An Agent-Based Model of the Housing Market Bubble in Metropolitan Washington, D.C. *

Rob Axtell   Doyne Farmer   John Geanakoplos   Peter Howitt
Ernesto Carrella   Ben Conlee   Jon Goldstein   Matthew Hendrey
Philip Kalikman   David Masad   Nathan Palmer   Chun-Yi Yang

May 25, 2014

Abstract

Several independent datasets concerning household behavior in the Washington, D.C. metropolitan area are combined to create an agent-based model of the recent housing market bubble and its aftermath. Comprehensive data on housing stock attributes, primarily from local government sources, are used as input to the model, as are administratively-complete data on household characteristics. Data covering all real estate transactions over the period 1997-2009 are used as targets for model output, including inventory levels and average days-on-market in addition to price statistics. Finally, a very large sample of mortgage service data (80-90 percent coverage of metro D.C.) serve as both input to the model (e.g., mortgage types, interest rates obtained) as well as target output (e.g., refinancing rates, foreclosure levels). The model consists of a large number of heterogeneous households who make rent or buy decisions, are matched to homes and commonly seek mortgages with which to purchase homes. These households have homogeneous rules of behavior but heterogeneous realized behavior since decisions depend on local household characteristics (e.g., size, composition, financials). This is a so-called agent-based computational model since each household and the banks originating mortgages are software agents while each home and each mortgage are software objects. The model is capable of running at full-scale with the metrpolitan D.C. housing market, over 2 million households. Overall, we find that certain empirically-grounded household decision rules are capable of generating a home price bubble much like what was observed during this time period. The model does not get the absolute bubble level and the timing of its bursting exactly right but does a good job on certain market 'internals' such as real estate sales, inventories and market tightness. Not in the model at present are fine-grained aspect of household decision-making (e.g., moving in advance of schools starting) and thus the model lacks certain well-known temporal phenomena like seasonality. Also, while the home purchase market is deeply represented in the model, few details of the rental market are present, a further weakness. These limitations and parameter sensitivities are described. Despite these flaws, we use the calibrated model to perform a few policy experiments. Our preliminary findings are that tighter interest rate policies would have done little to attenuate the price bubble, while limiting household leverage would have had a larger effect.

1 Introduction

Housing was at the heart of the recent financial crisis. [Reinhart and Rogoff, 2011] have shown that real estate bubbles and crashes characterize almost all financial crises. Understanding how household behavior leads to home price bubbles is therefore of central importance for developing policies to promote financial stability. Our goal is to study the behavior of house prices in the period leading up to and after the crisis, and to investigate the importance of various factors that are amenable to policy interventions, such as central bank interest rate policies, leverage limits, creditworthiness limits, removal of refinancing restrictions, writedowns of mortgage principal, and subsidies to first-time homeowners.

In order to study how the behavior of individual households can produce bubbles at the aggregate level we combine micro-data with agent-based computing. Agent-based models are computational models (see [Tesfatsion and Judd, 2006]) in which heterogeneous agents interact directly with one another and with their economic environment, following autonomous decision rules. Such models are used to study outcomes that emerge from complex interactions and that cannot be easily derived from the agents' objectives and behavioral rules. The earliest examples in the economics literature were abstract theoretical models such as [Schelling, 1971]'s model of segregation, in which highly segregated neighborhoods emerged from the interaction of individual location decisions that exhibited only a mild degree of aversion to living near people of different types. More recently, agent-based methodology has been used to create quantitative models of a specific kinds of social interactions and have proven useful for policy (e.g. the NASDAQ model of decimalization, or the Portland traffic model).

Here we present a quantitative agent-based model of a housing market that can be run at full-scale with a metropolitan area. We think that housing markets constitute an area where quantitative agent-based modeling is promising, for a number of reasons. Housing markets involve complex interactions between a host of different decisions made by households and banks. The former choose whether to buy or rent, how much to pay, how large a mortgage to choose and of what type, whether to invest in rental housing, how much rent to charge, whether to refinance a mortgage, whether to default on a mortgage, how much to save for a down payment, when to sell a house, and how much to ask for it. Banks choose what kinds of mortgages to offer, at what prices, allowing how much leverage, and to people of what kind of creditworthiness. Several authors believe that heterogeneity among different actors with respect to these decisions is important for making predictions even about aggregate outcomes. Indeed [Khandani et al., 2012] reports that quantitative agent-based models are already in use on Wall Street. Finally, it is now possible to calibrate such a model at a highly disaggregated level using rich data that cover all mortgages and almost all real estate transactions. A standard representative agent approach would be hard put to make use of such data without making extreme simplifying assumptions.

This paper is a first step in the direction of a quantitative agent-based model of the US housing market. To keep the project manageable, we have restricted attention to the market in the metropolitan area of Washington DC, over the period from 1997 to 2009. During that period there were about 2.2 million households in the area, and over 3 million mortgage loans were issued overall. We have calibrated our model using data on all of those mortgages and all related real-estate transactions—including sales, listings, delistings, and

changes in asking prices—that occurred through the various multiple-listing services in the area.

Our goals parallel those of the kinds of large-scale computational projects that cut across all the sciences today, from modeling all the chemical reactions and biological processes in a cell to simulating an entire brain by representing all neurons. Indeed, we live in the era of 1-to-1 computational instantiations of many complex systems, and agent-based computing is a way for economics to join this zeitgeist of digital synthesis. We focus on household behavior, taking as exogenous income, demographics, housing supply, and most of the behavior of banks. We study each of component of household behavior one by one, and code separate representations of them that combine in a dynamic model. After initializing the model with data to match the state of the market in 1997, it runs for some 100 years to eliminate initial transients to reach a quasi-steady-state that endogenizes correlations that are not directly observable from our data sources. Permitting the model to then run with data from the 1997-2009 period we find that the model output fits aggregate data reasonably well, particularly on the time series of home price appreciation. We achieve this fit despite estimating behavioral components of the model individually and separately from micro-data, and without modifying parameters to improve the model's overall performance.

This paper is organized as follows. Section 2 reviews the literature. Section 3 describes the sources of data we use in our estimations. Section 4 describes the model component-by-component, including the estimation procedure and data sources for each component. Section 5 describes the extent to which the model matches real data. Section 6 describes the results of several policy experiments we run to study the effects of leverage and interest rates on the housing market. Section 7 concludes.

2 Previous Work Modeling Housing Markets

Econometric models of the housing market go back at least to [Leeuw, 1971] who estimated an income elasticity of demand for housing in the range between 0.5 and 0.9. DSGE macro models that include a housing market were first constructed by [Iacoviello, 2005], who emphasized the importance of credit constraints for the propagation of macro shocks, and [Ortalo-Magné and Rady, 2006], who stressed down payment requirements as a motive for saving early in the lifecycle and developed theoretical propositions concerning overshooting to macro shocks. Neither of these DSGE models were primarily empirical in nature. More recently, [Kiyotaki et al., 2011] calibrated a DSGE model in which finance constraints in the housing market (à la Kiyotaki-Moore) have little effect on macro performance.

Several authors have examined the causes of the housing bubble and crash in the 2000s. [Geanakoplos, 2001], [Geanakoplos, 2010a], and [Geanakoplos, 2010b] argued that leverage and collateral, not interest rates, drove the economy in the crisis of 2007-2009, pushing housing prices and mortgage securities prices up in the bubble of 2000-2006, then precipitating the crash of 2007. [Glaeser et al., 2012] argued that leverage did not play an important role in the run-up of housing prices from 2000-2006. [Duca et al., 2011], on the other hand, argue that it did. [Haughwout et al., 2011] argue that leverage played a pivotal role. None of these authors has looked as closely at heterogeneity, used as rich a data set, or examined as broad a range of housing market variables as we do in this paper.

Other papers have looked at rich data sets to examine the housing market. [Mian and Sufi, 2009]

examine individual mortgage data and identify credit-worthiness by the zip code of the borrower, concluding that the housing bubble was indeed caused by an unprecedented expansion of loan availability to those living in "sub-prime" zip codes. [Landvoigt et al., 2012] use transactions data from public court records to calibrate a theoretical equilibrium model of heterogeneous housing behavior in the San Diego metropolitan area. They compared behavior in 2000 (at the start of the boom) to that in 2005 (near the peak) and concluded that the boom was caused by an expansion of credit to young buyers who drove up the price of less expensive starter homes, precipitating an increase in prices throughout the range of home qualities. Neither of these papers examined as broad a range of housing variables as we do in this paper.

Other authors have constructed agent-based housing models. One of the first was [Gilbert et al., 2009], who constructed a qualitative model of the UK market. That model did not attempt to undertake the micro-level empirical calibration that we are undertaking, and did not investigate the causes of the boom and crash of the 2000s.

3 Data

Models of house price dynamics are typically aggregate in nature, using some flavor of representative agents. Here we eschew such standard formulations and focus on building a model at the individual household level using data on individual households. There are four broad classes of data used by our model:

1. Information on households, their size and composition, income, savings and wealth: Some of these measures are directly available from Census and related Federal sources, some from the IRS public use sample, some inferred from PSID and CEX samples, and American Community Survey data on the rental market. The data we have are not 'matched' and so we have created a synthetic population of households that has the right unconditional properties and conditional relationships that seem approximately correct.

2. Comprehensive real estate transaction data: We have obtained from the local Multiple Listing Service (MLS) a purportedly complete set of transactional data covering each real estate purchase and sale in the Washington, D.C. metro area over the 1997-2009 period. These micro-data contain detailed information on individual home listing price, temporal adjustments to the listing price, sale price, some associated features of sale transactions (e.g., whether the seller pays closing costs), number of days on market, the characteristics of the buyer's financiing (e.g, mortgage type, duration, downpayment), tax information, and many related items that we do not use.

3. Housing characteristics: From local governments, typically at the county level, we have data on individual housing units, their characteristics, their tax assessments, and so on. Only some of these many facets that describe homes are used in the model. Specifically, ours is not a spatial model and so we do not use home addresses or location information of any kind.

4. Mortgage service data amounts to individual records of mortgage payment transactions. From CoreLogic, Inc. we have 80-90 percent coverage of the D.C. MSA,

including the timing of payments, overpayment amounts, delinquincies, number of mortgages by individual, second mortgage information, and so on. These data are supplemented by LoanPerformance data, which is particularly thorough on non-prime mortgages.

While CoreLogic has matched mortgage service and house location (address) data, these matchings were not provided to us due to privacy concerns, and we have made no attempt to do a proper matching. However, our model does produce an explicit matching of households to homes and while we do not expect this to be exact at the micro-level, a well-functioning model should produce realizations that are statistically similar to the single realization of the real world data.

4 The Model: Description and Estimation

The model consists of distinct populations of software objects. It is conventional to call objects that have purposive behavior 'agents' while those that merely hold data are 'objects'. Thus, households and banks in our model are agents while houses and mortgages are objects.

4.1 Model Agents and Objects

Households

Households are the main agents in our model. They buy, sell, and rent homes, earn income, accumulate wealth, age, and migrate. There were approximately 1.6 million households in the DC MSA in January 1997. Our code is capable of running at full-scale, with one agent for each household, although it can take many hours to run the model at this scale. It is often convenient to simulate the model at reduced scale, and the results to be described below have been generated with one agent for every ten real households. Households act monthly.

Banks

Our codebase has a user selectable number of banks. However, since we do not have detailed data on bank behavior all of the model simulations described below feature a single bank. The bank generates loans for households that want to buy homes. The bank balance sheet is not modeled: it can never go bankrupt. The bank also handles defaults and foreclosures.

Houses

Houses are bought, sold, rented, and lived in by households. Each house has a fixed "quality" drawn from the 1997 house price distribution. House quality impacts related variables such as asking price, price appreciation, and rent. The number of houses in the model comes from Census data, and for the specific simulations described below is scaled 10 to 1, as with households. Houses do not age or deteriorate in quality.

A special class of houses in the model are "apartments," houses that can't be bought or owned by households—only rented by them.

Loans

Loans are the mortgages taken by households to buy homes. They come in three types: fixed, adjustable-rate (ARM), and interest-only IO). They can be refinanced either for cash or for a lower rate. A household cannot sell its house for less than the total amount due on an associated outstanding loan. For simplicity we do not permit more than one loan on a house, although second and third loans are not uncommon in the data.

4.2 Model Execution

Each time step in the model corresponds to a month of real-time. Each month:

1. The model starts by aligning the agent population. New houses are built or destroyed (see section 4.11); new households are created and others migrate away or die (section 4.10).

2. The bank surveys all its loans and decides which loans to consider defaulted and which homes to foreclose (4.9).

3. Each household acts in turn, in random order. Each receives income (4.6), tries to pay its housing costs (4.7), lists or delists its houses (4.11), decides whether to refinance any of its loans (4.9), decides whether to buy or invest (4.5)—including deciding how much to pay (4.4) and applying for a loan (4.8)—and spends part of its income on non-housing consumption (4.7).

4. Buyers, sellers, investors, and renters are matched in the housing market (4.11).

4.3 Input Calibration

In agent-based computing each agent acts according to a set of rules. These rules are designed and implemented to be interchangeable. This allows for the study of counterfactuals.

We specify each set of rules independently, fitting the functional form and parameters of each rule or process to relevant data. This kind of independent micro-foundation in the data is typically called input calibration, since we are tuning model inputs (agent behaviors). We have conducted a preliminary calibration so that each agent's behaviors, and the model's piece-by-piece functionality, are reasonably consistent with reality. Emergent phenomena, and macro-level accuracy of the model, are genuine features of the model, not artifacts of target-fitting.

In the next sections we describe the standard rule-set used, describing the functionality of each module and summarizing the estimations we performed. We refer the interested reader to our Rulebook for a more detailed description of module-by-module estimations, module functionality, and nonstandard rules we implemented but are not using in the model.

4.4 Desired Expenditure Function

The house-buying process begins with a household setting a "desired expenditure," a total purchase price on a house. This number is the basis for the bargaining process over obtaining

a loan (4.8) and for eventual matching to a for-sale home (4.11). Using the Panel Study of Income Dynamics in odd years from 1999 through 2009, we estimate desired expenditure formation as a behavioral rule in which agents set their expenditure based on income and lagged home price appreciation. This rule is motivated by the rule of thumb that individuals spend "one third of income" on housing, but we estimate a functional form that better fits the data and captures heterogeneity in housing expenditures. This is the model's default expenditure rule. It is sublinear in income, reflecting that higher-income agents spend a smaller fraction of income on housing.

$$P = \frac{\varepsilon \times h \times \text{Income}^g}{(\tau + c + \text{LTV} \times \text{Prime Rate} - a \times \text{Lagged HPA})},$$

where

- P is desired house price

- Income is contemporaneous annual household income

- LTV is average loan-to-value of loans issued in the prior 3 months

- Prime Rate is the contemporaneous Freddie Mac Primary Mortgage Market Survey average 30-year prime rate

- Lagged HPA is the statewide, lagged (past-year) appreciation in Case-Shiller Home Price Index

- We estimate $g = 0.56, h = 38.8, a = 0.16$, where g captures the amount of income spent on housing, h is a scaling factor, and a controls the strength of the effect of HPA on expenditure

- ε is an additional heterogeneity term we estimate as $\log \varepsilon \sim \mathcal{N}(-0.13, 0.46^2)$

- τ is annual percent tax, insurance, and HOA fees

- c is annual percent maintenance costs

National Data and Descriptive Statistics

The panel format of the PSID data enables us to match incomes to housing expenditures and to incorporate intertemporal relationships such as home price appreciation as explanatory variables.

Because we only wish to study the formation of desired expenditure, only homeowners with recently purchased homes are relevant to our analysis. We subsetted the data to observations where the following conditions hold:

1. The household was a homeowner with a mortgage.

2. The household obtained its first loan in the same year as the year the observation was collected, or in the previous year.

(A further condition, that the household's total annual mortgage costs (on first and second mortgages) did not exceed one third of the house's value, was included as a safeguard against junk observations.) We further required that the data lie in between the minimum and maximum values reported in Table 4.4.1, except for maximum Annual Family Income, for which the maximum allowable annual family income of $750,000 was not achieved by observations meeting other requirements.

	Mean	Min	Q1	Median	Q3	Max
House Value	$228,487	$25,000	$110,000	$175,000	$290,000	$2,000,000
Annual Tax	$2,355	$50	$988	$1,700	$3,000	$45,000
Annual Insurance	$764	$0	$400	$600	$900	$8,000
Monthly Mort. Costs	$1,195	$0	$675	$1,026	$1,500	$7,000
Annual Fam. Income	$92,830	$10,000	$51,112	$76,113	$111,703	$719,400

Table 4.4.1: Summary Statistics of Subsetted PSID Data, 1999-2009 (N = 4,346)

Calibration Procedure

To calibrate the parameters in the rule, we have to express the rule in terms of observable data. We can write:

$$P = \frac{h \times \text{Income}^g}{\tau + c + LTV \times \text{Prime Rate} - a \times \text{Lagged HPA}}$$

$$P \times (\tau + c + LTV \times \text{Prime Rate} - a \times \text{Lagged HPA}) = h \times \text{Income}^g$$

$$P \times (\tau + c + LTV \times \text{Prime Rate}) = h \times \text{Income}^g + a \times \text{Lagged HPA} \times P$$

$$\text{Annual Housing Expenditure} = h \times \text{Income}^g + a \times \text{Lagged HPA} \times P.$$

This equation can be estimated using nonlinear least squares. We observe annual mortgage, tax, and insurance costs, along with house value, income, and HPA. Therefore we need to impute the part of annual expenditure on maintenance and HOA fees. We use the estimate

$$EXP = 12 \times \text{Monthly Mortgage Payments} + .05 \times P,$$

i.e., HOA fees, maintenance costs, taxes, and insurance are together approximately five percent of house value.

Results

We use nonlinear least squares and estimate $g = 0.56, h = 38.8, a = 0.16$. We plot in Figure 4.4.1 observed mean EXP/Income within percentiles of Income. We show estimated expenditure/Income in blue (for Lagged HPA fixed at its mean value).
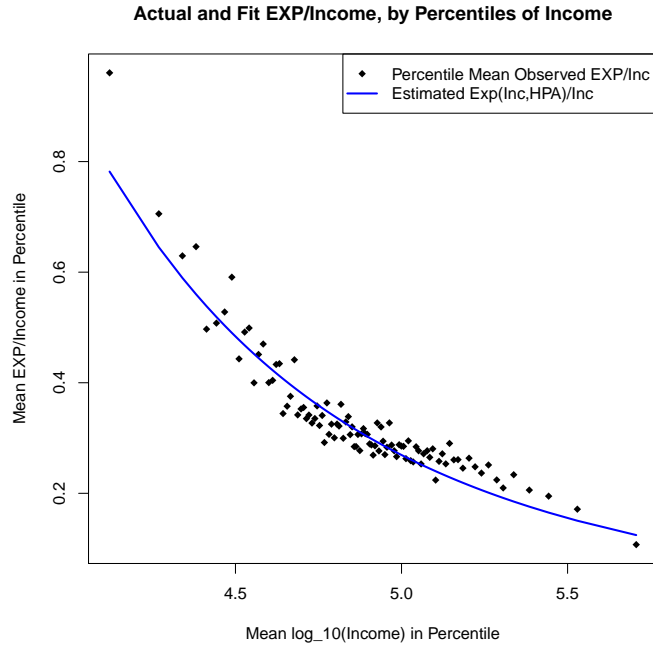
Actual and Fit EXP/Income, by Percentiles of Income

Figure 4.4.1: Fraction of Income Spent on Housing, Observed and Implied by Estimated Expenditure Rule

Heterogeneity

We examine the empirical distribution of ε, defined implicitly in the equation below as the quotient of observed housing expenditure and the expenditure implied in the right-hand-side functional form.

$$P = \varepsilon \times \frac{h \times \text{Income}^g}{\tau + c + LTV \times \text{Prime Rate} - a \times \text{Lagged HPA}}.$$

We plot in Figure 4.4.2 the cdfs of Normal and Log-Normal distributions with mean and variance matching those of the empirical distribution, after truncating observed values of $\varepsilon > 2$ to 2.
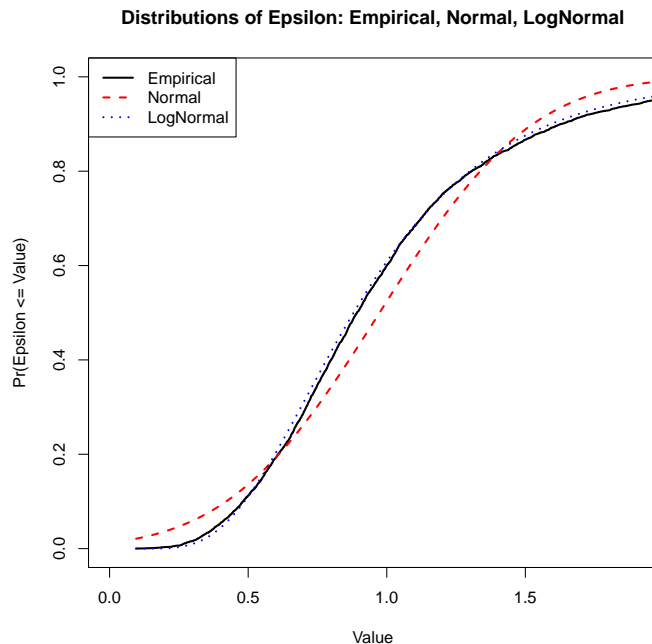
Distributions of Epsilon: Empirical, Normal, LogNormal

Figure 4.4.2: Empirical and Fitted Normal/Log-Normal Distributions of ε

The Log-Normal distribution shows a good match to the empirical distribution and exhibits a Kolmogorov-Smirnov statistic of 0.04. We fit a lognormal distribution to ε by maximum likelihood, yielding

$$\log \varepsilon \sim \mathcal{N}(-0.13, 0.46^2).$$

4.5 Buying and Investing

A household that has no home can try to become a buyer. A household that owns at least one unlisted home can try to become an investor.

Investors in the model are households that own multiple homes: they live in one and rent out the others, possibly selling them eventually. Only about 9% of the population are allowed to be investors, per [Chinco and Mayer, 2012].

To become a buyer, a household needs to be pre-approved for a mortgage. This process is described in section 4.8. Whether potential buyers and investors do actually buy depends on market dynamics, as explained in section 4.11

4.6 Income Process

We estimate a rule for assigning incomes to households each year according to a process motivated by [Carroll, 1992] but matching actual DC income data as reported to the IRS. The income growth process proceeds in three steps:

1. Each household is given a Carroll-Process "permanent" P and "total" income I. Both variables change every year as follows

$$\log I_t = \log(P_t) + \log V_t$$
$$\log P_{t+1} = \log G + \log P_t + \log N_{t+1},$$

where V_t is a transitory shock that determines a particular year's income for a household given its contemporaneous permanent income, N_t is a shock to permanent income, and G is a constant growth rate in permanent income. We assume $\log V_t, \log N_t$ are distributed normally with mean 0 and independently from each other and across time. We use PSID data to estimate during the model period $\sigma_{\log N} = 0.17$, $\sigma_{\log V} = 0.35$. We also take $G = 1.02$, but this is irrelevant because the Carroll process is only used to effect rank shuffling. Each household is assigned a new permanent and total income based on the processes above.

2. While a lognormal distribution of transitory income shocks is largely consistent with the data, it does not accurately capture the clumping of households with zero-income (e.g. the unemployed). To match that feature of the data, the Carroll process also incorporates a separate fixed probability each year of having zero income. We estimate this probability from PSID data.

3. Finally, those households that will not receive zero income are ranked by their total income realizations I_t, and we draw as many samples as such households from an empirical distribution of incomes in DC (based on IRS data). These empirically drawn incomes are sorted and assigned by rank to each household as its *true income* I, i.e. the income the household actually receives for the year.

We use the Carroll process as an intermediate step to accomplish rank-shuffling in a manner that is consistent with the data, while using empirically drawn actual realizations of income ensures that the money flow in the model (which we take to be exogenous) matches reality. Further, in order to dampen the jarring effect of aligning income, we uniformly distribute "activation months" across the population such that each household begins to receive its new income in a different month of the year; by the end of a model year every household has been assigned its new income.

4.7 Liquid Wealth Process

The household is credited its income every month. Income that it does not spend on either housing or consumption accumulates as liquid wealth. Liquid wealth is used to pay housing costs and for downpayments on loans. The household accrues no interest on its liquid wealth.

Paying Housing Costs

If the household is an owner, it pays 3 types of housing costs every month: mortgage payments, maintenance costs, and tax/insurance/HOA fees. Annual tax and related fees are a flat 2.5% of house quality times Case-Shiller index. Maintenance is $\eta\%$ yearly of

house quality, where $\eta \sim N(2.5, 0.7)$ is heterogenous across houses. (Mortgage costs are determined by the loan.) Unpaid taxes and maintenance accumulate. Unpaid loans become delinquent and risk being foreclosed (4.9). If the household owns multiple homes, it has to pay for each in turn.

A household that is renting has only to pay its rent.

Non-housing Consumption

Let consumption c be the dollar amount of households' non-housing expenditure. In this model, c flows out of the system. Given a household's monthly income I, consumption is simply:

$$\log_{10}(c) = a + b \log_{10}(I) + \epsilon,$$

where the parameters were estimated from 2010-2011 CEX micro-data. The a and b parameters are heterogeneous across households, drawn from a sample of 100 generated by bootstrapping a new sample against which to regress. ϵ is drawn empirically from the original regression, corrected by heteroskedasticity.

4.8 Loan Application & Approval Process

Households apply for a loan by telling the bank their desired expenditure P and desired LTV l^*, where l^* is drawn from an empirical distribution of LTV conditional on expenditure size (in bins of \$25,000) based on DC CoreLogic data. An example such a distribution is displayed in Figure 4.8.1
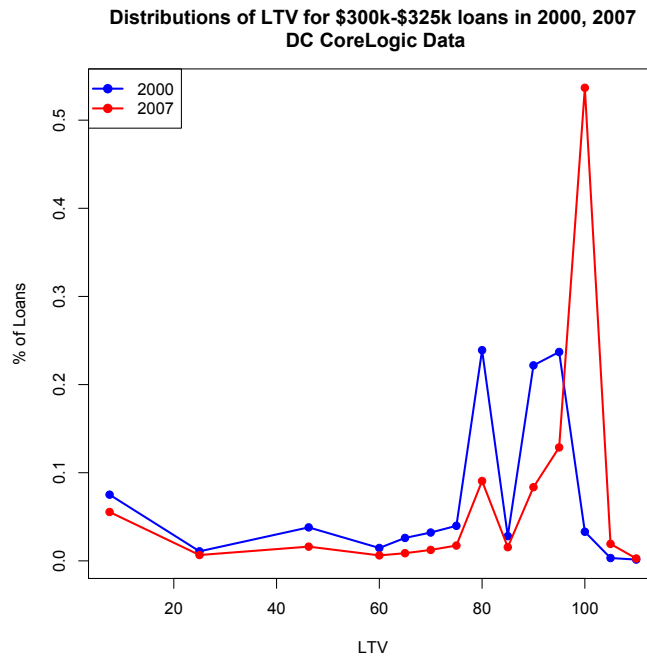

Figure 4.8.1: Distribution of LTV for \$300,000-\$325,000 loans, 2000 and 2007, DC CoreLogic Data

The bank bargains over LTV and debt-to-income (DTI) in order to keep issued loans within constraints that reflect empirical data. The bank's LTV constraint is set as the third-quartile of LTV of contemporaneous issued loans in DC, from CoreLogic data. In determining whether to issue a loan, the bank proceeds as follows:

1. Determine whether the applicant can afford closing costs if it buys the house.

2. If so, bargain over LTV so that it's lower than the bank's LTV limit while high enough that the household can afford the implied down payment.

3. If step 2 succeeds, compute a hypothetical DTI for a fixed-rate loan. Using the hypothetical DTI, select an actual loan rate and type from an empirical distribution based on CoreLogic data. There is one bin distribution of types for each modeled year, providing percentage loans of each type Fixed, ARM, or Interest-only, conditional on 10%-bins of DTI. Rates are drawn based on an empirical distribution of spreads from the prime rate conditional on loan type and 10%-bins of DTI, also based on CoreLogic data.

4. Compute the actual DTI based on the selected loan rate and type; re-bargain on LTV if the DTI is over the bank's DTI constraint.

Bargaining at point 2 proceeds as follows: defining $\bar{l}$ as the maximum of the bank's LTV constraint and $\underline{l}$ as the minimum LTV at which the household can afford the downpayment,

- If $\underline{l} > \bar{l}$, the application is immediately rejected.

- If $l^* \in [\underline{l}, \bar{l}]$, the application can continue

- If $\bar{l} < l^*$, reduce desired LTV to $l^* = \bar{l}$

- If $l^* < \underline{l}$, raise desired LTV to $l^* = \underline{l}$

Bargaining at point 4 adjusts LTV, if necessary, based on the DTI constraint rather than the LTV constraint. For a given type and rate of loan, DTI is determined as an implicit and increasing function of LTV,

$$d = d(l).$$

The bargaining proceeds as follows:

- If $d(\bar{l}) > \bar{d}$, the application is immediately rejected, as even the highest possible down-payment given LTV constraints cannot satisfy DTI constraints.

- If $d(l^*) \leq \bar{d}$, the application is successful.

- If neither of the above, change LTV to $l^* = d^{-1}(\bar{d})$.

If the loan approval process succeeds, the applicant is given a loan with principal

$$P \times (1 - l^*),$$

and is allowed to buy a house in the market that costs P or less. Notice that P, the household's desired expenditure, was set by the desired expenditure formation rule and was never bargained over.

4.9   Loan Transition Process

Loans in the model transition between being current, delinquent, defaulted, and liquidated; when they are completely paid off, they cease to exist. Current loans that are paid in full each month stay current; they are paid off completely after refinance, at the end of their 30-year term, or when a household sells the house. A loan that is not paid in full in a given month goes delinquent. Households fail to pay either due to insufficient funds or by deciding to strategically default on a loan. A loan delinquent two-or-more-months is considered "in default" by the bank which immediately tries to foreclose and liquidate the loan.

Refinance Decision

There are two kinds of refinancing available to the household: rate refi and cash-out refi. Rate refinancing in the model means getting a new loan from the bank of the same size as the current loan but with lower monthly payments. A household that cannot afford his monthly payment always tries to refinance. To rate refi the household must have enough liquid wealth to pay the $5000 refinance fee, must not be underwater, must not have defaulted, and must not have listed its house for sale. If all those conditions apply, the household tries a refinance with probability based on the moneyness of the loan, where the moneyness is defined as

$$\text{Moneyness} = \text{Current Prime Rate} - \text{Rate Paying}.$$

The probability of refinancing conditional on moneyness is an empirical distribution conditional on type of loan, based on CoreLogic data. For example, for fixed loans, the probability of refinancing in a given month is displayed in Figure 4.9.1
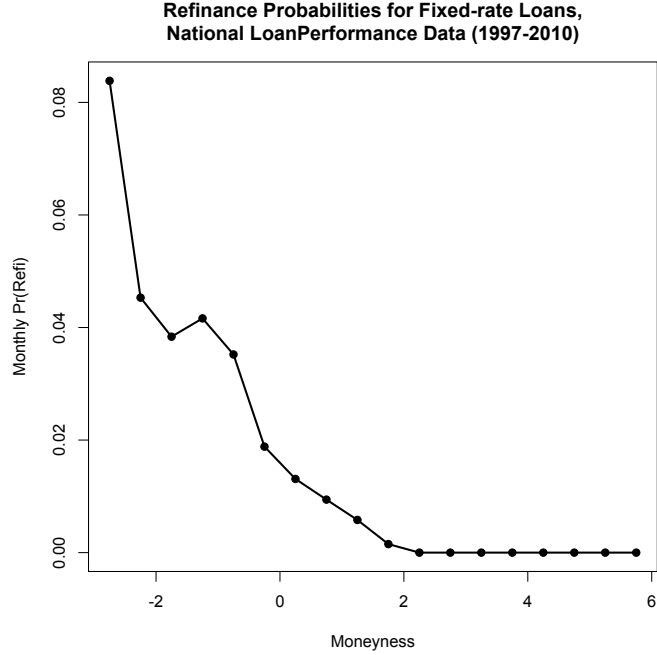
Figure 4.9.1: Monthly Refinance Probabilities for Fixed-rate Loans, National LoanPerformance Data (1997-2010)

If the household decides to refinance, the bank draws a new loan and offers it the Household (4.8). If the new loan has higher monthly payments than the old one, the household rejects it and doesn't refi.

If the household doesn't rate-refi, it can cash-out refi. A cash-out refi consists in obtaining a new loan with the same LTV as the old loan at its origination, paying back the old loan, and receiving the difference as a credit to liquid wealth. A household can cash-out refi only if the house is not listed and there would be a positive cash-flow after paying back the old loan and the \$5000 refinance fee. The household wants to cash-out refi with probability:

$$\text{Prob} = k \cdot \frac{\bar{l} - l}{\bar{l}},$$

where

$$\bar{l} = \text{Khandani Refinance Max CLTV}$$
$$k = \text{Khandani Refinance Linear Parameter}$$
$$l = \text{Current LTV of this Loan},$$

and k and $\bar{l}$ come from [Khandani et al., 2012].

Strategic Default

We model strategic default as a functional form specifying the probability of strategic default by an agent with a loan at given current combined loan-to-value (CCLTV). It is a well-known phenomenon among the non-Agency RMBS community that strategic default rates

are increasing in the CCLTV of the borrower. The percent chance of strategic default is

$$P(\text{Strategic Default}) = \max\{CCLTV - 0.5, 0\} \times (0.01 - 0.02 \times HPA),$$

where HPA is lagged (prior year) Case-Shiller home price appreciation. Only agents with $CCLTV \geq 50\%$ can strategically default. Coefficients were estimated using LoanPerformance data in DC (1997-2010).

Foreclosure & Liquidation

Banks succeed in liquidating a delinquent loan with a probability based on the number of months of delinquency. These probabilities are drawn from empirical liquidation rates in the same DC LoanPerformance data. Because data on old loans was limited, we artificially smooth the probability of liquidation success for very old loans. The monthly success probabilities and CDF of liquidation success are displayed in Figure 4.9.2.

Monthly Pr(Liq|#MonthsDQ), Modified & Original, Loan Performance Data

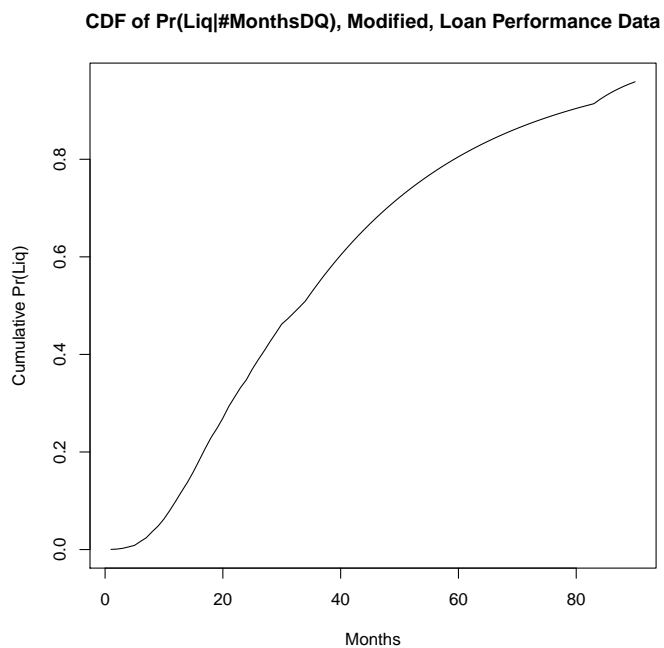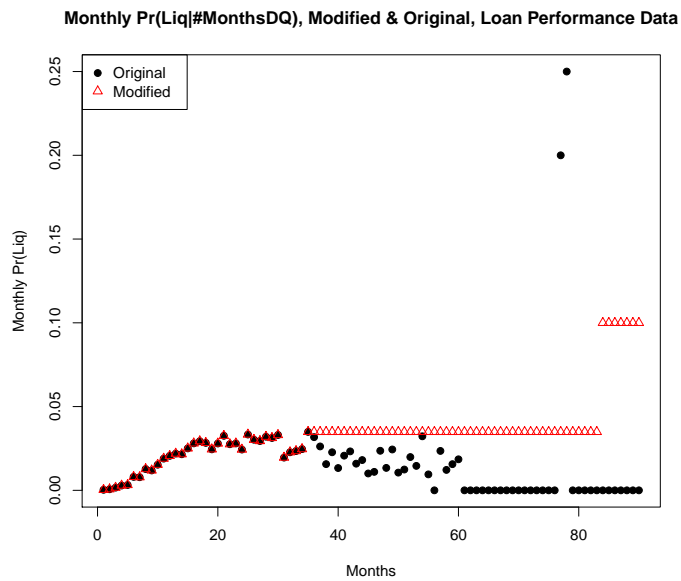**CDF of Pr(Liq|#MonthsDQ), Modified, Loan Performance Data**

Figure 4.9.2: Actual and Smoothed Delinquent Loan Liquidation Probabilities, DC Loan-Performance Data (1997-2010)

Once a home is foreclosed, the household is kicked out of the house, the loan is canceled and the house is listed on the market as a new house (4.11). A household that has been foreclosed upon is not allowed to apply for a mortgage for 7 years.

17

4.10 Household Demographics

The total number of households is pegged to Census data. Each month, households die according to probabilities drawn from an actuarial table. The houses of the deceased households are listed as new homes (4.11). The model then checks the population against Census Data. If there are not enough households, the model generates new households, assigning them random income and wealth and no home.

4.11 Housing Market

House Listing and Delisting

A household can list any of its homes on the market. The probability of listing is estimated from MLS data. Each household has a fixed probability of listing its home equal to:

$$\frac{\text{\# of new listings this month in the Data}}{\text{\# of non-listed houses in the model this month}}$$

If the household lists any of its homes, it chooses the list price to be:

$$\exp\left[0.22 + 0.99 \times \log(\bar{p}) + 0.22 \times \log(s) - 0.01 \times \log(\text{DOM}) + \epsilon\right],$$

where

$$
\begin{aligned}
s &= \text{ average sold price to OLP} \\
\text{DOM} &= \text{ average days on market} \\
\epsilon &= \text{random noise} \\
\bar{p} &= \text{average list price of 8 closest-quality homes sold in past 6 months}
\end{aligned}
$$

These parameters were estimated with MLS data.

For any house that was previously listed, the household can decide to delist it or change its price. The hazard rate of delisting is a function of months spent on market, estimated with MLS data. If the house is not delisted, the household uses a lookup table to decide whether to change its price. The lookup table is a function of time on market and number of previous price changes. If the household does want to change the price of the house, it reads a the change in markup from MLS data, again a function of time on market and number of previous price changes.

The house list price can never be below the total amount needed to payoff the mortgage.

House Supply

The total number of houses is pegged to Census data. At the beginning of each month, the model checks the difference between the number of houses in the model and the number implied by Census data. If there are not enough houses, the model "builds" them, drawing random qualities for new houses according to the original quality distribution. If there are too many, the model destroys some, preferring to destroy foreclosed or non-owned homes.

Houses that are just built are immediately listed for sale at fair value (quality times current Case Shiller index). Every three months the house is left unsold, its price drops 5%.

Apartments are never bought or sold, do not have an owner, and are always for rent.

Rental Market

Apartments and investor-owned houses are rented to households who cannot afford or choose not to buy a home. The rental price is set as a fraction of the quality of the house according to the following formula.

$$\text{Real Rental Growth} = \frac{\rho_{t+1}/CPI_{t+1}}{\rho_t/CPI_t} - 1 = \frac{\rho_{t+1}}{\rho_t} \times \frac{1}{1 + \pi_{t+1}} - 1 = -\beta(v_t - v^*) + \alpha,$$

where

- ρ_t is the ratio of rent to quality in year t

- π_{t+1} is the inflation rate $= \dfrac{CPI_{t+1}}{CPI_t} - 1$ [where CPI = Consumer Price Index]

- v_t is the contemporaneous in-model vacancy rate

- v^* is actual statewide average vacancy rate, computed from ACS.

- $\alpha = 0.02, \beta = 0.84$ are regression-estimated adjustment coefficients, estimated from ACS data from 2000-2010.

Renters are willing to spend up to 30% of income on rent.

Market Clearing

All pre-approved first-time buyers are ranked by their desired expenditure. The buyer with the highest expenditure acts first. The buyer chooses the highest quality house it can afford, and buys it. If the house was owned by another household, that household is credited the sale price, and uses the payment to pay off the old mortgage on the sold house. Then, the second-highest new buyer acts, and so on until there are no more buyers or no more houses for sale.

After all buyers have acted, if there are any houses still for sale, each investor is matched at random with one of them. The investor decides whether to buy it by computing the expected yield on the house:

$$\text{Yield} = \frac{\text{Rent} + \text{Appreciation} - \text{Housing Costs} - \text{Mortgage Payments}}{\text{Downpayment}}$$

Given the yield, the investor buys the house with probability

$$\frac{1}{1 + exp(-(Yield \times 24.0 - 4.5))}$$

The investor can buy the house for cash if it has enough liquid wealth; otherwise it has to apply for a loan (4.8).

Finally, all the buyers who failed to find a house and all the households who failed to get approved for a loan (4.5) try to rent a rentable house, i.e. an apartment or investor-owned home. Renters are sorted by willingness to pay rent, and matched to highest-quality affordable rentals as with buyers. They rent for a random number of months uniformly distributed from 1 to 12.

Any remaining households are considered homeless for the next month.

5 Model Output in Comparison to the Historical Data

In this section we compare the output from a typical run of our model, scaled down to one software agent for each real household, to actual housing price data on the Washington, D.C. MSA. Model sensitivity and lacunae are then briefly discussed.

5.1 Typical Results from a Single Run of the Model

Figure 5.1.1 shows the results of a typical single simulation of the model. Reading across and down, the first figure shows the actual Case-Shiller price index over time (dotted line) and the price index emanating from the model (solid line). Note that the model gets only about 2/3 of the actual price rise and has the bubble bursting some 18-24 months after the actual price peak. Average sale price (second panel) is similarly depressed in the model output and seems to display more high frequency fluctuations than the actual data. This may be a consequence of this run being made at 10:1 scaling, as previously mentioned. The ratio of sold price to original listing price (panel three) has approximatley the right shape, again with lots of variability. The number of active listings (first figure, second row) leaves much to be desired, as it appears to display a kind of seasonality that is not really present in the behavioral rules. The number of units sold (middle figure) is in rough agreement with the data. Days-on-market (last figure in the second row) has the right shape but too much volatility, perhaps again a consequence of the scale factor being used in this run of the model. Real estate market inventory (third row, first figure) has the right shape, perhaps unsurprising since the days-on-market is coming out about right, but again too much variability. The last two figures show that neither the home ownership rate nor the vacancy rate are well predicted by the model.
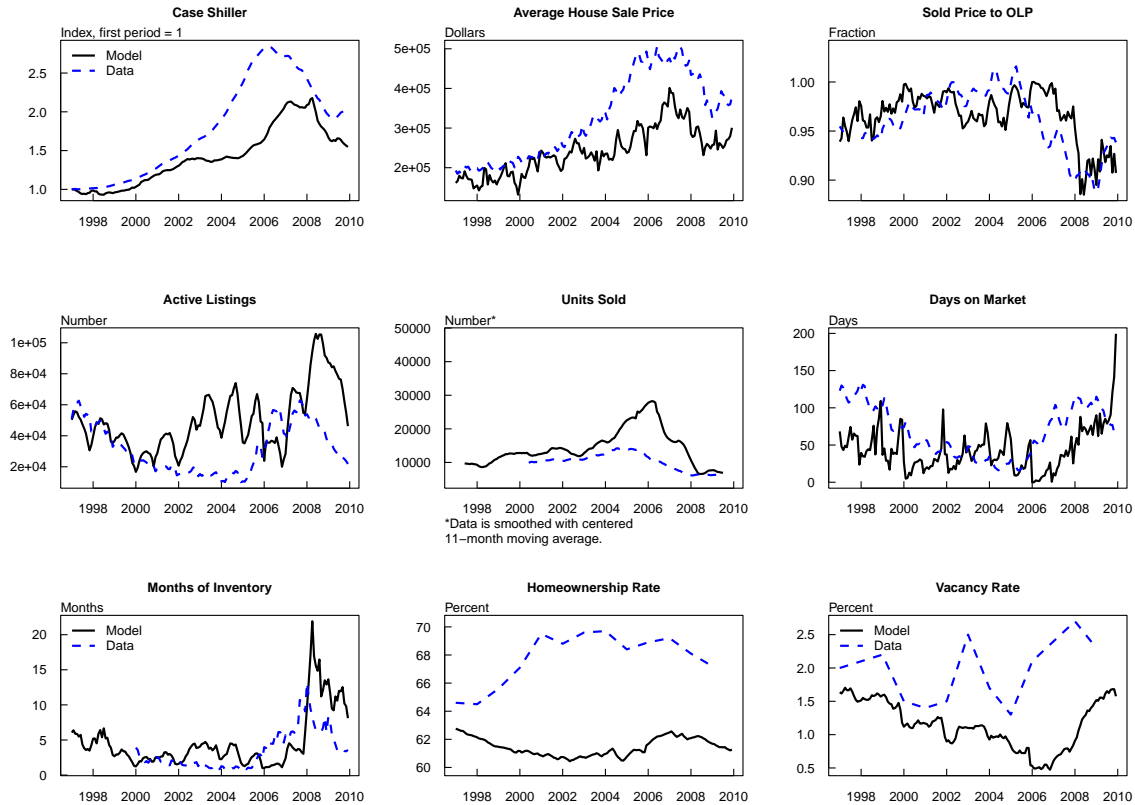
Housing Market Results

Figure 5.1.1: The results of a run with the calibrated model

Ovearall, we feel this represent a good start on building an empirically-accurate microeconomic model, but clearly there is much work to be done to improve its predictive power.

5.2 Limitations of the Current Model

Several facets of real housing markets have been left out of the current model, some of which may account for the inaccuracy of the model's output. Among the most significant of these are:

- Seasonality: in the model, households act the same way in January as they do in May. However, the housing market is actually very seasonal. In particular, there are more new listings in the spring, a feature which enters the model through the housing supply data. But demand in the model is non-seasonal, creating an artificial discrepancy that leads to an oversupply of houses in the spring that takes months to clear.

- Lending constraints and character of the boom: In the model we have a bubble and an increase in prices, but the home ownership rate actually decreases. In the model,

21

high prices drive some people out of the market, despite lowered banking constraints. We need to do further work to match real banking constraints over time.

- Role of real estate agents: Our household agents perform their own search for feasible homes to buy. In reality it is real estate agents who mediate this process to a great extent, alerting prospective buyers of new properties that have come on the market along with recent price reductions, while providing general information on market conditions. Such agents also advise sellers concerning reasonably sale price expectations, comparable properties on the market, and so on. In future versions of the model we may include such agents.

- Role of mortgage brokers: Many mortgages are originated not by prospective borrowers approaching banks but by specialist agents in the mortgage market who link home buyers to lenders. Such broker agents have more knowledge of financial products and bank standards than do typical homebuyers, and can therefore taylor mortgage products to individual household needs. Such agents could be included in future efforts with this model.

- There is no neighborhood or spatial structure in the extant model, and surely neighborhood amenities and commuting distances and durations are important facets of any household's home purchase decision. But it is unclear to use whether this matters significantly for the existence and character of home price bubbles. At any rate, we hope to be able to explore a version of our model having some simple kind of spatial representation, to see how it affects our basic results, e.g., producing stronger bubbles in some areas, weaker ones in others?

- Role of developers: New homes in the Washington, D.C. area are mostly the result of home builders realizing there is a market opportunity for certain kinds of homes in certain kinds of neighborhoods. At this point our model does not have sufficient fidelity to represent such decision-making properly, but the role of such marginal increases in the housing stock may be significant in determining price changes in housing markets. A future version of the model may include this class of agents.

- The extant model has been fit to data on one city only, and that city experienced neither the largest nor smallest bubble in the last decade. A considerable amount of the data required to instantiate the model for other cities has been acquired and we hope in the very near future to see if our basic model works for the medium-sized bubbles experienced by Boston and San Francisco. Then, with the model working for multiple cities, we hope to build versions both for cities that had larger bubbles (e.g., Phoenix and Las Vegas) and to those which experienced very little price rise (e.g., Dallas and various midwestern cities).

5.3   Sensitivity

The model results depend crucially on the changes in bank lending constraints through the boom and bust. Therefore, the way these constraints are defined matters. Figure 5.3.1 displays the effects of loosening LTV constraints on the Case-Shiller index in the model.
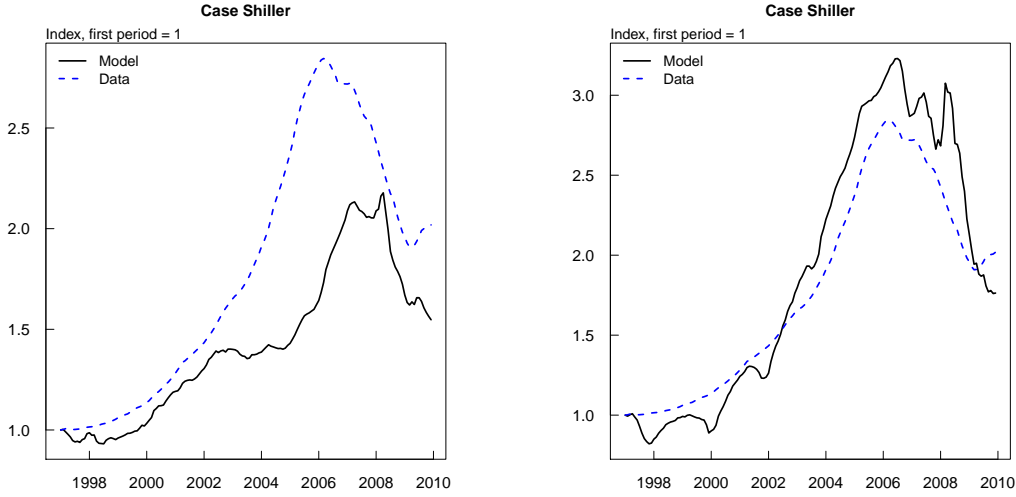
(a) Case-Shiller index, default LTV constraints (b) Case-Shiller, exaggeratedly loose constraints

Figure 5.3.1: Effects on Case-Shiller of loosening borrowing constraints.

Since the model is sensitive to LTV constraints, it is also sensitive to the amount of households' liquid wealth, since their liquid wealth determines what they can afford as downpayments. The model's 100-year burnout initialization period therefore makes the nonhousing consumption rule particularly important. Even a moderate increase in the annual savings rate can develop into a large cash reserve over the 100-year burnout period, which in turn negates the effect of the LTV constraints.

The model is also sensitive to the HPA-effect parameter a in the desired expenditure rule. This parameter determines to what extent agents modify their housing expenditure in response to past home price appreciation. Figure 5.3.2 shows Case-Shiller in the model with varying values of a, with the model default $a = 0.16$ in the middle.
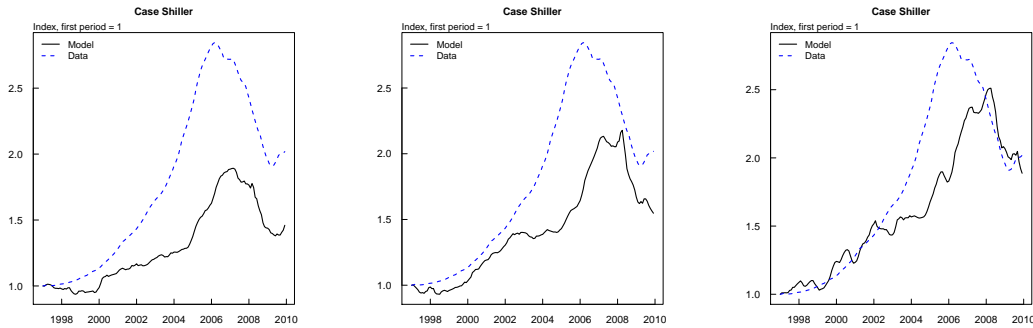

(a) Case-Shiller, $a = 0.08$ (b) Case-Shiller, $a = 0.16$ (c) Case-Shiller, $a = 0.24$

Figure 5.3.2: Effects on Case-Shiller of altering a, the HPA-effect parameter.

23

6 Policy Experiments

An important feature of agent-based models is that they allow systematic study of counterfactuals. We can investigate questions such as whether more stringent lending constraints or higher interest rates might have prevented the bubble. In Figure 6.0.3, we show output for the base case of the model (a) presented above but with interest rates fixed at 1997 levels (b), i.e., no low rates following the NASDAQ bubble collapse. Note that the bubble is attenuated but not eliminated. In (c) rates return to the base case but now leverage (LTV) is limited to 1997 levels, and the bubble is significantly reduced. The final panel (d) has both of these rate and leverage policies active. We see that more stringent lending constraints would have effectively eliminated the bubble.
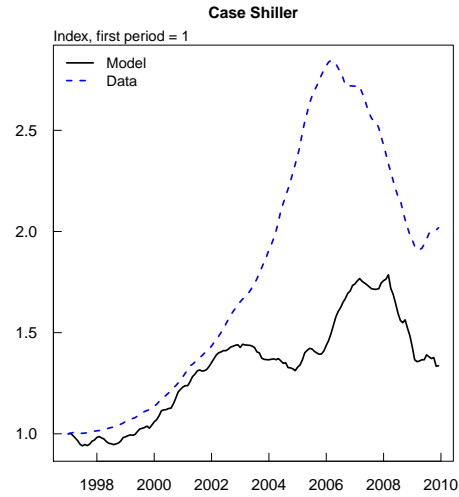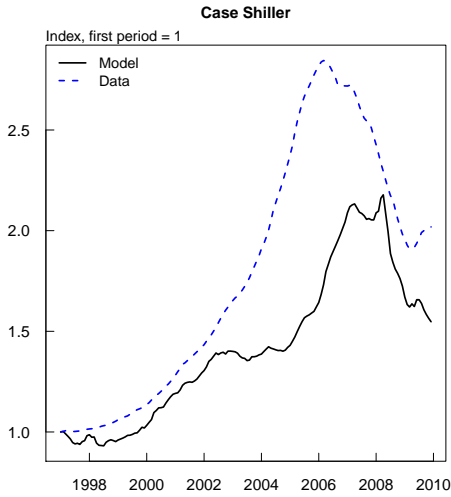
(a) The baseline simulated Case-Shiller index   (b) with interest rates fixed at 1997 values.
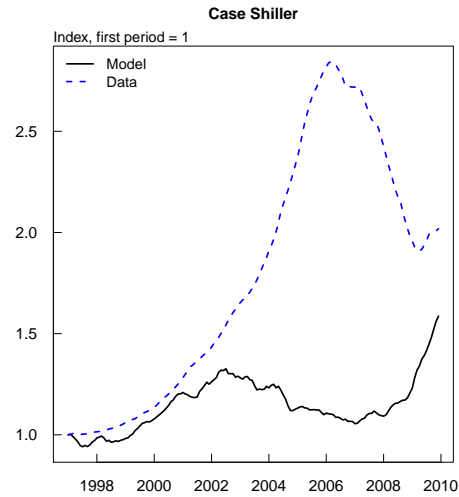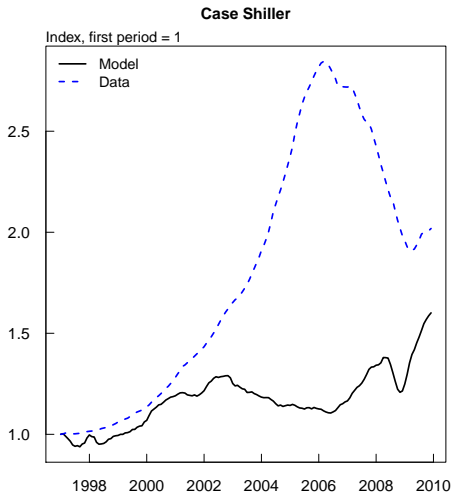
(c) with LTV constraints fixed at 1997 values   (d) with LTV constraints and interest rates fixed

Figure 6.0.3: Simulated counter-factuals under different policies

7   Conclusion

We have presented a micro-founded model of the housing market for the Washington, D.C. metro area that qualitatively reproduces the recent boom and bust. The model features explicit specifications of household behavior that have been independently estimated from certain micro data. Running the model yields outputs that can be directly compared to other micro data on real estate transactions, mortgage servicing, and so on. At this point, instead of being the 'last word' on this subject we suggest that it is rather the first foray

into a new an fertile paradigm of combining rich micro-data with a highly supple and flexible approach to computational modeling, to yield a new way of building housing market models. The model is in an active state of development at this time and there is much that we can do to improve its performance. For example, we can endogenize currently exogenous components such as housing construction and banking constraints. As the response of developers to changing home prices almost certainly plays a role in bubble dynamics, we need to make new home construction sensitive to home prices rather than setting housing construction using fixed census data. Banking constraints should also be modeled endogenously if possible. Doing so would help make policy experiments more realistic. As the government has more control over GSEs than it does over other financial agents, we need to study whether LTV limits would still dampen the bubble if they could be imposed only on a minority of lenders. We also need to study how the model performs in other cities; repeated good performance would legitimize the design of the model and our approach to estimation and provide further basis for conducting policy experiments. Finally, we need to consider whether this model could itself be embedded in a larger agent-based macroeconomic model, to provide a richer representation of households than is conventional in the current generation of DSGE models. Among other things, doing so would require that we further endogenize some components of the module, such as income and unemployment.

In summary, we have built a detailed computational representation of a specific housing market, using novel computing techniques (i.e., multi-agent systems) adapted to model economic processes (e.g., software objects as agents). We have demonstrated that such a model, when calibrated using micro-data, can do a reasonable job reproducing historical data. We believe we have only scratched the surface of the pregnant interface of comprehensive micro-data and agent-based modeling. Much work remains to be done in order to produce models that can simultaneously serve as explananations of housing dynamics, forecasting engines for future price movements, and decision support systems for policy-makers tasked with insuring that mortgage and housing markets operate in orderly fashion.

References

[Carroll, 1992] Carroll, C. D. (1992). The buffer-stock theory of saving: Some macroeconomic evidence. *Brookings Papers on Economic Activity*, 23(2):61–156.

[Chinco and Mayer, 2012] Chinco, A. and Mayer, C. (2012). Distant speculators and asset bubbles in the housing market. *Columbia Business School mimeo, April.*

[Duca et al., 2011] Duca, J. V., Muellbauer, J., and Murphy, A. (2011). Shifting credit standards and the boom and bust in US house prices. SSRN Scholarly Paper ID 1830979, Social Science Research Network, Rochester, NY.

[Geanakoplos, 2001] Geanakoplos, J. (2001). Liquidity, default and crashes: Endogenous contracts in general equilibrium. SSRN Scholarly Paper ID 284829, Social Science Research Network, Rochester, NY.

[Geanakoplos, 2010a] Geanakoplos, J. (2010a). The leverage cycle. *NBER Macroeconomics Annual*, 24(1):1–66.

[Geanakoplos, 2010b] Geanakoplos, J. (2010b). Solving the present crisis and managing the leverage cycle. SSRN Scholarly Paper ID 1539488, Social Science Research Network, Rochester, NY.

[Gilbert et al., 2009] Gilbert, N., Hawksworth, J. C., and Swinney, P. A. (2009). An agent-based model of the english housing market. In *AAAI Spring Symposium on Technosocial Predictive Analytics. Standford, California: AAAI*, pages 30–35.

[Glaeser et al., 2012] Glaeser, E. L., Gottlieb, J. D., and Gyourko, J. (2012). Can cheap credit explain the housing boom? NBER chapters, National Bureau of Economic Research, Inc.

[Haughwout et al., 2011] Haughwout, A., Lee, D., Tracy, J., and van der Klaauw, W. (2011). Real estate investors, the leverage cycle, and the housing market crisis. Technical Report 514, Staff Report, Federal Reserve Bank of New York.

[Iacoviello, 2005] Iacoviello, M. (2005). House prices, borrowing constraints, and monetary policy in the business cycle. *The American Economic Review*, 95(3):739–764.

[Khandani et al., 2012] Khandani, A. E., Lo, A. W., and Merton, R. C. (2012). Systemic risk and the refinancing ratchet effect. *Journal of Financial Economics*.

[Kiyotaki et al., 2011] Kiyotaki, N., Michaelides, A., and Nikolov, K. (2011). Winners and losers in housing markets. *Journal of Money, Credit and Banking*, 43(2-3):255–296.

[Landvoigt et al., 2012] Landvoigt, T., Piazzesi, M., and Schneider, M. (2012). The housing market(s) of san diego. Working Paper 17723, National Bureau of Economic Research.

[Leeuw, 1971] Leeuw, F. d. (1971). The demand for housing: A review of cross-section evidence. *The Review of Economics and Statistics*, 53(1):1–10.

[Mian and Sufi, 2009] Mian, A. and Sufi, A. (2009). The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis. *The Quarterly Journal of Economics*, 124(4):1449–1496.

[Ortalo-Magné and Rady, 2006] Ortalo-Magné, F. and Rady, S. (2006). Housing market dynamics: On the contribution of income shocks and credit constraints. *The Review of Economic Studies*, 73(2):459–485.

[Reinhart and Rogoff, 2011] Reinhart, C. M. and Rogoff, K. S. (2011). *A Decade of Debt*. Number 6222 in Peterson Institute Press: All Books. Peterson Institute for International Economics.

[Schelling, 1971] Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

[Tesfatsion and Judd, 2006] Tesfatsion, L. S. and Judd, K. L. (2006). Handbook of computational economics.