Housing Policies and the Homeownership Rate*

Marco Giacoletti† and Moritz Lenel‡

October 17, 2014

Abstract

In the 1990s, the US government implemented various policies to promote homeownership among low income households. This paper studies the effect of these policies both empirically and in a quantitative portfolio choice model. We document that homeownership rates increased during the housing boom (1995-2005) among young, highly educated households with high lifetime income. However, homeownership rates did not increase for households with low lifetime income. During the housing bust (2005-2012), homeownership rates fell among both groups, especially among households with low permanent income. To understand these stylized facts, we solve the portfolio problems of a large group of heterogeneous households. In the model, households can buy or rent houses of different qualities, and homeowners can take out a mortgage which they can refinance and default on. The model matches the dynamics of homeownership rates across demographic groups from 1995 to 2012. We find that young, educated households benefited more from the implemented policies than households with permanently low income.

1 Introduction

In November 1994, President Clinton directed the U.S. Department of Housing and Urban Development to develop “a detailed strategy that recommits America to homeownership, that will add millions of new homeowners by the end of this century.” He asked “to find ways to cut the costs and the regulations involved in buying a home. I want it to be simpler, less costly.”1 In 1995, he summarized the resulting strategy as “100 specific actions that address the practical needs of

*VERY PRELIMINARY AND INCOMPLETE.
†Stanford Graduate School of Business. Email: mgiacol@stanford.edu.
‡Department of Economics, Stanford University. Email: lenel@stanford.edu.
1See Clinton (1994).
people who are trying to build their own personal version of the American dream, to help moderate income families who pay high rents but haven’t been able to save enough for a downpayment, to help lower income working families who are ready to assume the responsibilities of home ownership but held back by mortgage costs that are just out of reach, to help families who have historically been excluded from home ownership.\textsuperscript{2}

In this paper, we differentiate between two groups of low income households. First, households with a low lifetime income, as measured by a low education level. Second, households with a low current income but a high expected lifetime income, namely young, highly educated households. In the empirical section, we document that the US homeownership rates of households with a low lifetime income did barely change between 1995 and 2005. At the same time, the homeownership rate among young households with a bachelor degree or more increased by more than 20\%. During the bust, homeownership rates declined for both groups. However, the decline was larger for households with low lifetime income. In fact, homeownership rates are more dispersed across educational groups in 2014 than they were in 1995.

To better understand the economic mechanisms behind these patterns, we develop a dynamic model for the optimal homeownership decisions of households. The model is used to assess, in a quantitative exercise, the welfare effects of housing policies such as mortgage rate subsidies and reductions in loan-to-income ratios for households with different age and income prospects. In the model households are risk averse, finitely lived and face the same lifetime income profiles that we find in the data respectively for low and high education levels. Households can rent or buy houses of different qualities. Moreover, subject to credit constraints, they can finance a house buy with a mortgage which they can refinance and default on. We choose households characteristics and housing market parameters to represent a "median" local Californian housing market. Given prices, our model matches the evolution of Californian homeownership rates across demographic groups.

[FINDINGS SUMMARY HERE]

To our knowledge, this paper documents for the first time the differential dynamics of the homeownership rate across lifetime income groups. Closely related are the empirical findings of Chambers et al. (2009). The authors study the increase in the US homeownership rate between 1994 and 2005 and consider its evolution within age and income groups. They find that the homeownership rate of households under the age of 35 increased by 5.7 percentage points, more than for any other age group. We contribute to this finding, by documenting that this increase is solely driven by highly educated households with a high expected lifetime income.

We are also building on Chambers et al. (2009) in our focus on housing policies. Their paper

\textsuperscript{2See Clinton (1995b).}
studies how much of the increase in the homeownership rate between 1994 and 2005 is due to non-demographic factor such as housing policies and mortgage innovations, and how much of it is due to changes in the income and age distribution. Our goal is similar, although this paper does not focus on the overall homeownership rate, but on the policy effects within age and education groups. Their theoretical framework studies mortgage contract innovations, while we concentrate on changes in mortgage rates and credit constraints. Different from us, the authors build a general equilibrium framework, although only housing services not houses are traded and there are therefore no house prices. In our theoretical model, households can buy and sell houses of different qualities and we can therefore study the differential effects of house price movements and housing policies. Finally, we have the advantage of time that allows us to also study the housing bust that followed after 2005. In order to do so we incorporate mortgage default into our model.

Our theoretical model builds on a large and growing strand of literature which studies owner-occupied housing as a consumption as well as an investment good. Early papers in that field, such as Flavin and Yamashita (2002) and Cocco (2005), focus on the effect of homeownership on the financial portfolio composition and in particular stock holdings. In our model, we consider only two financial assets, a risk-free savings account and a mortgage, and there is no risky financial asset such as stocks.

In that sense, our framework is similar to the one studied by Chen et al. (2013), who study the interaction between countercyclical labor income risk and mortgage refinancing. There, as here, mortgage origination is subject to a loan-to-value and a loan-to-income constraint and households can refinance and default. Different from us, the authors introduce a second debt instrument, a home equity line of credit (HELOC) which they model as one period debt. We instead allow for limited installment payments on the mortgage. Given that we are particularly interested in the interaction of lifecycle effects and housing, our agents are finitely rather than infinitely lived and differ in education.

In Chen et al. (2013), housing does not enter the utility function and non-homeowners are forced to pay a fixed rent every period. We introduce housing services in the utility function and both renters and homeowners can choose from a set of different house qualities. Instead of forming expectations over a single house price, households therefore consider the dynamics of a price vector. Nevertheless, our price process is related to the one in Chen et al. (2013), which we will discuss in more detail in section 5. Chen et al. (2013) structurally estimate their model and find that it quantitatively accounts for the dynamics of household leverage before, during and after the Great Recession. Our paper contributes to their findings by studying the differential dynamics across education groups and house qualities.

In this paper, we focus on the decisions of individual households, taking the evolution of prices as given. Other authors, such as Piazzesi and Schneider (2009), Favilukis et al. (2010) and Chatterjee
and Eyigungor (2011), have incorporated housing choice in equilibrium models. While not in the scope of this paper, an equilibrium version of our current model provides a natural and interesting extension. We will come back to this in section 8.

Very much related to our work is another equilibrium model by Landvoigt et al. (2012). The authors develop an assignment model to analyze the cross sectional dynamics of prices within a local housing market. They use their model to study the San Diego housing boom in the 2000s and find that cheaper credit for poor households affected, in particular, the lower end of the housing market. This effect can explain an observed price compression across qualities during that time. We build on their analysis by introducing different house types and modeling the differential price dynamics across quality levels. Our results complement their findings, which we will discuss in more detail in section 8.

In the next section, we provide a brief summary of the housing policies that were implemented in the 1990s. Section 3 documents the historical dynamics of homeownership rates across demographic groups, both in the US as a whole and within California. We then discuss our findings in light of the implemented housing policies and highlight the need for a quantitative model to better understand our empirical findings. Section 4 develops such a model, which we then implement numerically. We discuss our results in section 8 before section 9 concludes.

2 Housing Policies and the American Dream

In his foreword to the National Homeownership Strategy, President Bill Clinton noted in 1995 that “for millions of America’s working families throughout our history, owning a home has come to symbolize the realization of the American Dream.” He as well as his predecessor, George H. W. Bush, and his successor, George W. Bush, shared the view that for low and moderate income families realizing this dream had become too difficult since the 1980s. Their administrations implemented a variety of policies that were meant to increase the overall homeownership rate by targeting, in particular, low and moderate income as well as minority households. Without describing these policies in full detail, we argue in the following that they were implemented in order to reduce mortgage rates and relax credit constraints.

Since its beginnings after the Great Depression, US housing policy has focused on supporting the residential mortgage market. First by providing better funding to mortgage originators, then by providing mortgage insurance to lenders and finally by purchasing and securitizing mortgages. For the latter, the US government created over time three different government sponsored enterprises

---

4[SOURCE]
5For an overview on the history of US housing policies see.
(GSEs) which today are called Fannie Mae, Freddie Mac and Ginnie Mae. These three enterprises created the secondary market for mortgages and were the first issuers of mortgage backed securities (MBS) in the 1970s and 1980s. Ginnie Mae was always government owned since its creation in 1968 and restricted to only purchase and securitize mortgages insured by government agencies. Different from Fannie Mae and Freddie Mac, its debt was actually backed by the government. Fannie Mae was privatized in 1968 and Freddie Mac was owned by private lenders since its creation in 1970 but both are publicly chartered. Even though securities issued by Fannie Mae and Freddie Mac were explicitly not guaranteed by the government, the low yields of the issued securities suggest that investors have always assumed an implicit governmental guarantee.

In 1992 Congress revisited the charters of Freddie Mac and Fannie Mae and further strengthened the already existing statutory purpose “to provide ongoing assistance to the secondary market for residential mortgages (including activities relating to mortgages on housing for low- and moderate-income families involving a reasonable economic return that may be less than the return earned on other activities)”\(^6\). For the first time, both institutions also had to meet quantitative goals to show that they were fostering affordable housing. For example, 30% of all purchased mortgages had to qualify as low- and moderate-income mortgages. To qualify, mortgages had to be issued to families with an income below the median income in their area. Alternatively, they had to be originated to families in underserved areas, namely areas with a low median income or a large minority share. In the years that followed, the quantitative goals were increased and Fannie Mae and Freddie Mac expanded their activities. Legally, the GSEs were only allowed to purchase and securitize mortgages that conformed to certain standards. For example, the loan-to-value (LTV) ratio had to be less than 80%, the mortgage could not exceed a certain size and borrowers had to have strong credit. In order to grow further and meet their quantitative goals, Fannie Mae and Freddie Mac started buying mortgages with higher LTV ratio as long as the borrowers would buy private mortgage insurance for the share exceeding the 80% limit. Fannie Mae’s 2002 Annual Housing Activities Report, submitted to the US Department of Housing and Urban Debelopment (HUD), celebrates the introduction of innovative mortgage products such \textit{Flex 97}, \textit{Flex 100} and \textit{eZ Access}. While Flex 97 and Flex 100 stand for mortgages with 3% and 0% downpayment, the eZ Access product “allows lenders to qualify borrowers who may have less than perfect credit and limited available funds for down payment.” Lenders were also helped by an update to Fannie Mae’s underwriting software, the Desktop Underwriter 5.2. Newly introduced risk assessment capabilities “made it possible to deliver more approval recommendations and say ‘yes’ to more borrowers.”

These and similar innovations were implemented with the goal to bring homeownership to low- and

\(^6\)Pub. L. No. 102-550, 106 Stat. 3672, 3994, 4002 (October 28, 1992). Before the change this passage read: “to provide ongoing assistance to the secondary market for residential mortgages (including mortgages securing housing for low- and moderate-income families involving a reasonable economic return to the Corporation).”
moderate income households. We cannot say to what extent the GSEs were driven by politicians’ call to increase the homeownership rate and to what extent by shareholders’ call to increase profits. Fannie Mae’s annual report 2002 was aptly titled “As The American Dream Grows, So Do We.” It is even more difficult to quantify the effects of Fannie Mae’s and Freddie Mac’s activities on actual mortgage rates. Many authors study the rate spread between conforming mortgages that satisfy the GSEs purchase requirements and jumbo mortgages that do not. Before 2006, this spread was on average around 20 basis points but is unfortunately not sufficient to consider the spread directly, since loan characteristics of conforming and jumbo mortgages differ. Addressing the differences in credit risk, a variety of studies find an effect between 7bp-24bp. Naranjo and Toevs (2002) find that GSEs also have a spill-over effect on non-conforming loan rates. In our quantitative exercise, we give the GSEs the benefit of the doubt and assume that the rates of conforming mortgages are 50bp lower than their non-conforming counterparts. We furthermore assume that mortgage products such as Flex 97 allowed more households to benefit from this implicit 50bp subsidy.

In 1995, regulators also imposed quantitative goals on banks as they tightened compliance rules for the Community Reinvestment Act (CRA). Congress passed the CRA in 1977 in order to encourage banks to meet “the credit needs of its entire community, including low- and moderate-income neighborhoods.” Before 1995 lenders had to show that they were trying to originate mortgages in low and moderate income neighborhoods, but after the regulatory changes they had to prove that they were actually making these loans. Famously, banks were now evaluated according to their “use of innovative or flexible lending practices in a safe and sound manner to address the credit needs of low- or moderate-income individuals or geographies.”

It is again difficult to quantify the effects of these regulatory changes, but [MISSING TABLE] provides some suggestive evidence. Using HMDA data we calculate the reported income of successful mortgage applicants between in 2000 and 2005. We match this individual level data to Public Use Microdata Areas (PUMAs) as defined in the 2000 Census. We then sort PUMAs according to their mean reported income according to census 2000. Agarwal et al. (2012) also find that the CRA increased risky lending by studying the variation mortgage lending around CRA examinations.

3 Homeownership Rates

Following the implementation of the above policies, the US homeownership rate did in fact increase rapidly. While in 1995 about 64.6% of all households owned their home, this number steadily

---

812 U.S. Code §2903 - “Financial institutions; evaluation”
10See 2000 FDIC §345.
Table 1: HMDA income for newly issued mortgages vs Census income of new movers who own their house. Standard errors are in square brackets.

<table>
<thead>
<tr>
<th></th>
<th>1st Quartile</th>
<th>2nd Quartile</th>
<th>3rd Quartile</th>
<th>4th Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMDA 2000</td>
<td>52,349</td>
<td>59,475</td>
<td>83,942</td>
<td>141,666</td>
</tr>
<tr>
<td></td>
<td>[188]</td>
<td>[137]</td>
<td>[199]</td>
<td>[462]</td>
</tr>
<tr>
<td>HMDA 2005</td>
<td>69,481</td>
<td>78,334</td>
<td>103,192</td>
<td>156,092</td>
</tr>
<tr>
<td></td>
<td>[115]</td>
<td>[91]</td>
<td>[140]</td>
<td>[326]</td>
</tr>
</tbody>
</table>
| HMDA 2000:2005 Incr. (%) | 32.72 | 31.71 | 22.93 | 10.18

increased until 2005, when it peaked at 69.2%. Politicians quickly attributed this increase to the implemented housing policies. When the homeownership reached its first record high in 2000, President Clinton stated that “by providing record levels of homeownership loan assistance, increasing the availability of affordable housing, providing incentives to save for a home purchase, and maintaining our commitment to fiscal discipline that has kept interest rates low, we have worked to ensure that every family has the opportunity to own their own home.” Chambers et al. (2009) try to assess how much of the increase in the homeownership rate was due to such changes, and how much of it was driven by demographic changes. They conclude that 56% to 70% of the increase was not driven by demographic changes but rather policy changes and financial innovation.

This paper poses a different question. We ask to what extent policies that targeted low- and moderate income households reached their goals. We believe that the answer to that question will also help us to understand what happened after 2005, when the homeownership rate dropped from 69.2% back to 64.9% in 2013. Before that, we discuss the different possible definitions of income groups in the next section.

### 3.1 Income Groups

To evaluate whether low income households increased their homeownership rate, we have to define which households can be characterized as low income households. The easiest approach is to sort households into groups according to their current income. But is a 25 year old who earns $40,000 as poor or rich as a 50 year old who earns the same? Figure 1 plots the estimated after-tax earnings curves of households across different education groups. We find that lifetime income can vary dramatically between a 25 year old with Bachelor degree and a 50 year old with only a high school degree, even though their current incomes may not be that different. In this paper, we define income groups therefore by expected lifetime income and we will use attained education as a proxy for it.

---

11 Authors’ calculations based on the public sample of the March Supplement of the Current Population Survey.
Figure 1: Mincer curves for different education groups in California. For each age we plot the average net of taxes labor income from the Current Population Survey over the period between 1995 and 2013 and 95% confidence intervals. Income is measured in January 2000 dollars.

[MISSING HERE: WHICH INCOME DEFINITIONS USED IN POLICY IMPLEMENTATION]

3.2 US Homeownership Rates

The homeownership rate did in fact increase in the following years, from 68.2% in 1997 to 72% in 2005. But how successful were the attempts to bring houseownership to underserved population groups? Figure 2 plots the evolution of homeowner rates from 1990 to 2012 in the US across age and education groups. We use CPS data and focus on the age and education of household heads. In the first row we depict the homeownership rates of households with highschool degree or less, across five age groups. The second row does the same for households with some college but no bachelor degree and the third row depicts ownership rates for households heads with a bachelor degree or more.

We find that during the boom years from 1997 to 2005 the homeownership rate of 25 to 35 year olds with a bachelor degree or more increased from 47.1% to 58.8%. Households in which the households head had a highschool degree or less were not more likely to be homeowners in 2005 than they were in 1994. Our findings highlight the failure of those policies that were implemented in the years following the call of President Clinton. Instead of increasing homeownership rates of working class families, young academics were enabled to buy houses earlier in their lifetime.

After 2005 houseownership rates decreased for most age-education groups. But while 25 to 35
Figure 2: Homeownership rates across age and education groups (highschool degree (HS), bachelor degree (BA)). Data from the Current Population Survey. Education level and age of household heads.
year olds were about as likely to own a house in 2012 as in 1997 with a ownership rate of 49.1%,
the homeowner rate had dropped dramatically below the 1997 level for households with less than
a highschool degree. 40.7% of households with household heads in the age group 25 to 35 had
owned a house in 1997, in 2005 this rate was basically unchanged at 40.5%, but in 2012 the rate
had dropped to 32.5%. The numbers for 35 to 45 year olds show a similar pattern, going from
59.2% in 1997 to 61.0% in 2005 and dropping to 50.0% in 2012. Not only did the reduction in
credit standards and the policies implemented as part of the National Homeownership Strategy not
increase homeownership rates for households with low education, at the end of this boom and bust
cycle these household groups were the ones that saw their homeownership rates fall far below the
levels in 1997.

3.3 California Homeownership Rates

Our quantitative exercise focuses on California, not because we think that our results do not apply
to other US states, but because various legal differences in housing and mortgage markets within
the US make inter-state comparisons difficult. Californian housing markets were among those that
were affected the most, both in the boom and the bust, and they seem a natural starting point
for our analysis. The sample size of the Current Population Survey is not large enough to analyze
homeownership rates across demogaphic groups on a state level. We therefore use data from the
decennial census in 2000 and data from the American Community Survey starting in 2005.
Table 2 depicts the homeownership rates for the age and education groups in the years 2000, 2005
and 2012. For each home ownership rate we calculate standard errors following the methodology
suggested by the census. The group of households 25 to 35 year old with a bachelor degree or
more experiences a change in the homeownership rate of 8.2 percentage points from 2000 to 2005.
The increase for households in the same age group but with a high school degree or less is a mere
1.8 percentage points but the rate then drops from 21.0% in 2005 to 14.0% in 2012.
We observe the same patterns we have found on the national level. Young, well educated households
were much more likely to own houses in 2005 than in 2000, but for other groups the change in
homeownership rates was small and often negative. After the bust, households with low education
levels were the ones witnessing the largest decrease in homeownership rates with respect to 2000.

4 A Model of the Homeownership Rate

In order to better understand the effects of the housing policies discussed above, we develop in
the following a quantitative model of the homeownership rate. We consider a large group of

\[ \text{Appendix A describes this methodology.} \]
Table 2: Homeownership rates in California in %, Decennial Census 2000, American Community Survey 2005 and 2012. Standard errors are in square brackets.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HS or less</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-29</td>
<td>19.2 [1.01]</td>
<td>21.0 [0.90]</td>
<td>(1.8)</td>
<td>(7.0)</td>
</tr>
<tr>
<td>30-34</td>
<td>31.0 [1.01]</td>
<td>30.9 [0.75]</td>
<td>(-0.2)</td>
<td>(-11.8)</td>
</tr>
<tr>
<td>35-44</td>
<td>43.4 [0.73]</td>
<td>44.8 [0.60]</td>
<td>(1.4)</td>
<td>(-13.3)</td>
</tr>
<tr>
<td>45-54</td>
<td>52.0 [0.87]</td>
<td>54.1 [0.66]</td>
<td>(2.1)</td>
<td>(-8.5)</td>
</tr>
<tr>
<td>55-64</td>
<td>62.8 [0.98]</td>
<td>59.9 [0.76]</td>
<td>(-2.9)</td>
<td>(-4.0)</td>
</tr>
<tr>
<td>&gt;65</td>
<td>68.9 [0.64]</td>
<td>67.7 [0.55]</td>
<td>(-1.2)</td>
<td>(-3.2)</td>
</tr>
<tr>
<td>&gt;HS &amp; &lt;BA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-29</td>
<td>26.4 [1.23]</td>
<td>32.1 [1.23]</td>
<td>(5.7)</td>
<td>(-11.6)</td>
</tr>
<tr>
<td>30-34</td>
<td>41.5 [1.23]</td>
<td>42.5 [0.97]</td>
<td>(1.0)</td>
<td>(-13.7)</td>
</tr>
<tr>
<td>35-44</td>
<td>56.1 [0.77]</td>
<td>56.6 [0.78]</td>
<td>(0.5)</td>
<td>(-13.8)</td>
</tr>
<tr>
<td>45-54</td>
<td>67.0 [0.76]</td>
<td>66.5 [0.72]</td>
<td>(-0.4)</td>
<td>(-7.3)</td>
</tr>
<tr>
<td>55-64</td>
<td>75.7 [0.92]</td>
<td>73.7 [0.83]</td>
<td>(-2.0)</td>
<td>(-5.6)</td>
</tr>
<tr>
<td>&gt;65</td>
<td>79.1 [0.79]</td>
<td>79.2 [0.60]</td>
<td>(0.1)</td>
<td>(2.9)</td>
</tr>
<tr>
<td>BA and more</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-29</td>
<td>22.7 [1.16]</td>
<td>30.9 [0.93]</td>
<td>(8.2)</td>
<td>(-8.9)</td>
</tr>
<tr>
<td>30-34</td>
<td>43.9 [1.20]</td>
<td>48.7 [0.98]</td>
<td>(4.9)</td>
<td>(-10.4)</td>
</tr>
<tr>
<td>35-44</td>
<td>65.4 [0.77]</td>
<td>66.9 [0.76]</td>
<td>(1.5)</td>
<td>(-8.1)</td>
</tr>
<tr>
<td>45-54</td>
<td>76.7 [0.68]</td>
<td>77.5 [0.58]</td>
<td>(0.8)</td>
<td>(-5.0)</td>
</tr>
<tr>
<td>55-64</td>
<td>82.2 [0.82]</td>
<td>82.8 [0.61]</td>
<td>(0.6)</td>
<td>(-3.9)</td>
</tr>
<tr>
<td>&gt;65</td>
<td>82.0 [0.81]</td>
<td>81.7 [0.64]</td>
<td>(-0.3)</td>
<td>(-0.7)</td>
</tr>
</tbody>
</table>
heterogenous, finitely lived households which differ in age, education, income and wealth. Wealth is composed of savings, home ownership and possibly a mortgage. Given prices and price expectations, we will solve for the optimal portfolio choice of each household, including its homeownership and mortgage decision. By aggregating over household decisions, the model will yield a homeownership rate within each demographic group. The following subsection describes the individual portfolio problem. Section 7.1 describes the numerical implementation, while section 8 discusses the results.

4.1 The Household Portfolio Problem

We consider the portfolio problem faced by a finitely-lived household that consumes a numeraire good and housing services. The household receives a stochastic income stream in terms of the numeraire. The income level changes over the life-cycle. The households can save through a risk-free bank account. Housing services are obtained by either renting or buying a house. There is a discrete and finite set of available house qualities \( j \in \{1, \ldots, N\} \), delivering housing services \( h(1), h(2), \ldots, h(N) \). The household can rent all house qualities at rental prices \( q(j) \), but can only buy houses of quality \( j \geq N \). House prices are denoted by \( p_t(j) \). When buying a house, the household can, subject to credit constraints, obtain credit in the form of a mortgage and post the house as collateral. Mortgage origination is subject to a loan-to-value and a loan-to-income constraint. In the following periods, the household has to pay interest on the outstanding mortgage and can make limited down payments. Subject to the credit constraints and a refinancing fee, the agent can refinance the mortgage or increase it in order to extract equity. The household can also sell the house or default on the mortgage. When defaulting, the households loses home ownership and is excluded from the housing market and the higher quality rental market for a limited number of periods.

4.1.1 Preferences

We follow the preference specification of Piazzesi and Schneider (2009). The household is finitely lived, with death occurring with probability 1 at age \( a_{\text{max}} \). The planning horizon for the household is therefore \( T = a_{\text{max}} - a \), where \( a \) denotes its current age. We solve the model in discrete time and one period corresponds to one year in the data. The household consumes two goods, the numeraire good, that we call \( c \), and housing services, \( h \). The consumption of \( c_t \) units of the numeraire good and \( h_t \) units of housing services yields a consumption bundle of

\[
C_t = c_t^\delta h_t^{1-\delta}.
\]

Preferences over the stochastic stream of consumption bundles are given by a recursive utility
specification following Epstein and Zin (1989):

\[ U_{a,t} = \left( C_t^{1-1/\sigma} + \beta \pi_a E_t \left[ U_{t+1}^{1-1/\gamma} \right] \right)^{-\gamma}, \]

where \( \pi_a \) is the survival probability for the household over the current year. This choice of preferences allows to disentangle risk aversion \( \gamma \) from the coefficient of inter-temporal substitution, given by \( \sigma \).

### 4.1.2 Income

In year \( t \), the household receives an exogenous after-tax income \( Y_t \), given by

\[ Y_t = L_{e,a} \tilde{Y}_t. \]

The scaling factor \( L_{e,a} \) adjusts the income level according to the household’s age \( a \) and education level \( e \), where \( e = 1 \) denotes high school education, \( e = 2 \) some college attendance and \( e = 3 \) a bachelor degree or more. \( \tilde{Y}_t \) is stochastic and its log difference \( \Delta \tilde{y}_{t+1} = \log (\tilde{Y}_{t+1}) - \log (\tilde{Y}_t) \) follows the process

\[ \Delta \tilde{y}_{t+1} = g_y + \epsilon_{t+1}^y, \]

where the intercept \( g_y \) captures the average expected growth rate of income independent of individual life-cycle effects. Empirical studies usually find that the unit root hypothesis cannot be rejected for the personal income process and our specification builds on these findings.\(^{14}\)

### 4.1.3 Savings and Debt

Financial markets are incomplete such that the household cannot fully insure against its income risk. There are only two financial assets, a risk-free savings account and a mortgage. Savings in period \( t \) are denoted \( \theta_t \) and yield \( (1 + r)\theta_t \) in period \( t + 1 \). Only a homeowner can borrow money by taking out a mortgage. The house serves as collateral and the amount a household can borrow depends on the house price. The maximum mortgage size for an owner of house type \( j \) is given by

\[ \bar{m}_t = \bar{m}(Y_t, p_t(j)) = \min (\chi_{LTV} p_t(j), \chi_{LTI} Y_t), \]

with \( \chi_{LTV} \in [0, 1] \) and \( \chi_{LTI} \in [0, 1] \). \( p_t(j) \) denotes the current price for house type \( j \) which delivers services \( h_j \). The parameters \( \chi_{LTV} \) and \( \chi_{LTI} \) define respectively the maximum loan-to-value (LTV) and the maximum loan-to-income (LTI) ratio that the bank is willing to grant to the household.

\(^{14}\)See [SOURCES].
When originated at time $t$, the mortgage $m_t$ has to be smaller or equal than $\bar{m}_t$. The households is charged an origination fee $\kappa$ when taking out a mortgage.

After taking out a mortgage, the households enters the next period as a debtor. Unless the household defaults, it has to pay interest $\rho_{t+1}m_t$ in period $t+1$. The mortgage rate $\rho$ is a function of age and income in order to account for the fact that young and low-income households have little or bad credit history and are charged a higher rate. In particular,

$$
\rho_t = \rho(a_t, Y_t) = \begin{cases} 
\rho & \text{if } (a_t < \bar{a} \text{ or } Y_t < \bar{Y}_t) \\
\overline{\rho} & \text{if } (a_t \geq \bar{a} \text{ and } Y_t \geq \bar{Y}_t) 
\end{cases}.
$$

Mortgage interest payments are tax-deductible in the US and the household therefore receives a tax return $Z_t$ in our model. $Z_t$ is a function of the marginal tax rate $\zeta_t$ and therefore the level of the after-tax income $Y_t$. It is also a function of the pre-tax mortgage payment $\rho tm_{t-1}$, since the mortgage payment has to exceed the standard deduction $\bar{D}$ in order to reduce the tax bill of the homeowner. We therefore define

$$
Z_t = Z(a_{t+1}, Y_{t+1}, m_t) = \begin{cases} 
0 & \text{if } \rho tm_{t-1} \leq \bar{D}, \\
\zeta_t (\rho tm_{t-1} - \bar{D}) & \text{if } \rho tm_{t-1} > \bar{D}.
\end{cases}
$$

A detailed description of the calibration of $\zeta_t = \zeta(Y_{t+1})$ is given in Section 6, which discusses the model calibration.

The household can decide to repay a share of the outstanding mortgage. Any repayment which is larger than a fraction $\nu$ of the house value is subject to a prepayment penalty $\kappa$, which is equal to the origination fee. $\kappa$ is also charged as a refinancing fee when the households increases the size of its mortgage in order to extract equity. Equity extraction is subject to the same LTV and LTI constraints which were specified in Equation 1.

When the household decides to sell its house, it needs to fully repay the outstanding mortgage, including the required interest payments. The household also has to pay a fraction $\eta$ of the sales price in fees and real estate commissions.

Finally, the households can decide to default on the mortgage. In that case its owed interest and its outstanding mortgage are immediately reduced to zero and the household loses its house. We model the mortgage contract as a non-recourse loan, such that the bank cannot obtain any payment out of the household’s savings or income. We choose this institutional feature to model Californian housing markets, where state law mostly prohibits deficiency judgements. Defaulting households cannot access the housing market for a certain amount of time $\tau_{excl}$, which captures the fact that a default has a strong negative impact on a household’s credit score. We also assume that the
household can only rent low quality houses during that time.

4.1.4 Housing Qualities

As noted above, the household can choose its housing consumption from a discrete and finite set of available house qualities \( j \in \{1, ..., N\} \). Each housing quality \( j \) delivers housing services \( h(j) \). All of the house qualities are available for rent, but only houses of quality \( j \geq N \) are for sale. We denote the rental price of quality \( j \) by \( q(j) \) and the sale price by \( p_t(j) \). Note that only sale prices are assumed to be time varying. In order to make sure that every agent can afford some housing, we set \( q(1) = 0 \). In our numerical implementation, we set \( h(1) \) low enough such that households are unwilling to choose this house quality.

Even though the present model aims to capture the key elements of a household’s homeownership choice, there is a variety of factors that are not taken into account. For example, in our framework, agents only move in order to change their housing consumption. In the data, households are often required to relocate for work or family related reasons. We find that young and highly educated households are particularly likely to move for job related reasons, making homeownership less attractive to them. Another important simplification within our model is the fact that households can only invest in a risk-free bank account and a house. In the data, households can also invest in other risky assets, like stocks or corporate bonds. Our simplification is less restrictive for poor households and households with low education who show little to no stock market participation in the data. On the other hand, rich and highly educated households exhibit larger stock market participation in the data. Our model might overstate the attractiveness of housing for this group, since they choose from a large array of other risky assets in the data. Finally, households may value owning a home higher than renting a home, either because they derive utility from achieving the “American Dream” or because they feel more secure in an owned house. How much they do so, could again depend on social status and education. To capture these and other factors that might affect the trade-off between housing and renting, we introduce an education dependent *warm glow* of housing. In particular, a homeowner of education level \( e \) derives housing services \((1 + \omega_e)h(j)\) from its house of quality \( j \), while a renter only derives \( h(j) \).

4.1.5 The Optimization Problem

We now describe the intertemporal optimization problem faced by an household of age \( a \) and education level \( e \). Its current state is defined by its income \( y \), last period’s savings \( \theta \), its homeownership status and the size of its outstanding mortgage as well as current house prices \( p \). We denote the

---

\(^{15}\text{[SOURCE]}\)

\(^{16}\text{[SOURCE]}\) find that stock market participation is increasing in wealth and education.
current stochastic states of income and prices as \( s = \{y, p\} \). The expectation operator \( \mathbb{E} \) yields expectations over \( s' \) conditional on \( s \). In the next section we will discuss in more detail the assumed price process which governs the expectations over \( s' \).

The decision problem is solved recursively and for notational simplicity we define three different value functions: \( V_0 \) denotes the value function of a household not owning a house. \( V_D \) denotes the value function of household that is currently excluded from the housing market due to default. \( V (\cdot, j) \) denotes the value function of an owner of a house of quality \( j \).

**Default** For simplicity, consider first the household that is currently in default and that is therefore unable to buy a house. Let \( \tau_D \leq \bar{\tau}_D \) denote the the number of years the household has already spent in default. Its current wealth is summarized by the sum of the gross return on past savings, \( \theta (1 + r) \), and current income, \( y \). The household has to optimally allocate its current wealth to consumption, \( c \), rental quality, \( h(k) \), and savings, \( \theta' \).

\[
V_D (\theta, \tau_D, s, a) = \max_{\theta', k} \left\{ \begin{array}{ll}
\left( C^\psi + \beta \pi_a \mathbb{E} \left[ V_D (\theta', \tau'_D, s', a')^{1-\gamma} \right] \right)^{\frac{1}{1-\gamma}} & \text{if } \tau_D < \bar{\tau}_D, \ a < \bar{a} \\
\left( C^\psi + \beta \pi_a \mathbb{E} \left[ V_0 (\theta', s', a')^{1-\gamma} \right] \right)^{\frac{1}{1-\gamma}} & \text{if } \tau_D = \bar{\tau}_D, \ a < \bar{a} \\
C & \text{if } a = \bar{a}
\end{array} \right.
\]

subject to
\[
\begin{align*}
\theta' &= y + (1 + r) \theta - c - q(k), \\
C &= c^\delta h(k)^{1-\delta} \\
a' &= a + 1, \\
\tau' &= \tau + 1, \\
k &\in K_D \subset K,
\end{align*}
\]

where we define \( \psi = 1 - 1/\sigma \).

**Renter** Now consider an agent that is not in default anymore but does not own house. This household can either remain a renter or become a homeowner. We denote with \( V_{00} \) the value function of a renting household that remains a renter, and with \( V_{0j} \) the value function of a renter that becomes a homeowner.

\[
V_0 (\theta, s, a) = \max \left[ V_{00} (\theta, s, a), \ V_{0j} (\theta, s, a) \right]
\]
where

\[ V_{00} (\theta, \tau_D, s, a) = \max_{\theta', j} \left\{ \left( C^\psi + \beta \pi_a E \left[ V_0 (\theta', s', a')^{1-\gamma} \right]^{\frac{\psi}{\gamma}} \right)^\frac{1}{\psi} \right\} \]

subject to

\[ \theta' = y + (1 + r)\theta - c - q(j), \]
\[ C = e^{\delta h(j)} (1-\delta) \]
\[ a' = a + 1, \]
\[ j \in \{1, \ldots, N\} \]

and

\[ V_{0j} (\theta, \tau_D, s, a) = \max_{\theta', j', m'} \left\{ \left( C^\psi + \beta \pi_a E \left[ V (\theta', j', m', s', a')^{1-\gamma} \right]^{\frac{\psi}{\gamma}} \right)^\frac{1}{\psi} \right\} \]

subject to

\[ \theta' = y + (1 + r)\theta - c - p(j') + m' \geq 0, \]
\[ C = e^{\delta ((1 + \omega) h(j')) (1-\delta)} \]
\[ a' = a + 1, \]
\[ m' \leq \bar{m} (y, p(j')) \]
\[ j \in \{N, \ldots, N\} \]

**Homeowner** Finally, consider a household that enters the current period with homeownership and possibly a mortgage. \( V_{jj} \) denotes the value function of an agent that keeps its current house, \( V_{jj'} \) the value function of a household buying a new house, \( V_{j0} \) the value function of a household selling his house, and finally \( V_{jD} \) the value function of a defaulting household.

\[ V(\theta, j, m, s, a) = \max [V_{jj}, V_{jj'}, V_{j0}, V_{jD} | \theta, j, m, s, a] \] (6)

The maximization problem of \( V_{jj} \) is given by
\[
V_{jj}(\theta, j, m, s, a) = \max_{\theta', m'} \begin{cases}
\displaystyle \left( C^\psi + \beta \pi_a \mathbb{E} \left[ V(\theta', j, m', s', a')^{1-\gamma} \right] \right)^\frac{1}{\gamma} \quad & \text{if } a < \bar{a} \\
C \quad & \text{if } a = \bar{a}
\end{cases}
\]
subject to
\[
\begin{align*}
\theta' &= y + (1 + r)\theta - c - (1 + \rho)m + Z + m' - \kappa \mathbb{1}_{ref}, \\
C &= c^\delta ((1 + \omega)h(j))^{1-\delta} \\
a' &= a + 1, \\
m' &\leq \max [m, \bar{m}(y, p(j))] \\
\mathbb{1}_{ref} &= 1 \text{ if } (m' > m \text{ or } m' < (1 - \nu)m).
\end{align*}
\]

Notice that here and in the following \( \rho \) and \( Z \) are functions of age, income and mortgage size.

The maximization problem of \( V_{jj'} \) is given by
\[
V_{jj'}(\theta, j, m, s, a) = \max_{\theta', j', m'} \begin{cases}
\displaystyle \left( C^\psi + \beta \pi_a \mathbb{E} \left[ V(\theta', j', m', s', a')^{1-\gamma} \right] \right)^\frac{1}{\gamma} \quad & \text{if } a < \bar{a} \\
C \quad & \text{if } a = \bar{a}
\end{cases}
\]
subject to
\[
\begin{align*}
\theta' &= y + (1 + r)\theta - c - (1 + \rho)m + Z + m' + (1 - \eta)p(j) - p(j') \geq 0, \\
C &= c^\delta ((1 + \omega)h(j'))^{1-\delta} \\
a' &= a + 1, \\
m' &\leq \bar{m}(y, p(j')).
\end{align*}
\]

The maximization problem of \( V_{j0} \) is given by
\[
V_{j0}(\theta, j, m, s, a) = \max_{\theta', s', m'} \begin{cases}
\displaystyle \left( C^\psi + \beta \pi_a \mathbb{E} \left[ V_0(\theta', s', a')^{1-\gamma} \right] \right)^\frac{1}{\gamma} \quad & \text{if } a < \bar{a} \\
C \quad & \text{if } a = \bar{a}
\end{cases}
\]
subject to
\[
\begin{align*}
\theta' &= y + (1 + r)\theta - c - (1 + \rho)m + Z + (1 - \eta)p(j) \geq 0, \\
C &= c^\delta h(j)^{1-\delta} \\
a' &= a + 1.
\end{align*}
\]
And finally, the maximization problem of $V_{jD}$ is given by

$$V_{jD}(\theta, j, m, s, a) = \max_{\theta', k} \begin{cases} 
C^{\psi} + \beta \pi_a \mathbb{E} \left[ V_{D}(\theta', 1, s', a')^{1-\gamma} \right]^{1-\gamma} \quad & \text{if } a < \bar{a} \\
C \quad & \text{if } a = \bar{a} 
\end{cases}$$

subject to

$$\begin{align*}
\theta' &= y + (1 + r) \theta - c \geq 0, \\
C &= c^\delta h(k)^{1-\delta} \\
a' &= a + 1, \\
k \in K_D \subset K.
\end{align*}$$

(10)

5 The Dynamics of House Prices

5.1 Data

We collect monthly house price data from Zillow at the ZIP code level for the state of California over the period between January 1998 and March 2014.\textsuperscript{17} Zillow provides separate price series for houses with different numbers of bedrooms. Within each ZIP code, we collect the series of prices for two, three and four bedroom houses. We deflate nominal house prices using core CPI inflation and obtain real house prices in terms of January 2000 dollars. We then match available ZIP codes to the corresponding counties using the geo-correspondence files provided by the Missouri Census Data Center.\textsuperscript{18} For 36 out of the 58 Californian counties we are able to collect prices for ZIP codes covering at least 50% of the county population according to the 2000 decennial census. As we show in appendix B, excluded counties have low population, low population density and low housing unit density.

Two important features of our model are that house investments are lumpy and that households can rent or buy houses of different qualities. In practice, it is difficult to determine the level of housing services provided by a specific housing unit. Housing quality is potentially determined by a large number of house characteristics which cannot be easily observed or measured. In our analysis, we assume that within a specific county, differences in housing services can be summarized by two easily observable housing features: location and size. We use the data from Zillow to build for each of the 36 selected counties a cross section of house prices that differ across geographies (ZIP codes) and size (number of bedrooms). According to our assumption, the combination of ZIP code and

\textsuperscript{17}Accessed through \url{http://www.zillow.com/blog/research/data/}.

\textsuperscript{18}Accessed through \url{http://mcdc.missouri.edu/websas/geocorr12.html}. 
size identifies a specific house quality within each county. We furthermore assume that, within each county, ratios of 1998 average prices are proportional to the ratios of provided housing services. Note that, even if this methodology was providing accurate estimates of housing service ratios in 1998, these ratios might still change over time. We can think for example of industrial ZIP codes which gentrify, become residential areas and progressively offer higher housing services. A way to address our concerns about relative fluctuations in housing services across geographies is to aggregate ZIP codes in groups with similar price characteristics. For each house size in each county, we can form ZIP code "portfolios" or "tiers" based on quantiles of the price distribution across ZIP codes in the specific county. In particular, for each house size we form a low, a medium and a top tier, corresponding respectively to the lowest 33% of prices, to prices between the 33th and the 66th quantile and finally to prices above the 66th quantile. This assignment will be kept fixed for the entire period under analysis, and the house price for the specific tier will be calculated as the median house price across the ZIP codes assigned to that tier. Note that this approach is relatively robust, in the sense that it can be distorted only by systematic perturbations in the cross sectional ordering of house qualities across different ZIP codes. In the section on model calibration we will discuss the methodology that we use to calculate housing services and we will analyze how our assumptions compare to the evidence that we observe in rental data. The following plot shows average quarterly prices for three different house types. Prices in each tier are equal to the median across the 36 California counties that we use in our analysis. The left hand side panel reports real house prices in January 2000 dollars, while the right hand side panel shows prices normalized by their average level in the first quarter of 2000.

Compression in the cross section of prices between 2000 and 2005 is evident. The low tier of 2 bedroom houses nearly triples its price between 2000 and 2005, while the top tier of 4 bedroom houses experiences growth of "only" 80%. In the following section we will propose a statistical model for house price expectations. In order to estimate this model we need a longer time series than the ones published by Zillow. We therefore extend our historical data using other sources. The Federal Reserve Bank of St. Louis provides price time series of houses in different price tiers for several US cities. Using data from different tiers in Los Angeles, San Francisco and San Diego we are able to extend our time series backwards up to the first quarter of 1989. We can then go back to the first quarter of 1953 by using the Real Home Price Index by Robert J. Shiller. The merged quarterly time series is reported in the following figure.

5.2 Basic Model of Price Dynamics: Trend and Deviations

In order to solve their dynamic portfolio problem, households need to form expectations on the future evolution of house prices and on their volatility. Time series modeling of house prices and
Figure 3: The left hand side panel reports median house prices across the selected sample of 36 California counties. All figures are in January 2000 Dollars. The right hand side panel shows price series normalized by their average price in the first quarter of 2000.

house returns is challenging. As appears clearly from the figures in the previous section, these series are extremely persistent and show wide swings. The standard assumption used in most of the literature is that log house prices follow a unit root process and that there is cointegration relationship between house prices series and other macroeconomic fundamentals. However, Haurin, De Jong and Zhang (2013), reject the presence of a unit root and instead find that house price series are trend stationary with structural breaks or time variations in the trend parameters. They work with the long time series of real house prices provided by Shiller and a panel of MSA and state level real house prices. It appears that a key determinant of the difference between their results and previous research is the use of a longer sample, which extends through 2011 and therefore includes the most recent housing boom and bust. Since our quantitative exercise focuses on the period between 2000 and 2012, we take the evidence in Haurin, De Jong and Zhang seriously and rely on a trend stationary specification. However, household in our model are allowed to choose among different house types. This means that we also need to characterize expectations for an entire cross section of prices.

From our data we can obtain prices for different house sizes and tiers in each county. We can then calculate log prices for the $N$ different house types in the cross section as $p(n)_t = \log(P(n)_t)$ with $n \in \{1, ..., N\}$. The ordering of house types is increasing in the housing services provided. The house with apex 1 is the one offering the lowest level of housing services, while the one with apex
Figure 4: The figure shows a merged time series for quarterly house prices. Data sources are the Real Home Price Index Robert J. Shiller website (Q1 1953: Q1 1989), the Federal Reserve Bank of St. Louis (Q1 1989:Q1 1998, average of low tiers across available California cities) and Zillow (Q1 1998:Q3 2014, cheapest house type). Prices are normalized to match the level of the time series from Zillow in Q1 1998.
$N$ is the one offering the highest level. We start our analysis by focusing on a single price. We pick the house providing the lowest level of housing services and decompose its log price in a trend and a deviation component as follows:

$$p(1)_t = trend(1)_t + dev(1)_t$$

The dynamics of trend and deviation are:

$$trend(1)_t = \alpha_1 + g_{1,h} t$$
$$dev(1)_t = \phi_d dev(1)_{t-1} + x_t$$
$$x_t = \phi_x x_{t-1} + \sigma \epsilon_t$$

Where $\epsilon_t$ is a mean zero iid residual. When estimating the model we will assume $\epsilon_t \sim N(0,1)$. Note that the unconditional mean of the deviation component is constrained to be zero so that the long run expectation of the deviation is equal to zero and long run expected prices coincide with the trend. The deviation is modeled with an autoregressive component and a persistent shock $x_t$. This modeling choice allows for both long run mean reversion and short time momentum, as we show in the figure below. We plot the autocovariance function of the deviation component, which represents the covariance between the current shock $\epsilon_0$ and future deviations up to 50 periods ahead. Covariances are calculated for different values of $\phi_x$ and $\phi_d$, while $\sigma$ is set to 1. We can see that autocovariances are increasing for the first few periods, so that the impact of a current shock gets amplified in the short run. Then, mean reversion kicks in, making the autocovariance function decrease smoothly. The figure also shows that increasing $\phi_x$ when $\phi_d$ is fixed amplifies the rise in short term autocovariance and therefore in momentum. On the other hand, by decreasing $\phi_d$ while keeping $\phi_x$ fixed we push down and compress the entire profile of the autocovariance function, weakening momentum and making mean reversion faster.

Our analysis so far has been focused on decomposing a single log price in two factors. The next step is to model the cross section of prices. We impose that all house prices have the same trend growth, coinciding with the one estimated from the cheapest house. As we will show later, differences in trend growth across different house types are moderate, so this assumption does not set us too far apart from what is in the data. We then conjecture that the deviation factor extracted from the log price of the cheapest house is able to span a large fraction of the variation in the deviation components from the entire cross section of log house prices. We can then write the price of any house type $n$ as:
\[ p(n)_t = w^n_0 + w^n_dev_t + trend(1)_t + u^n_t \]

Where for the cheapest house \((n = 1)\) we have \(w^n_0 = 0\), \(w^n_1 = 1\) and residuals \(u^n_1\) are all identically equal to zero. The residuals \(u^n_t\) will in general not be equal to zero for \(n > 1\). However, we will show in the following sections that cross sectional pricing errors are very small for all the California counties under analysis. If we neglect the pricing errors, we obtain a parsimonious representation of the cross section of house prices. Note that this representation is potentially able to capture the compression and expansion of the spreads across different prices which we documented in the previous section. In fact, if the deviation factor loads more highly on cheapest houses, then house pricing cycles will be more extreme for these segments of the market.

### 5.3 Estimation and Down Weighting

We bring our model to the time series and cross section of house prices in each one of the 36 California counties selected in the previous section. For each county we build a time series of average quarterly prices for the cheapest house over the period between the first quarter of 1953 and the first quarter of 2014. The time series that we build in each county for different house types based on Zillow data go back only to the first quarter of 1998. Thus, as discussed in the previous
sections, we need to merge Zillow and Shiller datasets. This is of course problematic. One of the advantages of our data between 1998 and 2014 is the availability of a cross section of house prices for each county. In order to get a longer time series, we are forced to use California level data for low tier houses between 1989 and 1998 and then a national house price index for the period between 1953 and 1989. This means that the first part of the time series (up to 1998) will be identical for different counties. Moreover, until 1989 we will not be using data specific to California. On the other hand, fitting the model just to Zillow data will deliver parameter estimates which are completely driven by the observations from the last housing boom and bust. In appendix C we show how parameter estimates change when we exclude the sample between 1953 and 1989. Since the period before the 1980s is extremely quiet for house prices, while the one in the 2000s is extremely turbulent, sample choice clearly matters. Nonetheless, we believe that an interesting way to implement the model is to estimate parameters each quarter, using only price history available up to that point in time. In fact, when households form their expectations at a specific point in time, they can only observe house prices up to that point. Thus, even after considering all the caveats listed above, this exercise is still extremely interesting, since it allows to investigate the evolution of price expectations in real time, using the same price information available at that point in time to households. Haurin, De Jong and Zhang find structural breaks in the trend component. We do not have an easy way to include regime switching in our model. However, there is a simple naive way to deal with trend parameters drifting over time. We introduce in the recursive estimation of the model exponential down weighting of past observations. Under very restrictive assumptions (see for example McCulloch 2005), exponential discounting of past information based on a constant gain coefficient arises in models with time varying parameters. The coefficient \( \lambda \) is the constant gain coefficient, capturing the extent to which past information is down weighted. For \( \lambda = 0 \) we have no down weighting and the model is just estimated with OLS over an expanding window. As \( \lambda \) increases, down weighting becomes stronger.

When the trend component is estimated recursively, deviations can be calculated in two different ways. A first approach is to re-evaluate the entire history of deviations each time the model is estimated. A second one is to keep track of the deviations calculated in the past based on previous estimates of the trend coefficients. The current deviation is then calculated using contemporaneous parameter estimates. We believe that the second approach is more consistent with the behavior of an individual who is dealing with structural breaks in trend growth and we therefore do not update the deviations vector at each estimation.

In the following figures we show the cross sectional distribution of model parameters across California counties when the model is estimated respectively with data up to Q1 2000, up to Q1 2005 and up to Q1 2012. Even in the absence of down weighting (\( \lambda = 0 \)) we see relevant changes in the
distribution of parameters over time. In 2000 we estimate trend growth very close to zero in all counties. The median value of the autoregressive coefficient $\phi_d$ is slightly below 0.95, while the one for the momentum coefficient $\phi_x$ is close to 0.6. In 2005 estimated trend growth is larger, with median growth above 0.4 % per year. Moreover, estimates of $\phi_d$ and $\phi_x$ have increased massively. In particular, the autoregressive coefficient is larger than one for most counties. These results suggest that when $\lambda = 0$ the model is slow in raising the expected growth rate for the trend component of prices. Thus, it attributes most of the huge increase in prices between 2000 and 2005 to the deviation component. However, since the deviation from the trend is large and long lived, we end up with an explosive autoregressive coefficient and no mean reversion. This effect is however completely counteracted in 2012, when the distribution of $\phi_d$ shifts to values slightly lower than in 2000, while the trend growth rate raises even more reaching a median value larger than 0.9 % per year. The momentum component $\phi_x$ increases even more with a median value above 0.9. Given a median value of $\phi_d$ around 0.92, this delivers high momentum for expected house price fluctuations, as discussed in the previous section. The increase in the trend growth coefficient between 2005 and 2012 might appear counter-intuitive at first. However, this can be explained when looking at the change in the values of the linear trend intercepts $\hat{\alpha}$. We can see that the distribution of the intercepts slightly shifts to the left from 2000 to 2005 and again from 2005 to 2012. In order to keep up with the boom, the trend line has to rotate, lowering the intercept and increasing the slope. In 2005, most of the price history contains low pre-boom prices. The trend line starts rotating, but most of its fit is determined by the low price history. In 2012, history contains the entire boom-bust episode. The part of the sample with high prices is now larger and therefore the line has to rotate even more to accommodate the new data. This effect will disappear when we apply $\lambda > 0$, since this allow to downweight price history. When we introduce down weighting with $\lambda$ greater than 0, trend growth estimates become more reactive to price growth. This delivers higher estimated trend growth in 2005 and lower autoregressive coefficients for the deviation component. The larger $\lambda$, the more the boom in house prices is attributed to a structural change in the trend component. For $\lambda = 0.05$ estimates for $g_h$ swing widely. Annual trend growth has a median value close to -1 % in 2000, while in 2005 all estimates are larger than 0 and the median is larger than 4 %. In 2012 the median is close to 0, but there are counties for which estimated trend growth is smaller than -2 %. Nonetheless, momentum coefficients are increasing with the passing of time for all values of $\lambda$, even if estimates are slightly smaller when the downweighting is larger. Thus, perceived momentum in deviations becomes more and more important in later years, as an effect of the boom-bust cycle.

So far we have been focusing on the time series model for the cheapest house. However, in order to complete our framework, we need to deal with the entire cross section of prices in each county. As a first step, we need to decide which are the house types that we want to introduce in the model. We select among our house type indexes the 2 bedroom low tier house, which is perfectly spanned
Figure 6: Cumulative distribution functions across counties for parameter estimates from the recursively estimated time series model with $\lambda = 0$. Each county is weighted by its population according to the 2000 Decennial Census.

Figure 7: Cumulative distribution functions across counties for parameter estimates from the recursively estimated time series model with $\lambda = 0.025$. Each county is weighted by its population according to the 2000 Decennial Census.
by trend and deviation, the series for the 2 bedroom mid tier, the 3 bedroom mid tier house, the 4 bedroom mid tier house and the 4 bedroom top tier house. This leaves us with a cross section of 5 house types for each quarter and in each county. We then need to merge our data with the Shiller database. The price tiers based on Zillow data only extend back to Q1 1998. As discussed above, the Federal Reserve Bank of St. Louis provides low, medium and top tier price series for the cities of Los Angeles, San Diego and San Francisco. We build California level low, mid and top tier indices by averaging across the three cities. We then match the low tier index to the series of low and mid tier houses with 2 bedrooms, the mid tier to the 3 bedroom house and the top tier to the two 4 bedroom houses series. The trend component is restricted to be identical across all house types, as suggested in the previous section. While we do not formally test this restriction, we show in appendix C that trend coefficients estimated for higher quality houses are similar to the ones for lower quality houses. We then calculate for each house type in each county the deviation from the trend. Finally, county by county we project the deviation from the 2 bedroom low tier house on the deviations for other house types using as starting point Q1 1989, which is the first quarter for which house prices specifically from California are available. In the table below we report both loadings estimated over the full sample (Q1 1989 through Q1 2014) and loadings based on recursive estimation. All deviations are calculated for the case with $\lambda = 0$. We can clearly see how deviation loadings are decreasing in house quality. This result holds for both the full

Figure 8: Cumulative distribution functions across counties for parameter estimates from the recursively estimated time series model with $\lambda = 0.05$. Each county is weighted by its population according to the 2000 Decennial Census.
sample and the recursive estimates, with the exception of year 2000, when loadings are pretty
similar across house types, even if 4 bedroom houses have loadings sensibly smaller than one. The
deviation captures house price cycles. Thus, our stylized result is that houses of higher quality are
less exposed to cycles. Moreover, deviation loadings decrease for all house types over time (except
for the lowest quality house, where the loading is fixed to be equal to 1). This is a consequence
of the price compression across house types that has taken place during the period between 2000
and 2012. In fact, when $\lambda = 0$ most of the variation in prices during the last boom-bust episode is
attributed to the deviation component. The deviation is extracted from the price of the lowest
quality house and higher quality houses experienced a less extreme cycle. In the figure below we
compare for the different house type tiers median prices across counties against the median fitted
house prices based on the projection of trend and deviation. The fitted series are close to the actual
prices and capture their general tendencies, even if the fitted series seem to amplify the boom-bust
episode for higher quality houses. Mean absolute percentage errors calculated over the period from
Q1 1998 through Q1 2014 for the recursively estimated projections are equal to 3.53% of house
prices for the 2 bedroom mid tier house, 4.36% for the 3 bedrooms mid tier house, 4.44% for the 4
bedroom mid tier house and 6.50% for the top tier 4 bedroom house. It is clear that the model
prices better lower quality houses. Moreover, pricing errors appear to be persistent. While we
realize that our model is extremely stylized and has many limitations, we still believe it is able to
provide a satisfactory summary of the heterogeneity in house price fluctuations across different
tiers.

Table 3: Median estimates across counties of the loading of the deviation component for different
house types.

<table>
<thead>
<tr>
<th></th>
<th>2 bdm Low Tier</th>
<th>2 bdm Mid Tier</th>
<th>3 bdm Mid Tier</th>
<th>4 bdm Mid Tier</th>
<th>4 bdm Top Tier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sample: from Q1 1989 through Q1 2014</td>
<td>1.0000</td>
<td>0.8639</td>
<td>0.7931</td>
<td>0.7430</td>
<td>0.5770</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from Q1 1989 through Q1 2000</td>
<td>1.0000</td>
<td>1.0038</td>
<td>1.0004</td>
<td>0.9750</td>
<td>0.9721</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from Q1 1989 through Q1 2005</td>
<td>1.0000</td>
<td>0.9436</td>
<td>0.8636</td>
<td>0.8316</td>
<td>0.7440</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from Q1 1989 through Q1 2012</td>
<td>1.0000</td>
<td>0.8424</td>
<td>0.8010</td>
<td>0.7541</td>
<td>0.5921</td>
</tr>
</tbody>
</table>
6 Model Calibration

Our baseline calibration targets the data moments of Californian housing markets in the year 2000. The previous section described our estimation of a price process that is representative of expected local price dynamics in 2000. This section describes the calibration of other model parameters.

6.1 Preference Parameters

The discount factor is chosen to be $\beta = 0.9725$, in line with our calibration of the risk free rate which is discussed below. Risk aversion $\gamma$ is set to 5 and the IES to $\sigma = 0.5$. In Section 7.1 we discuss the effects of alternative preference specifications. The elasticity between housing and non housing consumption is chosen to be $\delta = 0.7$, which allows us to match the consumption expenditure shares on housing services in the Consumer Expenditure Survey of the Bureau of Labor Statistics. The education-dependent warm glow parameter $\omega_e$ will be chosen to match the homeownership rates in 2000. We will discuss this in more detail in the next section.
Table 4: Baseline Calibration, Year 2000; Preferences and Income Process.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preferences</strong></td>
<td></td>
</tr>
<tr>
<td>Discount factor $\beta$</td>
<td>0.9725</td>
</tr>
<tr>
<td>EZ Gamma $\gamma$</td>
<td>5</td>
</tr>
<tr>
<td>EZ Sigma $\sigma$</td>
<td>0.5</td>
</tr>
<tr>
<td>Cons. Shares $\delta$</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>Income Process</strong></td>
<td></td>
</tr>
<tr>
<td>Base growth $g_y$</td>
<td>0.02</td>
</tr>
<tr>
<td>Shock variance $\sigma_y$</td>
<td>0.15</td>
</tr>
<tr>
<td>Cov with housing shock</td>
<td>0</td>
</tr>
<tr>
<td>Lifecycle component $L_{e,a}$</td>
<td>CPS estimates</td>
</tr>
<tr>
<td>Base income $\bar{y}$</td>
<td>Census data</td>
</tr>
</tbody>
</table>

6.2 Income Dynamics and Income Taxes

In the baseline calibration income is uncorrelated with house price shocks, such that $\sigma_{\nu \epsilon} = 0$. The base log income is growing with the economy, such that for now $g_y = 0.02$.

As discussed in previous sections, we restrict the standard labor model which features permanent and transitory shocks by just allowing a permanent shock. The variance of the income shock is taken from Saporta (2014). He estimates $\text{var}(\nu) = 0.0225$ which is similar to Meghir and Pistaferri (2004) who find a $\text{var}(\nu)$ of 0.031 pooled over education groups. However, both authors include in their models transitory shocks as well.

Having defined the properties of the stochastic income component $\tilde{y}_t$, we have to also determine the life-cycle component $L_{e,a}$. To that extent we fit a polynomial to the Mincer curves depicted in Figure 1. [EXTEND DESCRIPTION]

[MOVE TO NUMERICAL EXERCISE] Income levels in the Census are gross of taxes. The Current Population Survey (CPS) reports for each household after credit income tax payments for both federal and state taxes, along with the labor income of the household and its financial income. For California households, we compute total income taxes as the sum of federal and state taxes and then subtract from them 33% of financial income, to match the California tax rate for these income sources. We are this way able to get an approximate estimate of labor income taxes for each household. We then use a polynomial to approximate the relationship between labor income
Table 5: Baseline Calibration, Year 2000; Housing and Financial Markets.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>House Price Process</strong></td>
<td></td>
</tr>
<tr>
<td>Trend intercept ( \alpha_{1,2000} )</td>
<td>4.8260</td>
</tr>
<tr>
<td>Trend growth (next years 10) ( g_h )</td>
<td>0.00016</td>
</tr>
<tr>
<td>Trend growth (years 11 and after) ( g_h )</td>
<td>0.02</td>
</tr>
<tr>
<td>AR(1) dev ( \phi_{dev} )</td>
<td></td>
</tr>
<tr>
<td>Momentum dev ( \phi_x )</td>
<td></td>
</tr>
<tr>
<td>Shock variance ( \sigma_\epsilon )</td>
<td>0.0210</td>
</tr>
<tr>
<td>Trend Price Ratios</td>
<td>1.0, 1.2853, 1.7478, 2.0988, 4.5755</td>
</tr>
<tr>
<td>Proj. Deviations</td>
<td>1.0, 1.0038, 1.0004, 0.9750, 0.9721</td>
</tr>
<tr>
<td><strong>Other Housing Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Rental dividends ( q )</td>
<td>0.065 ( \times ) ( p_{2000} )</td>
</tr>
<tr>
<td>Maintenance costs ( m )</td>
<td>0.02</td>
</tr>
<tr>
<td>Selling costs ( \eta )</td>
<td>0.08</td>
</tr>
<tr>
<td>Deduct. Thresh. ($1000) ( \hat{D} )</td>
<td>5</td>
</tr>
<tr>
<td><strong>Financial Markets</strong></td>
<td></td>
</tr>
<tr>
<td>Risk free rate ( r )</td>
<td>0.0275</td>
</tr>
<tr>
<td>Mort. Rate Low ( \rho )</td>
<td>0.0475</td>
</tr>
<tr>
<td>Mort. Rate High (age&lt;35, ( y&lt;\text{median} )) ( \bar{\rho} )</td>
<td>0.0525</td>
</tr>
<tr>
<td>Install limit ( \nu )</td>
<td>0.15</td>
</tr>
<tr>
<td>Ref. Cost ( \kappa )</td>
<td>0.008 ( \times ) ( \exp(trend(1)_{t}) )</td>
</tr>
<tr>
<td>Max LTV ( \chi_{LTV} )</td>
<td>0.80</td>
</tr>
<tr>
<td>Max LTI ( \chi_{LTI} )</td>
<td>3.5</td>
</tr>
<tr>
<td>Excl. Time Def. (years) ( \tau_{\text{excl}} )</td>
<td>3</td>
</tr>
</tbody>
</table>
taxes and labor income. This allows us to get for each household an estimate of its taxes given the income level. We can then calculate net income for each household. Details of the estimation can be found in the appendix.

Agents differ by their current income level as well as their life cycle component. For simplicity we normalize the life cycle component to 1 for 25 year old. The future evolution differs across education groups and is particular steep for agents with high education it is estimated. The mean income level of the 25 year old households changes for different education groups. We estimate from the Census an initial mean income level of 45,000$ for the young households with a bachelor degree and 30,000$ for households with high school at most. We measure income in annual 1000$ units, such that considering an agent with annual after tax income of 30,000$ requires to set \( \bar{y} = \log(30) \).

6.3 House Price Process

In this base line calibration we fit the time series model to quarterly data for each one of the available California counties over the sample between 1995 and 2000. The parameter \( g_h \) represents trend growth for house prices. We assume that households use the median parameter value across California counties as an estimate of trend growth for the first 10 years. They then assume that trend growth will be equal to the growth rate of income \( g_y \). The parameter \( \alpha_{1,2000} \) is equal to the trend component of the log price in the first quarter of 2000 for the 2 bedrooms low quality house. From the time series model we also obtain...

Cross section...

6.4 Other housing parameters

Each house type delivers a certain amount of housing dividends every period. We denote the vector of these services with \( h \). Housing services are taken to be a linear function of the estimated median 1998 prices.

The US department of Housing and Urban Development provides on its website \(^{19}\) data for average annual fair market rents across all US counties for houses with different numbers of bedrooms over the years from 1983 through 2014. We have already described how using Zillow data we are able to build price series for different house tiers in each California county and for different house sizes. We assume that the fair market rate applies to the median tier of each house size. This way, we are able to calculate price to rent ratios for houses with different number of bedrooms in

\(^{19}\)See http://www.huduser.org/portal/datasets/fmr.html.
different counties. We impose that the ratio stays fixed for houses of the same size but in different tiers. Finally, we assume that the household believes that rents will grow at the same rate as trend component of house prices. This allows for expected fluctuations in the price to rent ratio due to the deviation component.

Maintenance costs $m$ have to be paid each period when owning a house. Maintenance cost scale up with the size of the house, such that the actual costs are $m \cdot d$. There is data on the AHS on maintenance costs and insurance costs. They summed up to 0.7$ per square feet in the AHS 2003, or 700$ per $d_{hi}$ per year. We will therefore set $m = 700/(4 \cdot 1000) = 0.175$.

Real estate agent cost when selling is $\eta = 0.08$ as a share of the house value.

6.5 Financial Parameters

We set the interest rate $r^f$ to the real yield of one year treasuries and agents assume the rate to stay fixed. In the baseline calibration assume a annual real yield of 2.75%. The two mortgage rates are set to $\rho = 4.75\%$ and $\rho = 5.25\%$.

We assume the agent can make free installment payments of $\nu = 15\%$ per year. For everything above he has to pay a fee. Whenever the agent decides to repay more than that sum, the agent will have to pay a fee $\kappa$. We assume that this fee is a fixed cost. Scanning some mortgage websites it seems that a 1000$ is an appropriate guess. Given our measurement in 1000$ units, we set $\kappa = 1$.

The maximum loan-to-value ratio is set to $\chi_{LTV} = 0.8$ and the maximum loan-to-income ratio to $\chi_{LTI} = 3.5$.

7 Numerical Exercise

In order to bring our calibrated model to the data, we solve the portfolio choice problem for a large group of heterogenous households which represent the population of California. We take the characteristics of these households directly from the 2000 Census by extracting all Californian households from the public use sample. For each household, we observe age, gross income, education and the sample weight. We estimate a tax function using data from the Current Population Survey in order to transform gross income to net income. Details on that transformation can be found in the appendix. The Census does not contain data on financial wealth and we therefore use the Survey of Consumer Finance (SCF) to impute wealth characteristics. In particular, we sort households in the SCF into groups defined by age, education and income. For each group we estimate the distribution of wealth-to-income ratios. We then split each Census household into multiple fictitious households by assigning different values of wealth according to the estimated distribution of wealth-to-income ratios. Further details can again be found in the appendix.
Figure 10: Cross section of income for young Californian households in the 2000 decennial Census.

Figure 11: Cross section of wealth for young households, matched to data from the 2000 Survey of Consumer Finances.
For each age and education group, we solve the portfolio choice model over a grid of possible incomes and savings. The model features eight state variables (age, income, savings, home ownership, mortgage, default status, price deviation and momentum) and imposes accordingly some computational challenges. We use quadrature to simulate the stochastic variables and interpolate the respective value functions over up to 2500000 points. The solution is written in Fortran and solved using parallel computing on 12 cores.

7.1 Results

Having computed the value functions in the year 2000 for each age and education group, we can then directly solve the optimal portfolio choice for each fictitious household extracted from Census and SCF data. As a first exercise...match the data.

Table 6 compares the predicted homeownership rate with the data.

[INCOMPLETE]

8 Discussion

[INCOMPLETE]

9 Conclusions

[INCOMPLETE]
Table 6: Homeownership rates in California, 2000, Data vs Model

<table>
<thead>
<tr>
<th>Edu</th>
<th>Age</th>
<th>WG</th>
<th>Own. Rate</th>
<th>Mort Sh</th>
<th>Med LTV</th>
<th>Med LTI</th>
<th>90th LTV</th>
<th>90th LTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS</td>
<td>25-30</td>
<td>-</td>
<td>19.21 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>25-30</td>
<td>-</td>
<td>22.68 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\gamma = 5$, $\psi = 0.5$

<table>
<thead>
<tr>
<th>Edu</th>
<th>Age</th>
<th>WG</th>
<th>Own. Rate</th>
<th>Mort Sh</th>
<th>Med LTV</th>
<th>Med LTI</th>
<th>90th LTV</th>
<th>90th LTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS</td>
<td>25-30</td>
<td>0 %</td>
<td>8.94 %</td>
<td>71.30 %</td>
<td>44 %</td>
<td>1.77</td>
<td>77.50 %</td>
<td>2.45</td>
</tr>
<tr>
<td>HS</td>
<td>25-30</td>
<td>7 %</td>
<td>19.32 %</td>
<td>86.36 %</td>
<td>56 %</td>
<td>2.00</td>
<td>77.50 %</td>
<td>2.68</td>
</tr>
<tr>
<td>BA</td>
<td>25-30</td>
<td>0 %</td>
<td>25.94 %</td>
<td>79.55 %</td>
<td>49.50 %</td>
<td>2.14</td>
<td>71.50 %</td>
<td>3.22</td>
</tr>
<tr>
<td>BA</td>
<td>25-30</td>
<td>-1 %</td>
<td>22.30 %</td>
<td>76.63 %</td>
<td>46.86 %</td>
<td>2.03</td>
<td>67 %</td>
<td>3.26</td>
</tr>
</tbody>
</table>

$\gamma = 5$, $\psi = 1.5$

<table>
<thead>
<tr>
<th>Edu</th>
<th>Age</th>
<th>WG</th>
<th>Own. Rate</th>
<th>Mort Sh</th>
<th>Med LTV</th>
<th>Med LTI</th>
<th>90th LTV</th>
<th>90th LTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS</td>
<td>25-30</td>
<td>0 %</td>
<td>11.71 %</td>
<td>68.01 %</td>
<td>39 %</td>
<td>1.31</td>
<td>70 %</td>
<td>1.87</td>
</tr>
<tr>
<td>HS</td>
<td>25-30</td>
<td>6 %</td>
<td>18.58 %</td>
<td>79.14 %</td>
<td>50 %</td>
<td>1.49</td>
<td>69 %</td>
<td>2.03</td>
</tr>
<tr>
<td>BA</td>
<td>25-30</td>
<td>0 %</td>
<td>31.29 %</td>
<td>72.83 %</td>
<td>45.50 %</td>
<td>1.69</td>
<td>69.50 %</td>
<td>2.33</td>
</tr>
<tr>
<td>BA</td>
<td>25-30</td>
<td>-3 %</td>
<td>22.33 %</td>
<td>66.28 %</td>
<td>40.50 %</td>
<td>1.58</td>
<td>60.50 %</td>
<td>2.20</td>
</tr>
</tbody>
</table>

$\gamma = 2$, $\psi = 0.5$

<table>
<thead>
<tr>
<th>Edu</th>
<th>Age</th>
<th>WG</th>
<th>Own. Rate</th>
<th>Mort Sh</th>
<th>Med LTV</th>
<th>Med LTI</th>
<th>90th LTV</th>
<th>90th LTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS</td>
<td>25-30</td>
<td>0 %</td>
<td>3.12 %</td>
<td>82.92 %</td>
<td>38.65 %</td>
<td>2.18</td>
<td>80 %</td>
<td>3.50</td>
</tr>
<tr>
<td>HS</td>
<td>25-30</td>
<td>16.5 %</td>
<td>19.26 %</td>
<td>95.49 %</td>
<td>70 %</td>
<td>3.04</td>
<td>80 %</td>
<td>3.50</td>
</tr>
<tr>
<td>BA</td>
<td>25-30</td>
<td>0 %</td>
<td>12.61 %</td>
<td>78.24 %</td>
<td>68.23 %</td>
<td>3.42</td>
<td>80 %</td>
<td>3.50</td>
</tr>
<tr>
<td>BA</td>
<td>25-30</td>
<td>7.5 %</td>
<td>22.54 %</td>
<td>86.60 %</td>
<td>70 %</td>
<td>3.29</td>
<td>80 %</td>
<td>3.50</td>
</tr>
</tbody>
</table>

$\gamma = 2$, $\psi = 1.5$

<table>
<thead>
<tr>
<th>Edu</th>
<th>Age</th>
<th>WG</th>
<th>Own. Rate</th>
<th>Mort Sh</th>
<th>Med LTV</th>
<th>Med LTI</th>
<th>90th LTV</th>
<th>90th LTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS</td>
<td>25-30</td>
<td>0 %</td>
<td>6.33 %</td>
<td>75.18 %</td>
<td>36.31 %</td>
<td>1.88</td>
<td>80 %</td>
<td>3.41</td>
</tr>
<tr>
<td>HS</td>
<td>25-30</td>
<td>9.5 %</td>
<td>19.44 %</td>
<td>91.77 %</td>
<td>61 %</td>
<td>2.39</td>
<td>80 %</td>
<td>3.34</td>
</tr>
<tr>
<td>BA</td>
<td>25-30</td>
<td>0 %</td>
<td>16.18 %</td>
<td>75.46 %</td>
<td>58.13 %</td>
<td>3.05</td>
<td>80 %</td>
<td>3.50</td>
</tr>
<tr>
<td>BA</td>
<td>25-30</td>
<td>1.5 %</td>
<td>22.02 %</td>
<td>81.19 %</td>
<td>59.37 %</td>
<td>3.00</td>
<td>80 %</td>
<td>3.50</td>
</tr>
</tbody>
</table>
Figure 12: Cumulative distribution functions of house values owned by individuals with high school or less and age between 25 and 30 in 2000. Census data are shown by the histogram on the background, while model based moments for different calibrations are shown by the black dots.
Figure 13: Cumulative distribution functions of house values owned by individuals with bachelor degree or more and age between 25 and 30 in 2000. Census data are shown by the histogram on the background, while model based moments for different calibrations are shown by the black dots.
References


Appendix

A  Survey Standard Errors

Standard errors are determined by sampling error in the surveys. The Decennial Census is based on a 5% sample of the total population, while the American Community Survey is a 1% sample. Both surveys provide sample weights which correspond to the frequencies of the sampled observations in the population. Our estimate of the fraction of homeowners with age \(i\) and education level \(j\) is \(\hat{p}_{i,j}\). For the Decennial Census we calculate standard errors as:

\[
SE_{\text{census}} (\hat{p}_{i,j}) = DF \sqrt{\frac{19}{B_{i,j}} \hat{p}_{i,j} (100 - \hat{p}_{i,j})}
\]

The base \(B_{i,j}\) is obtained by cumulating the population frequencies (the sampling weights) of the observations used to calculated the fraction \(\hat{p}_{i,j}\), that is all sampled households whose householder has age \(i\) and education level \(j\). The scalar \(DF\) is the design factor, which adjusts the standard errors for distortion in the sampling of specific sub populations. Following the methodology suggested by the Census, we fix in our calculations \(DF = 3.4\).

The American Community Survey provides along with sampling weights 80 replica weights for each observations. Standard errors are then obtained as:

\[
SE_{\text{ACS}} = \sqrt{\frac{4}{80} \sum_{r=1}^{80} (\hat{p}_{i,j,r} - \hat{p}_{i,j})^2}
\]

Where \(\hat{p}_{i,j}\) is the estimated fraction of homeowners based on the sampling weights and each \(\hat{p}_{i,j,r}\) is the estimated fraction based on a different vector of replica weights.

B  Calibration
**B.1 California Counties**

Table 7: California Counties A to Ma. We assess population coverage according to 2000 Decennial Census. Counties selected for the estimation of house price dynamics are highlighted in bold.

<table>
<thead>
<tr>
<th>County Name</th>
<th>County FIPS</th>
<th>N. Zip</th>
<th>% Covered</th>
<th>Pop. per sqm</th>
<th>House Units per sqm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alameda</td>
<td>001</td>
<td>52</td>
<td>87.8 %</td>
<td>1,957</td>
<td>732</td>
</tr>
<tr>
<td>Alpine</td>
<td>003</td>
<td>3&gt;</td>
<td>-</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Amador</td>
<td>005</td>
<td>3</td>
<td>47.4 %</td>
<td>59</td>
<td>25</td>
</tr>
<tr>
<td><strong>Butte</strong></td>
<td>007</td>
<td>9</td>
<td>94.2 %</td>
<td>124</td>
<td>52</td>
</tr>
<tr>
<td>Calaveras</td>
<td>009</td>
<td>6</td>
<td>70.5 %</td>
<td>40</td>
<td>23</td>
</tr>
<tr>
<td>Colusa</td>
<td>011</td>
<td>3&gt;</td>
<td>-</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>Contra Costa</td>
<td>013</td>
<td>34</td>
<td>80.9 %</td>
<td>1,318</td>
<td>493</td>
</tr>
<tr>
<td>Del Norte</td>
<td>015</td>
<td>3&gt;</td>
<td>-</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td><strong>El Dorado</strong></td>
<td>017</td>
<td>10</td>
<td>89.2 %</td>
<td>91</td>
<td>42</td>
</tr>
<tr>
<td>Fresno</td>
<td>019</td>
<td>22</td>
<td>72.8 %</td>
<td>134</td>
<td>45</td>
</tr>
<tr>
<td>Glenn</td>
<td>021</td>
<td>3&gt;</td>
<td>-</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>Humbolt</td>
<td>023</td>
<td>3&gt;</td>
<td>-</td>
<td>35</td>
<td>16</td>
</tr>
<tr>
<td>Imperial</td>
<td>025</td>
<td>3&gt;</td>
<td>-</td>
<td>34</td>
<td>11</td>
</tr>
<tr>
<td>Inyo</td>
<td>027</td>
<td>3&gt;</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Kern</td>
<td>029</td>
<td>17</td>
<td>62.6 %</td>
<td>81</td>
<td>28</td>
</tr>
<tr>
<td>Kings</td>
<td>031</td>
<td>5</td>
<td>94.4 %</td>
<td>93</td>
<td>26</td>
</tr>
<tr>
<td>Lake</td>
<td>033</td>
<td>5</td>
<td>70.4 %</td>
<td>46</td>
<td>26</td>
</tr>
<tr>
<td>Lassen</td>
<td>035</td>
<td>3&gt;</td>
<td>-</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td><strong>Los Angeles</strong></td>
<td>037</td>
<td>163</td>
<td>92 %</td>
<td>2,344</td>
<td>806</td>
</tr>
<tr>
<td>Madera</td>
<td>039</td>
<td>6</td>
<td>72.2 %</td>
<td>58</td>
<td>19</td>
</tr>
<tr>
<td>Marin</td>
<td>041</td>
<td>12</td>
<td>73 %</td>
<td>476</td>
<td>202</td>
</tr>
<tr>
<td>Mariposa</td>
<td>043</td>
<td>3&gt;</td>
<td>-</td>
<td>12</td>
<td>6</td>
</tr>
</tbody>
</table>
Table 8: California Counties Me to Sh. Counties selected for the estimation of house price dynamics are highlighted in bold.

<table>
<thead>
<tr>
<th>County Name</th>
<th>County FIPS</th>
<th>N. Zip</th>
<th>% Covered</th>
<th>Pop. per sqm</th>
<th>House Units per sqm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mendocino</td>
<td>045</td>
<td>3&gt;</td>
<td>-</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>Merced</td>
<td>047</td>
<td>3</td>
<td>23.1 %</td>
<td>109</td>
<td>36</td>
</tr>
<tr>
<td>Modoc</td>
<td>049</td>
<td>3&gt;</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Mono</td>
<td>051</td>
<td>3&gt;</td>
<td>-</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Monterey</td>
<td>053</td>
<td>10</td>
<td>45.5 %</td>
<td>121</td>
<td>40</td>
</tr>
<tr>
<td>Napa</td>
<td>055</td>
<td>4</td>
<td>85.9 %</td>
<td>165</td>
<td>64</td>
</tr>
<tr>
<td>Nevada</td>
<td>057</td>
<td>8</td>
<td>99.0 %</td>
<td>96</td>
<td>46</td>
</tr>
<tr>
<td>Orange</td>
<td>059</td>
<td>48</td>
<td>58.8 %</td>
<td>3,606</td>
<td>1,228</td>
</tr>
<tr>
<td>Placer</td>
<td>061</td>
<td>17</td>
<td>60 %</td>
<td>177</td>
<td>76</td>
</tr>
<tr>
<td>Plumas</td>
<td>063</td>
<td>3&gt;</td>
<td>-</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Riverside</td>
<td>065</td>
<td>51</td>
<td>81.3 %</td>
<td>214</td>
<td>81</td>
</tr>
<tr>
<td>Sacramento</td>
<td>067</td>
<td>25</td>
<td>55.8 %</td>
<td>1,267</td>
<td>492</td>
</tr>
<tr>
<td>San Benito</td>
<td>069</td>
<td>3&gt;</td>
<td>-</td>
<td>38</td>
<td>12</td>
</tr>
<tr>
<td>San Bernardino</td>
<td>071</td>
<td>52</td>
<td>83 %</td>
<td>85</td>
<td>30</td>
</tr>
<tr>
<td>San Diego</td>
<td>073</td>
<td>72</td>
<td>91.1 %</td>
<td>670</td>
<td>248</td>
</tr>
<tr>
<td>San Francisco</td>
<td>075</td>
<td>22</td>
<td>97.6 %</td>
<td>16,634</td>
<td>7,421</td>
</tr>
<tr>
<td>San Joaquin</td>
<td>077</td>
<td>13</td>
<td>63.6 %</td>
<td>402</td>
<td>135</td>
</tr>
<tr>
<td>San Luis Obispo</td>
<td>079</td>
<td>11</td>
<td>83.4 %</td>
<td>75</td>
<td>31</td>
</tr>
<tr>
<td>San Mateo</td>
<td>081</td>
<td>25</td>
<td>98.5 %</td>
<td>1,575</td>
<td>580</td>
</tr>
<tr>
<td>Santa Barbara</td>
<td>083</td>
<td>11</td>
<td>54.5 %</td>
<td>146</td>
<td>52</td>
</tr>
<tr>
<td>Santa Clara</td>
<td>085</td>
<td>48</td>
<td>87.7 %</td>
<td>1,304</td>
<td>449</td>
</tr>
<tr>
<td>Santa Cruz</td>
<td>087</td>
<td>12</td>
<td>94.6 %</td>
<td>574</td>
<td>222</td>
</tr>
<tr>
<td>Shasta</td>
<td>089</td>
<td>8</td>
<td>90.2 %</td>
<td>43</td>
<td>18</td>
</tr>
</tbody>
</table>
Table 9: California Counties Si to Y. We assess population coverage according to 2000 Decennial Census. Counties selected for the estimation of house price dynamics are highlighted in bold.

<table>
<thead>
<tr>
<th>County Name</th>
<th>County FIPS</th>
<th>N. Zip</th>
<th>% Covered</th>
<th>Pop. per sqm</th>
<th>House Units per sqm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sierra</td>
<td>091</td>
<td>3&gt;</td>
<td>-</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Siskiyou</td>
<td>093</td>
<td>3</td>
<td>40 %</td>
<td>7</td>
<td>3.5</td>
</tr>
<tr>
<td><strong>Solano</strong></td>
<td>095</td>
<td>8</td>
<td>75.6 %</td>
<td>476</td>
<td>162</td>
</tr>
<tr>
<td>Sonoma</td>
<td>097</td>
<td>24</td>
<td>96.2 %</td>
<td>291</td>
<td>116</td>
</tr>
<tr>
<td>Stanislaus</td>
<td>099</td>
<td>11</td>
<td>70.3 %</td>
<td>299</td>
<td>101</td>
</tr>
<tr>
<td>Sutter</td>
<td>101</td>
<td>3&gt;</td>
<td>-</td>
<td>131</td>
<td>47</td>
</tr>
<tr>
<td>Tehama</td>
<td>103</td>
<td>3&gt;</td>
<td>-</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>Trinity</td>
<td>105</td>
<td>3&gt;</td>
<td>-</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td><strong>Tulare</strong></td>
<td>107</td>
<td>13</td>
<td>77.7 %</td>
<td>76</td>
<td>25</td>
</tr>
<tr>
<td>Tuolomne</td>
<td>109</td>
<td>7</td>
<td>90.3 %</td>
<td>24</td>
<td>13</td>
</tr>
<tr>
<td>Ventura</td>
<td>111</td>
<td>25</td>
<td>98.8 %</td>
<td>408</td>
<td>136</td>
</tr>
<tr>
<td>Yolo</td>
<td>113</td>
<td>3</td>
<td>60.5 %</td>
<td>166</td>
<td>61</td>
</tr>
<tr>
<td>Yuba</td>
<td>115</td>
<td>4</td>
<td>74.5 %</td>
<td>96</td>
<td>36</td>
</tr>
</tbody>
</table>
C House Price Model


In the following figure we show cumulative distribution functions of parameter estimates across the 36 counties under analysis. We consider both estimates from the entire sample (1953Q1:2014Q1) and from the part of the sample for which California data are available (1989Q1:2014Q1). Estimates of $g_h$, $\phi_d$, $\phi_x$ and $\sigma_x$ are in general larger when the time series start in 1989 rather than when they start in 1953. Furthermore, the dispersion of trend growth estimates across counties is larger when we use only the data starting from 1989. Thus, when we use the shorter sample, house prices seem to experience higher trend growth, more persistent deviations and stronger momentum in deviations.

Figure 14: Cumulative distribution functions across counties for parameter estimates from the time series model. Each county is weighted by its population according to the 2000 Decennial Census.

Several features of the data are driving the differences in parameter estimates between the longer and the shorter sample. First, time series are identical for all counties up to the first quarter of 1998. Thus, in the longer sample county heterogeneity affects a smaller fraction of the data, resulting in lower dispersion of the estimates. Moreover, in the US level data used for the period between 1953 and 1989, house price fluctuations are moderate and prices do not appear to trend upward. Note that this behavior might be due to the fact that these data are aggregated at a national level. The time series of California house prices over the same period might show a different behavior. On the
other hand, it is known that the recent housing boom and bust has been an impressively extreme cycle in housing prices. Thus, it is also reasonable to consider that the behavior of house prices might undergo structural changes in its trend and deviation components.

C.2 Model Performance: OOS Forecasting and Survey Data

How can we assess the expectations produced by the recurvise model? We propose here two criteria. The first one out of sample forecasting performance of the model. The second one is the comparison between model expectations and household surveys. For what concerns the first test, we analyze in the following table real house price returns RMSEs for out of sample forecasts generated by the model for different values of $\lambda$ and for predictive horizons of one quarter and one year over the period between Q1 1998 and Q1 2013. We compare these root mean squared errors (RMSE) against the ones obtained by forecasting with the recursively estimated historical mean of real returns. The house price under analysis is the one of the cheapest house in each county and we report the median and the 10th and 90th percentile of the RMSEe distribution across counties. It appears clearly that the model is able to deliver systematically smaller forecast errors than the historical mean. The overperformance is stronger for the horizons of one quarter and one year, while it appears weaker at the 3 years horizon. However, the choice of $\lambda$ affects the predictive performance. For the one quarter and one year horizons accuracy appears to be decreasing in the level of downweighting, with the best forecasting performance delivered by $\lambda = 0$. At the three years horizon the best performance is achieved by setting $\lambda = 0.025$, while the model with $\lambda = 0.1$ does worse than the historical mean. Nonetheless, models with $\lambda$ equal to 0, 0.01 and 0.05 seem to deliver very similar RMSE across counties.

For what concerns the comparison to survey, we collect from Case, Thompson and Shiller (2012) survey expectations for future returns on houses. The survey was conducted each year between 2003 and 2012. House buyers in the counties of Alameda, Middlesex, Milwaukee and Orange were asked about their expectations for annual price growth of houses for the following year and the following 10 years. Alameda and Orange are California counties and are also part of our sample. We can therefore test whether our model can match the survey expectations. In the following graphs we compare model based expectations for the cheapest house type in both Orange and Alameda county against the survey forecasts. The value of $\lambda$ is set to 0.025 for Orange and to 0.05 for Alameda county. For both counties, the model matches qualitatively well the survey expectations for the 1 year horizon. At the 10 year horizons the expected returns from the model are clearly below the ones reported in the survey until the bust of 2007. However, both model and survey expectations seems to follow a similar pattern.

Thus, the model delivers an acceptable out of sample forecasting performance and seems to be
Table 10: Out of Sample forecasting performance of the model. Median and quantile of RMSE across counties.

<table>
<thead>
<tr>
<th></th>
<th>RMSE 1 Quarter Horizon, sample from Q1 1998 through Q4 2013</th>
<th>RMSE 1 Year Horizon, sample from Q1 1998 through Q1 2013</th>
<th>RMSE 3 Year Horizon, sample from Q1 1998 through Q1 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hist Mean</td>
<td>λ = 0</td>
<td>λ = 0.01</td>
</tr>
<tr>
<td>Median</td>
<td>4.18%</td>
<td>1.97 %</td>
<td>2.12 %</td>
</tr>
<tr>
<td>Quant. 0.10</td>
<td>3.56%</td>
<td>1.64 %</td>
<td>1.78 %</td>
</tr>
<tr>
<td>Quant. 0.90</td>
<td>5.42%</td>
<td>2.68 %</td>
<td>2.77 %</td>
</tr>
</tbody>
</table>

Figure 15: Comparison between model and Case, Thompson and Shiller survey. Alameda county; λ = 0.05. Expected real returns are calculated using inflation expectations from the Federal Reserve Bank of Cleveland.
able to match some features of survey expectations. It is interesting to notice however that the best forecasting performance is achieved with no down weighting, while survey expectations are matched for a relatively high level of discounting. Higher Accuracy in forecasting is achieved by attributing most of the variation in house prices to the deviation component. However, surveys are more compatible with a model that features structural changes in trend growth and therefore higher variability in the trend component. Moreover, the level of discounting needed to match the survey data might be heterogeneous across different counties. While in this paper we will not further investigate the issue of the calibration of downweighting, we leave this issue to future research.

C.3 Differences in Trend Growth Across House Types

In this section we recursively estimate the deterministic trend for each house price using $\lambda = 0$. We report results for the 2 bedroom low tier house, the 3 bedroom mid tier house and the 4 bedroom top tier house in the figure below. For the period between 1989 and 2014, price series are build according to the methodology described in the main text. To make our series long enough, we are forced to use the national real house price index provided by Shiller for the period between Q1 1953 and Q1 1989. As we can see, trend growth appears to be lower for higher quality houses across all California counties. Nonetheless, differences in trend growth across house types appear in general
relatively small, especially when the sample is extended through Q1 2014 and more observations from the specific house type series become available. Moreover, changes in the distribution of growth rates over time are qualitatively very similar across house type. Since the fraction of the sample over which we can observe differences across house types is small, we believe it is conservative to assume the same trend growth rate for all house prices and to set this equal to the growth rate for the smallest house.

Figure 17: Cumulative distribution functions across counties of trend growth for different house types and at different points in time.