[4] Chapter 3 of this dissertation is an example of this literature, and also references many other examples of research dissecting the housing boom and bust. Examples of housing policy discussions include two events by different federal agencies titled "Conference on the Future of Housing Finance," first by HUD on August 17, 2010, then by the FDIC on October 25th and 26th 2010. Examples of policy academic policy proposals for the reform of the Government Sponsored Enterprises include Marron and Swagel (2010) and Ellen, Napier, and Willis (2010).

consequences, and policy responses to widespread foreclosure. Mian and Sufi (2009) and Mayer, Pence, and Sherlund (2009) are among the first papers to identify and quantify factors may have led to the default boom: the magnitude of price declines and increasing unemployment (Bhutta, Dokko, and Shan, 2011), increases in subprime lending , slackened underwriting standards, increased securitization (Keys, Mukherjee, Seru, Vig, 2010), and bankruptcy reform (Li, White, and Zhu, 2010). Other research documents a contagion effect in foreclosure (Harding, Rosenblatt, and Yao, 2009) and a negative impact of foreclosure on neighborhood housing prices (Schuetz, Been, Ellen, 2008) and household credit (Brevoort and Cooper, 2010).

Many of these papers claim as motivation that mortgage default, particularly when resulting in foreclosure, involves social costs and additional losses in addition to homeowner equity and creditor losses. This paper contributes to the literature that substantiates this motivation. Cui (2010) and Ellen, Lacoe, and Sharygin (2011), use detailed micro data of foreclosures and reported crime to document a relative increase in criminal activity in neighborhoods after increased foreclosure activity in Pittsburgh and New York City, respectively, while Schweitzer and Shane (2010) find suggestive evidence that housing price declines hindered the growth of small business. Closely related to this paper, Been, Ellen, Schwartz, Stiefel, and Weinstein (2011) document that New York City elementary students living in buildings entering foreclosure in the 2006-2007 were disproportionately black, qualified for reduced price lunch more often, and attended schools with significantly lower test scores. They find that children living in properties that received a foreclosed notice were more likely to move to a new school, particularly when the property went to auction. They also show that the test scores of the schools to which students move during their sample are on average lower than the schools they leave, but find little evidence that the decline is greater for students moving from buildings receiving a foreclosure than other movers with similar observable characteristics.

Our findings also contribute to the education literature which measures the importance of external (to the school) inputs to student outcomes. Carrell and Hoekstra (2010) find that students exposed to domestic violence score substantially lower on standardized tests and have increased incidence of misbehavior, while their peers also exhibited relative test score and behavior declines. A number of studies document a negative impact of parental job loss on child academic outcomes.[5] Using state-level job loss data, Ananat, Gassman-Pines, Francis, and Gibson-Davis (2011) document a negative relationship between job loss and state-level test scores, and note that their estimates indicate observed job losses could lead to an increased share of a State's schools failing to meet No Child Left Behind benchmarks.

Mortgage payments represent the largest recurring expenditure for homeowner's with mortgages, and the home equity that accumulates with mortgage payments (and in the past, home price appreciation) frequently represents a substantial

[5] Stevens and Schaller (2011) use Survey of Income and Program Participation samples prior to the housing bust to show that parental job loss negative impacts student outcomes; Rege, Telle, and Votruba (2011) document a negative impact on children's school performance for paternal job loss and a positive impact of maternal job loss using Norwegian register data.

share of household wealth.[6] When a borrower fails to make scheduled mortgage payments, the lender may file a Notice of Default, the first step in the legal process of transferring ownership from the borrower to the lender. The California Department of Real Estate recently published a "Homeowners Guide to Foreclosure in California" which provides an overview of the default process and describes the financial and emotional state of a typical homeowner in default. The guide begins with four examples of why a homeowner may fail to pay a mortgage payment: a change in the monthly payment of an adjustable rate (or low introductory payment) mortgage; job or income loss or other adverse financial event; divorce; the loan is "underwater" - the outstanding debt exceeds the value of the home (strategic default). When a borrower in California fails to pay the mortgage, although the bank may legally begin foreclosure proceedings almost immediately, there is typically some communication between the borrower and lender to determine whether the delinquency can be resolved.[7]

Because processing a foreclosure and discharging a recovered property is costly, lenders have an incentive to encourage loans to "cure" and return to payment. An additional reason that the foreclosure process does not begin immediately with delinquency in recent years is that "loan servicers," the companies who collect mortgage payments on behalf of lenders and process defaults and foreclosures were

[6] Consumer Expenditure Survey, 2009, Table 51.

[7] While there is some variation across States in the default foreclosure, the process we describe in California is largely representative of that across the country, with some differences in the time given to borrowers to "cure" a default at varying stages of the process.

understaffed and unable to process the increase volume of delinquency. During this period after a borrower misses payment and before the bank begins the legal foreclosure process, the delinquency may be resolved in a variety of ways. These include:

- Repayment of arrears and any late fees.
- An agreement between the borrower and lender to modify or restructure the terms and payment schedule of the existing loan. Loan modifications have been encouraged by government policy, but are fraught with moral hazard and the likelihood of recurring default.
- A refinance where the borrower finds a new lender who pays the outstanding loan, leaving the borrower with the terms of a new replacement mortgage. A loan with a longer term, for example, may result in lower payments that a borrower can manage. When home prices began falling however, fewer lenders were willing to initiate new loans against homes with more uncertain future value. Market declines have left many homes worth less as collateral than the balance of the outstanding mortgage.
- If the default is a cash flow problem, and the equity in the home is greater than the mortgage, the borrower may sell the home to pay off the loan and recover any residual equity.
- The bank may approve a "short sale" where the loan is sold to a third party for less than the outstanding mortgage and the bank agrees to forgive the difference between the recovered and outstanding amount.

- Offer a "deed-in-lieu" of foreclosure, where the borrower returns ownership of the property to the lender.
- Find another avenue for payment of loan, such as moving and renting out the home.

Our dataset does not allow us to observe these events. There are borrowers experiencing financial stress who remain in our "control" group because their delinquency does not progress to the legal foreclosure process that is publicly recorded.[8] The exclusion of these borrowers will dampen our estimates of effects on academic outcomes. Borrowers who are able to resolve their delinquency without triggering the foreclosure process are also likely different than those who do not. Repayment, for example, requires access to financial resources such as from extended family wealth, while other resolutions are more likely for borrowers more savvy or tenacious in their interactions with banks.

To gain leverage in the effort to return a delinquent mortgage to regular payment or take possession of the house that is collateral for the loan, lenders begin the legal foreclosure process by filing a Notice of Default with the local county recorder. Copies of this notice must also be mailed to the last known address of the property owner and physically posted on the property. This action begins a 90 day "reinstatement" or "cure" period in which the borrower may pursue any of the actions listed above to become current on payments or discharge the debt. After 90 days, the

---

[8] In future work, we plan to compare the impact of sales that are at a loss and potential "short" as another indicator of financial stress using our property transaction data described below to impute this outcome.

lender may file a "Notice of Sale" or "Notice of Trustee's Sale", which begins a minimum 20 day period prior to an auction of the property which is typically subject to a minimum bid, which if not satisfied results in the ownership of the property transferring to the lender.[9] While some new owners have begun accepting tenancy arrangements with former owners after trustees it is typically the case that the former owner is required to move, by eviction if necessary, following the trustee's sale date.

The filing of a Notice of Default is unambiguous evidence of family (or owner, for rented properties) financial stress. At a minimum, the owner has lost enough equity in the property to decide that the reputation and transaction costs of defaulting on the loan are less severe than the cost of continuing to pay off the debt.[10] The Guide to Homeowners consoles the prospective defaulter:

> Losing your home through foreclosure is a traumatic experience that usually occurs at a time when you already are facing significant financial, and even physical and psychological stress. It is understandable why some homeowners make poor choices when facing foreclosure.

We argue that this potential for home loss via foreclosure that begins with delinquency and a subsequent "Notice of Default" will disrupt the family attention and home environment that are inputs to a children's education. This argument is consistent with standard models of the determinants of children's attainments, as outlined for example

---

[9] In February of 2009, the California Foreclosure Prevention Act (CFPA) extended the reinstatement period by 90 days for loans held by lenders or servicing agents that did not have an "approved comprehensive loan modification program."

[10] Even for "strategic defaulters" who have the wherewithal to continue payment but choose not too because of negative equity, there remains a negative social stigma to mortgage default (Guiso, Sapienza, Zingales, and Macelli, 2009).

in Haveman and Wolfe (1995). Of course, failing to make mortgage payments temporarily increases financial resources relative to making the payment. For short spells of delinquency, we expect any positive impacts from this windfall to be small. However, in some cases, due to legal actions or mortgage servicer backlogs the time that families may postpone the financial ramifications and enjoy the immediate financial benefit of delinquency has protracted to six months or longer.[11]

A disruption in the home environment due to family financial stress may also spill over to affect a student's peers. While Carrell and Hoekstra's work on domestic violence provides an extreme example of this argument, the less offensive stress of financial hardship at home may still promote distraction or disruptive behavior that degrades peer's learning environment.

The social cohesion and the quality of the physical maintenance of the neighborhood in which a child lives are also inputs to a child's academic outcomes. Because they are likely to move, neighbors in default may exhibit less commitment to the neighborhood and engage in less beneficial social behavior, and cohesion can suffer from increased turnover. Meanwhile, vacated properties tend towards blight and neglect, and have been shown to correlate to increased crime (Cui, 2010; Ellen, Lacoe, and Sharygin, 2011). We anticipate that negative social spillovers of default from both neighborhoods and classrooms to negatively affect student academic outcomes.

---

[11] We find some suggestive evidence of improvement in student outcomes in the school year subsequent to a Notice of Default relative to the year of the default, but additional data and analysis are necessary to confirm this artifact.

While our exposition of the default process has focused on homeowners, landlords can also fail to meet mortgage obligations. Indeed, in our data, just more than half of the Notices of Default in the San Diego Unified School District during our sample period do not receive an owner occupant tax exemption (our indicator of owner occupancy). Aside from failing to pay the rent or otherwise imposing a financial burden on the owner, tenants are not responsible for the default, and in many cases are unaware of the financial distress attached to the property until the foreclosure notice is posted. Tenant protections during foreclosure have increased in response to the wave of defaults, with the right to remain in the property extended from 30 to 60 days after a new owner takes possession of the property.[12] It is common for a bank or new owner to evict tenants from buildings in preparation for the resale of the property, although banks have increasingly begun negotiating with tenants to make alternative arrangements. While the default and foreclosure likely does not result from the tenant's financial situation, there remains the stress of unexpected relocation. Banks and new owners often offer to pay moving expenses for tenants who will move immediately, which can buffer the cost of forced relocation. Tenants remain legally liable for rent payments, although anecdotally it is difficult for delinquent landlords to enforce collection. On balance, we expect that there may be a smaller, shorter duration negative impact of owner default to student's living in rental properties.

1.3 Empirical approach

[12] In some states, this protection extends to the end of the term of any existing lease agreement.

We begin by describing the incidence of mortgage default in the San Diego Unified School District (SDUSD). We compare average individual, teacher and class, school, and neighborhood (census block group) characteristics for students who experience a mortgage default episode in a given year to those who do not. Our unit of observation throughout is a student in a given school year. In addition to comparing sample means for a large set of variables, we describe the incidence of default by regressing an indicator variable for default on our large set of characteristics. Because we expect owner-occupants in single family homes to be more affected by mortgage default, we also make these comparisons for this subpopulation.

We rely on the panel structure of our data to identify the causal effect of mortgage default on student academic outcomes. The key empirical difficulty in identifying this effect is the possibility of unobservable student, family, or school attributes that are different for students who do and do not experience mortgage default that may correlate with student outcomes. We want to know the effect of default on students who experience default, or

$$E[Outcome_{1i}|Default_i = 1] - E[Outcome_{0i}|Default_i = 1],$$

where the ith student's academic outcome, in our case math and reading test score and absence rate, differs by whether they experience a default,

$$Outcome_i = \begin{cases} Outcome_{1i} \text{ } if \text{ } Default_i = 1 \\ Outcome_{0i} \text{ } if \text{ } Default_i = 0 \end{cases}$$

and $Default_i = 1$ indicates that a student experienced a default episode. Of course, we are unable to observe what a student's outcome would have been in the absence of a default episode for students who default.

Our comparison of the incidence of default highlights that there are substantive differences in observable characteristics, schools, and neighborhoods of students experiencing default. These differences suggest that unobservable inputs to academic outcomes, such as ability or parental involvement, may also differ across students who do and do not experience mortgage default. As such, observed differences in academic outcomes for the two groups, even after controlling for observable characteristics, are the combined result of the mortgage default episode and any differences in unobservable test score inputs.

To estimate the difference between observed outcomes for students who default and the unobservable counterfactual of their scores in the absence of default, we assume that differences between a student's average test scores over the entire sample period and her score in the year of a default episode is attributable to the default after controlling for time-varying observable test score inputs. More specifically, we assume that

$$E\left[Outcome_{0i} \middle| \alpha_i, \widetilde{\boldsymbol{X}}_{it}, \boldsymbol{Z}_{it}, Default_i = 1\right] = E\left[Outcome_{0i} \middle| \alpha_i, \widetilde{\boldsymbol{X}}_{it}, \boldsymbol{Z}_{it}, Default_i = 0\right]$$

where α_i is a students' time-invariant attributes, $\widetilde{\boldsymbol{X}_{it}}$ includes time-varying observable student attributes, and $\boldsymbol{Z}_{it}$ includes teacher, school, and neighborhood characteristics.

To implement this assumption we estimate the empirical model

$$Outcome_{it} = \beta_0 + \gamma Default_{it} + \widetilde{\boldsymbol{X}}_{it}\beta + \mathbf{Z_{it}}\delta + \alpha_i + \varepsilon_{it} \quad (1)$$

as a fixed effects panel regression. The estimated coefficient on the default episode indicator, $\hat{\gamma}$ is the effect of mortgage default on student outcomes unless students experiencing default episodes would have otherwise experienced different trends in observed outcomes which are not explained by time-varying inputs. If our assumption is incorrect and students experiencing defaults would systematically change position in the within-grade and cohort distribution of test scores (we normalized scores by grade and year) or rate of absenteeism even without default, our estimates of the causal effect of the financial stress that results in default will be biased in the direction of the systematic difference. If this is the case, our results are still informative in measuring the combined effect of the default episode and whatever secular bias may occur due to differing time-invariant characteristics.

We explore the possibility of heterogeneous impact of default episodes by augmenting equation (1) by interacting $Default_{it}$ with teacher experience, grade, and census block group median income.

As an alternative specification we consider $PostDefault_{it}$ in place of $Default_{it}$, where

$$PostDefault_{it} = \begin{cases} 1 & \text{if } Default_{i\tau} = 1 \text{ for } \tau \leq t \\ 0 & \text{otherwise} \end{cases}$$

This specification assumes that rather than having a one year impact, a default episode imposes a permanent shift in outcomes. Note that the coefficient will be identified by students that have outcomes measured prior to default. In practice, over 50 percent of our default spells include two years where $PostDefault_{it}$ is "turned on", and 19 percent include the maximum in our data of 3 years.[13]

1.4 Data

To estimate the effect of mortgage default on student outcomes and describe the population experiencing default, this paper uses individual student records from San Diego Unified School District (SDUSD) administrative databases, publicly available legal notices of mortgage default, and county tax assessor records of property characteristics and sales. Property boundary maps, school catchment area maps, census boundary maps, and Census 2000 neighborhood characteristics data are used to merge datasets, group records, and characterize neighborhoods.

We obtain our outcome variables of math and reading test scores and absenteeism rates from an anonymized "student-level" dataset that links individual students' records longitudinally. The data include all SDUSD students' academic experience and demographic characteristics. We focus on students enrolled in elementary grades two through five. The California Standards Test (CST) is given in each grade each year starting in second grade. We use math and test scores separately, standardized within school year and grade, through fifth grade. We focus on the elementary population in part because elementary students typically have a primary "home room" classroom and teacher, which represents a natural peer group and which is reported in the student record. The other outcome measure we recover from the

[13] As additional data becomes available we will be able to estimate more flexible post-default models.

student record is the percent of school days the student was absent, which, like Babcock and Betts (2009), we argue is outside of the elementary school student's direct control.

The Notice of Default that begins the legal process of foreclosure on the delinquent mortgage and repossession of the mortgaged home is a public document. A San Diego area business publication, the San Diego Daily Transcript, extracts the details of each notice and makes them available to subscribers to its website as a searchable directory. These public listing are the source of our mortgage default data.[14] In this paper, we use the date of the default and assessor's parcel number included for the property collateralizing the mortgage in default. The parcel number is used to merge the default record to county assessor's property map, record of property characteristics, and property sales file. The default date is used to identify the school year in which a "default episode" occurred. We categorize the timing of a default episode slightly differently with respect to the default date for single family owner occupied and renter occupied or multi-unit properties (as indicated in the county assessor data described below). For single family owner occupied properties, we categorize a default as occurring during a given school year if the default date occurs prior to the August 31st following the end of that school year. This represents 90 days after the first of June ending the school year, a conservative estimate of the time since

[14] Notice of Default listings for properties with five or more units are not included in the San Diego Daily Transcript database. Alternate data sources would be required to determine the effect of owner financial stress on tenants in these larger properties. We do note, however, that our estimates find no impact on student outcomes for renters in single family or 2-4 unit buildings where the landlord becomes delinquent on the property's mortgage.

the homeowner stopped making payments. The California Standards Test is administered during a two to three week window centered around the day on which 85 percent of the instructional year is completed, which for a typical school would fall near the end of April or beginning of May. A single family owner occupant default would have been struggling to meet mortgage obligations by at least this time, so that stress associated with defaults occurring later than this date was present during the school year. For renter occupied single family properties and multi-unit properties, we use the actual filing date of the default notice, rather than the imputed delinquency 90 days earlier. A Notice of Default is physically posted on the property on this date, and may be the first indication to a tenant that the property in which they live is in financial disarray. It is at this point that tenants are certainly aware that they may be pressured to move, evicted at the end of their lease, and at least must navigate a change of ownership of their home.

A second SDUSD administrative dataset of student home addresses is necessary to connect students to properties. The available dataset lists the home address on file for each student in the district in the fall of each school year beginning from 2001-2002 to 2009-2010. The addresses are linked to the student records by the anonymized student ID numbers.[15] These students addresses are matched to property records by house number and street name, with successful matches obtained for 84 percent of student addresses.[16] A successful match requires that the postal address

[15] Confidentiality of student identity was strictly maintained throughout the research process.
[16] Four percent of addresses were successfully matched, but to parcels zoned for non-residential use. An audit of these addresses using Google Maps verifies that the student's

listed for the student matches the property physical address included in county property records (described below). Some addresses fail to match because units in multi-family properties in San Diego can have unique street numbers rather than apartment numbers, whereas the entire property has a single parcel number and a single associated street address in the county records. To our knowledge there is no publicly available crosswalk between properties and all of their constituent valid street addresses. Our second step to matching student record addresses to parcels relies on geocoding methods. Each address is plotted on the map using standard geocoding software and assigned to the parcel in which the geocoded point fell. Points that map to non-residential parcels are discarded as incorrect matches. An additional 6.3 percent of addresses fall within parcels with multiple residential units, while an additional 2 percent of the geocoded points fall in single family parcels. Ultimately, 89 percent of student addresses are successfully assigned to a residential county parcel number.

The combination of individual student records with home addresses to notice of default and other property records is a key contribution of this paper. We note, however, that the addresses appear at annual intervals and that an examination of property transactions indicates that addresses on file are sometimes not updated immediately when students move.[17] We update the frequency of student addresses to

---

address on file is indeed associated with a non-residential building such as a school, church, business, or warehouse.

[17] For example, a single family owner occupied property where a student is listed for a number of consecutive school years is sold. The student's address remains unchanged in the subsequent school year address database, but is then changed to the address of a single family owner occupied home purchased within days of the sale of the prior property.

school quarters or the summer using the detailed school enrollment information from the student records as well as county assessor's property transactions as follows. For a change in addresses from one year to the next where a change in catchment areas associated with the addresses coincides with a change in the school attended recorded in the student records, we update the timing of the addresses move to that of the school move.[18] When an owner occupied property listed as a student's address for multiple years is sold (purchased) near the end (beginning) of the listing spell for the student, we update the timing of the address change to correspond to the date of the property transaction. We are conservative in our adjustments in assigning as a summer move any case where differing property transaction dates create ambiguity as to whether the move occurred during a school year, and we do not adjust address changes by more than two school years for owner occupied transitions and by more the one school year for any transition.

An individual property boundary map for all parcels in San Diego County is used in the geocoding process linking student addresses to properties described above, as well as to determine the elementary school catchment area and census block in which each parcel lies by superimposing these respective maps on the parcel map. The parcel map is publicly available from SanGIS, a San Diego City and San Diego County joint agency that maintains a geographic database of the area.[19] The map

[18] Parcel catchment areas are determined using GIS methods. Elementary school boundary maps obtained from SDUSD for each school year are superimposed on the county parcel map, allowing the assigned school for each parcel to be determined.
[19] Data was retrieved from www.sangis.org in January 2010.

embeds characteristic data for each parcel including the current land use (single family, multi-family, retail, etc.), owner occupancy status, and number of units. The characteristics information was verified using the 2003 and 2008 vintages of the San Diego County Assessor's property characteristics databases, which include additional property characteristics not recorded with the map. Of these, we use property square footage to calculate an average square feet per unit variable by dividing the property square footage by the number of units. In some instances the data vintages differed on recorded characteristics, as homes were constructed or remodeled or as a result of data errors. In this study we use the "land use" variable in conjunction with the number of units to classify the property as Single Family, Condo, 2-4 Family, or 5+ family. These variables are very stable across the data vintages (except, of course for newly built properties). We also rely on the owner occupancy category. San Diego property owners may claim at most one property as a primary residence.[20] When owner occupancy status changes between data vintages, we impute the date of change to the last sales date prior to the subsequent data vintage. Because we anticipate a greater impact of mortgage default on owner occupants, we are conservative in updating this variable in favor of indicating non-owner occupancy. We make similarly conservative updates to substantive square footage changes. We also join the data to the publicly

[20] According to the Assessor's website, http://arcc.co.san-diego.ca.us/faq.aspx#Exemptions, applying for owner occupancy status exempts $7,000 of assessed property value from property taxes for an annual savings of approximately $70.

available Assessor's records of all property transactions, which include sales date and price to determine student address changes as described above.[21]

Based on this property data, 39% of approximately 394,000 housing units within the SDUSD boundaries are single family residences, 27% are in mutli-family buildings with at least 5 units, 21% are condominiums, and 11% are in properties with 2-4 units, and 2% are classified as mobile homes, time shares, or other.[22] The data report that 40% of units in the SDUSD boundaries are owner occupied, with 74% of single family residences and 50% of condos owner occupied. Tenure measures differ on aggregate from shares reported in the American Community Survey (ACS) for the school district, which reports a lower 47% percent owner occupancy rate across all housing types. The single family detached rate of 75% closely matches the 74% rate in the County data. In all, based on the ACS estimates for the SDUSD geography, it appears that owner occupancy is somewhat underreported in the County data for units in multi-unit buildings.

The assessor characteristics and notice of default records are used to determine six mutually exclusive categories of default episodes in a given school year:

[21] The sales data begins in 1983 and includes all sales through December 2010.

[22] These shares differ somewhat from the 2005-2009 American Community Survey estimated shares for the district, which are based on a total unit count that is 23,232 units more than that reflected in the Assessor data. While the classifications are not directly comparable across dataset, some of the differences in calculated shares appear to arise from the differences in how condominium units are reported.

1. Single family owned: a Notice of Default is filed for the mortgage on the single family property with an owner occupant tax emption within 90 days of June 1st.
2. Condo owned: a Notice of Default is filed for the mortgage on a condo, mobile home, or time share before August 31st.[23]
3. Single family not owned: a Notice of Default is filed respective to a single family property with no owner occupant tax exemption before June 1st.
4. Condo not owned: a Notice of Default is filed respective to a single family property with no owner occupant tax exemption before June 1st.
5. 2-4 Unit owned: a Notice of Default is filed respective to a 2-4 unit property with an owner occupant tax exemption before June 1st.
6. 2+ Unit not owned: a Notice of Default is filed respective to a property with 2 or more units and no owner occupant tax exemption before June 1st.

In addition to providing our educational outcome variables, the SDUSD student records provide a wealth of student, teacher, and school characteristics that we use to describe the incidence of mortgage default on SDUSD elementary students and as regression controls. Administrative characteristics include the student's grade, the school they attend, and their primary teacher. Demographic characteristics include ethnicity, English learner status, and for nearly three in four students their parent's level of education. Teacher characteristics include the number of years taught, and whether the teacher has a Bachelor's or advanced degree, while class characteristics

[23] Mobile homes and time shares make up a negligible share of properties. All results are robust to alternate groupings of these property types.

include the number of classmates. School characteristics include percentages of reduced price lunch qualifiers, English learners, and students of each ethnicity. Because we match student addresses to geographic location, we are able to also add descriptive neighborhood variables from the 2000 census (at the census block group level) to each student record with a matched address. There are 1084 census block groups represented in our data (compared to 161 schools with elementary students), with an average population of about 1600.

We use our merged dataset to calculate additional variables relevant to our analysis of spillover effects from classroom peers and neighborhood defaults. For each school year and for each property, we calculate the percent of the rest of the properties in the census block experiencing any category of default, as well as the percent experiencing a single family owned default. Similarly for each student-year observation we use the teacher identifier to group students into classes and calculate the percent of classmates (omitting the student) exposed to any type of default as well as those exposed to a single family owned default.

1.5 The incidence of mortgage default among SDUSD elementary students

The distribution of property types in which SDUSD students live and the overall sample frequency of any type of default is reported in Table 1.1. Single family homes are the predominant housing type in the district, and have the highest sample default rate of 2.1%. Most defaults occur in the latter half of our panel with the highest rate of students experiencing default in the 07-08 year, when over one in twenty students living in single-family, owner-occupied homes experienced a default and more than one in ten students living in single-family renter-occupied homes experienced a default, as shown in the rightmost column of Table 1.2.

Students experiencing a mortgage default in a given school year differ on average from those who do not on a variety of observable dimensions.[24] Among single family owner occupied properties, the homes are on average 270 square feet smaller. Also reported in Table 1.3, default is associated with a higher incidence of a recent (summer) move as well as a move during the school year. Students are also more likely to attend multiple schools in a given year. These averages are consistent with the results found by Been, Ellen, Schwartz, Stiefel, and Weinstein (2011) in New York City described above.

Summary statistics of our student characteristics variables, reported in Table 1.4, reveal that Hispanic students, the most common ethnicity among SDUSD, and blacks are disproportionately overrepresented in both the any and single-family default groups. It is possible that this reflects higher levels of subprime lending among minorities, a documented characteristic of the housing credit boom (Mayer and Pence 2008). Defaults and related foreclosures on subprime loans are often described as the "first wave" of the foreclosure increase. Prime loans to more qualified borrowers have recently begun to default, and data for subsequent school years may reveal a more

[24] Because students experiencing default make up a relatively small proportion of our sample, sample means for the No Default groups are nearly identical to means for all observations. We report only the No Default statistics to condense the presentation.

nuanced pattern of default and race and ethnicity. Parental education levels are also lower on average for students experiencing default in single-family owned housing, with almost twice the rate of less than high school and 1.5 times as the rate of high school grads. Again, this may be related to trends in subprime lending which defaulted early in the lower tiers of the housing market where prices rose and fell more dramatically. Differences were more muted for those in the full sample.

Students in higher grades are also more likely to experience a default. We suspect that this is an artifact of the relationship between the life-cycle timing of buying a home which led parents to purchase homes during the housing market boom and their children started school, whereas prices were falling as students in younger grades in our sample period with high defaults reached the age when their parents would have made a home purchase.[25]

We next compare a variety of census block group neighborhood characteristics of students experiencing default to their peers. As shown in Table 1.5, average tract group median income is lower by $3,000 among all types of default, and $13,300 for single family owned defaults. On average, the neighborhoods where students experiencing default live have historically (2000 census) higher poverty and unemployment rates, with more military employment and higher historical (2000 census) unemployment. Differences in racial composition and adult educational attainment of the student's neighborhoods parallel the differences found for the student's own characteristics.

[25] Additional analysis of our sales data is necessary to provide further evidence for this idea.

Figure 1 depicts the spatial concentration of defaults in the district in the 2008-2009 school year. Differences in the demographic makeup of the southern portion of the district, where, within many school catchment areas more than 1 in 10 properties experienced a default, and the northern and coastal areas are consistent with the averages we find in our data.

While students experiencing mortgage default in our sample period live in socioeconomically more disadvantaged neighborhoods, as shown in Table 1.6, their teachers do not differ substantially from their peers, with the exception of teacher experience among single family owned defaulters, which is one year lower. In the full student sample, teacher education levels are actually higher for students with any default relative to their peers, although no difference emerges in the sample of students in single-family owned homes. However, the average number of classmates is higher for any and single family owned defaulters, and a greater percentage of those classmates also experience default, consistent with the spatial concentration of mortgages depicted in Figure 1. This pattern is also observed in our school level demographic characteristics where, for example, average rate of reduced price lunch eligibility is over 70 percent for students experiencing any default, and on average a third of students in schools attended by students with a single-family owned default are English learners as compared to a fifth for those with no default.

To summarize, when compared to the rest of our sample, students in a given year who experience a mortgage default are disproportionately Hispanic and black, have less educated parents, and live in socioeconomically disadvantaged neighborhoods. Our findings are consistent with others in the literature on subprime lending and default incidence and earlier work in New York City discussed above. While teacher qualifications are comparable across groups, students experiencing default attend larger classes in schools with more English learners and students qualified for reduced price lunch. As data on subsequent school years with more defaults in other segments of the housing market becomes available, these comparisons may reveal a more nuanced story.

1.6 Mortgage default and academic outcomes

The differences in mean student characteristics for students who experience mortgage default are reflected in the summary statistics of the outcome variables we use to examine the effect of default on student outcomes. Our first two outcomes are California Standards Test scores for math and reading are standardized by grade and cohort for all students in the SDUSD administrative records. The standardized score represents a student's distance measured in standard deviations from the cohort-grade mean.[26] Our third outcome is the percent of days a student is absent during the school year, drawn from administrative attendance records.

As reported in Table 1.7, whereas students living in single family owned homes score 0.38 and 0.43 standard deviations above their cohort grade mean in math and reading respectively, those experiencing a default score -0.05 and -0.01 standard deviations below average. Means of absence rate are higher by more than half a

[26] The mean score in our sample is greater than zero because, to facilitate our panel data estimation approach, we drop observation for students with less than three years of test scores from our data.

percent for students experiencing any type of default, while single family owned defaults have an 0.8 percentage point higher rate of absenteeism than those in single-family owned houses without a default.

Some of the differences in these unconditional means are explained by observable student, teacher, neighborhood, and school characteristics. Tables 1.9, 1.10, and 1.11 present regression coefficients for math, reading, and percent of days absent for four empirical specifications. Each column reports coefficients for indicator variables for each type of Notice of Default as defined above. The dependent variable in columns (1) and (3) - (5) is the *level* of the given outcome variable, while column (2) regresses first differences in the outcome variable on the default group indicators and controls. These latter "gains" models are included as alternate specifications and estimated coefficients are similar to those in our preferred fixed effects specifications. In column (1), which includes the largest set of control variables but no group fixed effects, statistically significant conditional correlations persist of -0.195, -.0144, and 0.758 for math, reading, and percent absence respectively. Column (3) replaces census tract group control variables with census tract group fixed effects. Coefficient estimates change very little from (1) to (3) for all three outcomes, suggesting our set of control variables approximates the more flexible census tract fixed effects well. Column (4) includes teacher fixed effects and shrinks the significant coefficients on single family owned default towards zero for math and reading scores, which suggests that our teacher experience and education controls do not fully capture outcome differentials attributable to teacher heterogeneity. Column (5) includes both teacher and census tract fixed effects in place of teacher and tract characteristics. For these models, a statistically significant negative conditional correlation persists after controlling for observable characteristics for single family and condo owned (marginally significant) and single-family rented defaults in math, while the same variables show a positive relationship with rate of absenteeism, while for reading scores, only single-family owned default remains significant.

The marked differences in observable characteristics of students by whether they experience a default suggests that these students may also differ on unobservable inputs to education outcomes. We isolate the causal effect of the financial stress by estimating the student fixed effects model of equation (1) outlined above. As reported in Tables 1.12-1.14, after controlling for unobservable time-invariant student characteristics, observable time-varying student characteristics, and class, teacher and neighborhood characteristics, students experiencing a single family owned default score a 0.048 standard deviations lower on math and are absent 0.233 percentage points more frequently, both statistically significant at the five percent revel. We also estimate that reading tests scores have a negative coefficient associated with single-family default, but the effect is not significant. Because we find causal effects only for single family owned defaults, we focus on these default group for our remaining estimations.

The effect size on math scores is meaningful. A drop of 0.05 standard deviations in math scores is enough to move a student from the 50th percentile to the 48th percentile in a single year. This effect could of course compound across students within a year (through peer effects) and across years for the individual student. We investigate both issues below.

Another way to put a drop of 0.05 standard deviations in perspective is to compare to the estimated effect of other reforms. Clotfelter, Ladd and Vigdor (2007) estimate that in North Carolina reducing class size by 5 students is associated with gains in achievement of 0.01 to 0.015 of a standard deviation. Thus the impact of a default is quite high relative to the ability of policymakers to boost test scores through class size reduction.

In addition to finding that single-family owner-occupied default adversely impacts individual students, we find evidence consistent with our expectation that students are adversely affected as the proportion of classroom peers experiencing single family owned default increases and by neighborhood decline associated with increased rates of mortgage default. We estimate a -0.00183 coefficient on math scores for the percent of classmates with a single-family owned home entering default in that year. This variable has a minimum and median of zero for the entire sample over all years, but evaluated at the mean of 4.17 for the greater than zero subsample, gives a -0.008 decline in math scores associated with peer default. Similarly for the neighborhood peer group, an increase in the share of block group neighbors experiencing default decreases math scores, with a coefficient of -0.00565. At the mean of nonzero observations of this variable of 2.3%, this corresponds to a -0.013 standard deviation decline. Reading scores show a similarly sized negative relationship with neighborhood foreclosure rate, while absence rates are positively related to peer default rates: evaluating the coefficients at the mean among observations with greater than zero peers experiencing defaults indicates an increase of 0.059 and 0.042 percentage points in percent days absent.

Put together, these two distinct type of peer effects, evaluated at the mean for those with defaults in their peer group, implies a drop of 0.02 standard deviations in math scores. The foregoing estimate of the impact of class size on math achievements indicates that to counteract this drop one would need to lower class size by seven to ten students.

We further examine the effect of default by interacting single-family owned default with neighborhood median income, teacher experience, and grade level. Coefficients for regressions including these interaction terms are reported in Tables 1.16-1.18. We measure no differential effect of neighborhood resources as measured by median income for any of our outcome measures. Our estimated negative effect on math scores is exacerbated by -0.170 standard deviations for students with teachers in their first three years of teaching. This is a very large effect and deserves further investigation. We find no evidence that reading scores or attendance rates among students residing in defaulting homes are differentially higher when they also have inexperienced teachers. Interacting default with grade level suggests that our estimates are largely driven by impacts on students in the older grades. These results suggest that the disruption of the financial stress of default may have a larger impact in the education production function of older students.

Our last set of empirical results, reported in Tables 1.19 and 1.20, examine our specification and reliance on student fixed effects. In Table 1.19, we report regressions that allow the effect of foreclosure to persist beyond the initial year. Columns (4) through (6) include an indicator for default in the current or any prior year, rather than the current year alone. The estimates give evidence of continued negative impact beyond the first year, and particularly that reading scores exhibit a decline when the post default periods are combined. In reporting these results, we note that most defaults occur at the end of our sample, and additional years of data will allow a greater focus on long term impacts. An alternative explanation for our estimated effect of default is that unobservable student characteristics are related to a secular decline in relative outcomes for students that also happen to experience default. We investigate the possibility of a trend in the regressions reported in Table 1.20 by regressing our outcome variables on subsequent defaults, limiting our sample to pre-default observations. We find no statistically significant effects of future default on the year prior or two year prior outcomes, evidence against a prevalent downward trend for defaulting students.

The estimated effect on both math and reading scores two years after a default are quite striking, with predicted drops of 0.16 standard deviations in math and 0.08 standard deviations in reading. Assuming a normal distribution, affected students who started at the 50th percentile are predicted to drop to the 47th percentile in math and the 44th percentile in reading two years after a default.

In summary, we find that students who experience default are different on a wide array of observable characteristics. They are more often Hispanic or black, have less educated parents, attend schools with greater poverty levels, and live in poorer neighborhoods. Students living in properties that receive a notice of default have lower math and reading test scores and higher absence rates than their peers after controlling for a varied and detailed set of observable individual, school, teacher, and neighborhood characteristics. To estimate the effect of default on education outcomes, we control for unobservable time-invariant student inputs with student level fixed effects. We find a -0.048 decrease in math scores, no relative change in reading scores, and an increase in absence in the year of the default. Teacher inexperience substantively exacerbates this math score effect. Specification checks suggest that there may be persistence in the effects we estimate, with reading scores declining in post default years taken together relative to pre-default years. Limiting our sample to pre-default years finds no evidence of pre-default declines in outcomes. This is an important test as it shows that there is no placebo effect, in the sense that we are capturing trends that were underway regardless of default.

1.7 Conclusion

Borrowers' inability and unwillingness to meet mortgage obligations in the aftermath of the housing market collapse and subsequent recession continues to be a stress on families, the housing market, and the economy. This paper presents some of the first empirical estimates of how this financial stress permeates other aspects of family and community wellbeing using an extensive dataset of student academic records, property records, and geographic information for San Diego Unified School District.

Notices of Default in San Diego Unified School District during our sample period were disproportionately experienced by students with lower test scores, that were Hispanic or black, had less educated parents, and lived in neighborhoods with lower median incomes. These observable differences motivate our use of student-level fixed effects regressions to control for time-invariant unobservable student characteristics that may influence our outcomes and be correlated with default. We find that students living in single family owner occupied homes that enter into default experience declines in math scores in the year of the default and increased absenteeism in the year of the default relative to their own average outcomes in all years after controlling for neighborhood, teacher, and school inputs and time varying individual characteristics. We find evidence that the impact of a default for a given student more than doubles two years after the default, and becomes quite large in terms of its impact on a student's ranking in the achievement distribution. Absences respond only in the year of the default, however.

We find no evidence that renters in default experience measurable declines in outcomes relative to their sample average, although like single family defaulters, they differ in average observable outcomes and characteristics from their peers. Our results speak directly to the large literature on peer effects. Rare to this literature, we are able to identify peer effects working both through the classroom and the immediate neighborhood. We find moderate but statistically significant effects of increased default rates among a student's classmates and in a student's neighborhood, providing evidence of consequences of default that are external to the borrower's household.

The impact of the housing market collapse on children and the external effects on neighbors and classmates that we document support a role for policy to prevent future housing market instability. Our finding that the negative impact on outcomes is exacerbated for students with inexperienced teachers suggests a possible avenue for schools to counteract the negative response. Additional support to inexperienced teachers with pupils experiencing out-of-school stress may help mitigate the most severe negative impacts.

Our research also informs the discussion of the usefulness of student micro data in teacher and school evaluation. Housing market and family financial distress, which we show to influence academic outcomes, are stresses outside of principal and teacher influence. Teacher and school evaluations based on student outcome data may face a difficulty in dealing with external shocks to student achievement. The federal No Child Left Behind law requires states to evaluate schools based on the percent of students, overall and by numerically significant subgroup, who are deemed proficient on the state test. Further, the recent Race to the Top competition for federal education dollars incentivized states to draw up plans to evaluate teachers to some extent based on the test scores of their students. Models of teacher effectiveness may potentially misattribute local housing market or economic decline fluctuations to teacher or school effectiveness. Unfortunately, with double digit rates of all US mortgages still in delinquency or foreclosure, negative impacts beyond the housing market we describe may continue to present a challenge for families and educators.

I thank Julian Betts, coauthor of the research presented in this chapter. It is with his permission that I include our research in this dissertation.

Table 1.1: Property type of SDUSD student homes

	Full Sample		No Default		Default	
Property Type	N	% of all types	N	% of type	N	% of type
Single Family	86,651	50.6	84,861	97.9	1,790	2.1
Condo	19,059	11.1	18,794	98.6	265	1.4
2-4 Family	24,153	14.1	23,790	98.5	363	1.5
5+ Family	36,390	21.5	36,340	99.9	50	0.1
No Valid Address	5,065	3.0				
Total	171,318					

Sample includes student by school year observations from 2001-2009 with at least three math and reading test scores for students in second to fifth grade.

Table 1.2: Default type by year

School Year	No Default	Single Fam Owned (% of type)	Condo Owned (% of type)	Single Fam Rented (% of type)	Condo Rented (% of type)	2-4 Fam Owned (% of type)	2 + Family Rented (% of type)	Total	Total % with any default
01-02	14,476	23	2	30	5	0	8	14,544	0.47%
		0.51%	0.45%	1.37%	0.43%	0.00%	0.17%		
02-03	22,263	31	4	47	1	2	8	22,356	0.42%
		0.43%	0.57%	1.34%	0.05%	0.30%	0.11%		
03-04	29,308	36	3	40	1	1	5	29,394	0.29%
		0.37%	0.31%	0.86%	0.04%	0.12%	0.05%		
04-05	28,962	57	2	50	5	6	5	29,087	0.43%
		0.57%	0.21%	1.09%	0.21%	0.68%	0.05%		
05-06	25,955	101	20	83	8	2	16	26,185	0.88%
		1.08%	2.40%	1.98%	0.41%	0.26%	0.19%		
06-07	21,592	210	24	125	44	17	61	22,073	2.18%
		2.62%	3.10%	3.53%	2.71%	2.65%	0.88%		
07-08	16,259	325	37	330	53	32	125	17,161	5.26%
		5.08%	6.13%	11.51%	4.57%	6.34%	2.33%		
08-09	10,034	161	24	157	35	17	90	10,518	4.60%
		4.03%	6.38%	8.75%	4.99%	5.74%	2.81%		
Total	168,849	944	116	862	152	77	318	171,318	1.44%
		1.59%	2.06%	3.15%	1.13%	1.53%	0.57%		

Sample includes student by school year observations with at least three math and reading test scores for students in second to fifth grade.

Table 1.3: Descriptive variable summary statistics by default: housing and moves

	Full Sample			**Single Family Owned**	
		Housing characteristics+			
Owner Occupied	No Default	Default		No Default	Default
N	163785	2468 *			
Mean	0.42	0.46			
StDev	0.49	0.50			
Square Feet					
N	140836	2427	N	58291	933
Mean	1269.63	1246.95	Mean	1643.49	1373.14 *
StDev	644.94	505.77	StDev	668.02	469.04
Min	250.75	340.5	Min	323	450
Max	10285	4120	Max	10285	3669
		Move Type			
Summer	No Default	Default		No Default	Default
N	168849	2469	N	58366	933
Mean	0.13	0.19 *	Mean	0.08	0.16 *
StDev	0.34	0.39	StDev	0.27	0.36
School year					
N	168849	2469	N	58366.00	933.00
Mean	0.01	0.02 *	Mean	0.00	0.01 *
StDev	0.10	0.15	StDev	0.06	0.11
		School Change			
Between school years	No Default	Default		No Default	Default
N	168849	2469	N	58366.00	933.00
Mean	0.10	0.16 *	Mean	0.05	0.07 *
StDev	0.31	0.37	StDev	0.22	0.25
During school year					
N	168849	2469	N	58366.00	933.00
Mean	0.08	0.08	Mean	0.06	0.12 *
StDev	0.27	0.27	StDev	0.24	0.33

Summary statistics calculated over indicated groups for all observations where available.
* t statistic indicates statistically significant difference of means (with a p-value < .05)
+Housing characteristics are for the property where the student lived at the start of the school year, before any within-year move.

Table 1.4: Student descriptive variable summary statistics by default (continued)

	Full Sample			Single Family Owned	
Student Characteristics					
Grade	No Default (N=168,849)	Default (N=2,469)		No Default (N=59,299)	Default (N=933)
Mean	3.47	3.93*	Mean	3.51	3.91*
StDev	1.07	0.97	StDev	1.08	0.96
Min	2	2	Min	2	2
Max	5	5	Max	5	5
School years in data					
Mean	3.58	3.56*	Mean	3.62	3.56*
StDev	0.51	0.51	StDev	0.49	0.51
Min	3	3	Min	3	3
Max	5	5	Max	5	5
English Learner					
Mean	0.33	0.37*	Mean	0.18	0.30*
StDev	0.47	0.48	StDev	0.38	0.46
Ethnicity					
White	No Default	Default		No Default	Default
Mean	0.24	0.11*	Mean	0.39	0.14*
StDev	0.43	0.31	StDev	0.49	0.35
Black					
Mean	0.13	0.16*	Mean	0.09	0.13*
StDev	0.33	0.36	StDev	0.28	0.33
Asian					
Mean	0.18	0.17	Mean	0.23	0.24
StDev	0.38	0.38	StDev	0.42	0.43
Hispanic					
Mean	0.45	0.56*	Mean	0.28	0.48*
StDev	0.50	0.50	StDev	0.45	0.50
Other					
Mean	0.01	0.00	Mean	0.01	0.01
StDev	0.09	0.07	StDev	0.10	0.10

Table 1.4: Student descriptive variable summary statistics by default (continued)

Full Sample			**Single Family Owned**		
		Parent Education			
Less than High School	No Default	Default		No Default	Default
Mean	0.16	0.15	Mean	0.07	0.13*
StDev	0.37	0.35	StDev	0.26	0.34
HS Grad					
Mean	0.18	0.20*	Mean	0.14	0.21*
StDev	0.38	0.40	StDev	0.35	0.40
Some College					
Mean	0.15	0.15	Mean	0.17	0.17
StDev	0.36	0.36	StDev	0.38	0.37
College					
Mean	0.14	0.10*	Mean	0.22	0.12*
StDev	0.35	0.30	StDev	0.42	0.33
Beyond College					
Mean	0.09	0.04*	Mean	0.17	0.05*
StDev	0.29	0.21	StDev	0.37	0.22
Missing/Not reported					
Mean	0.27	0.36*	Mean	0.22	0.32*
StDev	0.44	0.47	StDev	0.48	0.47

Summary statistics calculated over indicated groups for all observations.
* t statistic indicates statistically significant difference of means (with a p-value < .05)

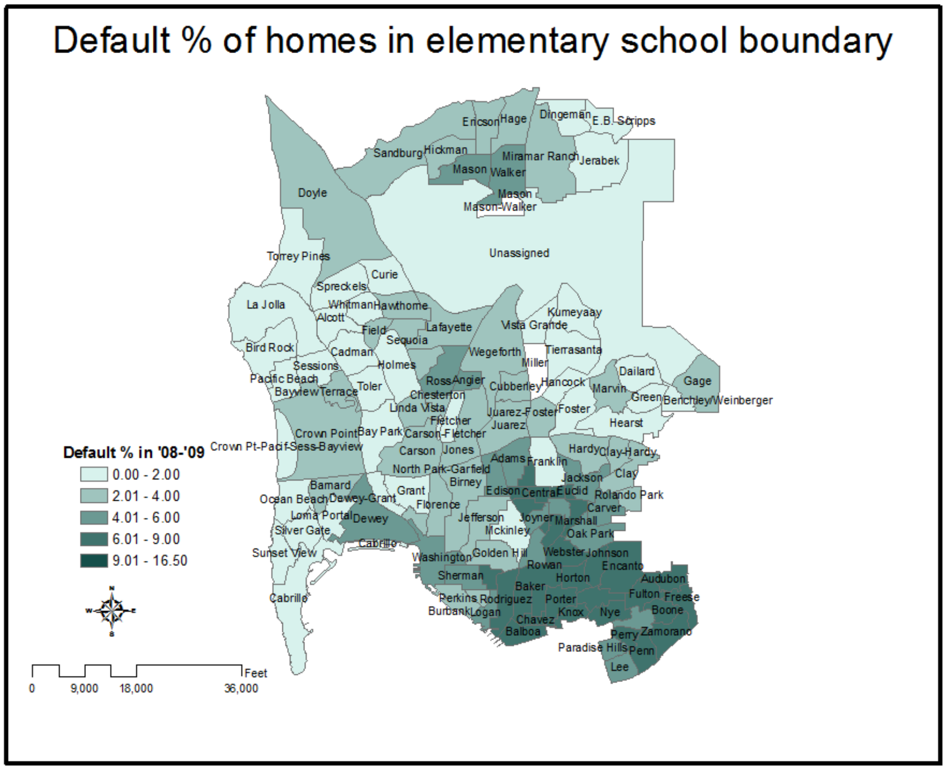

Figure 1.1: Percentage of properties in school boundary experiencing a default during the 2008-2009 school year.

Table 1.5: Neighborhood descriptive variables summary statistics (continued)

	Full Sample			Single Family Owned	
	Block Group Characteristics				
Median Income	No Default	Default		No Default	Default
N	165,071	2,467	N	58,263	933
Mean	46,630	43,627*	Mean	62,618	49,288*
StDev	25,949	19,791	StDev	27,573	19,581
Min	12,500	1,250	Min	0	0
Max	200,001	142,351	Max	200,001	142,351
Sample Default Rate					
N	165,078	2,467	N	58,263	933
Mean	0.018	0.024*	Mean	0.016	0.023*
StDev	0.011	0.010	StDev	0.011	0.010
Min	0	0	Min	0	0.0026
Max	0.11	0.09	Max	0.11	0.06
% Below the Poverty Line					
N	165,071	2,466	N	58,258	932
Mean	19.57	19.15	Mean	10.94	14.52*
StDev	15.71	15.35	StDev	11.51	12.95
Min	0	0	Min	0	0
Max	74.05	65.84	Max	65.84	55.87
1999 % Unemployed					
N	165,073	2,466	N	58,258	932
Mean	4.37	4.72*	Mean	3.40	4.15*
StDev	2.99	2.96	StDev	2.60	2.59
Min	0	0	Min	0	0
Max	34.82	30.04	Max	28.60	23.32
1999 % Armed Forces					
N	165,073	2466	N	58,258	932
Mean	2.54	1.75*	Mean	1.78	1.95*
StDev	6.87	2.65	StDev	2.43	2.55
Min	0	0	Min	0	0
Max	100	36.11	Max	34.26	19.90
% Less than High School					
N	165,073	2,466	N	58,258	932
Mean	28.65	31.90*	Mean	18.53	26.27*
StDev	23.04	21.63	StDev	18.21	19.48
Min	0	0	Min	0	0.61
Max	78.50	78.50	Max	78.50	77.24

Table 1.5: Neighborhood descriptive variable summary statistics (continued)

	Full Sample			Single Family Owned	
% High School Grad	No Default	Default		No Default	Default
N	165,073	2466	N	58258	932
Mean	9.23	9.87*	Mean	8.18	9.87*
StDev	4.24	3.76	StDev	4.21	4.02
Min	0	0	Min	0	0
Max	47.41	47.41	Max	47.41	47.41
% Some College					
N	165073	2466	N	58258	932
Mean	29.45	30.25*	Mean	32.02	32.77*
StDev	11.15	11.12	StDev	9.42	10.33
Min	5.24	5.24	Min	5.24	5.24
Max	77.27	57.48	Max	58.27	57.48
% Advanced degree					
N	165073	2466	N	58258	932
Mean	8.50	5.49*	Mean	12.29	6.45*
StDev	9.54	6.45	StDev	10.92	6.95
Min	0	0	Min	0	0
Max	68.15	44.92	Max	68.15	42.16
% Non Citizen					
N	165073	2466	N	58258	932
Mean	19.69	20.10	Mean	12.89	16.74*
StDev	13.64	12.52	StDev	10.17	10.63
Min	0	0	Min	0	0
Max	52.34	52.34	Max	52.34	49.91
% With Children					
N	165071	2466	N	58258	932
Mean	42.77	44.80*	Mean	38.91	42.65*
StDev	17.31	14.36	StDev	14.51	13.41
Min	0	3.60	Min	0	6.94
Max	100.00	87.50	Max	87.22	80.70
% Single Parent					
N	165071	2466	N	58258	932
Mean	13.29	14.40*	Mean	9.17	11.89*
StDev	9.55	9.06	StDev	7.47	8.41
Min	0	0	Min	0	0
Max	43.49	43.49	Max	43.49	43.49

Table 1.5: Neighborhood descriptive variable summary statistics (continued)

	Full Sample			Single Family Owned	
% White	No Default	Default		No Default	Default
N	165073	2466	N	58258	932
Mean	0.35	0.25*	Mean	0.46	0.29*
StDev	0.30	0.25	StDev	0.31	0.26
Min	0	0	Min	0	0
Max	0.99	0.96	Max	0.99	0.96
% Black					
N	165073	2466	N	58258	932
Mean	0.11	0.15*	Mean	0.10	0.15*
StDev	0.12	0.14	StDev	0.13	0.14
Min	0	0	Min	0	0
Max	0.82	0.82	Max	0.82	0.82
% Hispanic					
N	165073	2466	N	58258	932
Mean	0.34	0.38*	Mean	0.22	0.31*
StDev	0.26	0.26	StDev	0.22	0.24
Min	0	0	Min	0	0
Max	0.98	0.98	Max	0.98	0.97
% Pacific Islander					
N	165073	2466	N	58258	932
Mean	0.006	0.008*	Mean	0.007	0.010*
StDev	0.015	0.017	StDev	0.017	0.020
Min	0	0	Min	0	0
Max	0.196	0.181	Max	0.196	0.181
% Native American					
N	165073	2466	N	58258	932
Mean	0.0039	0.0034*	Mean	0.00303	0.00332
StDev	0.0069	0.0062	StDev	0.00596	0.00601
Min	0	0	Min	0	0
Max	0.1071	0.0514	Max	0.05433	0.05142
% Asian					
N	165073	2466	N	58258	932
Mean	0.15	0.17*	Mean	0.17	0.20*
StDev	0.16	0.17	StDev	0.18	0.19
Min	0	0	Min	0	0
Max	0.82	0.78	Max	0.82	0.78

Summary statistics calculated over indicated groups for all observations.
* t statistic indicates statistically significant difference of means (with a p-value < .05)

Table 1.6: Teacher, class, and school descriptive variable summaries (continued)

	Full Sample			Single Family Owned	
Teacher and class characteristics					
# of Years Taught	No Default	Default		No Default	Default
N	164,772	2,441	N	57,170	923
Mean	13.58	13.56	Mean	15.01	13.78*
StDev	9.80	9.11	StDev	10.06	9.21
Min	1	1	Min	0	1
Max	45	43	Max	45	41
Has Bachelor's Degree					
N	168,849	2,469	N	58,366	933
Mean	92.00	94.53*	Mean	93.60	94.11
StDev	27.14	22.74	StDev	24.47	23.57
Has Advanced Degree					
N	168,849	2,469		58,366	933
Mean	54.71	59.48*	Mean	59.43	61.74
StDev	49.59	26.14	StDev	49.03	48.34
Number of Classmates					
N	168,849	2,469	N	58,366	933
Mean	24.45	26.14*	Mean	24.78	26.31*
StDev	5.77	5.61	StDev	5.88	5.73
Min	15	15	Min	15	15
Max	50	49	Max	50	49
% of Classmates with Default					
N	168,849	2,469	N	58,366	933
Mean	1.44	4.63*	Mean	1.52	4.29*
StDev	3.22	5.42	StDev	3.32	5.19
Min	0	0	Min	0	0
Max	33.33	27.78	Max	33.33	27.78

Table 1.6:Teacher, class, and school descriptive variable summaries (continued)

Full Sample			Single Family Owned		
School Characteristics					
% Reduced price lunch	No Default (N=168,849)	Default (N=2,441)		No Default (N=58,366)	Default (N=933)
Mean	62.81	70.29*	Mean	47.34	64.04*
StDev	30.91	27.29	StDev	29.38	27.07
Min	2.52	2.52	Min	2.52	2.52
Max	100.00	100	Max	100.00	100.00
% English Learners					
Mean	28.21	37.29*	Mean	19.60	32.08*
StDev	24.60	21.97	StDev	18.51	19.42
Min	0.00	0	Min	0.00	0.00
Max	100.00	82.17	Max	100.00	81.57
% White					
Mean	24.46	16.01*	Mean	33.46	18.73*
StDev	23.49	19.34	StDev	25.40	20.45
Min	0.00	0.00	Min	0.00	0.00
Max	100.00	81.10	Max	81.10	81.10
% Black					
Mean	13.49	16.13*	Mean	12.98	16.65*
StDev	11.81	13.54	StDev	12.37	13.78
Min	0.00	0.29	Min	0.00	0.29
Max	98.41	98.41	Max	98.41	96.11
% Hispanic					
Mean	42.29	48.68*	Mean	32.10	42.14*
StDev	26.82	26.40	StDev	22.55	25.05
Min	0.00	0.95	Min	0.00	2.09
Max	100.00	97.88	Max	100.00	96.95
% Asian					
Mean	16.60	17.94*	Mean	20.18	21.11
StDev	15.49	16.64	StDev	17.09	17.98
Min	0	0.00	Min	0.00	0.00
Max	71.02	71.02	Max	71.02	71.02
% Native American					
Mean	0.48	0.418*	Mean	0.552	0.477*
StDev	0.59	0.494	StDev	0.619	0.531
Min	0.00	0.000	Min	0.000	0.000
Max	9.81	4.410	Max	5.410	4.410
% Pacific Islander					
Mean	0.7	0.83*	Mean	0.73	0.89*
StDev	0.7	0.75	StDev	0.70	0.74
Min	0	0.00	Min	0.00	0.00
Max	5.19	4.48	Max	4.52	4.12

Summary statistics calculated over indicated groups for all observations.
* t statistic indicates statistically significant difference of means (with a p-value < .05)

Table 1.7: Outcome variable summary statistics by default; all property types

	Full Sample (N=171,318)	No Default (N=168,849)	Default (N=2,469)	t test of means (Default==No Default)	
Math Score				t stat	p value
Mean	0.084+	0.087	-0.125	10.810	0.000
St dev	0.970	0.971	0.864		
Min	-3.132	-3.132	-2.503		
Max	4.609	4.609	3.248		
Read Score					
Mean	0.081	0.084	-0.122	10.567	0.000
St dev	0.964	0.965	0.854		
Min	-3.199	-3.199	-2.490		
Max	5.449	5.449	3.978		
% of Days Absent					
Mean	3.732	3.724	4.285	-7.598	0.000
St dev	3.638	3.629	4.150		
Min	0.000	0.000	0.000		
Max	33.330	33.330	29.440		

The mean of standardized scores differs from zero because the standardization was done using all students in the district records, while our sample requires that students have scores for three years.

Table 1.8: Outcome variable summary statistics by default; single family owned

	Full Sample (N=59,299)	No Default (N=58,366)	Default (N=933)	t test of means (Default==No Default)	
Math Score				t stat	p value
Mean	0.376	0.383	-0.055	13.476	0.000
St dev	0.988	0.988	0.871		
Min	-2.679	-2.679	-2.503		
Max	4.609	4.609	3.248		
Read Score					
Mean	0.419	0.426	-0.006	13.531	0.000
St dev	0.968	0.968	0.858		
Min	-2.441	-2.441	-2.344		
Max	4.986	4.986	3.340		
% of Days Absent					
Mean	3.346	3.333	4.171	-7.875	0.000
St dev	3.224	3.209	3.946		
Min	0.000	0.000	0.000		
Max	33.330	33.330	29.440		

Table 1.9: Baseline regressions of math score and gains on default type

	(1)	(2)	(3)	(4)	(5)
	Math	GainMath	Math	Math	Math
Default Type					
Single Fam Owned	-0.195***	-0.0409*	-0.201***	-0.155***	-0.155***
	(0.0259)	(0.0224)	(0.0259)	(0.0240)	(0.0240)
Condo Owned	-0.134*	-0.0498	-0.148*	-0.120*	-0.125*
	(0.0775)	(0.0710)	(0.0767)	(0.0712)	(0.0691)
Single Fam Rented	-0.0621**	0.00930	-0.0691***	-0.0334	-0.0357
	(0.0265)	(0.0242)	(0.0265)	(0.0253)	(0.0253)
Condo Rented	0.104	-0.0277	0.0839	0.142**	0.114*
	(0.0707)	(0.0591)	(0.0707)	(0.0667)	(0.0665)
2-4 Family Owned	0.0293	0.00280	0.0306	0.0656	0.0746
	(0.0946)	(0.0870)	(0.0949)	(0.0885)	(0.0889)
2+ Family Rented	-0.0587	-0.0765**	-0.0675	-0.0777*	-0.0806**
	(0.0428)	(0.0350)	(0.0428)	(0.0409)	(0.0408)
Group Fixed Effects					
Teacher FE	No	No	No	Yes	Yes
Tract FE	No	No	No	No	Yes
Tractgroup FE	No	No	Yes	No	No
Controls					
Student Characteristics	Yes	Yes	Yes	Yes	Yes
Property Type	Yes	Yes	Yes	Yes	Yes
Move Type	Yes	Yes	Yes	Yes	Yes
School Change	Yes	Yes	Yes	Yes	Yes
Student Demographics	Yes	Yes	Yes	Yes	Yes
Class Size and Teacher Experience	Yes	Yes	Yes	Yes	Yes
Teacher Education	Yes	Yes	Yes	No	No
School Demographics	Yes	Yes	Yes	Yes	Yes
Census blockgroup demographics, sample default rate	Yes	Yes	Yes	No	No
Constant	Yes	Yes	Yes	Yes	Yes
Observations	171,318	120,807	171,318	171,318	167,545
R-squared	0.271	0.008	0.270	0.157	0.165

Robust standard errors in parentheses

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

OLS regression for student years of SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.10: Baseline regressions of reading score and gains on default type

	(1)	(2)	(3)	(4)	(5)
	Read	GainRead	Read	Read	Read
Default Type					
Single Fam Owned	-0.144***	-0.0211	-0.149***	-0.110***	-0.109***
	(0.0245)	(0.0199)	(0.0245)	(0.0239)	(0.0240)
Condo Owned	-0.0808	-0.00311	-0.0922	-0.0573	-0.0587
	(0.0654)	(0.0537)	(0.0649)	(0.0641)	(0.0626)
Single Fam Rented	-0.0397	0.0157	-0.0460*	-0.0216	-0.0192
	(0.0252)	(0.0218)	(0.0252)	(0.0237)	(0.0237)
Condo Rented	0.0560	-0.00729	0.0443	0.0827	0.0618
	(0.0622)	(0.0510)	(0.0621)	(0.0592)	(0.0589)
2-4 Family Owned	0.0280	0.0280	0.0292	-0.00166	0.00350
	(0.0808)	(0.0716)	(0.0819)	(0.0755)	(0.0762)
2+ Family Rented	-0.0434	-0.0136	-0.0521	-0.0474	-0.0455
	(0.0390)	(0.0309)	(0.0391)	(0.0371)	(0.0369)
Group Fixed Effects					
Teacher FE	No	No	No	Yes	Yes
Tract FE	No	No	No	No	Yes
Tractgroup FE	No	No	Yes	No	No
Controls					
Student Characteristics	Yes	Yes	Yes	Yes	Yes
Property Type	Yes	Yes	Yes	Yes	Yes
Move Type	Yes	Yes	Yes	Yes	Yes
School Change	Yes	Yes	Yes	Yes	Yes
Student Demographics	Yes	Yes	Yes	Yes	Yes
Class Size and Teacher Experience	Yes	Yes	Yes	Yes	Yes
Teacher Education	Yes	Yes	Yes	No	No
School Demographics	Yes	Yes	Yes	Yes	Yes
Census blockgroup demographics, sample default rate	Yes	Yes	Yes	No	No
Constant	Yes	Yes	Yes	Yes	Yes
Observations	171318	120589	171318	171318	167545
R-squared	0.371	0.007	0.369	0.214	0.221

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
OLS regression for student years of SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.11: Baseline regressions of days absent on default type

	(1)	(2)	(3)	(4)	(5)
	Absent	Change Absent	Absent	Absent	Absent
Default Type					
Single Fam Owned	0.758***	0.362***	0.781***	0.746***	0.753***
	(0.129)	(0.112)	(0.129)	(0.129)	(0.130)
Condo Owned	0.586	-0.281	0.641*	0.726*	0.763**
	(0.371)	(0.432)	(0.369)	(0.372)	(0.375)
Single Fam Rented	0.450***	0.148	0.479***	0.358**	0.359**
	(0.142)	(0.143)	(0.142)	(0.147)	(0.147)
Condo Rented	0.466	0.552	0.596*	0.402	0.430
	(0.333)	(0.381)	(0.334)	(0.323)	(0.326)
2-4 Family Owned	-0.0856	0.680	-0.0876	-0.121	-0.139
	(0.510)	(0.437)	(0.509)	(0.498)	(0.497)
2+ Family Rented	-0.00515	-0.00330	0.0445	-0.0103	-0.00997
	(0.240)	(0.254)	(0.240)	(0.243)	(0.241)
Group Fixed Effects					
Teacher FE	No	No	No	Yes	Yes
Tract FE	No	No	No	No	Yes
Tractgroup FE	No	No	Yes	No	No
Controls					
Student Characteristics	Yes	Yes	Yes	Yes	Yes
Property Type	Yes	Yes	Yes	Yes	Yes
Move Type	Yes	Yes	Yes	Yes	Yes
School Change	Yes	Yes	Yes	Yes	Yes
Student Demographics	Yes	Yes	Yes	Yes	Yes
Class Size and Teacher Experience	Yes	Yes	Yes	Yes	Yes
Teacher Education	Yes	Yes	Yes	No	No
School Demographics	Yes	Yes	Yes	Yes	Yes
Census blockgroup demographics, sample default rate	Yes	Yes	Yes	No	No
Constant	Yes	Yes	Yes	Yes	Yes
Observations	171318	150755	171318	171318	167545
R-squared	0.056	0.005	0.054	0.040	0.048

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
OLS regression for student years of SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.12: Math score student fixed effects regressions

	(1)	(6)	(7)
	MathScore	MathScore	MathScore
Default Type			
Single Fam Owned	-0.195***	-0.0479***	-0.0481**
	(0.0259)	(0.0186)	(0.0187)
Condo Owned	-0.134*	-0.0736	-0.0748
	(0.0775)	(0.0590)	(0.0576)
Single Fam Rented	-0.0621**	0.0234	0.0201
	(0.0265)	(0.0206)	(0.0207)
Condo Rented	0.104	0.0261	0.0314
	(0.0707)	(0.0451)	(0.0466)
2-4 Family Owned	0.0293	-0.0256	-0.0182
	(0.0946)	(0.0808)	(0.0813)
2+ Family Rented	-0.0587	-0.0115	-0.0194
	(0.0428)	(0.0293)	(0.0296)
Group Fixed Effects			
Student FE	No	Yes	Yes
Tract FE	No	No	Yes
Controls			
Property Type	Yes	Yes	Yes
Move Type	Yes	Yes	Yes
School Change	Yes	Yes	Yes
Time varying student characteristics	Yes	Yes	Yes
Time invariant student characteristics	Yes	No	No
Class Size and Teacher Characteristics	Yes	Yes	Yes
School Demographics	Yes	Yes	Yes
Census blockgroup demographics, sample default rate	Yes	Yes	No
Constant	Yes	Yes	Yes
Observations	171318	171318	167545
R-squared	0.271	0.005	0.009

Robust standard errors in parentheses

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

OLS, (1), and fixed effects, (2) and (3), linear regression with student level fixed effects for SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.13: Reading score with student fixed effects

	(1)	(6)	(7)
	ReadScore	ReadScore	ReadScore
Default Type			
Single Fam Owned	-0.144***	-0.0136	-0.0144
	(0.0245)	(0.0160)	(0.0160)
Condo Owned	-0.0808	0.000329	-0.000899
	(0.0654)	(0.0481)	(0.0474)
Single Fam Rented	-0.0397	0.0244	0.0233
	(0.0252)	(0.0183)	(0.0185)
Condo Rented	0.0560	0.0278	0.0306
	(0.0622)	(0.0380)	(0.0381)
2-4 Family Owned	0.0280	0.0331	0.0382
	(0.0808)	(0.0638)	(0.0631)
2+ Family Rented	-0.0434	0.0340	0.0326
	(0.0390)	(0.0276)	(0.0274)
Group Fixed Effects			
Student FE	No	Yes	Yes
Tract FE	No	No	Yes
Controls			
Property Type	Yes	Yes	Yes
Move Type	Yes	Yes	Yes
School Change	Yes	Yes	Yes
Time varying student characteristics	Yes	Yes	Yes
Class Size and Teacher Characteristics	Yes	Yes	Yes
School Demographics	Yes	Yes	Yes
Census blockgroup demographics, sample default rate	Yes	Yes	Yes
Constant	Yes	Yes	Yes
Observations	171318	171318	167545
R-squared	0.371	0.007	0.009

Robust standard errors in parentheses

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

OLS, (1), and fixed effects, (2) and (3), linear regression with student level fixed effects for SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.14: Percent absent with student fixed effects

	(1)	(6)	(7)
	Absent	Absent	Absent
Default Type			
Single Fam Owned	0.758***	0.215**	0.233**
--	(0.129)	(0.0986)	(0.0992)
Condo Owned	0.586	0.0381	0.139
	(0.371)	(0.351)	(0.317)
Single Fam Rented	0.450***	0.195	0.166
	(0.142)	(0.127)	(0.126)
Condo Rented	0.466	0.203	0.188
	(0.333)	(0.278)	(0.283)
2-4 Family Owned	-0.0856	-0.126	-0.0245
	(0.510)	(0.389)	(0.387)
2+ Family Rented	-0.00515	-0.309	-0.272
	(0.240)	(0.207)	(0.207)
Group Fixed Effects			
Student FE	No	Yes	Yes
Tract FE	No	No	Yes
Controls			
Property Type	Yes	Yes	Yes
Move Type	Yes	Yes	Yes
School Change	Yes	Yes	Yes
Time varying student characteristics	Yes	Yes	Yes
Class Size and Teacher Characteristics	Yes	Yes	Yes
School Demographics	Yes	Yes	Yes
Census blockgroup demographics, sample default rate	Yes	Yes	Yes
Constant	Yes	Yes	Yes
Observations	171318	171318	167545
R-squared	0.056	0.004	0.009

Robust standard errors in parentheses

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

OLS, (1), and fixed effects, (2) and (3), linear regression with student level fixed effects for SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.15: Peer impacts (Contiued)

VARIABLES	(1) Mathscore	(2) Mathscore	(3) Mathscore	(4) Mathscore
Single family owned default	-0.0481***	-0.0475**	-0.0518***	-0.0512***
	(0.0186)	(0.0186)	(0.0187)	(0.0187)
% of classmates with SF owned default		-0.00183**		-0.00159*
		(0.000875)		(0.000888)
% of blockgroup with SF default default			-0.00565***	-0.00555***
			(0.00122)	(0.00122)
Student FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	171318	171318	164370	164370
R-squared	0.005	0.005	0.005	0.005
Number of student fixed effects	48820	48820	47717	47717

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Student level fixed effects regressions for SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

VARIABLES	(1) Readscore	(2) Readscore	(3) Readscore	(4) Readscore
Single family owned default	-0.0140	-0.0140	-0.020	-0.0201
	(0.0160)	(0.0160)	(0.0161)	(0.0161)
% of classmates with SF owned default		0.000236		0.000480
		(0.000757)		(0.000766)
% of blockgroup with SF default			-0.00620***	-0.00623***
			(0.00105)	(0.00104)
Student FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	171318	171318	164370	164370
R-squared	0.007	0.007	0.005	0.005
Number of student fixed effects	48820	48820	47717	47717

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Student level fixed effects regressions for SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.15: Peer impacts (continued)

VARIABLES	(1) %Absent	(2) %Absent	(3) %Absent	(4) %Absent
Single family owned default	0.213**	0.208**	0.248**	0.267***
	(0.0986)	(0.0986)	(0.0985)	(0.0992)
% of classmates with SF owned default		0.0165***		0.0141***
		(0.00431)		(0.00434)
% of blockgroup with SF default			0.0193***	0.0184***
			(0.00662)	(0.00661)
Student FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	171318	171318	164370	164370
R-squared	0.004	0.004	0.004	0.004
Number of student fixed effects	48820	48820	47717	47717

Robust standard errors in parentheses

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

Fixed effects linear regression with student level fixed effects for SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.16: Tractgroup median income, teacher experience, and grade level interacted with default, math

	(1)	(4)	(10)	(7)
VARIABLES	mathscore	mathscore	mathscore	mathscore
Single family owned default	-0.0481***	-0.0531	-0.0356*	0.0654
	(0.0186)	(0.0507)	(0.0190)	(0.0629)
SF Owned*Tract Group Median Income (10k)		0.00101		
		(0.00933)		
Tract Group Median Income(10k)		0.00201		
		(0.00324)		
SF Owned*Inexperienced Teacher			-0.170**	
			(0.0785)	
Inexperienced Teacher			-0.0335***	
			(0.00572)	
SF Owned * Third Grade				-0.0512
				(0.0733)
SF Owned * Fourth Grade				-0.120*
				(0.0682)
SF Owned * Fifth Grade				-0.179**
				(0.0704)
Third Grade				0.00240
				(0.00357)
Fourth Grade				0.0230***
				(0.00562)
Fifth Grade				-0.000635
				(0.00602)
Student Fixed Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	171318	171318	171318	171318
R-squared	0.005	0.005	0.005	0.006
Number of studid_all	48820	48820	48820	48820
Student FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

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

Fixed effects linear regression with student level fixed effects for SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.17: Tractgroup median income, teacher experience, and grade level interacted with default, reading

VARIABLES	(2) readscore	(5) readscore	(11) readscore	(8) readscore
Single family owned default	-0.0140	0.0149	-0.0200	0.0974*
	(0.0160)	(0.0398)	(0.0167)	(0.0560)
SF Owned*Tract Group Median Income (10k)		-0.00588		
		(0.00765)		
Tract Group Median Income(10k)		0.00110		
		(0.00272)		
SF Owned*Inexperienced Teacher			0.0784	
			(0.0545)	
Inexperienced Teacher			-0.00807*	
			(0.00487)	
SF Owned * Third Grade				-0.0777
				(0.0641)
SF Owned * Fourth Grade				-0.101*
				(0.0610)
SF Owned * Fifth Grade				-0.181***
				(0.0622)
Third Grade				0.00902***
				(0.00306)
Fourth Grade				0.0335***
				(0.00474)
Fifth Grade				0.0218***
				(0.00521)
Student FE	Yes	Yes	Yes	Yes
Conrols	Yes	Yes	Yes	Yes
Observations	171318	171318	171318	171318
R-squared	0.007	0.007	0.006	0.007
Number of studid_all	48820	48820	48820	48820
Student FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

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

Fixed effects linear regression with student level fixed effects for SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.18: Tractgroup median income, teacher experience, and grade level interacted with default, reading

VARIABLES	(3) abspcnt	(6) abspcnt	(12) abspcnt	(9) abspcnt
Single family owned default	0.213**	-0.0621	0.255**	-0.245
	(0.0986)	(0.281)	(0.103)	(0.289)
SF Owned*Tract Group Median Income (10k)		0.0561		
		(0.0487)		
Tract Group Median Income(10k)		-0.0153		
		(0.0154)		
SF Owned*Inexperienced Teacher			-0.547	
			(0.348)	
Inexperienced Teacher			0.111***	
			(0.0301)	
SF Owned * Third Grade				0.198
				(0.334)
SF Owned * Fourth Grade				0.553*
				(0.334)
SF Owned * Fifth Grade				0.676**
				(0.341)
Third Grade				-0.126***
				(0.0169)
Fourth Grade				-0.0453
				(0.0295)
Fifth Grade				-0.0354
				(0.0323)
Student FE	Yes	Yes	Yes	Yes
Conrols	Yes	Yes	Yes	Yes
Observations	171318	171318	171318	171318
R-squared	0.004	0.004	0.004	0.004
Number of studid_all	48820	48820	48820	48820
Student FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

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

Fixed effects linear regression with student level fixed effects for SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.19: Alternate specifications of default impact

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Math	Reading	%Abst	Math	Reading	%Abst	Math	Reading	%Abst
Single family owned default	-0.0481***	-0.0140	0.213**				-0.0639***	-0.0355**	0.267**
	(0.0186)	(0.0160)	(0.0986)				(0.0204)	(0.0175)	(0.113)
Post SF owned default				-0.0590***	-0.0426**	0.246**			
				(0.0205)	(0.0176)	(0.114)			
Single family default in prior year							-0.0153	-0.0698***	0.130
							(0.0295)	(0.0251)	(0.164)
Single fam. default 2 years ago							-0.158***	-0.0761*	0.325
							(0.0460)	(0.0403)	(0.236)
Student FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	171318	171318	171318	171318	171318	171318	171318	171318	171318
R-squared	0.005	0.007	0.004	0.005	0.006	0.004	0.005	0.006	0.004
Number of studid_all	48820	48820	48820	48820	48820	48820	48820	48820	48820
Student FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

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

Fixed effects linear regression with student level fixed effects for SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

Table 1.20: Future default regressed on pre-default sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Math	Reading	%Abst	Math	Reading	%Abst	Math	Reading	%Abst
Any default in subsequent year	0.00360	0.00804	0.0537						
	(0.0161)	(0.0136)	(0.0899)						
SF owned default in subsequent year				-0.0318	-0.00512	0.156			
				(0.0295)	(0.0239)	(0.149)			
SF owned default in two years							-0.00254	0.0217	0.136
							(0.0341)	(0.0307)	(0.184)
Student FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	117859	117859	117859	117859	117859	117859	73379	73379	73379
R-squared	0.006	0.007	0.004	0.006	0.007	0.004	0.007	0.008	0.006
Number of studid_all	48678	48678	48678	48678	48678	48678	48666	48666	48666

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Fixed effects linear regression with student level fixed effects for SDUSD 2nd through 5th graders. Suppressed coefficients are available from the authors on request.

References

Ananat, E.O., A. Gassman-Pines, D.V. Francis, and C.M. Gibson-Davis, "Children Left Behind: The Effects of Statewide Job Loss on Student Achievement," *NBER Working Paper* No. 17104, 2011.

Babcock, P. and J. R. Betts, "Reduced-class distinctions: Effort, ability, and the education production function," *Journal of Urban Economics,* 65(3), 314–322, 2009.

Been, V., S. Chan, I. G. Ellen, and J. Madar., "Decoding the foreclosure crisis: Causes, responses, and consequences," *Journal of Policy Analysis and Management,* 30(2), 388-396, 2011.

Been, V., I. G. Ellen, A. E. Schwartz, L. Stiefel, and M. Weinstein., "Does losing your home mean losing your school?: Effects of foreclosures on the school mobility of children," *Regional Science and Urban Economics,* 41(4), 407-414, 2011.

Bhutta, J. D., J. Dokko, and H. Shan, "Consumer Ruthlessness and Strategic Default During the 2007-2009 Housing Bust," *SSRN eLibrary,* 2011.

Brevoort, K. P. and C. R. Cooper, "Foreclosure's Wake: The Credit Experiences of Individuals Following Foreclosure," *SSRN eLibrary* 2010.

The California State University Real Estate & Land Use Institute, "A Homeowner's guide to foreclosure in California," California Department of Real Estate, 2010.

Carrell, S. E. and M. L. Hoekstra, "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids," *American Economic Journal: Applied Economics,* 2(1), 211-228, 2010.

Clotfelter, C. T., H. Ladd, and J. L. Vigdor, "How and why do teacher credentials matter for student achievement?", NBER Working Paper No. 12828, 2007.

Cui, L. "Foreclosure, Vacancy and Crime," *SSRN eLibrary,* 2010.

Ellen, I. G., J. Lacoe, and C. Sharygin, "Do Foreclosures Cause Crime?" Furman Center for Real Estate and Urban Policy Working Paper, 2011.

Ellen, I. G., J. Napier Tye, and M. A. Willis, "Improving U.S. Housing Finance through Reform of Fannie Mae and Freddie Mac: Assessing the Options" NYU

Furman Center for Real Estate and Urban Policy and What Works Collaborative, 2010.

Gerardi, K., S. L. Ross, and P. Willen, "Understanding the foreclosure crisis," *Journal of Policy Analysis and Management,* 30(2), 382-388, 2011.

Guiso, L., P. Sapienza, L. Zingales, V.D. Macelli, "Moral and Social Constraints to Strategic Default on Mortgages," NBER Working Paper No. 15145, 2009.

Harding, J. P., E. Rosenblatt, and V. W. Yao, "The contagion effect of foreclosed properties," *Journal of Urban Economics,* 66(3), 164-178, 2009.

Haveman, R and B. Wolfe, "The Determinants of Children's Attainments: A Review of Methods and Findings," *Journal of Economic Literature,* 33(4), 1829-1878, 1995.

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig, "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans*," *Quarterly Journal of Economics,* 125(1), 307-362, 2010.

Koedel, C. and J. R. Betts, "Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique," *Education Finance and Policy,* 6(1), 18-24, 2011.

Li, W., M. J. White, and N. Zhu, "Did Bankruptcy Reform Cause Mortgage Default to Rise?" NBER Working Paper No. 15968, 2010.

Marron, D.B. and P. Swagel, "Whither Fannie and Freddie? A Proposal For Reforming The Housing GSEs," *Economic Policies for the 21st Century,* economics21.org, 2010.

Mayer, C. J. and K. Pence, "Subprime Mortgages: What, Where, and to Whom?" NBER Working Paper w14083, 2008.

Mayer, C. J., K. Pence, and S. M. Sherlund, "The rise in mortgage defaults," *Journal of Economic Perspectives,* 23(1), 27-50, 2009.

Mian, A. and A. Sufi, "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis*," *Quarterly Journal of Economics*, 124(4), 1449-1496, 2009.

Rege, M., K. Telle, and M. Votruba, "Parental Job Loss and Children's School Performance," *The Review of Economic Studies,* forthcoming, 2011.

Schuetz, J., V. Been, and I. G. Ellen, "Neighborhood effects of concentrated mortgage foreclosures," *Journal of Housing Economics.* 17(4), 306-319, 2008.

Schweitzer, M. E. and S. Shane, "The effect of falling home prices on small business borrowing," *Economic Commentary*, 2010-18, 2010.

Stevens, A. H, and J. Schaller, "Short-run effects of parental job loss on children's academic achievement," *Economics of Education Review,* 30(2), 289-299, 2011.

Chapter 2

Understanding the Solar Home Price Premium: Electricity Generation and "Green" Social Status

2.1 Introduction

On a per-capita basis, California has the most installed residential solar capacity in the United States. Solar homes are expensive. It can cost $30,000 to install such a system. Several state and federal programs actively subsidize this investment. Judged on strictly efficiency criteria (foregone electricity expenditure per dollar of investment), solar panels may be a bad investment. Borenstein (2008) finds that the cost of a solar photovoltaic system is about 80 percent greater than the value of the electricity it will produce.

Solar panels bundle both investment opportunities (the net present value of the flow of electricity they generate) and conspicuous consumption opportunities (that it is common knowledge that your home is "green"). Kotchen (2006) provides a theoretical analysis of the case in which individuals have the option of consuming "impure" public goods that generate private and public goods as a joint product. Outside of the Toyota Prius, solar homes are perhaps the best known "green products" sold on the market.

The owner of a solar home faces low electricity bills and, if an environmentalist, enjoys the "warm glow" for "doing his duty" and producing minimal greenhouse gases (Andreoni 1990). Because the presence of solar panels on most roofs is readily apparent, the solar home owner knows that others in the same community know that the home owner has solar panels. This community level re-enforcement may further increase the demand for this green product. This "observability" is likely to be even more valued in an environmentalist community (i.e a Berkeley) than in a community that dismisses climate change concerns. The recent political divide between Democrats and Republicans over climate change mitigation efforts (see Cragg, Zhou, Gurney and Kahn 2011) highlights that in conservative communities solar panels may offer less "warm glow" utility to its owners.

We examine two facets of solar purchases in this paper. Our primary empirical contribution is to provide new hedonic marginal valuation estimates for a large sample of solar homes based on recent real estate transactions in San Diego County. We test the robustness of our results using data from Sacramento County. We document evidence of a solar price premium and find that this premium is larger in environmentalist communities. In most mature housing markets, we expect that the econometrician knows less about the market than the decision makers. In the case of solar panels, our interactions with professionals in the field suggests that these professionals have little basis for estimating the pecuniary benefits of solar installation. Our second empirical contribution is to document what types of people, in terms of education, political ideology and demographic attributes do and do not live in solar homes. Most hedonic studies that use sales data (rather than Census data) know very little about the household who actually lives in the home, but we can observe household characteristics for a single year.

Our hedonic study contributes to two literatures. The real estate hedonics literature explores how different housing attributes are capitalized into home prices. Solar installation can be thought of as a quality improvement in the home. Recent studies have used longitudinal data sets such as the American Housing Survey (which tracks the same homes over time) to study how home upgrades such as new bathrooms and other home improvements are capitalized into resale values (Harding, Rosenthal and Sirmans 2007, Wilhelmsson 2008). A distinctive feature of solar panels is that on a day to day basis they have no “use value” as compared to a new bathroom or kitchen. Solar panels reduce your household’s need to purchase electricity but from an investment standpoint they represent an intermediate good that indirectly provides utility to households. For those households who derive pleasure from knowing that they are generating their own electricity, the solar panels will yield “existence value”. Such households will recognize that they have reduced their greenhouse gas emissions and thus are providing world public goods. In their local communities, such households may be recognized by neighbors for their civic virtue. Households who take pride in engaging in “voluntary restraint” will especially value this investment (Kotchen and Moore 2008).

A recent literature in environmental economics has examined the demand for green products. Most of these studies have focused on hybrid vehicle demand such as

Kahn (2007), Kahn and Vaughn (2009) and Heutel and Muehlegger (2010) or the diffusion of solar panels across communities (Dastrup 2010 and Bollinger and Gillingham 2010). By using hedonic methods to estimate the price premium for green attributes our study shares a common research design with several recent studies that have used hedonic methods to infer the "green product" price premium such as Delmas and Grant's (2010) study the demand for organic wine, Eichholtz, Kok, and Quigley's (2010) work on the capitalization of Energy Star and LEED status for commercial buildings, and Brounen and Kok's (2010) investigation of the capitalization of residential energy efficiency when Dutch homes are certified with regards to this criterion.

2.2 The hedonic pricing equilibrium and the make versus buy decision over solar installation

A household who wants to live in a solar home can either buy such a home or buy another home that does not have solar panels and pay a contractor to install these solar panels. This option to "make" versus "buy" should impose cross-restrictions on the size of the capitalization effect. Consider an extreme case in which all homes are identical and there is a constant cost of $c to install solar panels. By a no arbitrage argument, in the hedonic equilibrium, we would recover a price premium of "c" for the solar homes. Over time, any supply innovations that lead to a lower installation cost or higher quality of the new solar panels would be immediately reflected in the hedonic price premium.

In reality, homes are differentiated products that differ along many dimensions. No home has a "twin". The non-linear hedonic pricing gradient is such that different homes are close substitutes at the margin (Rosen 2002). Since at any point in time the same home is not available with and without solar panels, there is no reason why the hedonic solar capitalization must equal the installation cost.

We recognize that the investment decision in solar has an option value component. Households may be uncertain about how much electricity the solar panels will generate, the future price of electricity and future price declines in quality adjusted solar systems. In a standard investment under uncertainty problem, it can be rational to delay and not exercise the option. Households may also be uncertain about what the resale value of their house would be if they install solar. All of these factors, as well as the household's power needs and its ideology, will influence demand for solar panels.

On the supply side, there are two sources of solar homes. There are existing homes whose owners have installed solar panels in the past and are now selling their home. In contrast, the second set of solar homes is produced by developers of new homes who will compare their profit for building a home with and without solar panels. Such developers are likely to have invested more effort in the basic marketing research of determining the market for this custom feature.

2.3 Empirical specification

We employ both a hedonic and a repeat sales approach to assess the extent to which solar panels are capitalized into home prices. The hedonic specification decomposes home prices by observable characteristics for all transactions while flexibly controlling for spatial and temporal trends. Solar panels are included as a home characteristic and average capitalization is measured as the coefficient on the solar panel variable. The repeat sales model controls for average appreciation of properties from one sale to the next within each census tract, with an indicator for installation of panels between sales.

Hedonic approach

Our first approach to measuring the capitalization of solar panels in home sales is to decompose home prices by home characteristics and neighborhood level time trends. We interpret the average difference between the log price of homes with solar panels and those without after controlling for observable home characteristics and average neighborhood prices in each quarter as the average percent contribution to home sales price of solar panels. The baseline equation we estimate in our hedonic specification is

$$\log\left(\mathrm{Price}_{ijt}\right) = \alpha \mathrm{Solar}_{it} + \beta \mathrm{X}_i + \gamma_{jt} + \varepsilon_{ijt} \qquad (1)$$

where Price_{ijt} is the observed sales price of home i in census tract j in quarter t. The variable Solar_{it} is an indicator for the existence of a solar panel on the property and α is the implicit price of the panels as a percentage of the sales price -- our measure of the extent of capitalization. Home, lot, and sale characteristics are included as X_i.

We allow for the differential capitalization across geographic areas of home and lot size by interacting the logs of these observable characteristics with zip code level indicator variables.[27] Additional characteristics contained in X_i are the number of bathrooms, the number of times the property has sold in our sales data, the number of mortgage defaults associated with the property since 1999, indicators for the building year, if the property has a pool, a view, and is owner occupied, and month of the year indicators to control for seasonality in home prices. In equation (1), we are imposing a constant solar capitalization rate across time and space.[28]

We control for housing market price trends and unobserved neighborhood and location amenities with census tract-quarter fixed effects, γ_{jt}. Allowing different appreciation patterns for different geographies is critical because these different geographical appreciation patterns are correlated with the incidence of solar panel installation.

Any hedonic study is subject to the criticism that key explanatory variables are endogenous. While we have access to a detailed residential data set providing numerous controls, we acknowledge that there are plausible reasons why the solar panel dummy could be correlated with unobserved attributes of the home.

---

[27] There is substantial variation in climate and other local amenities across the three counties in our data sets. Our specification allows a home or lot of a given size on the temperate coast near the beach to be valued by the market differently than the same size home or lot in the inland desert region.

[28] Recent changes in the federal tax incentives for solar may affect the solar price capitalization. On October 3, 2008 the President signed the Emergency Economic Stabilization Act of 2008 into law. The bill extends the 30% ITC for residential solar property for eight years through December 31, 2016. It also removes the cap on qualified solar electric property expenditures (formerly $2,000), effective for property placed in service after December 31, 2008 http://www.clarysolar.com/residential-solar.html. We do not have enough observations to determine whether the law has affected the size of the solar capitalization effect.

Our OLS capitalization estimate of α measures the average differential in sales price of homes with solar panels and homes without panels in the same census tract selling in the same quarter after controlling for differences in observable home characteristics. Interpreting the hedonic coefficient estimate as the effect on home price of solar panels requires assuming that the residual idiosyncratic variation in sales prices (ε_{ijt} in our framework), solar panel installation and unobservable house attributes are uncorrelated. This assumption is invalid if homeowners who install solar panels are more likely to make other home improvements that increase sales prices of their homes than their neighbors who do not install. We investigate how this might influence our capitalization estimate by estimating (1) with a control for whether a home improvement is observed in building permit data available for a large subset of San Diego County. Alternatively, homes with solar panels may be homes of higher unobserved quality. We explore whether these homes command a time-invariant premium by including an indicator for whether a home will have panels installed at some point in the future relative to a particular sale.

We allow the capitalization of panels to vary over system size and neighborhood characteristics by interacting our solar indicator variable in equation (1) with a linear term including the characteristic. Our estimating equation becomes:

$$\log(\mathbf{Price}_{ijt}) = \alpha_0 \mathbf{Solar}_{it} + \alpha_1 \mathbf{N} * \mathbf{Solar}_{it} + \beta \mathbf{X}_i + \gamma_{jt} + \varepsilon_{ijt} \quad (2)$$

.

The value of installed solar panels may be influenced by factors beside the financial implications of installation, and we estimate equation (2) using a number of proxies for other factors. Households may have preferences for the production technology used to generate the electricity they use if they are concerned about their individual environmental impact or value their own energy independence. A desire to appear environmentally conscious may increase the value of solar, because it is a visible signal of environmental virtue. Our proxies for environmental idealism and the social return to demonstrating environmental awareness are the percent of voters registered as Green party members in the census tract and the Toyota Prius share of registered vehicles in the zip code. For comparison, we estimate capitalization variation by Democratic party registered voter share and the pickup truck share of registered vehicles in the zip code. We also examine solar panel capitalization by census tract log median income and percent of college graduates.

Repeat sales approach

A second approach to measuring the average additional value to a home sale of solar panels is to average the additional appreciation of a single home from one sale to the next (repeat sales) when solar panels are installed between sales. We interpret the average differential in the appreciation in consecutive sales of properties where solar was installed between sales and other properties in the same census tract with no installation between consecutive sales as the average capitalization of solar panels in home sales. The baseline equation we estimate for our repeat sales specification is

$$\log\left(\frac{\text{Price}_{ij(t+\tau)}}{\text{Price}_{ijt}}\right) = \tilde{\alpha}\Delta\text{Solar}_{i(t+\tau)} + T_{j(t+\tau)} + \tilde{\varepsilon}_{ij(t+\tau)} \qquad (3)$$

where $\mathbf{Price}_{ij(t+\tau)}$ and $\mathbf{Price}_{ijt}$ are consecutive sales of the same property i in neighborhood j occurring τ quarters apart where the first sale is in period t. The variable $\Delta\text{Solar}_{i(t+\tau)}$ is an indicator for the installation of solar panels at a property between sales (after t but before $t + \tau$). Census tract specific time effects are included as the vector $T_{j(t+\tau)}$, with remaining idiosyncratic property appreciation measured as $\tilde{\varepsilon}_{ij(t+\tau)}$.

Our repeat sales GLS capitalization estimate, $\tilde{\alpha}$, of the capitalization of solar panels in housing prices measures the average additional appreciation of homes with solar installed between sales beyond that measured by the housing price indexes of their respective census tracts. Interpreting $\tilde{\alpha}$ as the effect of panel installation on subsequent sales price requires the assumption that idiosyncratic price appreciation of homes is not correlated with solar panel installation. Again, this will not be the case if unobserved changes in properties are correlated with solar panel installation.[29]

2.4 San Diego County data

[29] Our hedonic and repeat sales approaches are related. Since differencing consecutive observations on the same property i in equation (1) results in equation (3), both methods estimate the same parameter for the average capitalization of solar panels, $\alpha = \tilde{\alpha}$. An advantage of the repeat sales approach is that this differencing controls for unobservable time-invariant housing characteristics, in addition to the observable X_i, that may be correlated with solar installations. The census tract-quarter time effects, $T_{j(t+\tau)} = \gamma_{i(t+\tau)} - \gamma_{it}$, are jointly estimated as quarterly repeat sales price indexes for each census tract using standard GLS procedures to account for the dependence of the idiosyncratic error $\tilde{\varepsilon}_{ij(t+\tau)}$ on τ, the number of quarters between sales.

Our hedonic analysis utilizes single family home sales records occurring between January 1997 and early December 2010 in San Diego County. For our sample of repeat sales of single family homes in which solar was installed between sales we use first sales beginning as early as January of 1990. When we restrict our analysis to homes for which we know the home square footage, the number of bedrooms and bathrooms, the year the house was built or most recently underwent a major remodeling, whether the property has a pool, whether the property has a view, and if the property is subject to a lower tax because it is owner occupied, we obtain 364,992 sales records for the hedonic analysis and 80,182 records for the repeat sales analysis.[30] The Data Appendix provides details on the variables.

We control for the home observable characteristics mentioned above as well as lot size, the number of times the property has transacted in our dataset and the number of public mortgage default notices associated with the property. We view the latter as proxies for idiosyncratic home quality. We also control for neighborhood characteristics. We use the percent of voters in each census tract who are Green Party registrants as a measure of the level of environmentalism in the neighborhood. We use the Toyota Prius share of registered automobiles from zip code totals of year 2007 automobile registration data as a proxy of the neighborhood prevalence of both the level of environmentalism and of displayed environmentalism.[31] We use the percent

[30] The building year is not recorded for 1,681 properties, 46 of which are matched to solar panel installations.
[31] See Kahn (2007) for a discussion on the Green Party and party membership as an identifier of environmentalists.

registered Democrats and vehicles classified as trucks from the respective summary datasets as comparison measures. We control for year 2000 census tract median income and average census tract education levels as percent of the over age 25 population who are college graduates. We also control for census tract specific time effects.

We know which homes have solar panels from administrative records from four incentive programs which have subsidized residential solar panel systems in San Diego County (details about these programs are given in the Data Appendix). These programs cover virtually all solar installations in San Diego County, as we have confirmed with conversations from industry experts.

The solar systems consist of solar panels installed on the property, typically on the roof, which are connected to the electricity grid, meaning the home draws electricity both from the panels and from standard utility lines and the panels supply electricity to the local infrastructure when production exceeds consumption at a given home. We use a dataset of the administrative records from these programs to determine the presence of solar panels on a property being sold as well as the installation of panels between sales.[32]

We know, for each installation, the address of the property, size of the system in terms of kilowatt production potential, and date completed. Most installations also include information on the cost of the system and the amount subsidized by the respective program. We successfully match installation records to 6,249 single family

[32] Federal tax credits allow homeowners to recover 30% of the costs of a system, but we do not have access to tax return data as an additional source of installation detail.

homes by address to public San Diego County Assessor property records for installations through early December 2010.[33]

We assign each home in our sample to one of four mutually exclusive and exhaustive categories. At the time the home was sold, the home can 1) already have solar panels installed (329 observations); 2) concurrently have installed solar panels (73 observations); 3) have solar panels installed in the future but be sold without solar panels at the time of the specific sale (3,433 observations); and, 4) not have solar panels as of Winter 2010. In the regressions, this fourth category will be the omitted category.[34] We use the date of installation of each system to determine how many homes in the same census block had solar panels installed for each month of our sample.

We use building permit data to examine whether homeowners who install solar panels also make other improvements to their homes more often than their neighborhoods, thus potentially biasing our estimate of the home price premium for solar panels. Our building permit reports begin in 2003 for San Diego City, the largest permit issuing jurisdiction in San Diego County, and for Escondido, a smaller municipality in our sample area. We define a “major renovation” as one referencing a

---

[33] We match nearly 90% of installation records, and have verified that many unmatched records are business or multifamily addresses. Match quality was verified by inspecting publicly available aerial photographs (www.bing.com/maps) of the installation addresses for the existence of solar panels for a subset of the records.

[34] An additional 50 transactions with an existing solar systems occurred within the year following a public mortgage default notice or sometimes attendant notice of trustee's sale. These are excluded from the analysis here. Including them, along with an indicator for a sale following default for all observations does substantively alter our results.

kitchen, bath, HVAC, or roof with an associated value greater than $1,000 and a "high value" renovations as one with an associated value greater than $10,000.

Summary statistics for San Diego

Table 1 shows that compared to homes sold without solar, those sold with solar are bigger, have more bedrooms and bathrooms, and are more likely to have a view and a pool, among various other characteristics. We thus need to control for observable home characteristics as well as census tract location in our empirical specification so that our regressions are comparing sales prices of homes with solar panels to sales of similar homes in the same census tract.

Neighborhoods where solar panels have been installed are richer, whiter, more educated, have more registered Democrats, and have larger homes than the 103 of 478 census tracts where no solar was installed during period covered by our data (see Table 2). Our empirical analysis exploits the gradation in these differences across neighborhoods to examine how capitalization in home price varies with ideological and demographic characteristics.

2.5 Who lives in solar homes?

Most hedonic real estate studies have detailed information about the home, its sales price, location and physical attributes but they know little about the marginal buyer who chose to pay the sales price to live there. For the city of San Diego in 2009, we have information for registered voters on their age, education, political party of registration, and contributions to environmental, political, and religious organizations.[35] These data enable us to investigate what types of people self select into solar homes.

We estimate linear probability models using the full stock of City of San Diego homes in the year 2009. We regress a dummy variable indicating whether the home has solar panels on various household characteristics, including the number of voters in conservative (Republican, American, and Libertarian) and liberal parties (Democrat, Peace and Freedom, and Green), whether the two oldest registered voters in the household contribute to environmental, political, and religious organizations, the highest education level of the two oldest registered voters, the age of the oldest registered voter in the household, whether a child is present, the highest imputed income (based on census block data and the age of the household) of the two oldest registered voters in the household, and census tract fixed effects.

We find that households in which everyone is a registered liberal and in which the household contributes to environmental organizations are much more likely to be in solar homes controlling for education, imputed income, the age of the oldest registered household member, and whether any children are present in the household (see Table 3). When everyone in the household is a registered liberal (and also controlling for contributions to organizations) the probability of being in a solar home increases by 0.002, an 18 percent increase from the base of 0.011. When the household contributes to environmental organizations (and controlling for party

[35] Our data are from www.aristotle.com. We merged by street address to each home. We were able to match 90% of the sample.

registration) the probability of being in a solar home increases by 0.006, a 55 percent increase.

Education, age, and income were also predictors of living in a solar home. Those with a college education have a 0.003 greater probability of living in a solar home than those with less than a high school education and those with a graduate degree have a 0.006 greater probability of living in a solar home. This represents roughly a 27-55% increase in the probability of living in a solar home. Households living in a solar home are also most likely to be those where the oldest voter was born after 1950 (relative to being born before 1950) and households with imputed income above the 70th percentile compared to households with imputed income between the 50th and 60th percentile (results not shown).

We have shown that environmentalists, the college-educated, baby-boomers and later generations, and richer households paid the hedonic premium to live in solar homes. We next estimate the size of these hedonic premia.

2.6 Estimation results

Tables 1 and 2 showed that large nice homes in rich white neighborhoods are more likely to have solar than small homes in poor minority neighborhoods. Our estimated solar coefficient is the average premium for a large nice home with solar (in a rich white neighborhood) relative to the other homes *in the same neighborhood* after flexibly controlling for observable differences between the two homes. Because the hedonic regressions based on equation (2) contain census tract by quarter fixed effects, the coefficient picks up the price premium for a home with solar relative to homes in the same tract. Similarly, our repeat sales approach measures the average additional increase in price between sales for homes with solar installed between sales relative to other homes in the neighborhood because we are fitting census tract specific repeat sales indexes.

Hedonic estimates

All of our hedonic specifications estimate the capitalization of solar panels in observed property sales while controlling for housing characteristics, and census tract/quarter fixed effects. We find that solar panels add 3.6% to the sales price of a home after controlling for observable characteristics and flexible neighborhood price trends (see Table 4). This corresponds to a predicted $22,554 increase in price for the average sale with solar panels installed.[36] Homes which do not yet have solar installed but will at some subsequent time in our sample have no associated premium, indicating that our measured solar effect is not attributable to unobserved, time-invariant differences in these homes. Homes in which the solar installation was done "concurrently" receive a statistically insignificant capitalization rate of 2.8 percent, probably because they are a combination of two types of installations. If the installation was done before the sale (for example, for new developments or contract remodels) then the price will be capitalized in the sales price. If the installation was done after the sale, the home owner probably added the panels. Unfortunately, we

[36] We convert the coefficient estimate to a dollar amount by differencing the predicted sales price from our estimated model with our solar indicator equal to one and zero and all other characteristics equal to the mean values of all other homes with solar.

cannot distinguish between these two cases because we do not have the precise date of installation.

We estimate the solar premium to be 1% higher if other homes in the same census block have previously installed panels, but the coefficient is not statistically different from zero. We observe a decreasing return to additional system size, a positive relationship between the capitalization rate and Prius penetration, Green party registration share, Democrat registration share, median income, and education, as well as a negative relationship between capitalization and truck ownership. Controlling for building permit activity in a subsample of our data suggests that the solar panel addition rather than unobserved home improvements are responsible for the measured price premium.

The returns to solar investment based on the San Diego estimates

Table 5 compares this predicted increase in price of $22,554 to four different measures of costs of solar panels. The first potential comparison is the average total cost of the systems, which is $35,967.[37] However, this amount does not include subsidies which lowered the effective price to homeowners to about $20,892. Although we do not know the value to the homeowners of federal tax credits for each installation, this comparison suggests that, on average, homeowners fully recover their costs of installing solar panels upon sale of the property. Another measure of the value of panels is the average cost of adding panels during the quarter in which the home was sold. We calculate this value for each quarter in our data, and for our sales the

[37] All dollar amounts are adjusted to 2010 dollars using the "All items less shelter" consumer price index from the Bureau of Labor Statistics.

average of this replacement cost measure is \$30,858 before and \$21,047 after subsidies. Buyers purchasing homes with pre-installed solar panels are paying less than the cost of a new system. However, the 30% tax credit lowers this replacement cost measure net measure to \$14,733, below our estimated capitalization value.

We use our hedonic estimates of equation (3) to test for heterogeneous impacts of solar installation across communities and structure attributes. First we include the log of the size in watts (maximum production capacity) of the solar system, $\mathrm{N} = \log(Watts_{it})$ as a measure of the expected energy production from the system. Although a larger system by definition produces more electricity, because of the structure of electricity rates and the valuation of electricity produced under California's "net metering" system, we do not expect capitalization to increase proportionately with system size. For excess generation, households may opt in to the net metering system that compensates them for electricity returned to the grid at (currently) between \$0.171 and \$0.275/kWh depending on the time of day, but the compensation is capped at the total of their annual electric bill and households face typically higher time of use prices for any electricity purchased from the utility.[38] The combined effect of the rate structure and net metering is that electricity produced by residential solar panels in excess of their annual electricity consumption is essentially

[38] Consumer electricity prices in San Diego County are tiered by monthly consumption, with each household allocated a geography specific baseline amount of electricity (from 9.6 kWh along the coast to 16.4 kWh per month in the inland desert during the summer) at a relatively low price (currently \$0.039/kWh during the summer months) with an up to five fold increases for above baseline consumption (the top of four tiers is \$0.197/kWh during the summer for all consumption over 200% of the baseline). Households pay for electricity use in excess of what is produced by the panels at any given point in time.

donated to the utility. While households may value larger systems for other reasons, additional financial incentives to installing capacity decrease with system size.[39]

Allowing capitalization to vary by neighborhood characteristics demonstrates that the addition to a home's market value from solar panels varies across neighborhoods by environmental ideology, income, and education levels. The estimated coefficients on the linear solar term are jointly statistically significant in each neighborhood variable specification, as listed in Table 6. In each case, the capitalization of solar panels follows a pattern that would be predicted by the measure of environmental ideology, income, or education. Neighborhoods with relatively high Prius concentrations, Green party and Democrat registrant share, and median income capitalize solar panels at a higher value, while in neighborhoods with a large share of trucks, panels provide less of a premium to home sales.

Our final hedonic specification suggests that our estimates are not driven by unobserved home upgrades besides solar panel installation (see Table 8). Our capitalization estimate of 6.2% in the smaller subsample of San Diego City and Escondido is robust to the inclusion of our building permit measures. Our estimates suggest that remodeling a kitchen or bath or replacing a roof or HVAC system has a small impact on price, while high value renovations with costs similar to solar panels are estimated to have a similar value on home prices.

---

[39] Because of these institutional factors, estimated or actual household specific expected electricity demand is necessary for a complete accounting of the financial benefit of installing a system as a function of system size, and is beyond the scope of this paper.

Repeat sales estimates

The results of our hedonic specification are largely replicated in our repeat sales approach. All of the presented results are based on three stage GLS estimates, with observations in the final stage weighted based on time between sales, and controlling for jointly estimated census tract level repeat sales indexes.[40] Our average capitalization estimate of 3.6% (see Table 8) implies that installing solar panels leads to an increase of $20,194 from the first to the second sale when the average price of the first sale is $558,100. Households who install panels thus recuperate more than their costs in subsequent sales even though our estimated value remains below our "replacement cost" measure of solar value. Our estimate of the contribution of system size to the capitalization rate suggests an anomalous large negative relationship. Neighborhood characteristics estimates in the repeat sales framework also indicate that the capitalization of solar panels depends on local preferences and incomes (results not shown).

2.7 Capitalization of soalr homes: evidence from Sacramento County

We examine the robustness of our capitalization estimates using data on 90,686 single family home transactions in Sacramento County between January 2003 and November 2010. We believe that this is a 100% sample of all homes transacted in this period in the county. For each of these homes, we observe its sales date and sales

[40] OLS estimates of solar capitalization that do not correct for time between sales do not vary greatly from our GLS estimates.

price and its physical attributes. We are also able to identify every single family home in Sacramento County that has solar panels as of November 2010 and that was sold at least once between January 2003 and November 2010. For each of these 620 homes, we know the solar system's installation date. Using the information on the installation date and the sales date, we are able to partition these homes into four mutually exclusive and exhaustive categories. A home can either not have solar panels, or it can have solar panels already installed at the time of the sale (true for 256 observations), concurrently have installed solar panels (52 observations), or in the future this same home will have solar panels installed but it does not have solar panels at the time of the specific sale (312 observations).[41] We also define a "solar" street as a street where at least two homes adjacent to each other have solar panels. These streets are more likely to be new developments and solar installation is cheaper when done on all homes in a new development.

We find that the premium for solar homes in Sacramento is 4 percent (see Table 9), similar to the premium for solar homes in San Diego (see Table 4). We find an even larger capitalization of 7 percent for a solar home in Sacramento that is not on a solar street and a smaller one of 3 percent when it is on a solar street.

2.8 Conclusion

This study used a large sample of homes in the San Diego area to provide some of the first capitalization estimates of the resale value of homes with solar panels

[41] For the "concurrent" set of homes, we do not know if the home had solar panels when it was sold. Either the new home buyer installed solar panels after purchase or the developer installed solar panels.

relative to comparable homes without solar panels. Although the residential solar home market continues to grow, there is little direct evidence on the market capitalization effect. Using both hedonics and a repeat sales index approach we find that solar panels are capitalized at roughly a 3% to 4% premium. This premium is larger in communities with more registered Prius hybrid vehicles and in communities featuring a larger share of college graduates.

Our new marginal valuation estimates inform the debate led by Borenstein (2008) on whether expenditure on residential solar is a “good investment.” His analysis, consistent with those taken by others in the literature, treats residential solar installations as a ‘pure’ investment good judged in terms of upfront cost and power generation. Our evidence suggests that similar to other home investments such as a new kitchen, solar installation bundles both investment value and consumption value. Some households may take pride in knowing that they are producers of “green” electricity and “warm glow” may triumph over present discounted value calculations in determining a household’s install choice.

I thank Josh Graff Zivin, Matthew Kahn, and Dora Costa, coauthors of the research presented in this chapter. It is with his permission that I include our research in this dissertation.

Table 2.1: San Diego summary statistics and mean comparisons for solar and no solar home sales

	Sales with no solar	Sales with solar	No solar - solar
Variable	Mean *Std Dev*	Mean *Std Dev*	Difference in means *Pr(|T|>|t|)*
Sale price (2000 $s)	427,047 *380,536*	667,645 *426,980*	-240,599 *0.000*
Square feet	1,984 *961*	2,512 *1,124*	-528 *0.000*
Bedrooms	3.39 *0.89*	3.76 *0.86*	-0.37 *0.000*
Baths	2.37 *0.88*	2.86 *1.00*	-0.48 *0.000*
View	0.30 *0.46*	0.36 *0.48*	-0.06 *0.020*
Pool	0.18 *0.38*	0.33 *0.47*	-0.15 *0.000*
Acres	0.40 *1.51*	0.88 *2.56*	-0.49 *0.001*
Owner occupied	0.70 *0.46*	0.69 *0.46*	0.02 *0.531*
Building year*	1978 *19.5*	1983 *20.9*	-5.56 *0.000*
Sales since 1983	2.76 *1.39*	2.60 *1.19*	0.17 *0.012*
Defaults since 1999	0.29 *0.62*	0.22 *0.51*	0.07 *0.018*
System cost (2000 $s)$^{+}$		27,790 *17,245*	
System size (kW)		3.37 *2.23*	
Incentive amount^{+}		11,930 *8,301*	
Observations	364,663 (*363,504)	329 ($^{+}$307)	

Table 2.2: San Diego neighborhood summary stats and comparison by solar penetration

Variable	Neighborhoods with no solar Mean *Std Dev*	Neighborhoods with at least one solar Mean *Std Dev*	No Solar - Solar Difference in Means *Pr(|T|>|t|)*
Average square footage	1,278 *326*	1,822 *535*	-544 *0.000*
Average acreage	0.22 *0.44*	0.44 *0.88*	-0.22 *0.000*
Percent with pools	3.01 *3.73*	15.01 *11081*	-12.00 *0.000*
Percent Green Party	0.50 *0.50*	0.52 *0.45*	-0.02 *0.709*
Percent Democrat	47.38 *9.42*	35.63 *8.95*	11.75 *0.000*
Median income ($1000s)	30.35 *11.97*	55.86 *22.85*	-25.51 *0.000*
Percent White	26.73 *22.70*	60.85 *23.67*	-34.13 *0.000*
Percent Owner Occupied	53.89 *18.21*	72.87 *8.95*	-18.99 *0.000*
Percent College Grads	13.54 *13.33*	31.19 *17.95*	-17.66 *0.000*
Percent Prius*	0.39 *0.03*	0.39 *0.03*	0.002 *0.993*
Percent Truck*	51.83 *8.23*	45.61 *6.92*	6.21 *0.126*
Observations	89 (*6)	496 (*89)	

*Auto data variables reported at the zip code level, all others are census tract averages

Table 2.3: Correlates of living in a solar home in the city of San Diego in 2009

	Full Sample			Aristotle Sample	
Dependent Variable: Dummy=1 if solar home	Mean	Coefficient (Std Error)	Coefficient (Std Error)	Mean	Coefficient (Std Error)
Home has solar panels	2,282			1,272	
Conservative (all HH voters)	0.703			0.405	
Liberal (all HH voters)	0.199	0.002***	0.002**	0.399	0.002**
		(0.001)	(0.001)		(0.001)
Mixed Conservative and Liberal	0.0111	0.005	0.005*	0.022	0.005
		(0.003)	(0.003)		(0.003)
Other Party	0.0866	0.000	0.000	0.174	0.000
		(0.001)	(0.001)		(0.001)
Less than high school	0.0337			0.067	
High school grad	0.103		0.001	0.205	0.001
			(0.001)		(0.001)
Some College	0.125		0.000	0.249	0.000
			(0.001)		(0.001)
College Grad	0.127		0.003**	0.253	0.003**
			(0.001)		(0.001)
Post graduate	0.0859		0.006***	0.171	0.006***
			(0.001)		(0.001)
Household has contributed to environmental organizations	0.0404		0.005***	0.080	0.005***
			(0.002)		(0.002)
political organizations	0.246		-0.001	0.490	-0.001
			(0.001)		(0.001)
religious organizations	0.0289		0.001	0.058	0.001
			(0.002)		(0.002)
Census Tract Fixed Effects		Y	Y		Y
Observations		202,864	202,864		100,943
R-squared		0.012	0.013		0.010

Estimated from a linear probability model. Additional controls include the age of the oldest registered voter in the household, whether a child is present in the household, the highest imputed income of the two oldest registered voters in the household, and an indicator for being in the Aristotle data base. A conservative is registered as Republican, American, or Libertarian Party. A liberal is a registered as Democrat, Peace and Freedom, or Green Party. Robust standard errors in parentheses. The symbols *, **, and *** indicate significance at the 10, 5, and 1 percent level, respectively.

Table 2.4: San Diego Hedonic OLS regression estimates of log sales price on solar panels

Dependent variable: Log(SalePrice)	Baseline	Neighborhood	System Size
	Coefficient *(Std Error)*	Coefficient *(Std Error)*	Coefficient *(Std Error)*
Solar	0.036*** *(0.010)*	0.031** *(0.014)*	0.043 *(0.137)*
Solar will be installed	0.004 *(0.003)*	0.004 *(0.003)*	
Solar concurrently installed	0.028 *(0.021)*	0.028 *(0.021)*	
Solar home in solar block		0.010 *(0.020)*	
Log Size (watts) * Solar			-0.001 *(0.017)*
Joint significance of solar terms			F Stat = 6.60, Prob > F = 0.001
Log(Acres)†	0.074*** *(0.003)*	0.074*** *(0.003)*	0.074*** *(0.003)*
Swimming Pool	0.050*** *(0.001)*	0.050*** *(0.001)*	0.050*** *(0.001)*
View	0.049*** *(0.001)*	0.049*** *(0.001)*	0.049*** *(0.001)*
Log(SquareFoot)†	0.432*** *(0.003)*	0.432*** *(0.003)*	0.432*** *(0.003)*
Bathrooms	0.024*** *(0.001)*	0.024*** *(0.001)*	0.024*** *(0.001)*
Constant	9.385*** *(0.012)*	9.385*** *(0.012)*	9.385*** *(0.012)*
Census tract quarter fixed effects (578 tracts, 56 quarters)	30,426	30,426	30,426
Observations	364,992	364,992	364,992
Sales with solar	329	329	329
R^2 within; overall	0.64; 0.34	0.64; 0.34	0.64; 0.34

Significant at *** 1% and ** 5% levels; † Zip code specific variation in these coefficients is also estimated; Building vintage, mortgage default frequency, sales frequency, owner occupancy tax status, and month in year of sale are included in all regressions, with coefficient estimates available from the authors by request.

Table 2.5: Predicted value of solar from hedonic estimates and comparison sample values

Predicted added value of solar at mean characteristics of sales with solar	\$22,554; (\$5.65/watt)
Average total (before subsidy) system cost of solar for solar sales	\$35,967; (\$9.02/watt)
Average net (after subsidy) system cost of solar for solar sales	\$20,892; (\$5.24/watt)
Average mean total (before subsidy) system cost of all systems installed during quarter of home sale (replacement cost)	\$30,858; (\$7.74/watt)
Average mean net (after subsidy) system cost of all systems installed during quarter of home sale	\$21,047; (\$5.28/watt)

All values adjusted to 2010 dollars

Table 2.6: Hedonic OLS regression estimates of log price on solar panels with neighborhood characteristic interaction

	Prius Share	Truck Share	Green Share	Dems Share	Log Med Income	College Grads
Variable	Coeff. *(S.E.)*	Coeff. *(S.E.)*	Coeff. *(S.E.)*	Coeff. *(S.E.)*	Coeff. *(S.E.)*	Coeff. *(S.E.)*
$Solar_{ijt}$	-0.002 *(0.022)*	0.198*** *(0.078)*	0.031** *(0.014)*	-0.027 *(0.047)*	-0.156 *(0.277)*	-0.022 *(0.026)*
$NbhdVar_j$ * $Solar_{ijt}$	0.076** *(0.038)*	-0.004** *(0.002)*	0.009 *(0.022)*	0.002 *(0.002)*	0.017 *(0.025)*	0.001* *(0.0005)*
Joint significance of solar terms - F Stat; (Prob > F)	8.77; (0.000)	8.90; (0.000)	6.69; (0.001)	7.55; (0.001)	6.84; (0.001)	8.09; (0.000)
Home characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Census tract quarter fixed effects (578 tracts, 56 quarters)	29,697	29,697	30,420	30,420	30,420	30,420
Observations	349,108	349,108	364,985	364,985	364,985	364,985
Sales with solar	319	319	329	329	329	329
R^2 within; overall	0.64; 0.33	0.64; 0.33	0.64; 0.34	0.64; 0.34	0.64; 0.34	0.64; 0.34

***,**,* Significant at 1%, 5%, 10% levels, respectively

Table 2.7: Hedonic OLS regression estimates of solar on log price with building permits

Variable	Baseline Coefficient *(Std Error)*	Major renovation Coefficient *(Std Error)*	High value renovation Coefficient *(Std Error)*	Any Permit Coefficient *(Std Error)*
$Solar_{ijt}$	0.062*** *(0.016)*	0.062*** *(0.016)*	0.060*** *(0.016)*	0.062*** *(0.016)*
Building $Permit_{ijt}$		0.025*** *(0.007)*	0.056*** *(0.005)*	-0.036*** *(0.001)*
Home characteristics	Yes	Yes	Yes	Yes
Census tract quarter fixed effects (578 tracts, 51 quarters)	13,416	13,416	13,416	13,416
Observations	136,389	136,389	136,389	136,389
Sales with solar	122	122	122	122
Sales with permit		725	1,411	20,324
Sales with solar and permit		4	12	25
R^2 within; overall	0.57; 0.31	0.57; 0.31	0.57; 0.31	0.57; 0.32

***Significant at the 1% level

Table 2.8: Repeat sales GLS regression estimates of log of sales price ratio on added solar

Variable	Baseline Coefficient *(Std Error)*	System Size Coefficient *(Std Error)*
$\Delta Solar_{ijt}$	0.036** *(0.018)*	0.611** *(0.277)*
Log Size (watts) * $\Delta Solar_{ijt}$		-0.073** *(0.035)*
Joint significance of solar terms		F Stat = 4.36, Prob > F = 0.013
Census tract specific HPIs	110	110
Observations	80,182	80,164
Sales with solar	160	160
R^2	0.76	0.76

**Significant at the 5% level

Table 2.9: Sacramento Hedonic OLS regression estimates of log sales price on solar panels

Dependent Variable: Log(Sale Price)	Mean	Baseline Coefficient (Std Error)	Street Coefficient (Std Error)
Solar	0.003	0.04	0.073
		(0.014)***	(0.026)***
Solar will be installed	0.003	0.009	0.009
		(0.013)	(0.013)
Solar concurrently installed	0.001	0.024	0.065
		(0.030)	(0.041)
Solar home on solar street			-0.046
			(0.030)
Log(acres)	-1.803	0.156	0.156
		(0.002)***	(0.002)***
Swimming Pool	0.116	0.076	0.076
		(0.002)***	(0.002)***
Log(Square Foot)	7.365	0.559	0.559
		(0.004)***	(0.004)***
Bathrooms	2.201	0.018	0.018
		(0.002)***	(0.002)***
Constant		8.523	8.523
		(0.028)***	(0.028)***
Year Built Dummies		Y	Y
Zip Code/Year/Month Dummies		Y	Y
Observations		90686	90686
Sales with solar		265	265
R^2		0.852	0.852

*** indicates significantly different from 0 at ***1% level. Regressions include year built dummies. Average sales prices is $305,178.

2.9 Data Appendix

Solar panel installations

California's Emerging Renewables Program subsidized solar panel installations as early as 1999 and supported almost all installations through 2007, when it was replaced as the primary State subsidy regime by the California Solar Initiative, which continues today.[42] Over 95% of the systems in our data are installed under these two programs. The New Solar Homes Partnership aims to encourage developers to include solar on new properties, and accounts for less than 1% of installations in our data. These programs are administered in areas of California serviced by public utilities, including San Diego County. A final program supported solar panel installations on rebuilding projects during 2005 to 2007 following wildfires in San Diego County.

Property records

The San Diego County Assessor maintains public records of characteristics and transactions of all property in the county for tax assessment purposes. We use a corresponding publicly available map file (GIS shapefile) of the boundaries of all county properties to determine the acreage of the lot on which each home is built. We also obtain information on the number of times the property has transacted in our dataset and the number of public mortgage default notices associated with the property.[43] Homes are grouped spatially using the county property map and census tract and zip code boundary maps to assign each parcel number to the respective

---

[42] http://www.gosolarcalifornia.org/about/gosolar/california.php

[43] Default data is matched by parcel number from public records published online by the San Diego Daily Transcript.

geography in which its property lies.[44] We use these groupings to construct spatial and temporal controls as well as for matching a home to the characteristics of its census tract and zip code. The assessor also maintains a record of each property transaction in the county. The date, sales price, and parcel number identifier of all single family home sales since 1983 is publicly available from these records, which form the dataset which is our source for sales prices and dates.

Our building permit data begin in 2003 for San Diego City and for Escondido. In San Diego City, building permits are required for "all new construction" including for "repair or replacement of existing fixtures, such as replacing windows." Permits are also required for changes to a home's “existing systems”; for example, moving or adding an electrical outlet requires a permit."[45] A permit is not required “wallpapering, painting or similar finish work” and for small fences, decks, and walks.[46]

Neighborhood characteristics

We use voter registration summary statistics for each San Diego County Census tract in the year 2000 from the Berkeley IGS (see http://swdb.berkeley.edu/), zip code level automobile registration summary statistics from 2007, and 2000 Census tract level demographic as sources of descriptors of San Diego neighborhoods over which solar panel capitalization may vary. The voter registration summary files report

[44] Maps were retrieved from www.sangis.org.

[45] Although not all improvements may be completed with a permit, as long as homeowners who install solar panels are not less likely than others to obtain permits for other improvements, including permitting activity in our capitalization regressions should provide evidence of the extent of bias due to unobserved home improvements and maintenance in our capitalization estimates.

[46] http://www.sandiego.gov/development-services/homeownr/hometips.shtml#whendo

the total number of registrants by political party affiliation for each census tract in California. From these reports we calculate the percent of voters in each tract who are Green Party registrants. Similarly, we calculate the Toyota Prius share of registered autos from zip code totals of year 2007 automobile registration data (purchased from R.L Polk). We likewise calculate the percent registered Democrats and vehicles classified as trucks from the respective summary datasets. We obtain reported census tract median income and the percent of the over age 25 population who are college graduates from the 2000 Census.

References

Andreoni, J., "Impure altruism and donations to public goods: a theory of warm-glow giving," *The Economic Journal*, 100(401), 464–77, 1990.

Bagwell, L.S. and B.D. Bernheim, "Veblen effects in a theory of conspicuous consumption," *American Economic Review,* 86(3), 349–73, 1996..

Becker, G. S. "A note on restaurant pricing and other examples of social influences on price," *Journal of Political Economy,* 99(5), 1109-1116, 1991..

Borenstein, S., "The market value and cost of solar photovoltaic electricity production," UCEI Working Paper CSEM WP 176, 2008.

Bollinger, B. and K. Gillingham, "Environmental preferences and peer effects in the diffusion of solar photovoltaic panels," Stanford Working Paper, 2010.

Brounen, D. and N. Kok, "On the economics of energy labels in the housing market," *Journal of Environmental Economics and Management,* forthcoming, 2011.

Chung, E. and E. Fischer, "When conspicuous consumption becomes inconspicuous: the case of migrant Hong Kong consumers," *Journal of Consumer Marketing,* 18(6), 474-87, 2001.

Costa, D. L., and M.E. Kahn, "Why has California's residential electricity consumption been so flat since the 1980s?: a microeconometric approach," NBER Working Paper No. 15978, 2010.

Cragg, M.I, Y. Zhou, K. Gurney and M.E Kahn, "Carbon Geography: The Political Economy of Congressional Support for Legislation Intended to Mitigate Greenhouse Gas Production," NBER Working Paper No 14963, 2011.

Dastrup, S. R., "Factors influencing the consumer adoption of solar panels in San Diego," Unpublished Manuscript, 2010.

Delmas, M. and L. Grant, "Eco-labeling strategies and price-premium: the wine industry puzzle," *Business & Society,* 20(10), 1-39, 2010.

Eichholtz, P., N. Kok, and J.M. Quigley, "Doing well by doing good? green office buildings," *American Economic Review*, 100(5), 2010.

Harding, J., C. F. Sirmans, and S.S. Rosenthal, "Depreciation of housing capital, maintenance, and house price inflation: estimates from a repeat sales model," *Journal of Urban Economics*, 61(2), 193-217, 2007.

Heutel, G., and E. Muehlegger, "Consumer learning and hybrid vehicle adoption," HKS Faculty Research Working Paper Series RWP 10-013, 2010.

Kahn, M.E., "Do greens drive Hummers or hybrids? Environmental ideology as a determinant of consumer choice," *Journal of Environmental Economics and Management,* 54(2), 129-145, 2007.

Kahn, M.E. and R. Vaughn, "Green market geography: The spatial clustering of hybrid vehicles and LEED registered buildings," *B.E. Journal of Economic Analysis and Policy*, 9(2), 1-22, 2009.

Kotchen, M. "Green markets and private provision of public goods," *Journal of Political Economy*, 114(4), 816-845, 2006.

Kotchen, M and M. Moore, "Conservation: From Voluntary Restraint to a Voluntary Price Premium," *Environmental & Resource Economics*, 40(2), 195-215, 2008.

Rosen, S. "Markets and diversity," *American Economic Review*, 92(1), 1-15, 2002.

Veblen, T., "The theory of the leisure class: an economic study of institutions," 1899, Repr., Kila, MT: Kessinger, 2004.

Wilhelmsson, M. "House price depreciation rates and level of maintenance," *Journal of Housing Economics,* 17(1), 88-101, 2008.

Chapter 3

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

3.1 Introduction

The value of the U.S. housing stock fell $4.4 trillion dollars from 2006 to the first quarter of 2009,[47] leading to considerable turmoil in financial markets across the world. This fall is widely believed to be one of the primary contributing factors to the subsequent financial crises and recession. 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.[48] Casual inspection of the magnitude of the change in housing prices across metropolitan areas suggests that the changes are anything but uniform. Across 358 Metropolitan Statistical Areas (MSAs) the magnitude of the fall varies from essentially zero to over sixty percent. Figure 3.1 illustrates this variation in a map of the MSAs. In this paper, we investigate whether it

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

[48] 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).

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*. While no claims of causality are being made, the statistical relationships we document give insight into what went wrong in many housing markets. In this sense, our paper is much less ambitious than several working papers currently in circulation or recently published (e.g.,Glaeser, Gyourko and Saiz, 2008; Glaeser and Gyourko, 2007; Himmelberg, Mayer and Sinai, 2005) that attempt to explain the entire dynamics of housing price changes. Our contribution is to present a set of stylized facts focusing on differences in the recent housing price declines across all U.S. 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 when our exercise is repeated with the data prior to 2000. The repetition of these relationships across the decades of available data suggests a predictive relationship of local housing variables and the size subsequent price declines. It would be possible to incorporate these metropolitan level indicators into mortgage lending policy and portfolio risk assessment.

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 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.

Our paper is related to two literatures. The first describes and models housing price dynamics. While we do not model price movements, this strand of research informs our examination of relationships between the magnitude of recent price declines and predetermined market factors in both the choice of factors to explore and the discussion of hypotheses for the observed relationships. Market-level price trends historically exhibit short run positive serial correlation and long run mean reversion (e.g. Case and Shiller, 1989). Capozza et al. (2002) estimate variability by market conditions in this mean reversion and serial correlation for 62 metropolitan areas and estimate adjustment rates to fundamental values based on 1979 to 1995 data. They find that construction costs and income and population growth are associated with higher autocorrelation and real construction costs are positively related to serial correlation. In calibrating a model of housing market dynamics, Glaeser and Gyourko (2007) document that most variation in housing prices and construction is local, not national. Malpezzi and Wachter (2005) discuss the role of speculation in the context of efficient housing markets and demonstrate in a simulation that speculation contributes to boom and bust cycles when supply is inelastic. Glaeser, Gyourko, and Saiz (2008) build on the topographic supply elasticity work of Saiz (2008) and emphasize the importance of supply constraints in the trajectory of housing bubbles both in a dynamic model and in the 1980s to 1996 booms and busts experienced many metropolitan areas. They find that more inelastic places had shorter booms with more severe mean reversion, although on average "the elasticity was uncorrelated with either price or quantity changes during the bust." They note the "the fact that highly elastic places had price booms is one of the strange facts about the recent price explosions" and note potential continued price declines in inelastic cities where prices remained above their calculated costs. Ortalo-Magne and Rady (2006) provide an example of incorporating demand constraints into a model, with credit constrained first time buyers, housing trade-ups, and income shocks producing price overreactions relative to income growth.

The second literature, to which our paper more closely belongs, is the description and analysis of the recent housing, building, and credit boom and bust and subsequent recession. Our paper focuses particularly on describing variability in the magnitude of single family housing price declines using a variety of predetermined variables which are examined individually elsewhere in this growing literature. Examples include discussions of the mortgage bubble such as Hendershott, Hendershott, and Shilling (2010) which focuses on the role of Fannie Mae and Freddie

Mac and the securitization of low quality loans. Pavlov and Wachter (2011) introduce a model wherein aggressive mortgage instruments magnify the real estate cycle and the effects of demand shocks. They find empirically that regions with higher concentrations of these instruments experienced larger price increases and, consistent with our results, subsequent declines. Mian and Sufi (2010) draw on their collection of papers analyzing microeconomic credit and housing data to argue that expanded credit supply was at the heart the crisis preceding the "Great Recession". An example of a special issue devoted to describing particular causes and consequences of this episode is the symposium "The Mortgage Meltdown, the Economy, and Public Policy" in the B.E. Journal of Economic Analysis and Policy (Gabriel, Quigley, and Rosenthal 2009) which includes forecasts of the future of mortgage finance, analyses of the housing market regulatory environment, and the impact on urban neighborhoods of the foreclosure crises. Our paper contributes to this growing literature by establishing a set of descriptive facts relating the magnitude of metropolitan area level price housing price declines to a number of predetermined market characteristics in both a simple linear regression framework and a flexible nonparametric setting.

3.2 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%.[49] 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 3.2 shows two groups of geographically close metropolitan statistical areas with disparate house price experiences.

We look at several alternative hypotheses to motivate our set of covariates that might predict the magnitude of price declines. The first takes the position that the run-up in housing prices post 2000 was in some ways artificial and not driven by (or well in excess of) growth in underlying economic fundamentals. This suggests that the larger the run-up in housing prices the larger the likely fall, consistent with the established eventual mean reversion of housing prices.[50] Closely related to this alternative would be variants that revolve around changes in the shape of the

---

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

[50] There are more nuanced, but related, versions of this hypothesis focusing on demand and supply factors that may render some markets more volatile than others. For example, housing supply may adjust slowly to demand fluctuations due to regulatory, geographic, or local housing industry capacity constraints. Markets may also differ in expectations of future economic growth and the speed of housing supply responses. Glaeser, Gyourko, and Saiz (2008) provides a start on identifying variation in supply factors. Differences in the speed of housing price adjustment after a price peak are predicted by these models and appear to be a fruitful avenue for future research.

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.

3.3 Data

We take as our dependent variable, *%DropHPI*, the percentage drop in a market's real housing price index (HPI) defined as the percent decline from its highest point between the first quarter of 2000 and the third quarter of 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 each of the 363 metropolitan areas, 358 of which we are able to match to other data sources described below.[51] We use the OFHEO HPI rather than some of its competitors, such as the well-known Case-Shiller index, largely because it is available for all US metropolitan areas.[52] There is considerable variation in *%DropHPI*. It ranges from 2.39 to 61.72 with a mean of 13.35 and a standard

---

[51] 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.

[52] The average correlation between the OFHEO and the Case-Shiller indexes for the 20 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. Our two HPI based variables, *%DropHPI* and *%GainHPI,* calculated for the 20 Case-Shiller cities have correlations of 0.996 and 0.931 with the same variables calculated using the OFHEO series. Leventis (2008) finds that the inclusion of appraisals (refinance value), differences in weighting procedures and in sample composition account for much of the differences in the geographies covered by both indexes. Private data providers such as CoreLogic are starting to enter the market with more detailed data.

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 3.3.[53]

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.[54] *%GainHPI* can be viewed as representing a shift in a market's home price distribution. Figure 3.4 shows the predictive relationship of *%GainHPI* and *%DropHPI*.[55] Changes in the shape of market price

---

[53] Comparing the 20 cities in the Case Shiller index to their OFHEO counterparts, *%GainHPI* varies little across the two indexes, while *%DropHPI* is on average 2.6 times greater using the Case Shiller index. The coefficient on *%GainHPI* as a predictor of *%DropHPI* in a stacked univariate regressions combining the data sources, while greater when using Case Shiller data, are statistically indistinguishable.

[54] 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.

[55] 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.

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).[56] 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*.[57]

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*; population; and a list of demographic variables. Percentage change variables are constructed for median income, owner occupancy rates, percent retired, and population using the parallel variables in the ACS.[58] 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

---

[56] 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.

[57] 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 SandP/Case-Shiller indexes.

[58] Demographic variables include percent of households in poverty, racial composition, average education levels, and percent retired (imputed as the over 65 population not in the labor force).

2000 Census housing stock.[59] 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*.[60]

Each of our predictor variables is predetermined with respect to *%DropHPI* in the sense that its realization occurred before housing price declines began in earnest. However, we do not claim that we are examining exogenous variation to determine the causal impact of the predictors. Indeed, the variables we examine are determined jointly by underlying supply and demand processes in the housing, labor, and credit markets. This interdependence is suggested by the correlation matrix presented in Table 3.1, which shows that many of our key predictor variables are indeed highly correlated. Our linear regression analysis describes the additional predictive ability of each variable when controlling for the others, while our nonparametric approach uses cross validation to choose the best predictors from the set while allowing for more robust nonlinear relationships. The measured relationships should thus be interpreted as reduced form partial correlations.

---

[59] That is: $\%ExcessPermits = \left(\frac{Population_{2005}}{Population_{2000}} - 1\right) - \frac{\sum_{2000}^{2005} Permits}{Housing\ Stock_{2000}}$. 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.

[60] While the HMDA measure does not capture many of the non-traditional mortgage products that are of interest in evaluating this boom, it is a measure available at the MSA level, and other researchers have found correlation among the HMDA measure and alternative loans. 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 3.2, we present summary statistics for the 358 MSAs for which all data is available.[61] 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

[61] 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.

housing market changes is an "affordability index," typically a comparison of median prices to median incomes. Figure 3.5 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.[62] 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 and Summers, 2008) or of geographic supply constraints (Glaeser, Gyourko and Saiz, 2008; Saiz, 2008) that likely interact in important ways with our *%ExcessPermits* measure. 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).

3.4 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

---

[62] 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.

explanatory variables. We also estimate models based entirely on geography and demographic variation. Estimates are provided in Table 3.3; all are OLS estimates with White standard errors.[63]

The results in Table 3.3 provide several insights. First, one variable, the magnitude *%GainHPI,* explains 55% of the variance in the drop (column 1) and the elasticity[64] [evaluated throughout at the vector of mean covariate values] from the full model (column 6) implies that two thirds of the post first quarter 2000 gain is given up conditional on the other covariates. 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). As additional variables reflecting contemporaneous changes in metropolitan area housing markets are added to the model, the parameter estimate on *%GainHPI* is remarkably stable, suggesting that this relationship is largely orthogonal to the other explanatory variables.[65] Glaeser, Gyourko, and Saiz document an "enormous mean reversion" from

---

[63] 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.

[64] We report elasticities as statistical calculations describing the magnitude of the conditional correlations measured as opposed to a parameter of a structural relationship, which we are not estimating.

[65] 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.

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.[66]

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).[67] In the full model (column 6), 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

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

[67] Alone (in a regression without *%GainHPI*) the shape variables explain 3% of the variance.

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 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.[68] Figure 3.6 plots these tiered indexes for each of the 17 cities for which they are

The variable measuring the share of high priced loans in a market, *%Subprime*, has the expected positive sign. Its associated elasticity is also sizeable at .49. This is consistent with work that has been done looking at subprime and other nontraditional loans and the financial crisis, notably Mian and Sufi (2009b). 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

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

major implications for government policy toward loans of this type.[69] The parameter estimate for *%ΔOwnOcc* is negative and significant in models controlling for demographic makeup and location (column 6). 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%].

Our next set of findings focus on variables intended to capture changes in housing demand and initial price, income, and ownership levels. Our demand proxies are income and population growth. Change in median income is an important predictor of price declines. The parameter estimates for *%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. Our measure of changes in pressure on the local housing stock, *%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. The initial levels of prices, incomes and ownership have weaker relationships with price declines. The level of income

---

[69] 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 and Sufi 2009a) for an in depth analysis of the contribution of household borrowing during the price run-ups to the subsequent rise in mortgage default.

variable, *MedInc($lk),* 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 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($10K),* enters positively but is not 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.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.

Our final stylized fact is that Census division, demographics, and MSA size 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 6 unreported coefficients), 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). An MSA's demographic makeup adds predictive power relative to Census division and MSA size alone (columns 7 and 8), but individual coefficients are insignificant after conditioning on housing market variables and Census division (column 6; individual coefficients not reported).[70]

In contrast, a model with just the population size indicators and Census divisions (column 7) presents a much different picture. This model explains 35% of the variance. The two population size indicators are highly significant and positive (2.97 and 3.89, for the middle level and large size categories,[71] 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. Adding geographic, demographic, and size variables to the model with housing variables

[70] When demographics are added to the housing market variables without Census division controls, *%Black* has a negative and significant association, while *%HighSchool* has a positive and significant association.
[71] An F-test fails to reject that these coefficients are equal.

increases the variance explained by 8%. In this full model, MSA level housing market characteristics maintain their significance in describing price decline. In contrast, adding housing indicators to the division, demographics, and population size model explains an additional 29% 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.[72] Our decision to go with the largest possible sample size in terms

---

[72] An examination of the residuals from our model suggests that there may be a localized phenomena associated with large deviations. The first of these is centered on Detroit and several nearby MSAs where we under predict %DropHPI. This is plausibly a result of the collapse of the Detroit-based automobile sector. The second is in the Central Valley of California, centered on Merced where there was hope of the rapid expansion of a new

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.

We note that our estimates allow for multiple interpretations of the underlying process generating price declines. Our basic linear framework estimates partial correlations which indicate, for example, that increased subprime lending share is associated with greater price declines in the recent downturn after conditioning on a variety of market variables that were jointly predetermined with respect to the downturn. This is consistent with various underlying causes such as heterogeneous local economic conditions not captured by our changing income measure, differences in local credit market structure, and differences in household and investor expectations about local housing price dynamics. The likelihood of multiple underlying processes generating the correlations we document is also apparent for other variables. 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 perhaps much smaller geographic areas and adapted to include the additional contributing factors suggested by the

---

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.

stylized facts above. Increased access to improved microdata will aid the research process of determining the contribution of various underlying causal mechanisms for each factor. For example, large datasets with detailed characteristics of individual loans and the securitization process are being used to examine the loan origination process (Keys, Mukherjee, Seru, and Vig, 2010) and housing price movements are being measured for finer geographies than MSAs. This type of data and improved fine geography measures of supply constraints may prove fruitful in distinguishing the causal mechanisms of the relationships we have established.

Table 1 underscores that our descriptive variables are interrelated and intuition suggests that the interaction of some factors during the housing boom may affect subsequent declines. Nonlinearities also likely exist in the relationships between our predictor variables and *%DropHPI.* Figure 3.8 plots the residuals from our preferred linear specification (column 4 of Table 2) against three of our key descriptor variables. It is apparent that nonlinear relationships between our variable of interest and our descriptors remain in the residual. Rather than augmenting our linear framework with higher order terms and interactions, we turn to a nonparametric descriptive approach.

3.5 Nonparametric Estimates

To address the apparent nonlinearities and interactions in the observed relationships between the magnitude of price declines and our predetermined housing market variables , 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. [73] The variables used in the selected model are listed in Table 3.4, 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

[73] 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.

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 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 3.9.

Threshold values are also estimated 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 3.10 and 3.11 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 4 in Table 3.3) and presented fully in Table 3.4.

To explore the importance of predictors based on variable interactions, we allow for two way interactions in the MARS procedure.[74] The resulting model (MARS 2) utilizes 23 basis functions constructed from 10 variables with a generalized cross

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

validated R-squared of 0.90. The variables are listed with predictive importance scores in Table 3.6. 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. 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 3.12 and Figure 3.13, which illustrate that simultaneous high levels of risky lending, overbuilding, and rapid price appreciation predict the greatest price declines. The full model is presented in Table 3.7.

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

3.6 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.[75] 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*. We utilize variables from the 1990 Census[76] 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

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

[76] 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.

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 3.14 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 (column 1 of Table 3.3). 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 that historically when real prices have declined, 1/3 of the prior 5 years of appreciation has been lost.[77] 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

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

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 observations prior to 2000 where relative to the previous five years, real prices appreciated at a rate of higher than 4.3% a year,[78] the rate our nonparametric model selects as a kink point above which price increases predict the magnitude of subsequent declines. As reported in Table 3.8, 951 (58%) of these 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.[79] 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

[78] Calculated using the median 7 years from 2000 to peak for *%GainHPI*.

[79] An additional 62 MSAs dropped at least 10% but had pre-peak growth rates less than 5.4%.

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 regard to MSA level risk. As an example, at the time of this writing, 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.[80] Similarly, HUD handbooks of mortgage insurance guidelines as well as GSE underwriting guides are

[80] 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.

silent on local market level risks.[81] 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 and Barberio 2009).

The value to a lender of a mortgage depends on housing price trends through prepayment and default risks (Downing, Stanton and 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

[81] 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.

based on local market conditions. The observed uniformity in pricing and underwriting presents a puzzle for future research.

3.7 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* 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*.[82] From the vantage point of financial markets, the fundamental question is why there

[82]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.

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.

I thank Richard Carson, coauthor of the research presented in this chapter. It is with his permission that I include our research in this dissertation.

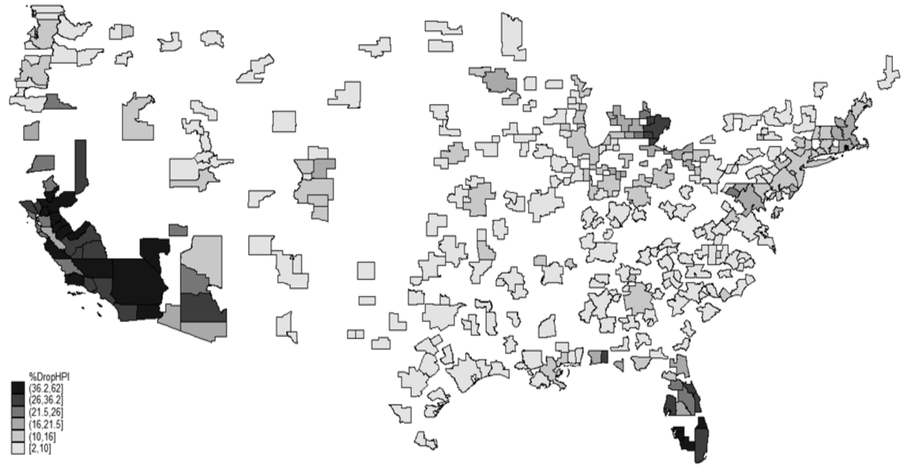

Figure 3.1: Variation in HPI declines for US MSAs

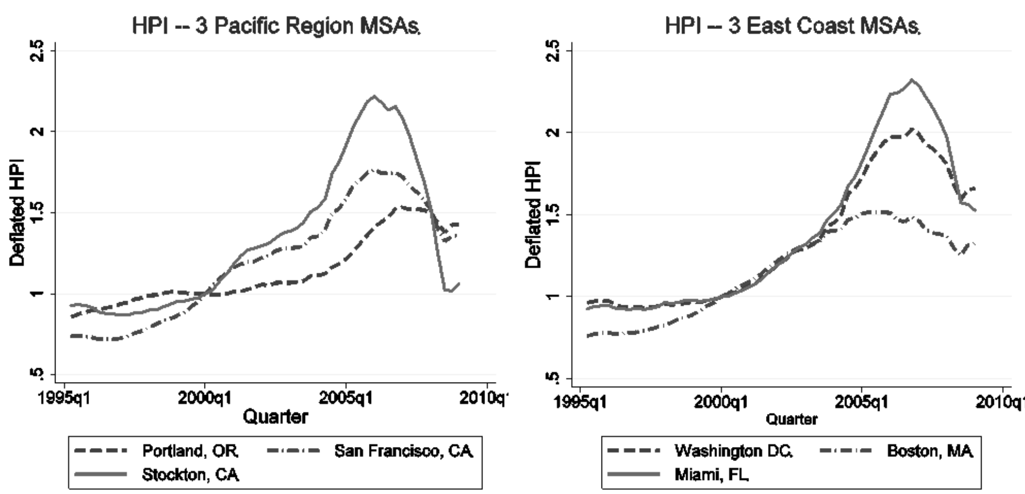

Figure 3.2: Examples of HPI variation Across MSAs

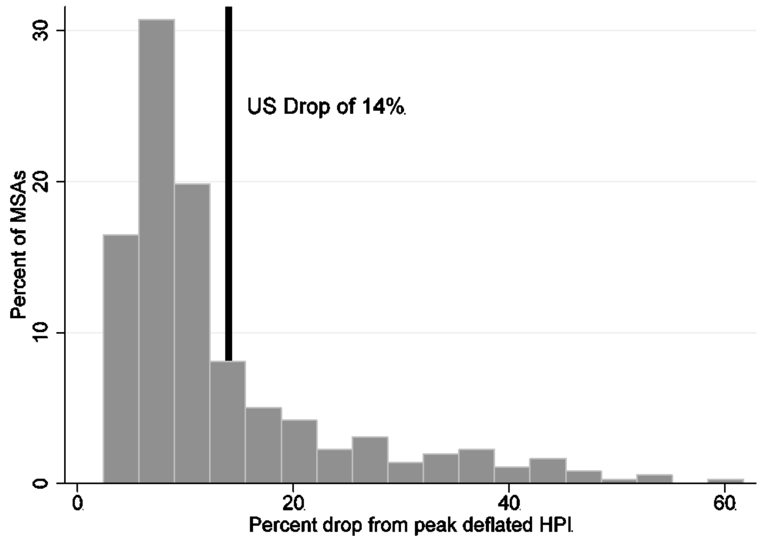

Figure 3.3: Histogram of *%DropHPI*

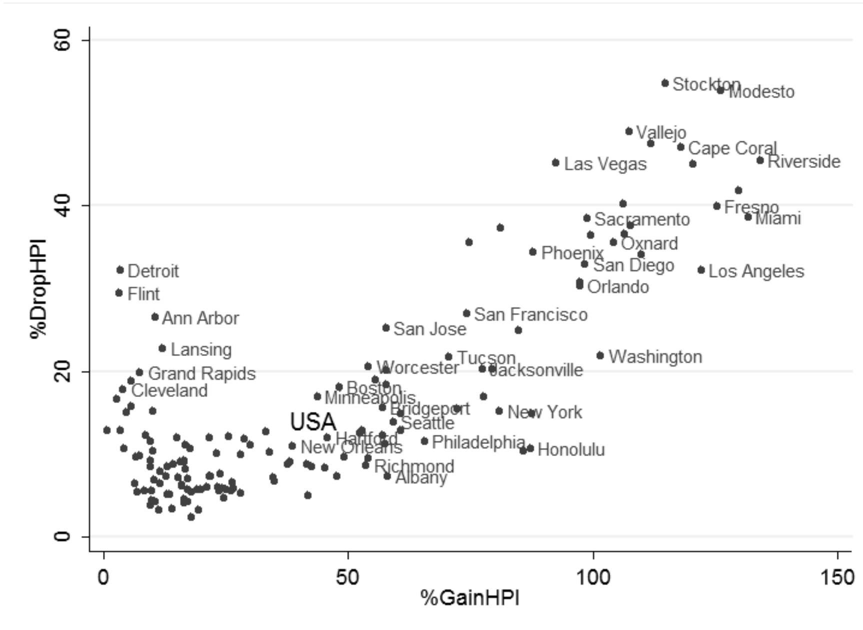

Figure 3.4: ***%DropHPI*** **against** ***%GainHPI***

Table 3.1: Correlation matrix of key predictor variables

	%GainHPI	*%Excess Permits*	*MedInc ($1K)*	*%ΔMedInc ($1K)*	*%Subprime*	*%ΔPop*
%DropHPI	0.74	0.13	0.24	0.27	0.34	0.33
%GainHPI	1	-0.06	0.19	0.50	0.14	0.37
%ExcessPermits		1	0.03	-0.16	-0.07	0.11
MedInc($1K)			1	-0.07	-0.24	0.11
%ΔMedInc($1K)				1	-0.02	0.18
%Subprime					1	0.09

Table 3.2: 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			

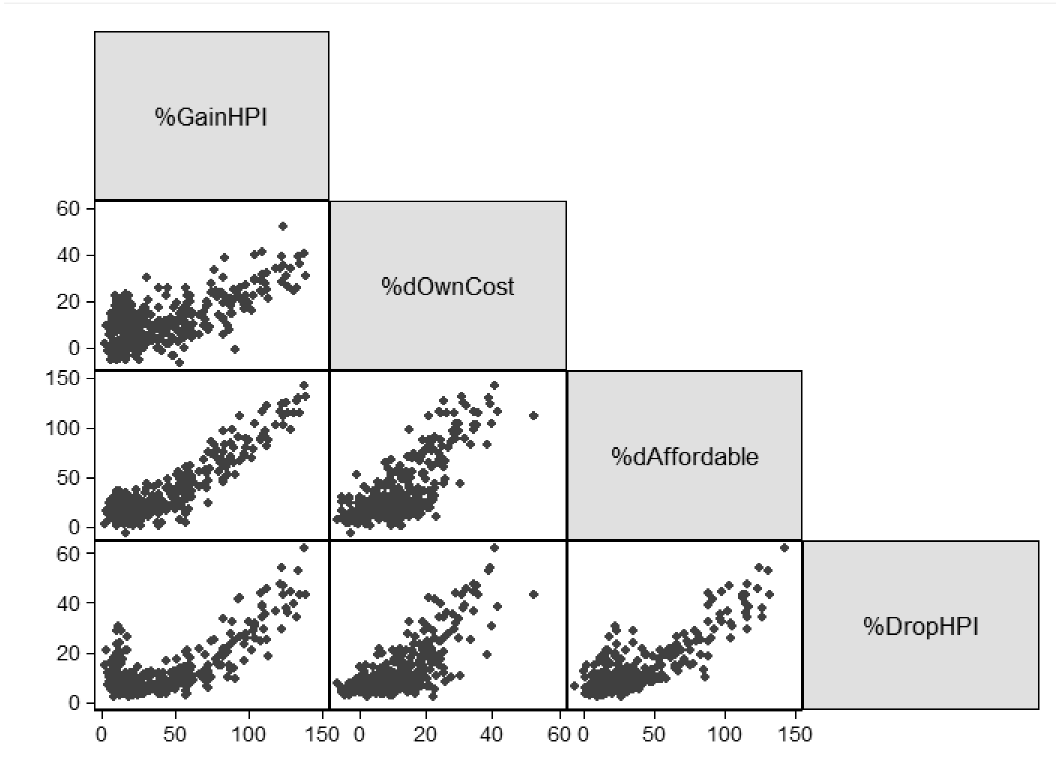

Figure 3.5: Scatterplots of price increase variables and *%DropHPI*

Table 3.3: Housing variables regression results

	(1)	(2)	(3)	(4)	Elasticity[‡]	(6)	Elasticity[‡]	(7)	(8)
%GainHPI	0.24***	0.25***	0.25***	0.26***	0.74	0.24***	0.67		
	(0.02)	(0.02)	(0.01)	(0.02)	(0.05)	(0.02)	(0.05)		
%ΔMedHPI		0.38***	0.31***	0.33***	0.04	0.22***	0.03		
		(0.05)	(0.04)	(0.05)	(0.01)	(0.05)	(0.01)		
%ΔIQHPI		-0.09***	-0.08***	-0.05*	-0.07	-0.04	-0.04		
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)		
%ExcessPermits			0.57***	0.50***	0.19	0.24*	0.09		
			(0.13)	(0.14)	(0.05)	(0.13)	(0.05)		
%Subprime			0.28***	0.30***	0.50	0.29***	0.49		
			(0.05)	(0.05)	(0.08)	(0.06)	(0.10)		
%ΔOwnOcc			-0.18	-0.16	0.02	-0.45***	0.06		
			(0.12)	(0.149)	(0.02)	(0.15)	(0.02)		
%ΔMedInc($1k)				-0.18***	-0.24	-0.09	-0.12		
				(0.06)	(0.08)	(0.06)	(0.07)		
%ΔPopulation				0.02	0.01	0.13**	0.06		
				(0.05)	(0.03)	(0.05)	(0.03)		
MedInc($1k)				0.052	0.16	0.25	0.74		
				(0.098)	(0.29)	(0.15)	(0.46)		
MedHPrice($10k)				0.22*	0.18	0.13	0.10		
				(0.12)	(0.10)	(0.15)	(0.12)		
%OwnOcc				0.20**	0.94	-0.08	-0.36		
				(0.08)	(0.37)	(0.10)	(0.45)		
Demographics[†]						Yes		Yes	
Census Division/ Size						Yes		Yes	Yes
Constant	4.31***	4.54***	-4.62	-19.73**		-18.11		-30.78	25.29***
	(0.61)	(0.72)	(1.48)	(5.42)		(14.35)		(19.55)	(2.40)
Observations	358	358	358	358		358		358	358
R-squared	0.55	0.63	0.70	0.73		0.81		0.52	0.35

Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; [†]Demographic variables are listed in Table 2; [‡]Reported at mean.

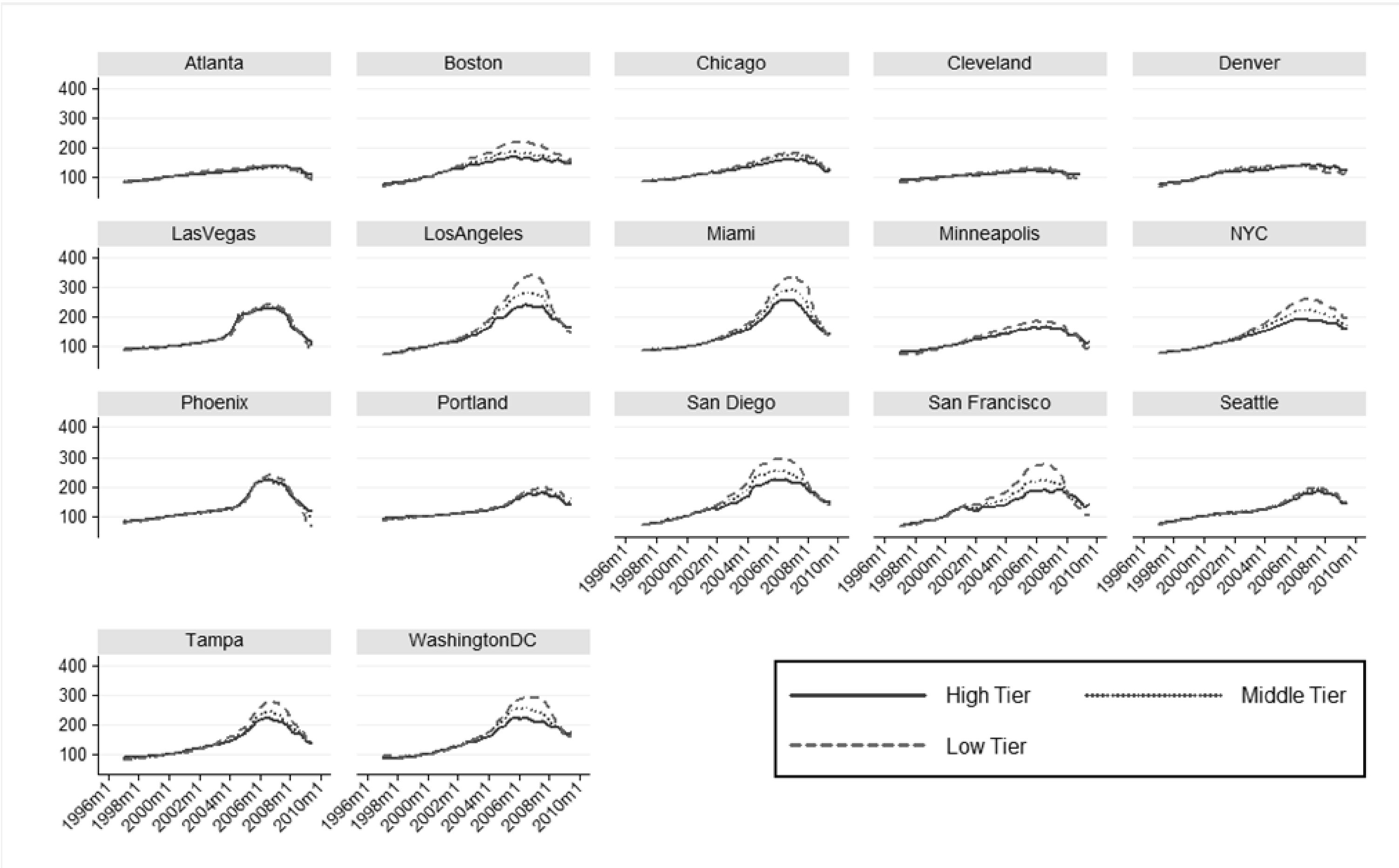

Figure 3.6: SandP/Case Shiller Tiered HPI by City

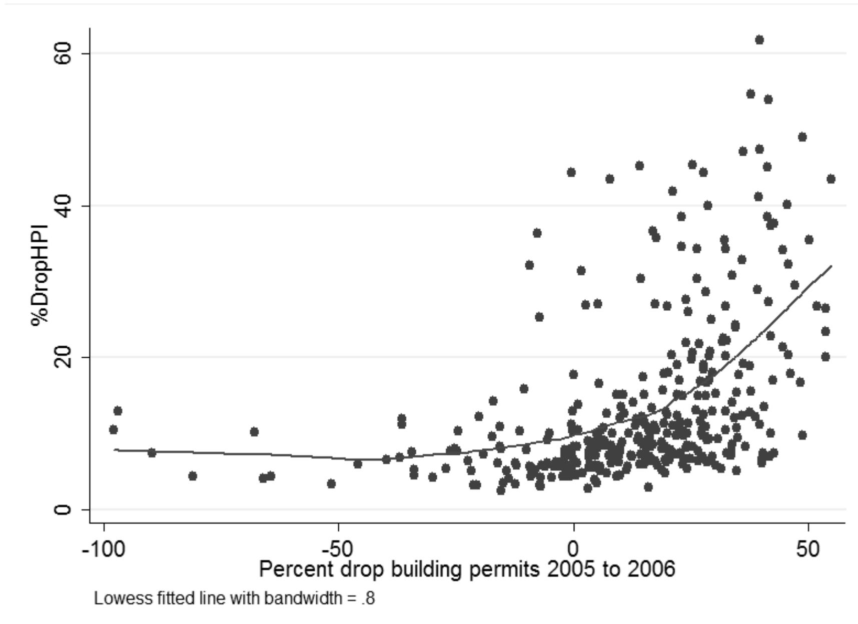

Figure 3.7: Percent decline in building permits 2005-2006 and %DropHPI

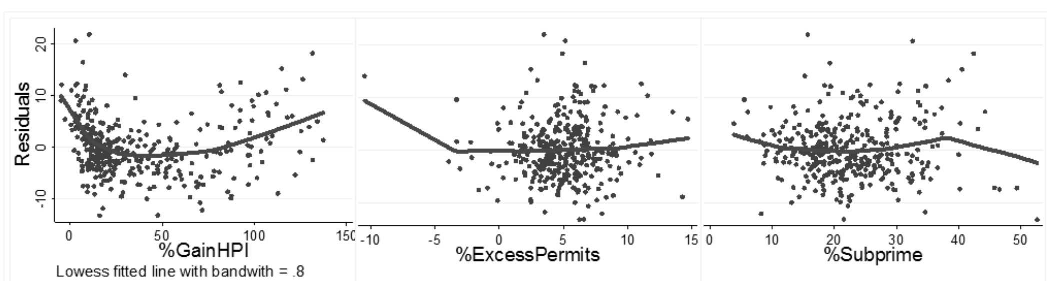

Figure 3.8: Residuals and key predictor variables

Table 3.4: 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	|||

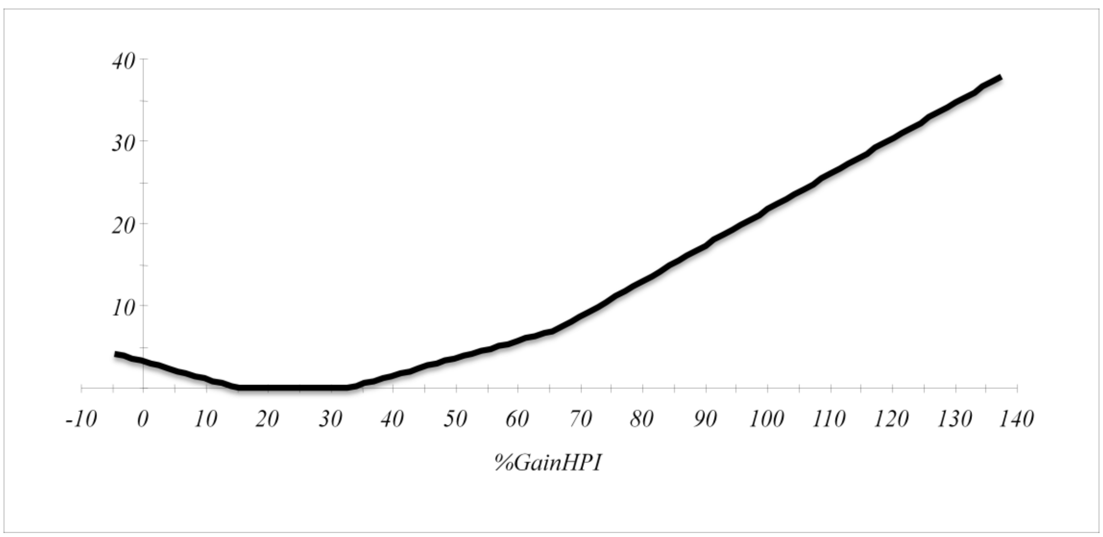

Figure 3.9: MARS 1 predicted response of %DropHPI to %GainHPI

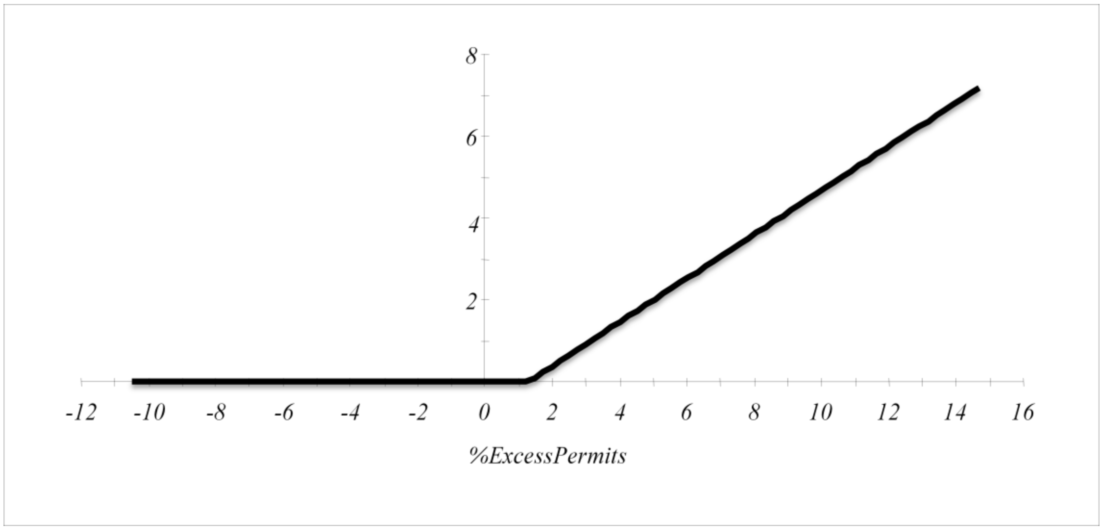

Figure 3.10: MARS 1 predicted response of of *%DropHPI* to *%ExcessPermits*

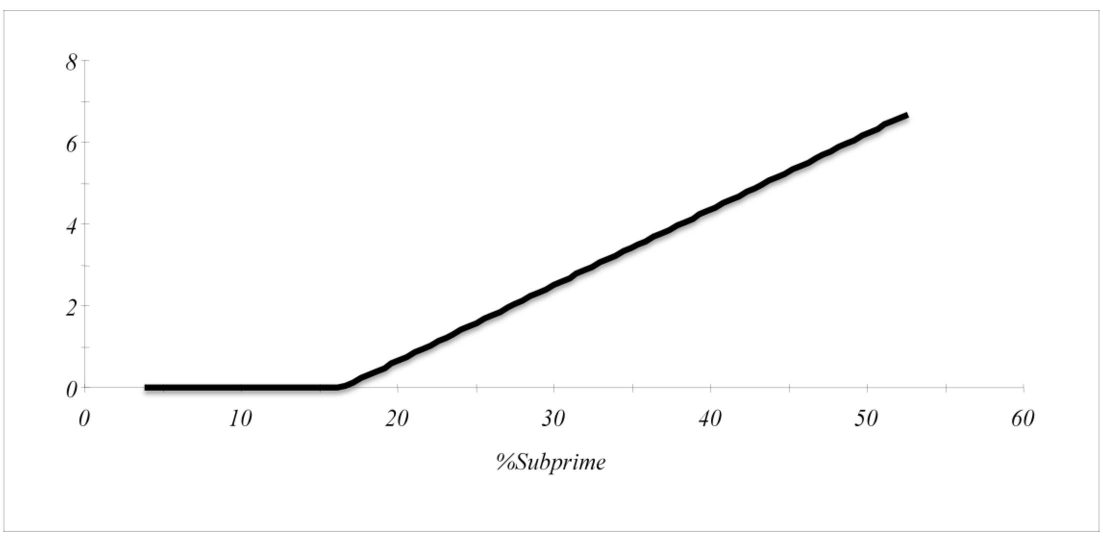

Figure 3.11: MARS predicted response of %DropHPI to %Subprime

Table 3.5: Mars 1 basis functions and coefficients predicting *%DropHPI*

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

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 above.

Model includes census division subsets: 1 New England, East North Central, Mountain, Pacific, 2 West North Central

Table 3.6: 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

Table 3.7: 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, *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

The 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 above.

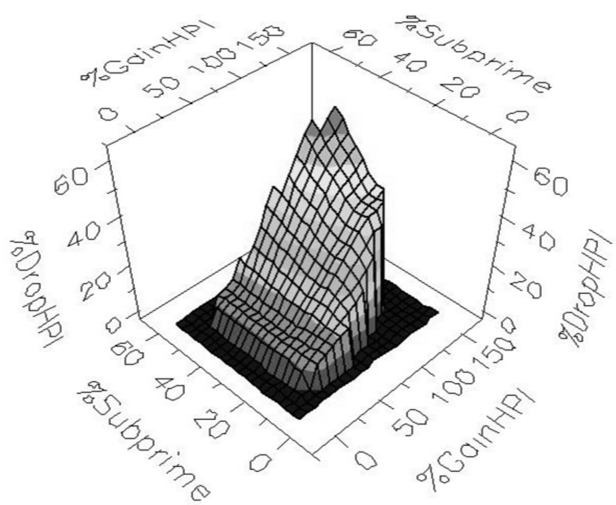

Figure 3.12: MARS2 predictive response of *%DropHPI* to *%GainHPI* and *%Subprime* interaction

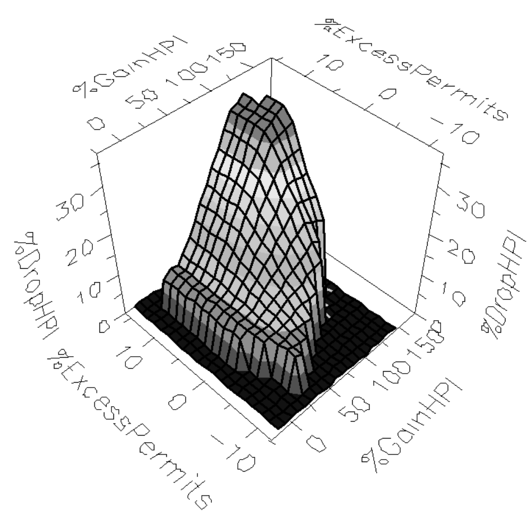

Figure 3.13: MARS 2 predicted response of *%DropHPI* to *%ExcessPermits* and *%GainHPI* Interaction

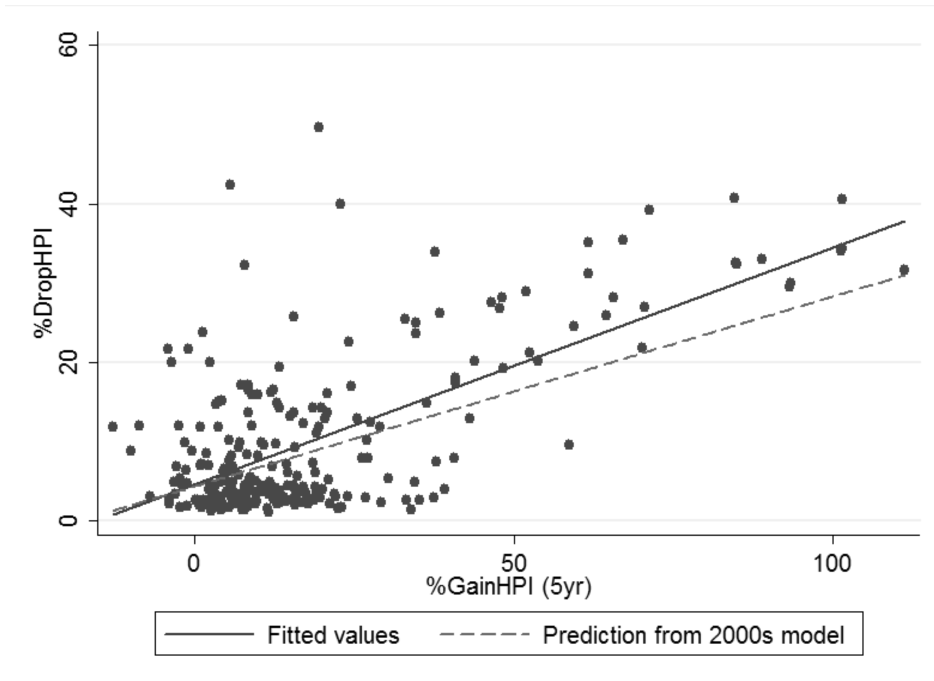

Figure 3.14: Pre-2000 HPI drops and prediction line of 2000s model

Table 3.8: 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			

Table 3.9: 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			

Summary Statistics for price declines and 1990 census variables for 243 HPI peaks in our pre-2000 sample of MSAs.

Table 3.10: Regression Results for Pre 2000 *%DropHPI* -- Housing Variables

	(1)	(4)	Elasticity at Mean	(6)	(7)	(8)
%GainHPI	0.30***	0.23***	0.42	0.17***		
	(0.02)	(0.03)	(0.07)	(0.04)		
MedInc($1k)		0.21	0.70	0.76**		
		(0.18)	(0.58)	(0.35)		
MedHPrice($10k)		0.16	0.13	-0.35		
		(0.28)	(0.23)	(0.39)		
%OwnOcc		-0.32**	-1.87	-0.15		
		(0.14)	(0.79)	(0.15)		
Demographics/Size				Yes	Yes	
Census Division				Yes	Yes	Yes
Constant	4.60***	16.41**		9.84	3.47	30.40
	(0.59)	(6.97)		(13.18)	(14.36)	(2.97)
Observations	243	243		243	243	243
R-squared	0.44	0.49		0.60	0.56	0.37

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Regression of *%DropHPI* subsequent to pre-2000 peak on prior 5 year appreciation and 1990 census variables for 243 MSAs level peaks.

3.8 Appendix: Historical Comparisons

Price gains and drops for historical housing cycles and demographic variables from the 1990 Census are used to repeat our linear 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 3.10 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 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 1.11 and compare well to our estimates for current declines. Our price increase variable, *%GainHPI*, alone captures 44% of the variation in historical price declines. 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.

References

Brueckner, J.K. “Growth Controls and Land Values in an Open City,” *Land Economics,* 66(3), Private Markets, Public Decisions: An Assessment of Local Land-Use Controls for the 1990s, 237-248, 1990.

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

Capozza, D.R., P.H. Hendershott, C. Mack, and C.J. Mayer, “Determinants of Real House Price Dynamics,” NBER Working Paper No. 9262, 2002.

Case, K.E. and R.J. Shiller, “The Efficiency of the Market for Single-Family Homes,” *The American Economic Review,* 79(1), 125-137, 1989

Downing, C., R. Stanton, and N. Wallace, “An Empirical Test of a Two-Factor Mortgage Valuation Model: How Much Do House Prices Matter?” *Real Estate Economics,* 33(4), 681-710, 2005.

Engle, R., P. Navarro, and R. Carson, “On the Theory of Growth Controls,” *Journal of Urban Economics,* 32, 269-269, 1992.

Friedman, J.H., “Multivariate Adaptive Regression Splines,” *The Annals of Statistics,* 19(1), 1-141, 1991.

Gabriel, S.A., J.M. Quigley, and L.A. Rosenthal, “The Mortgage Meltdown, the Economy, and Public Policy,” *The B.E.Journal of Economic Analysis and Policy,* 9(3), Symposium, 2009.

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

Glaeser, E.L., J. Gyourko, and A. Saiz, “Housing Supply and Housing Bubbles,” *Journal of Urban Economics,* 64(2), 198-217, 2008.

Goetzmann, W.N., L. Peng, and J. Yen, “The Subprime Crisis and House Price Appreciation,” NBER Working Paper w15334, 2009.

Gyourko, J., A. Saiz, and A. Summers, “A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index,” *Urban Studies,* 45(3), 693-729, 2008.

Hendershott, P., R. Hendershott, and J. Shilling, "The Mortgage Finance Bubble: Causes and Corrections," *Journal of Housing Research*, 19(1), 1-16, 2010

Himmelberg, C., C. Mayer, and T. Sinai, "Assessing High House Prices: Bubbles, Fundamentals and Misperceptions," *Journal of Economic Perspectives,* 19(4), 67-92, 2005.

Kelly, A.J., "Skin in the Game': Zero Down Payment Mortgage Default," *Journal of Housing Research,* 17(2), 75-99, 2008.

Keys, B.J., T. Mukherjee, A. Seru, V. Vig, "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *Quarterly Journal of Economics*, 125(1), 307-362, 2010.

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

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

Malpezzi, S. and S. Wachter, "The Role of Speculation in Real Estate Cycles," *Journal of Real Estate Literature*, 13(2), 141-164, 2005.

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

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

Mian, A.R. and A. Sufi, "House Prices, Home Equity-Based Borrowing, and the U.S. Household Leverage Crisis," NBER Working Paper w15283, 2009a.

Mian, A.R. and A. Sufi, "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *Quarterly Journal of Economics,* 127(4), 1449-1496, 2009b.

Mian, A.R. and A. Sufi "The Great Recession: Lessons from Microeconomic Data," *American Economic Review,* 100(2), 51-56, 2010.

Ortalo-Magne, F. and S. Rady, "Housing Market Dynamics: On the Contribution of Income Shocks and Credit Constraints," *Review of Economic Studies,* 73(2), 459-485, 2006.

Pavlov A.D., and S. Wachter, "Subprime Lending and Real Estate Prices," *Real Estate Economics*, forthcoming, 2010.

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

Sirotic, A., H. Somerville, and V. Barberio, "Resilogic™:U.S. Residential Mortgage Loss Model Criteria," www.fitchratings.com, 2009.

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

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