

**GETTING AT SYSTEMIC RISK VIA AN AGENT-BASED MODEL
OF THE HOUSING MARKET**

By

**John Geanakoplos, Robert Axtell, Doyne J. Farmer, Peter Howitt,
Benjamin Conlee, Jonathan Goldstein, Matthew Hendrey,
Nathan M. Palmer, and Chun-Yi Yang**

March 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1852



**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281**

<http://cowles.econ.yale.edu/>

Getting at Systemic Risk via an Agent-Based Model of the Housing Market¹

by
John Geanakoplos
(Yale University, Santa Fe Institute, Ellington Management Group)

Robert Axtell (George Mason University)
Doyne J. Farmer (Santa Fe Institute)
Peter Howitt (Brown University)
Benjamin Conlee (Ellington Capital Management)
Jonathan Goldstein (George Mason University)
Matthew Hendrey (George Mason University)
Nathan M. Palmer (George Mason University)
Chun-Yi Yang (George Mason University)

Abstract

Systemic risk must include the housing market, though economists have not generally focused on it. We begin construction of an agent-based model of the housing market with individual data from Washington, DC. Twenty years of success with agent-based models of mortgage prepayments give us hope that such a model could be useful. Preliminary analysis suggests that the housing boom and bust of 1997-2007 was due in large part to changes in leverage rather than interest rates.

Key Words:

Agent Based Models, Housing Prices, Boom and Bust, Leverage, Interest Rates, Foreclosures, Systemic Risk

JEL Classification Codes:

E3, E31 E32, E37, E44, E63, R2, R20, R21, R23, R28, R3, R30, R31, R38

¹ This material is based upon work supported in part by grants from the National Science Foundation (No. 0965673), the Institute for New Economic Thinking, and the Alfred P. Sloan Foundation. We gratefully acknowledge the support of Ellington Management Group, CoreLogic, and Real Estate Business Intelligence, LLC.

Macroeconomists were completely surprised by the financial crisis of 2007-2009. Up until then they had concentrated on macroeconomics with a capital M: global imbalances, interest rates, monetary policy. Few foresaw the central role the housing market and mortgage securities would play in the crash. It is now universally recognized that real estate and housing bubbles played an important role in the most recent financial crisis and in many others besides. No monitoring of systemic risk can pretend to be satisfactory if it does not include a model of the housing market.

Not only were the Macroeconomists looking at the wrong markets, they might have been looking at the wrong variables. John Geanakoplos (2003, 2010a, 2010b) has 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. Geanakoplos has also argued that the best way out of the crisis is to write down principal on housing loans that are underwater (see Geanakoplos-Koniak (2008, 2009) and Geanakoplos (2010b)), on the grounds that the loans will not be repaid anyway, and that taking into account foreclosure costs, lenders could get as much or almost as much money back by forgiving part of the loans, especially if stopping foreclosures were to lead to a rebound in housing prices.

There is, however, no shortage of alternative hypotheses and views. Was the bubble caused by low interest rates, irrational exuberance, low lending standards, too much refinancing, people not imagining something, or too much leverage? Leverage is the main variable that went up and down along with housing prices. But how can one rule out the other explanations, or quantify which is more important? What effect would principal forgiveness have on housing prices? How much would that increase (or decrease) losses for investors? How does one quantify the answer to that question?

Conventional economic analysis attempts to answer these kinds of questions by building equilibrium models with a representative agent, or a very small number of representative agents.

Regressions are run on aggregate data, like average interest rates or average leverage. The results so far seem mixed. Edward Glaeser, Joshua Gottlieb, and Joseph Gyourko (2010) argue that leverage did not play an important role in the run-up of housing prices from 2000-2006. John Duca, John Muellbauer, and Anthony Murphy (2011), on the other hand, argue that it did. Andrew Haughwout et al (2011) argue that leverage played a pivotal role. In our view a definitive answer can only be given by an agent-based model, that is, a model in which we try to simulate the behavior of literally every household in the economy. The household sector consists of hundreds of millions of individuals, with tremendous heterogeneity, and a small number of transactions per month. Conventional models cannot accurately calibrate heterogeneity and the role played by the tail of the distribution.

Though the Haughwout et al (2011) model moves significantly in the direction of recognizing heterogeneity, only after we know what the wealth and income is of each household, and how they make their housing decisions, can we be confident in answering questions like: How many people could afford one house who previously could afford none? Just how many people bought extra houses because they could leverage more easily? How many people spent more because interest rates became lower? Given transactions costs, what expectations could fuel such a demand? Once we answer questions like these, we can resolve the true cause of the housing boom and bust, and what would happen to housing prices if principal were forgiven.

Conventional thinking suggests that an agent-based model of the housing market is an impossibly ambitious task. We would need too much data. It all depends on arbitrary behavioral rules, each of which depends on too many parameters to estimate reliably. Without the discipline of equilibrium, expectations cannot be pinned down. As the world changes, what seemed like appropriate behavioral rules will be revealed to be crazy. Against this we have the basic argument that the agent-based approach brings a new kind of discipline because it uses so much more data. Aside from passing a

basic plausibility test (which is crucial in any model), the agent-based approach allows for many more variables to be fit, like vacancy rates, time on market, number of renters versus owners, ownership rates by age, race, wealth, and income, as well as the average housing prices used in standard models. Most importantly, perhaps, one must be able to check that basically the same behavioral parameters work across dozens of different cities. And then at the end, one can do counterfactual reasoning: what would have happened had the Fed kept interest rates high, what would happen with this behavioral rule instead of that.

The real proof is in the doing. Agent-based models have succeeded before in simulating traffic and herding in the flight patterns of geese. But the most convincing evidence is that Wall Street has used agent-based models for over two decades to forecast prepayment rates for tens of millions of individual mortgages.

In this paper we describe the prepayment modeling Geanakoplos directed on Wall Street as head of fixed income research at Kidder Peabody and then as head of research at Ellington Capital Management. We shall see how the Kidder model began as a conventional aggregate model and evolved into an agent based model, even before detailed loan level information became available. Some of the criticisms of agent-based modeling can be seen to have merit; the model that worked in the 1990s needed substantial changes after 2000. But the approach was still generally superior to aggregate modeling.

Next we turn to the more ambitious housing model. So far we only have data on the greater Washington DC area, not the whole country. But that includes over 2.2 million households. Furthermore our modeling is still at a preliminary stage. But results we have obtained so far suggest that leverage, and not interest rates, played the dominant role in the housing boom and bust from 1997-2009.

I. The Mortgage Prepayment Problem

There are approximately 55 million first mortgages extant in the United States today. In each mortgage the borrower is obliged to make a monthly coupon payment, calculated either as a fixed

percentage of the original balance, or as a variable percentage of the original balance, for a period of years, often 30, though occasionally 15 or even shorter. Nearly every single one gives the borrower an option each month to pay off the remaining balance (or prepay) in lieu of all future payments. Prepayments radically change the cash flows and valuation of the mortgages to the holder. Partly to reduce this risk, starting in the 1970s, mortgages were aggregated into pools; shares of the pools were then sold as securities in a process called securitization. Typically pools consisted of a large number of individual mortgages issued at approximately the same time, with approximately the same payment conditions (e.g. fixed rate mortgages of about 8% issued in the first half of 1986).

IA. The Conventional Approach

Mortgage prepayment modeling was pioneered by academics, and then taken up in a series of remarkable working papers by various Wall Street investment banks throughout the late 1980s and early 1990s. Typical of those models was one at Kidder Peabody. Since the purpose of pooling was to diversify the risks inherent in any homeowner's decision, it seemed perfectly natural to the pioneers of prepayment modeling to use statistical methods to predict the aggregate prepayment of each pool directly, even though those prepayments are the sum of individual homeowner prepayments. An apparently decisive argument for this approach was that it was much simpler to predict one number than 10,000 numbers, and that the individual outcomes were not available in the data anyway. In the first Figure the solid line is an example of a prepayment history (expressed at an annualized rate) up to 1999 on the pool of 1986 Fannie Mae 8% coupons. Homeowners do not prepay optimally. Historical aggregate prepayments are never exactly 100% or 0%; in fact, they are rarely more than 10% in any month. So the conventional approach looks for common sense *patterns* in the data, and tries to refine the common sense by estimating parameters. It is rather like an agent based approach with only one agent per pool.

Macroeconomists in the early 1980s would predict prepayments one year ahead according to how they thought the macro-economy was going, and whether that meant interest rates were going up or down. Starting in the late 1980s it became clear that one had to make ten year predictions or more, and that these had to be *conditional* predictions: if interest rates and housing prices follow such and such a path, then prepayments will do such and such. Is it imaginable that standing in 1986 in Figure 1, one could predict the future path of prepayment rates, if one knew the future path of interest rates and housing prices?

The conventional model essentially reduced to estimating an equation with an assumed functional form for prepayment rate

$$\text{Prepay}(t) = f(\text{age}(t), \text{seasonality}(t), \text{old rate} - \text{new rate}(t), \text{burnout}(t), \text{parameters}),$$

where old rate – new rate is meant to capture the benefit to refinancing at a given time t , and burnout is the summation of this incentive over past periods. Mortgage pools with large burnout tended to prepay more slowly, presumably because the most alert homeowners prepay first.

IB. The Agent-Based Model

Note that the conventional prepayment model uses exogenously specified functional forms to describe *aggregate* behavior directly, even when the motivation for the functional forms, like burnout, is explicitly based on heterogeneous individuals. By contrast, the new prepayment model Geanakoplos and his team developed at Kidder Peabody in the early 1990s and then refined many times at Ellington beginning in 1995 starts from the *individual* homeowner and in principle follows every single individual mortgage. It produces aggregate prepayment forecasts by simply adding up over all the individual agents. Each homeowner is assumed to be subject to a *cost c* of prepaying, which include some quantifiable costs such as closing costs, as well as less tangible costs like time, inconvenience, and psychological costs. Each homeowner is also subject to an *alertness* parameter **a**, which represents the probability the agent is

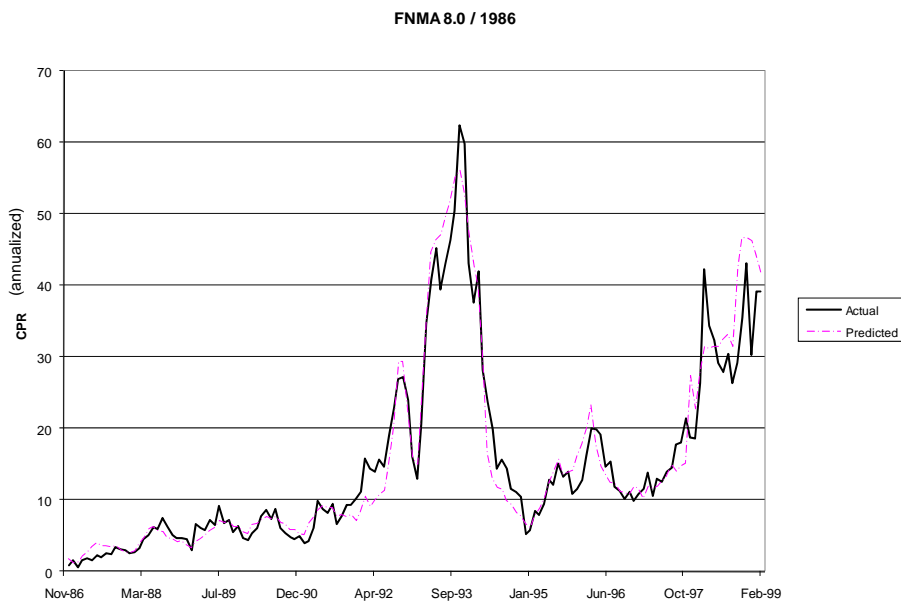
paying attention each month. The agent is assumed aware of his cost and alertness, and subject to those limitations chooses his prepayment optimally to minimize the expected present value of his mortgage payments, given the expectations that are implied by the derivatives market about future interest rates. Agent *heterogeneity* is a fact of nature. It shows up in the model as a distribution of costs and alertness, and turnover rates. Each agent is characterized by an ordered pair (c,a) of cost and alertness, and also a turnover rate t denoting the probability of selling the house. The distribution of these characteristics throughout the population is inferred by fitting the model to past prepayments. The effects of observable borrower characteristics can be incorporated in the model (when they become available) by allowing them to modify the cost, alertness, and turnover.

To capture burnout with a parameterized curve, one must have already identified the phenomenon and build it into the curve. In contrast, burnout is a natural consequence of the agent-based approach; there is no need to add it in afterwards. The agents with low costs and high alertness prepay faster, leaving the remaining pool with slower homeowners, automatically causing burnout. The same heterogeneity that explains why only part of the pool prepays in any month also explains why the rate of prepayment burns out over time.

The model has a number of other parameters, including the seasonality and age parameters as in the conventional model. The world also does not stand still; technology improves, and people become more sophisticated. The “*smart factor*” quantifies the rate at which costs decline and alertness increases over time as prepayment behavior becomes more rational. The smart factor allows the model to fit for more than a decade, as seen in the graph. The model also requires parameters to capture contagion: as prepayments increase, alertness rises in subsequent months, presumably because people are more likely to learn about the advantages of prepaying from friends who have just done so.

With a relatively small number of parameters one can closely fit thousands of data points, including all coupons issued either by Fannie or Freddie each year since 1986. As an example of how the model fits, consider the model fit (estimated on data up until early 1996) on the FNMA 1986 8% coupon. The in-sample fit is very good, for this coupon and the others, but so is the out of sample fit over three years later. It is hard to imagine another macroeconomic series with so many ups and downs that can be fit as well in and out of sample. We see that an agent based approach to prepayments can retrospectively fit a long history and even make reasonable (conditional) predictions several years into the future.

Figure 1. Conditional Prepayment Rate Projections Can Be Very Reliable in an Agent-Based Model



In the first decade of 2000 a new feature appeared, the cash out refi. Homeowners began to refinance their mortgages even when interest rates barely dropped, or rose, in order to get bigger loans. Fannie and Freddie began to extend loans to a larger group of homeowners whose behavior was significantly different. Homeowners actually began to default. The agent-based model needed a substantial new effect, depending crucially on house price appreciation, which had not been foreseen in

the model of the 1990s. This illustrates the standard criticism of agent based models, that behavioral rules eventually become inappropriate (or in need of revision) as the world changes. On the other hand, vastly more individual data became available, making the agent-based approach unavoidable.

II. The Agent-Based Housing Model

The goal of our current project is to try to apply this kind of agent-based thinking to retrospectively understand the housing boom and crash of 1997-2009. For the analysis presented here, we do counterfactual experiments to see what would have happened to housing prices if interest rates or leverage had been held constant.

One advantage of the agent-based approach is its ability to incorporate vast amounts of data and produce numerous outputs that can be matched to other data. We have already collected and incorporated into our model data on household demographics, economic conditions, housing stocks, loan characteristics, and housing market behaviors. We use other data sources to corroborate the outputs of our model.

In principle we are hoping to collect data on every housing unit in the greater Washington, DC area, whether owner occupied or a rental. We rank all housing units according to their (commonly perceived) value based on historically observed sale prices. We collected data on every individual by race, income, wealth, age, marital status, household position and so on. We match demographic trends, using historical data on population size, death rates, and migration patterns—and soon we will incorporate race and marital status as well. Individual incomes are chosen each year to match detailed IRS income data from the Washington DC area. Wealth is chosen to match the PSID figures. The permanent income process of every individual follows the Christopher Carroll (1997) design.

The model is initialized by making an initial guess at matching households to houses and running the model for a number of periods using 1997 data until it settles down into a more or less steady state.

Then our simulation period of 1997 to 2009 begins. New households form, either through aging or immigration, and decide whether to rent or own based on their current conditions and demographic factors. Houses are voluntarily sold when people randomly decide to move (according to historical averages) or die or are foreclosed on. Homeowners can also default (perhaps involuntarily) and refinance depending on their individual economic situations.

When a homeowner decides to put his house on the market, he lists an offering price that is somewhat above (on average a bit less than 10%) the “fair market” value for his house as computed by comparison to other recent sales. Every few months that a house does not sell, the homeowner reduces the asking price by some percent (e.g., 3-5%). The actual values of the markup, markdown, and markdown frequency are chosen from distributions whose parameters will be matched to actual seller behavior as recorded in the Washington DC Area MLS records.

Buyers (who may be immigrants, or young people coming of age, or people who randomly decide to move and stay within the Washington area) have a desired annual expenditure and desired down payment. We suppose that the typical buyer wants to spend a third of his permanent income per year on housing, where the payments are a combination of mortgage payments, direct expenses (like upkeep, taxes, depreciation) less a term capturing the expected income from the house price rise. For simplicity we suppose that the direct expenses are a fixed fraction (five percent) of the purchase price, that the mortgage expenses are the annual coupon payments, and that the return is last year’s house price appreciation multiplied by a constant (which we also take to be 10%) to capture the idea that homeowners don’t think prices will rise (or fall) forever. We then add some noise to get a distribution of debt/income. New buyers have a desired loan-to-value (LTV) ratio that we take to be a distribution depending on their incomes and wealth, corresponding to historical data on actual LTV.

After expressing their desires, homeowners apply for loans, which they get if they satisfy prevailing LTV and debt/income requirements. If their loans are rejected, they apply again by altering their choices until they are approved or determine they cannot get a loan. We match the empirical loan mix (using the categories fixed, ARM, and interest only loans) by drawing each home owner a loan type based on his debt/income ratio (and in the future FICO). Once a buyer has a loan preapproval, he goes to the market and bids on the highest-value house he thinks he can afford given his desired annual expenditure and the LTV for which he been approved.

Figure 2 shows how the model fits house prices and other housing market indices from 1997 – 2010. The model captures a large fraction of the boom and bust. It also matches a number of other features of the housing market, like time on market and so on, but not everything (ownership percentage). Notice that if we freeze leverage (LTV) at constant levels, the boom gets dramatically attenuated, and the bust disappears. If we freeze interest rates, but allow leverage to vary historically, the boom and bust reappear, though perhaps shrunk by 12-15%. If we freeze both leverage and interest rates the rise in housing is similar to the modest rise when interest rates were allowed to fall, driven by the increase in income over the 13 year period. Thus leverage, not interest rates, seems to be the important factor driving the 1997 – 2010 boom and bust.

III. Bibliography

Carroll, Christopher (1997) “Buffer-Stock Savings and the Life-Cycle/Permanent Income Hypothesis” *Quarterly Journal of Economics*, CXII,1,1-55.

Duca, John, V., John Muellbauer and Anthony Murphy (2011) “Shifting Credit Standards and the Boom and Bust in U.S. House Prices,” Federal Reserve Bank of Dallas, Research Department.

Geanakoplos, John “Liquidity, Default, and Crashes, Endogenous Contracts in General Equilibrium,” *Advances in Economics and Econometrics: Theory and Applications, Eighth World Conference, Volume II*, Econometric Society Monographs, 2003, pp.170–205.

Geanakoplos, John (2010a) “The Leverage Cycle,” in D. Acemoglu, K. Rogoff, and M. Woodford, eds., *NBER Macroeconomics Annual 2009*, Vol. 24, University of Chicago Press, Chicago, 2010, pp. 1–65.

Geanakoplos, John (2010b) “Solving the Present Crisis and Managing the Leverage Cycle,” *Federal Reserve Bank of New York Economic Policy Review* (August), pp.101–131.

Geanakoplos, John and Susan Koniak (2008) “Mortgage Justice is Blind” New York Times Op-Ed.

Geanakoplos, John and Susan Koniak (2009) “Matters of Principal” New York Times Op-Ed.

Glaeser, Edward, Joshua Gottlieb, and Joseph Gyourko (2010) “Can Cheap Credit Explain the Housing Boom?,” Harvard Working Paper.

Haughwout, Andrew, Donghoon, Lee, Tracy, Joseph and van der Klaauw, Wilbert (2011) “Real Estate Investors, the Leverage Cycle, and the Housing Market Crisis,” Federal Reserve Bank of New York Staff Report.

Figure 2: Housing Market Results for Washington, DC

