A Real Options Approach to Housing Investment

Chris Downing
Nancy Wallace *

May 8, 2000

*This paper represents the views of the authors and does not necessarily represent the views of the Federal Reserve System or members of its staff. Please address correspondence to (Downing): Federal Reserve Board, Mail Stop 89, Washington, DC 20551. Phone: (202) 452-2378. Fax: (202) 452-5296. E-Mail: cdowning@frb.gov. (Wallace): Haas School of Business, University of California at Berkeley, Berkeley, CA. Phone: (510) 642-4732. Fax: (510) 643-1420. E-Mail: wallace@haas.berkeley.edu.

Abstract

In this paper, we focus on investments by existing homeowners to improve their homes. In our model, the value of a house is equal to the expected net present value of a perpetual stream of service flows emanating from the attributes of the house. An important innovation in our model is that the set of house attributes evolves over time according to the investment decisions of the homeowner. The homeowner's decisions to invest in house attributes are modeled as real options. The homeowner compares the value of an additional attribute, net of the value of the opportunity to invest in the future, to the cost of the investment when deciding whether or not to invest. Our model of investment embeds a multi–factor term structure model and a general model of the evolution of service flows. We employ numeric simulations to explore the properties of the investment model, and to motivate our empirical test of the model.

The main contribution of this paper is an empirical test of whether or not observed homeowner investment behavior is consistent with the real option theory of investment. Using a nationally representative panel from the American Housing Survey, we test two implications of the real option theory. First, we test whether investment is more likely when the spread between the return to housing and the cost of capital is wide. Second, we test whether greater spread volatility depresses investment. Our empirical results indicate that observed homeowner investment behavior is consistent with these implications of the theory, even after controlling for business cycle, aging, tenure and for–sale influences.

1 Introduction

Homeowner remodeling decisions are an important factor in the evolution of housing supply in the United States. The Home Improvements Research Center, of Tampa, Florida, estimates that expenditures for home improvement products in 1999 were about $163 billion.[1] Despite the significance of these expenditures, there is a lack of either empirical or theoretical research that focuses on homeowner investment decisions. The usual approach to modeling these decisions has been to focus on the importance of life-cycle effects (Ortalo–Magne and Rady (1999)), or on the possible relationship between housing tenure decisions and housing investment (Goetzmann and Spiegel (1995)). Outside of these papers, most housing investment research has focused on the production of new housing by professional developers (Capozza and Sick (1991), Williams (1993), and Capozza and Li (1994)).

In this paper, we focus on investments by existing homeowners to improve their homes. In our model, the value of a house is equal to the expected net present value of a perpetual stream of service flows. The service flow from a house in a given period is modeled hedonically, as a function of a set of attributes describing the structure. An important innovation in our model is that the set of house attributes evolves over time according to the investment decisions of the homeowner.

The pricing model is a dynamic, continuous–time extension of the static house price model appearing in Poterba (1984) and Rosen and Topel (1988), and nests standard hedonic regression models, such as those found in Case and Quigley (1991) and Mills and Simenauer (1996). The model developed here differs in at least three important respects from the previous models. First, the earlier models are static, in the sense that the state variables do not evolve stochastically over time; we explicitly allow for stochastic state variables. Second, in the earlier models, service flows are not explicitly modeled; we model service flows hedonically. Finally, the evolution through time of house attributes is a stochastic process controlled by the homeowner, in contrast to the previous models, where changes in house attributes are exogenously determined.

The homeowner's decisions to invest in house attributes (for example, to add square footage) are modeled as real options. In keeping with the overall

[1] The Joint Center for Real Estate at Harvard University estimates that, in 1995, expenditures for home improvements totaled just under $150 billion dollars, or about two percent of total GDP (Joint Center for Housing Studies, Harvard University (1999)).

model of house value, each attribute of a house is assumed to generate a flow of services over time, so that the value of an additional unit of an attribute is the expected net present value of the stream of services that it provides. The homeowner compares the value of an additional unit of an attribute, net of the value of the opportunity to invest in the future, to the cost of the investment when deciding whether or not to invest. We consider a generalization of the basic real option model discussed in Dixit and Pindyck (1994).

The main contribution of this paper is an empirical test of whether or not observed homeowner investment behavior is consistent with the real option theory of investment. Using panel data from the American Housing Survey, we test two implications of the real option theory. First, we test whether the decision to invest is sensitive to the spread between the return on housing and the user cost of capital. Consistent with the real option theory, we find that homeowners are more likely to invest when the spread between the return to housing and the cost of capital is wide. Second, we test whether the decision to invest is sensitive to the volatility of the spread. An implication of our model is that the value of waiting to invest increases when the net return on the investment is more volatile. Thus, *ceteris paribus*, we should observe less investment during periods of high return volatility. Our empirical results are consistent with this prediction, as well.

The paper is organized as follows. In the next section, we present the house pricing model, we review the real option theory of investment, and we discuss the results of some pricing experiments calibrated to real data. Section three presents our main econometric results, and the final section concludes.

2 House Prices and Investment

A general, continuous–time model of house prices serves as the framework for our model of housing investment. The pricing model is an extension of the models appearing in Poterba (1984) and Rosen and Topel (1988). Following these papers, houses are priced in equilibrium as the expected present value of future service flows, where the flows are discounted at the user cost of capital. Service flows are modeled hedonically, as a function of the attributes of the house. The attributes of the house evolve endogenously according to the investment decisions of the homeowner. For simplicity, we consider the case of a homeowner's decision to make a single addition to his or her home; we focus on the testable implications of this model that are common to all real option models.

In our model, a house is described by a set of K attributes, with $a_{k,t}$ denoting the units of attribute k at time t. Each attribute generates a flow of services accruing to the homeowner, defined as the rental rate of the attribute. We assume that rental rates are set in a perfectly competitive market, subject to random demand shocks:

$$\pi_t = x_t' D(q_t). \tag{1}$$

Here π_t is a $K \times 1$ vector of rental rates, x_t is a vector of exogenous demand shocks, q_t is total supply, and $D(q_t)$ is the inverse demand function, with $D' < 0$. The demand shocks are assumed to evolve stochastically through time as geometric Brownian motions:

$$dx_t = \alpha_x x_t dt + \sigma_x x_t dW_{x,t}, \tag{2}$$

where the W_x are independent standard Wiener processes. [2] The demand shocks can be interpreted as reflecting changes in tastes for house attributes, shocks to income, changes in migration patterns, or any of a host of other possibilities.

The dynamics of rents are determined by the demand shocks, as well as by changes in total supply, as can be seen by applying Itô's Lemma to (1):

$$d\pi_t = \alpha_x \pi_t dt + \sigma_x \pi_t dW_{x,t} + \frac{D'(q_t)}{D(q_t)} \pi_t dq_t. \tag{3}$$

In what follows, we assume that rental dynamics due to changes in total supply, given by the term $\frac{D'(q_t)}{D(q_t)} \pi_t dq_t$, are such that they are absorbed into the terms α_x and σ_x. In other words, we assume that the growth rate in rents, and the volatility in rents, impound all of the effects of changes in the total supply of attributes. As a result, rents evolve according to a geometric process.

The user cost of capital is defined as the after–tax risk–free spot rate of interest, adjusted to reflect depreciation and costs of repair. We assume that the instantaneous risk–free rate evolves as follows:

$$\begin{aligned} dr_t &= \alpha_r(r_t, \sigma_t)dt + \sigma_r(r_t, \sigma_t)dW_{r,t}, & (4) \\ d\sigma_t &= \alpha_\sigma(r_t, \sigma_t)dt + \sigma_\sigma(r_t, \sigma_t)dW_{\sigma,t} & (5) \end{aligned}$$

[2] Correlated demand shocks are a natural extension of the model. For our purposes here, correlated shocks introduce an additional degree of complexity that is unnecessary.

where $W_{r,t}$ and $W_{\sigma,t}$ are standard Wiener processes, independent of one another, as well as of the $W_{x,t}$. The system of equations (4) and (5) forms a stochastic volatility model of the spot rate. [3] The user cost of capital is given by:

$$\rho_t = \delta + \mu + (1 - \tau)(\theta + r_t) \tag{6}$$

where

$$\begin{aligned}
\delta &\equiv \text{Rate of depreciation,} \\
\mu &\equiv \text{Expenditures for maintenance and repairs} \\
& \text{as a fraction of house value,} \\
\tau &\equiv \text{Marginal income tax rate,} \\
\theta &\equiv \text{Property tax rate.}
\end{aligned}$$

In general, it is not possible to instantaneously trade attributes, and it is also not possible to short sell attributes, so valuation approaches based on arbitrage arguments are highly unrealistic. Instead, we use the equilibrium approach, which is now more or less standard in the literature on real options. For simplicity, both in the development of the model and in the interpretation of the comparative statics experiments to follow, we assume risk–neutrality when pricing attributes and options on attributes. This assumption can be relaxed by making the appropriate risk–adjustments to the drift rates of the underlying stochastic processes when pricing the assets.

Let $V_t(\pi_t, \rho_t, \sigma_t)$ denote the value of an attribute. The instantaneous total return on the attribute is given by: [4]

$$dV_t + \pi_t dt, \tag{7}$$

where dV_t is the capital gain and $\pi_t dt$ is the instantaneous rental flow.

Applying Itô's Lemma, and taking the appropriate expectation, the expected total return is given by:

$$\left[\frac{\mathcal{D}V_t + \pi_t}{V_t}\right] dt, \tag{8}$$

[3]The functions $\alpha.$ and $\sigma.$ in the system (2)-(5) are assumed to satisfy the technical conditions sufficient for the system to have a unique solution (see Karatzas and Shreve (1991)).

[4]In what follows, the arguments to functions will often be omitted, in the interest of notational simplicity.

where the operator $\mathcal{D}$ is defined as:

$$\mathcal{D}=\frac{1}{2}\mathrm{tr}\left[\sigma_x\sigma_x'\right]\frac{\partial^2}{\partial\pi^2}+\frac{1}{2}\sigma_\rho^2\frac{\partial^2}{\partial\rho^2}+\frac{1}{2}\sigma_\sigma^2\frac{\partial^2}{\partial\sigma^2}+\sum_{k=1}^{K}\alpha_{x_k}\frac{\partial}{\partial\pi_k}+\alpha_\rho\frac{\partial}{\partial\rho}+\alpha_\sigma\frac{\partial}{\partial\sigma}. \quad (9)$$

In equilibrium, the expected total return to the attribute must equal the instantaneous user cost of capital, producing:

$$\mathcal{D}V_t-\rho_tV_t+\pi_t=0. \quad (10)$$

Two more conditions render a well–posed partial differential equation (pde) that characterizes the value of the attribute:

$$\lim_{\rho_t\to\infty}V_t(\pi_t,\rho_t,\sigma_t) = 0, \quad (11)$$

$$V_t(0,\rho_t,\sigma_t) = 0 \quad (12)$$

Condition (11) says that, as the user cost of capital approaches infinity, all future values become worthless. Because zero is an absorbing barrier for the rental flow process, the attribute becomes worthless if rents hit zero, as stated in condition (12).

Alternatively, we can write the value of an attribute as the expected discounted present value of future rental flows:

$$V_t=E_t\int_t^\infty\pi_s e^{-\int_t^s\rho_\nu d\nu}ds. \quad (13)$$

Duffie (1996) has a sketch of the proof that (13) solves (10)-(12); Karatzas and Shreve (1991), and Chung and Williams (1990) contain more detail.

The price of a house is a function of the attributes and their associated rental streams:

$$H_t(a_t,\pi_t,\rho_t,\sigma_t)=E_t\int_t^\infty h(a_s,\pi_s)e^{-\int_t^s\rho_\nu d\nu}ds, \quad (14)$$

where $h(\cdot)$ is a smooth function. [5] The evolution of house prices thus depends

[5]Note that the expectation is taken under the assumption of risk–neutrality. If this assumption were relaxed, it would be necessary to compute the expectation after first risk–adjusting the drifts of the underlying stochastic processes.

on the dynamics of rents, as well as the evolution of the attribute vector over time.[6]

A homeowner is free to invest in his or her home at any time. Thus, one way the house attribute vector evolves over time is according to the investment decisions of the homeowner. [7] Let $F_{k,t}(\pi_t, \rho_t, \sigma_t)$ denote the value of the homeowner's option to invest in one unit of attribute k, where the cost of the investment is a constant dollar amount I. More complicated problems might be constructed, such as options to add multiple attributes, uncertain investment costs, and the like. However, we focus on the testable implications of our model that hold true under all of these extensions. Dropping the subscript k for notational simplicity, we arrive at the pde that characterizes $F_t(\cdot)$ by following the same steps that we used to find the pde that characterizes the attribute pricing function.

The total return to holding the option is equal to the capital gain less the value of forgone service flows. By Itô's Lemma, and after taking the appropriate expectation, the expected total return is given by:

$$\left[\frac{\mathcal{D}F_t - \pi_t}{F_t}\right] dt, \tag{16}$$

where $\mathcal{D}$ is defined in (9) above. Setting the expected total return equal to the user cost of capital, we have:

$$\mathcal{D}F_t - \rho_t F_t - \pi_t = 0, \tag{17}$$

subject to the infinite user cost boundary condition:

$$\lim_{\rho_t \to \infty} F_t(\pi_t, \rho_t, \sigma_t) = 0, \tag{18}$$

[6]If we assume that the attributes are fixed exogenously, and we use, as a first approximation to the true value, a linear specification for $h(\cdot)$, we can write:

$$H = \sum_{k=1}^{K} a_k V_{k,t}, \tag{15}$$

using (13). By projecting a cross–section of house prices H observed at time t onto the space of attributes a_k, one can estimate the attribute values $V_{k,t}$. This is the hedonic regression approach for constructing house price indices. Case and Quigley (1991) construct a hedonic model that allows attributes to change over time. However, in their model, attributes change according to an unspecified exogenous process.

[7]We ignore catastrophic events such as fire and flood.

and the zero rental flow condition:

$$F_t(0, \rho_t, \sigma_t) = 0. \tag{19}$$

The continuation region $\mathcal{C}$ is defined as the set of π, ρ and σ values for which immediate exercise of the investment option is not optimal:

$$\mathcal{C} = \left\{ (\pi_t, \rho_t, \sigma_t) \in \Re^3_+ : \mathcal{D}F_t - \rho_t F_t - \pi_t = 0, \text{ and } F_t > V_t - I \right\}$$

The exercise region $\mathcal{E}$ is defined as the set of values for which it is optimal to invest immediately:

$$\mathcal{E} = \left\{ (\pi_t, \rho_t, \sigma_t) \in \Re^3_+ : \mathcal{D}F_t - \rho_t F_t - \pi_t \leq 0, \text{ and } F_t = V_t - I. \right\}.$$

The location of the exercise region $\mathcal{E}$ is endogenous, and is recovered as part of the problem of solving the pde, an issue we return to below.

The time at which the system first passes into the region $\mathcal{E}$ is defined as the "optimal exercise policy." The distribution of first passage times induces a distribution on the time of a unit increase in a, and thus provides, at least conceptually, a technology by which to evaluate the expectation in (14). We'll briefly revisit this problem below, but the full solution is left for future work. Here we focus on establishing some results for the pricing problems defined by (10)-(12) and (17)-(19).

In order to solve the pricing problems, we make the following additional specializations. Suppose that the system of equations defining the instantaneous user cost process is given by: [8]

$$\begin{aligned} d\rho_t &= \gamma_\rho(\phi_\rho - \rho_t)dt + \sigma_t\sqrt{\rho_t}dW_{\rho,t} && (20) \\ d\sigma_t &= \gamma_\sigma(\phi_\sigma - \sigma_t)dt + \xi\sqrt{\sigma_t}dW_{\sigma,t}. && (21) \end{aligned}$$

The system (20)-(21) can be justified on empirical grounds. It is similar to the system shown by Andersen and Lund (1997) to closely describe US Treasury bill dynamics. More generally, the system is consistent with the growing empirical support for the view that a multi–factor model is required in order to capture all of the dynamics of the short rate.

According to the Bureau of Economic Analysis, housing rents in the United States as a whole grew on average by about 3.5 percent each year

[8]For simplicity, in the numerical calculations in this section, we assume that the constants δ,μ,τ,θ are equal to zero. In our numerical experiments, we use r and ρ, and the terms "interest rate" and "user cost," interchangeably.

between 1988 and 1999, with a standard deviation of about 2.0 percent. Thus, a realistic calibration of the rental process is given by:

$$d\pi_t = 0.035\pi_t dt + 0.02\pi_t dW_{\pi,t}. \tag{22}$$

Our parameterization of the interest rate system is taken from Andersen and Lund (1997).[9] We compute prices under the following specification of the term structure processes:

$$d\rho_t = 0.16(0.0595 - \rho_t)dt + \sigma_t\sqrt{\rho_t}dW_{\rho,t} \tag{23}$$
$$d\sigma_t = 1.04(0.04 - \sigma_t)dt + 1.89\sqrt{\sigma_t}dW_{\sigma,t} \tag{24}$$

The long–term mean in the interest rate is roughly six percent, with a standard deviation of about thirty percent of the long–term mean when volatility equals its long–term mean of four percent. The volatility process exhibits substantial deviation about the mean, but its rate of mean reversion is such that volatility shocks are relatively short–lived.

Even with these specializations, the partial differential equations that characterize asset values do not admit closed–form solutions. Using a finite–difference algorithm, we numerically solve for attribute values and option prices.[10] Figure 1 displays attribute values under the rent and interest rate system specified by (22)-(24). The figure displays the values on a grid of rental and user cost values, for a "slice" of the state–space along the volatility axis at $\sigma = 0.04$. For a dollar of rental flow ($\pi = 1$), the attribute is worth approximately \$95 when the user cost is at its long-run mean ($\rho = 0.0595$). The attribute value rises as the interest rate falls or the rent flow increases, and vice–versa.

Using the attribute values displayed in figure 1, we next solve for the value of the option to invest. We compute the values of an option that is at–the–money when the rental flow is one dollar ($\pi = 1$), and the user cost and volatility variables are at their long–run means ($\rho = 0.0595$ and $\sigma = 0.04$).

[9] The system estimated by Andersen and Lund (1997) is slightly different from the system that we use here. Our volatility process is a square–root process, while the original paper used a constant–diffusion process. Nevertheless, the parameterization remains realistic.

[10] See Ames (1977), Press, Teukolsky, Vetterling and Flannery (1994) and Smith (1996) for information on the finite difference approach to solving partial differential equations. Hull and White (1990) discuss some of the nuances of valuing derivative securities using the finite difference approach.

Figure 2 overlays the option values and the attribute prices of figure 1. In regions of the state space near the exercise region (high value of π, low values of ρ), the option value as a proportion of the attribute value approaches the value $\frac{V-95}{V}$, because the option is worth its payoff, $V - 95$, at exercise. The option value is large relative to the attribute value even in regions well away from the exercise region. For example, near the point $\pi = 0.5$, $\rho = 0.12$, the option is worth roughly seventy percent of the attribute ($\frac{F}{V} \approx 0.7$).

It is interesting to explore the effects of stochastic volatility on option values. Figure 3 overlays two surfaces: the surface defined by taking a slice of the state space at $\sigma = 0.04$, and a relatively high–volatility surface defined at $\sigma = 0.085$. In general, higher volatility increases the value of the option to invest, because the option is a convex function of interest rates. The effect of higher volatility is greatest when the spot rate is low and the rental rate is high. This is because points in this region are closest to the exercise boundary. When the rental rate is very low (the option is way out of the money), increased volatility has less effect on the value of the option to invest.

The first passage times from points in the continuation region to points in the exercise region are directly related to the probability of a change in the attribute (investment). Figure 4 shows the approximate locations of the exercise and continuation regions for the portion of the state space defined by $1.5 < \pi < 5$, $\rho < 0.1$, and $0 < \sigma < 0.06$. From any point in the continuation region (points unenclosed by the curves), for a fixed level of volatility, a reduction in the spot rate, and/or a high return on the attribute, increases the probability of observing an investment over a given time horizon, and vice–versa. In fact, it is clear from the figure that, given a level of volatility, the spread between the rental rate and the spot rate is sufficient to characterize the probability of observing an investment. If the spread widens, then either the rental rate rose (and attribute values rose), or the spot rate fell, or both – in any case, the probability of observing an investment over some future time interval increases because the system has moved closer to the exercise region. When the spread narrows or turns negative, then either the rental rate fell, or the spot rate rose, or both – in any case, the probability of observing an investment is relatively lower. In the following section, we test whether, after controlling for volatility, the probability of observing investment is positively related to the spread between housing returns and the user cost.

When σ takes on a high value, the volatility of the spread of the rental rate to the user cost is higher than when σ takes on a low value.[11] From figure 4, we see that the effect of greater volatility in the spread is to reduce the probability of observing investment over any time horizon, holding the spread constant. This is because, as σ increases, the exercise region contracts inward. Thus, the distance to the exercise region increases as σ increases, holding π and ρ fixed. In the next section, we test whether, after controlling for the spread, the probability of observing an investment is inversely related to the level of volatility.

[11]In this case, the higher volatility in the spread is due to higher interest rate volatility; the qualitative results do not change if the volatility is due to higher volatility in rental rates, or higher volatility in both variables.

Figure 1: Attribute Values

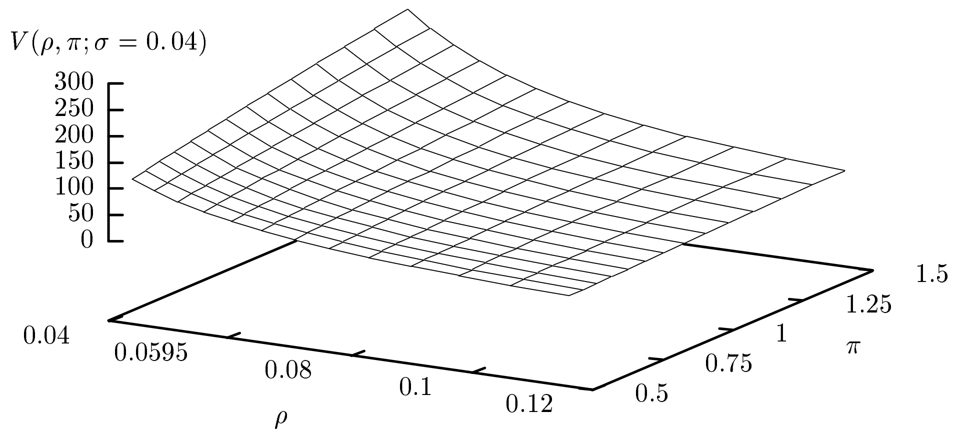


Figure 2: Attribute and Option Values

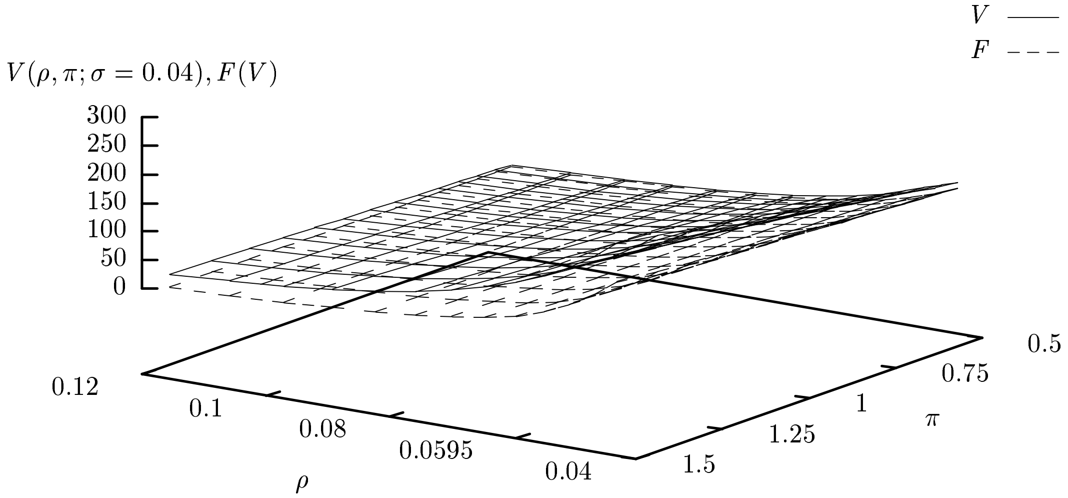
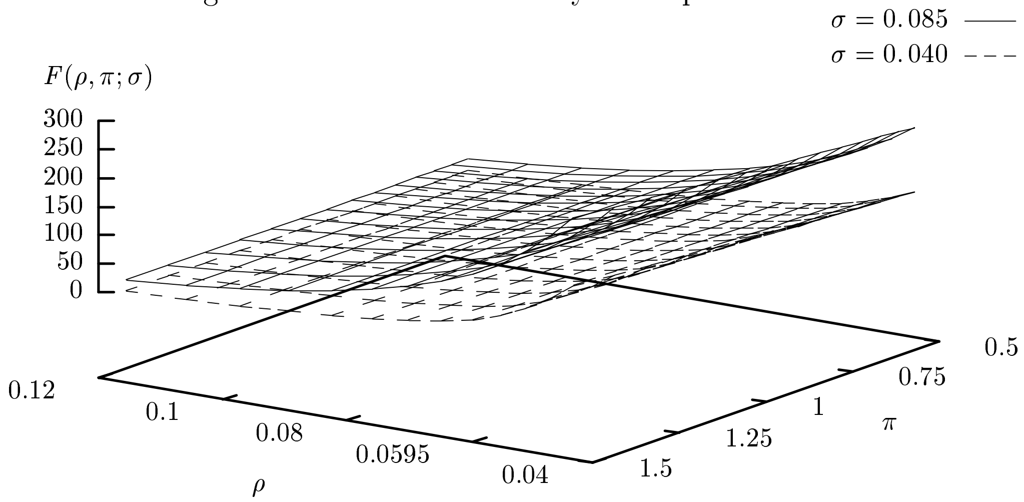


Figure 3: Stochastic Volatility and Option Values



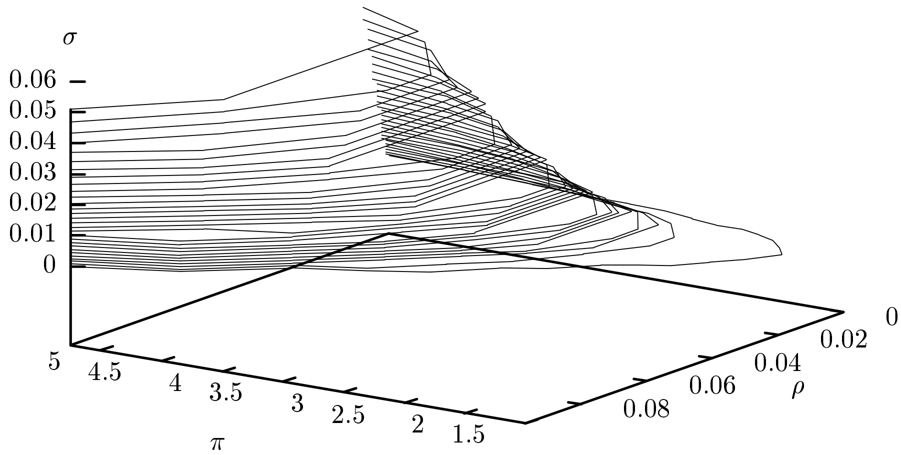


Figure 4: Selected Exercise and Continuation Regions

3 Empirical Tests and Results

We test the predictions of the real option model using panel data from the American Housing Survey (AHS). The AHS is an ongoing, biennial survey of randomly selected homes in the United States. The survey is sponsored by the U.S. Department of Housing and Urban Development and conducted by the Census Bureau. Approximately 53,000 interviews are conducted in each biennial survey from 1985 through 1997. In order to maintain the anonymity of respondents, geographic information is suppressed in the public–use data for many of the interviews. We are able to determine the SMSA for 14,477 houses in large metropolitan areas, for a total of 64,398 observations. In appendix A, we discuss the dataset at more length. [12]

Our test strategy is as follows. First, we construct a binary dependent variable that indicates if investment occurs during a two–year survey period. Using repeat sales house price indices and computations of the user cost of capital using data from the surveys, we compute spreads between the annualized return to housing and the user cost of capital. Finally, modeling the probability of investment with a logistic distribution, we test whether the spread and the volatility of the spread predict investment.

In our application, it is important that we maintain as long a time series as possible, in order to reveal how investment behavior responds as spreads and spread volatilities change over time. For this reason, we focus on a generic measure of investment. The 1995 and 1997 surveys collected detailed information on the types, timing, and costs of house additions. Unfortunately, the degree of detail declines as one moves back through the earlier surveys, forcing a tradeoff between the length of the time series and the detail of information on investments. All of the survey waves collected information on whether or not, over the previous two years, the homeowner built any new additions. Using this information, we code our dependent variable as follows:[13]

$$Y = \begin{cases} 0 & \text{No additions in previous 2 years} \\ 1 & \text{One or more additions in previous 2 years.} \end{cases} \tag{25}$$

[12] The American Housing Survey is complex in scale and scope. Interested readers are referred to the AHS web site (`http://www.huduser.org/datasets/ahs.html`) for full details.

[13] See appendix A for detail on the construction of all of the variables that we use in this section.

The types of additions that might be included are new bedrooms, new bathrooms, kitchens, and other "big–ticket" items that are at least partially irreversible, and which have a major impact on the value of the home.

We don't observe the exact time at which an investment occurs. Thus, for each property, we compute a measure of the average spread between the annualized return on housing less the user cost of capital, where the average is taken over the year of the survey and the previous year. To be precise, if we let $h_{n,t}$ be the house return in year t for houses in the same SMSA as house n, and $\rho_{n,t}$ be the user cost of capital for homeowner n in year t, then we measure the spread as $s_{n,t} = h_{n,t} - \rho_{n,t}$. The variable $SPREAD$ is measured for house n in survey year t as:

$$SPREAD_{n,t} = \frac{s_{n,t} + s_{n,t-1}}{2}. \tag{26}$$

We also construct a measure of the volatility of the spread over the period. First, we compute a longer–run average spread measure, given as:

$$\bar{s}_n = \frac{1}{T_n} \sum_{t=1985}^{1997} I(s_t) s_t, \tag{27}$$

where $I(s_t)$ is an indicator that is one if the spread measure is non–missing, and zero otherwise, and T_n is the number of non-missing spread observations for house n. The volatility of the spread for house n in survey year t is measured as follows:

$$VOLATILITY_{n,t} = \frac{|s_{n,t} - \bar{s}_n| + |s_{n,t-1} - \bar{s}_n|}{2}. \tag{28}$$

The variable $VOLATILITY$ thus measures the average absolute deviation of the spread from the long–run average spread for the property.

The annualized returns on housing, $h_{n,t}$, are computed as the percent changes in house prices, where the levels of house prices are measured using repeat sales house price indices obtained from Fannie Mae and Freddie Mac. Separate indices for each of nearly 150 different SMSAs are used, in order to introduce a high degree of cross–sectional variation in house returns.

The user cost of capital, $\rho_{n,t}$, is based on the annualized 30–day Treasury bill return. We use the 30–day Treasury bill so as to maintain consistency with our continuous–time model. For house n at time t, we form the user cost of capital as follows:

$$\rho_{n,t} = UPKEEP_{n,t} + (1 - (FMAR_{n,t} + SMAR_{n,t})) * (PTAX_{n,t} + r_t). \tag{29}$$

The variable $UPKEEP$ measures annual expenditures, as a percentage of total house value, for maintenance and upkeep. The variable $FMAR$ is the federal marginal tax rate, computed using household income and federal tax schedules, and $SMAR$ is the state marginal tax rate, computed similarly. The variable $PTAX$ is the amount of property tax paid by the individual, as a percent of total house value. The user cost variable thus varies in the cross–section at the level of the house, and varies over time at annual frequency.

The investment indicator, house returns, and user cost variables form the core of our empirical model. However, we also control for other factors that might, as alternatives to our real option model, explain the incidence of housing investment. We introduce controls for business–cycle, tenure and on–market effects, defined below. Table 1 shows the labels, definitions, and univariate statistics for all of the independent variables.

In order to control for possible business–cycle effects, we include a measure of per capita income, labeled $INCOME$, which is measured at the SMSA level. Blanchard and Katz (1992) argue that if states grow at different rates because they are differentially attractive to firms, then high growth will be associated with low unemployment and higher wages. Their empirical results indicate that there is a strong co–movement of per capita income and employment in the United States. House prices are shown to have short-run sensitivity to shocks to employment levels, however, in the long-run house prices increase as employment levels increase. These results suggest that house prices and housing investment would be positively correlated with per capita income. We use INCOME to proxy for this pro-cyclical effect.

Goetzmann and Spiegel (1995) propose that housing investment tends to cluster either at the time a family moves into a home (what we call the effects of *tenure*), or just before a home is put up for sale (the effects of being *on–market*). The AHS data allow us to test both of these alternative explanations against our real option model. To control for the effects of tenure, we include the variable $RECMOVER$, which is a dummy variable that takes the value one when the respondent indicates that the family moved in at some time during the previous year. In our sample, we identify 3,771 observations as recent movers, or approximately six percent of the observations. To control for the effects of time–on–market, we include the variable $FORSALE$. The variable $FORSALE$ is a dummy variable that takes the value one when the respondent indicates that the house is up for sale or rent. In our sample, we only observe 16 properties that are up for sale at the time of the survey.

It is possible that older buildings require more renovation and updating than newer buildings. In order to control for possible age effects on the probabilities, we include the AGE variable. The AGE variable measures the age of the structure in years. From table 1, we see that the "average" house is 38 years old, but that there is substantial deviation in ages in the sample. In fact, the structures range in age from newly built all the way up to 80+ years old.

The probability of observing a sequence of investment decisions by an individual is modeled with the mixed–logistic distribution (McFadden and Train (1998), Revelt and Train (1999)). We assume that the probability of observing the sequence of investment decisions y by individual n is given by:

$$P_n(y|\mu, \Sigma) = \int P_n(y|\beta) g(\beta|\mu, \Sigma) d\beta, \tag{30}$$

where $P_n(y|\mu, \Sigma)$ is the mixed–logistic investment probability, $P_n(y|\beta)$ is the probability of observing y conditional on the coefficient vector β, and $g(\beta|\mu, \Sigma)$ is the distribution of coefficient vectors in the population. We assume that $g(\beta|\mu, \Sigma)$ is a multivariate normal distribution with mean vector μ and covariance matrix Σ. Furthermore, we assume that the investment decisions are independent, so that:

$$P_n(y|\beta) = \prod_{i=1}^{T} L_n(y_i, i|\beta), \tag{31}$$

where:

$$L_n(y_i = 1, t|\beta) = \frac{e^{\beta X_{n,t}}}{1 + e^{\beta X_{n,t}}}, \tag{32}$$

is the probability of observing an investment at time t by homeowner n. In equation (30), we integrate over the population density $g(\beta|\mu, \Sigma)$ because we do not observe β. Our aim is to estimate the coefficients μ and Σ.

The mixed logit model has the advantage over standard logit that it does not exhibit the independence from irrelevant alternatives. Moreover, McFadden and Train (1998) show that any choice model can be approximated by a mixed–logit model with an appropriate choice of the density for β.[14] The real power of the mixed–logit model, however, lies in the ability to compute the expected value of β, *for each individual*, conditional on the

[14] Here we have chosen the normal density on an ad–hoc basis. We leave for future research the determination of the best density choice for modeling the probability of investment under the continuous–time model of the previous section.

individual's observed investments and characteristics. We exploit this feature of the model below.

An economic motivation for using the mixed–logit specification is the possibility that the effects of the explanatory variables are heterogeneous in the sample, due to unobserved (and perhaps difficult–to–measure forms of) heterogeneity among homeowners. In particular, homeowners might be differentiated by their degree of financial sophistication. We might expect that individuals with a high level of financial sophistication would tend to make investment decisions that align closely with the predictions of the real option model. On the other hand, individuals who are prone to leaving unexploited their financial opportunities would exhibit a pattern of investment decisions that are inconsistent with the predictions of the model. To the extent that the parameters of the logit specification reflect the degree to which individuals behave as the real option model predicts, we should then expect them to be distributed in the population in a way that reflects the level of financial sophistication of homeowners.

The integral in equation (30) does not have a closed form, and so it is approximated through simulation. To simulate the investment probability, we make R draws from the multivariate normal density with mean μ and covariance matrix Σ. For each draw β_r, we compute:

$$\tilde{P}_n(y|\beta_r) = \prod_{i=1}^{T} L_n(y_i, i|\beta_r) \tag{33}$$

and the results are averaged over the R draws. The simulated probability is therefore given by:

$$\tilde{P}_n(y|\mu, \Sigma) = \sum_{r=1}^{R} \tilde{P}_n(y|\beta_r). \tag{34}$$

The simulated log–likelihood function is:

$$SLL = \sum_{n=1}^{N} \log(\tilde{P}_n(y|\mu, \Sigma)). \tag{35}$$

The values $\hat{\mu}$ and $\hat{\Sigma}$ that maximize (35) are maximum–likelihood estimates of μ and Σ. Any of a variety of methods for maximizing non–linear multivariate functions can be used to calculate $\hat{\mu}$ and $\hat{\Sigma}$; we make use of the panel mixed–logit estimator using Halton sequences developed in Train (1999).[15]

[15]Employing Halton sequences in the numerical integrations substantially reduces the

As a familiar point of comparison for our main results, we first estimate a standard logit model, which nests within the mixed–logit specification under the restriction that $\Sigma = 0$. Table 2 displays the results. In general, the parameters are estimated with a good deal of precision, which is not surprising given the large sample size.[16] The results suggest that neither $RECMOVER$ nor $FORSALE$ are important for explaining the investment probabilities, given that the coefficient estimates are insignificantly different from zero. The rest of the variables are highly significant.

The marginal effects of the regressors, computed at the sample means for $SPREAD$, $VOLATILITY$, $INCOME$ and AGE, and $RECMOVER = 0$ and $FORSALE = 0$, are shown in table 3.[17] As can be seen from the table, a wider spread between housing returns and the user cost of capital increases the probability of observing investments. A one percentage point increase in the spread (one sixth of one standard deviation) results in approximately a one third percentage point increase in the probability of investment. Moreover, greater spread volatility tends to depress investment. A one percentage point increase in volatility (one fourth of one standard deviation) reduces the probability of investment by two thirds of a percentage point. As we would expect, higher per capita income tends to spur investment. In sum, the results support the real option theory of investment.

The estimated coefficients for the mixed–logit specification are displayed in table 4. For each variable, the estimated standard deviation is labeled with the prefix $STDEV$. For example, the estimated standard deviation of $SPREAD$ is labeled $STDEVSPREAD$. The second column of data displays the t–statistic, corrected for simulation variance using the method discussed in McFadden and Train (1998). The substantial increase in the value of the log–likelihood suggests that the mixed–logit specification produces a much better fit of the observed investment patterns than does the standard logit model.[18]

---

time required to compute (34). For more details, the reader is referred to the source noted in the text.

[16] In order to check the robustness of the results, we re–estimated the model with all of the coefficient except the intercept restricted to zero. The restriction was rejected in a likelihood ratio test at better than the 99.5 percent significance level. The log–likelihood under the restriction is $-22,728$.

[17] The effect on Λ of a change in x_i is given by $x_i\Lambda(1-\Lambda)$.

[18] A formal likelihood ratio test of the null hypothesis that all of the standard deviations of the coefficients (the $STDEV$ coefficients) are zero is rejected at better than the 99.5 percent significance level.

The patterns of significance and the magnitudes of the estimated standard deviations are interesting, because the $STDEV$ coefficients tell us something about how the coefficients are distributed in the population. The estimated standard deviation of the $SPREAD$ coefficient is 0.65 and insignificant, which, when compared with the mean 3.28, suggests that the effects of a wider spread are homogeneous throughout the population. On the other hand, $STDEVVOLATILITY$ is large relative to the coefficient on $VOLATILITY$, and both are highly significant. This suggests substantial heterogeneity in the population. In fact, we see that for a small portion of the population, the $VOLATILITY$ coefficient may well be positive, suggesting that these homeowners invest in ways that are inconsistent with the real option model. The effects of $INCOME$ are constant in the population, as are the effects of AGE.

The calculation of the marginal effects for the mixed–logit model is complicated by the distributions of the random coefficients. Now, not only do we have to pick the values of the independent variables at which to compute the marginal effects, but we must also pick the values of the coefficients. A natural choice is the mean of the continuous independent variables, and the estimated means of the coefficients. As before, we set the $RECMOVER$ and $FORSALE$ dummy variables to zero. Table 5 displays the marginal effects. The results are similar to the standard logit case. The most important difference is the change in the marginal effect of $VOLATILITY$. The increase in the marginal effect stems from the higher value of the estimated mean of the coefficient compared to the estimate under the standard logit model.

It is interesting to consider the implications of the estimated standard deviations of the coefficients in terms of the marginal effects. In table 6, we display the results of computations of the marginal effects at two different values for $VOLATILITY$. In the second column ($+\sigma$), we show the marginal effects computed for a one standard deviation increase in the $VOLATILITY$ coefficient, where we have used the estimated standard deviation of the coefficient to arrive at the new value. The third column ($-\sigma$) shows the same computation using a one standard deviation decrease. The marginal effect for a one standard deviation increase in the coefficient on $VOLATILITY$ remains slightly negative (-0.0571). For a one standard deviation decrease, the marginal effect is of course highly negative. The marginal effects are not normally distributed, despite the fact that we have assumed that the coefficients are normally distributed in the population. Nevertheless, the results show that for all but a small portion of the population, the real option model appears to do a good job of explaining the observed investment.

As noted earlier, the mixed–logit specification allows one to compute the expected value of the coefficients for each individual, given the individual's sequence of observed investments. In other words, having estimated the population values for μ and Σ, we can estimate where in the coefficient distribution each individual falls. Placing each individual in the coefficient distribution allows us to look for clusters of individuals in different regions of the coefficient space. In particular, we would like to try and identify the common characteristics, if any, of the individuals in the upper tail of the distribution of the $VOLATILITY$ coefficient.

The density $g(\beta|\mu, \Sigma)$ gives the distribution of the coefficients in the population. We would like to identify where each individual's β lies in the distribution g. Let $h(\beta|y, \mu, \Sigma)$ denote the density of β conditional on the individual's sequence of investments y, μ and Σ. If follows by Bayes' rule that:

$$h(\beta|y, \mu, \Sigma) = \frac{P(y|\beta)g(\beta|\mu, \Sigma)}{P(y|\mu, \Sigma)}. \tag{36}$$

Thus, the expectation of β, given y, μ and Σ, is:

$$E[\beta|y, \mu, \Sigma] = \int \beta h(\beta|y, \mu, \Sigma) d\beta. \tag{37}$$

We simulate the value of $E[\beta|y, \mu, \Sigma]$ as follows:

$$\hat{E}[\beta|y, \mu, \Sigma] = \frac{\sum_{i=1}^{R} \beta_r P(y|\beta_r)}{\sum_{i=1}^{R} P(y|\beta_r)}. \tag{38}$$

The key thing to realize is that, once we have estimates of μ and Σ, we have all of the information we need to compute equation (38). We can think of this as an importance–sampling algorithm, where the importance sampling weights are given by $\frac{P(y|\beta_r)}{\sum_{i=1}^{R} P(y|\beta_r)}$. A draw of β_r that is relatively close to the true β for the individual (and thus is relatively likely to generate his or her sequence of observed investments), will produce a high value for $P(y|\beta_r)$, and thus receive a relatively heavy weight.

Figure 5 displays a plot of each individual's expected $VOLATILITY$ coefficient against average $INCOME$. The averages of $INCOME$ are taken over the non–missing observations for each individual. It is clear that low income individuals tend to have expected $VOLATILITY$ coefficients that are more positive than high income individuals. As indicated by the estimated value for $STDEVVOLATILITY$, there is a lot of dispersion in the expected values, especially for low income individuals. In fact, a small proportion of low income individuals have expected values that are positive. We take this to be a clear indication of differences in financial sophistication among individuals. The key point, however, is that for the vast majority of the sample, the expected value of the coefficient on $VOLATILITY$ is highly negative, indicating that the real option model is useful for thinking about the timing of housing investment by existing homeowners.

Label	Definition	Units	Mean	Standard Deviation
SPREAD	House return less user cost	%	-0.010	0.058
VOLATILITY	Volatility of $SPREAD$	%	0.044	0.036
INCOME	Per capita income (SMSA level)	$K	21.547	5.118
RECMOVER	Indicator of recent move–in	0/1	0.059	0.235
FORSALE	Indicator of house on market	0/1	0.002	0.016
AGE	Age of house	years	38.103	20.793

Table 1: Independent Variables

Variable	Coefficient Estimate	t–statistic
Constant	-5.4361	-70.4512
SPREAD	3.6602	11.9720
VOLATILITY	-8.3329	-16.8516
INCOME	0.1590	55.9483
RECMOVER	0.0641	1.1560
FORSALE	0.7937	1.4497
AGE	0.0017	2.4174
Log–Likelihood	-20,233	
N	64,398	

Table 2: Standard Logit Estimation Results

Variable	Coefficient Estimate
Constant	-0.4307
SPREAD	0.2900
VOLATILITY	-0.6603
INCOME	0.0126
RECMOVER	0.0051
FORSALE	0.0629
AGE	0.0001

Table 3: Standard Logit Marginal Effects

Variable	Coefficient Estimate	t–statistic
Constant	-5.67893	-65.86519
SPREAD	3.28327	10.41481
STDEV SPREAD	0.64901	0.88876
VOLATILITY	-13.28119	-20.50591
STDEV VOLATILITY	12.75752	18.03191
INCOME	0.17284	52.37571
STDEV INCOME	0.00010	0.59464
RECMOVER	0.03155	0.42796
STDEV RECMOVER	0.26704	1.02602
FORSALE	1.11035	2.08858
STDEV FORSALE	0.11589	0.72360
AGE	0.00163	2.26201
STDEV AGE	0.00010	0.19437
Log–likelihood	20,151	
N	64,398	

Table 4: Mixed–Logit Estimation Results

Variable	Coefficient Estimate
Constant	-0.3927
SPREAD	0.2270
VOLATILITY	-0.9184
INCOME	0.0120
RECMOVER	0.0022
FORSALE	0.0768
AGE	0.0001

Table 5: Mixed Logit Marginal Effects

Variable	$+\sigma$	$-\sigma$
Constant	-0.6195	-0.2380
SPREAD	0.3581	0.1376
VOLATILITY	-0.0571	-1.0911
INCOME	0.0189	0.0072
RECMOVER	0.0034	0.0013
FORSALE	0.1211	0.0465
AGE	0.0002	0.0001

Table 6: $VOLATILITY$ and Marginal Effects

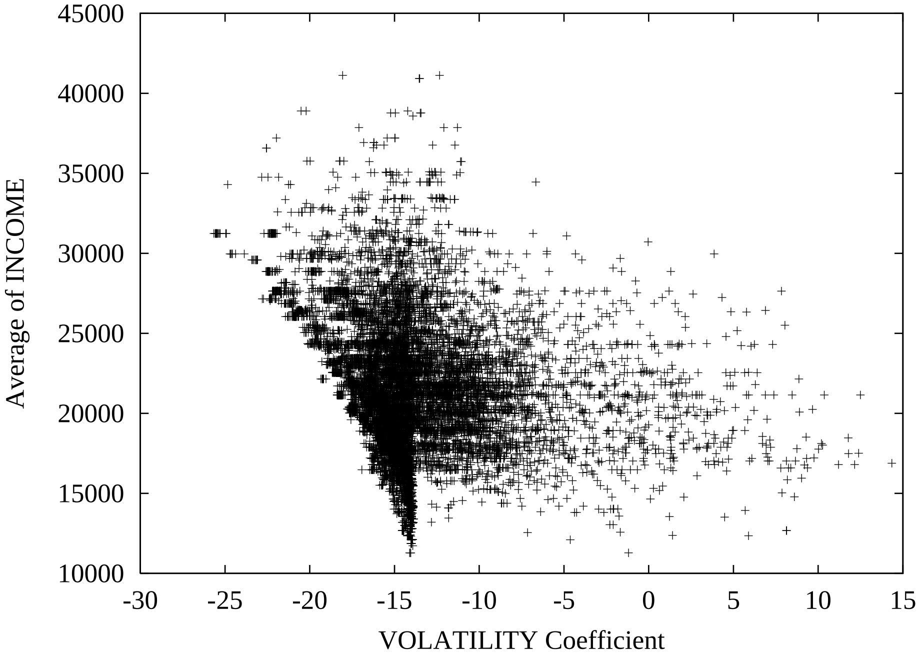


Figure 5: Expected $VOLATILITY$ Coefficients and Average $INCOME$

4 Conclusion

In this paper, we constructed a continuous–time hedonic house price model. The main innovation of the model was to allow the set of house attributes, from which service flows emanate to the homeowner, to evolve over time as a result of investment decisions by the homeowner. We modeled the investment decision using real option theory, under which the homeowner compared the present discounted value of an additional house attribute, net of the value of the option to invest in the future, to the cost of the attribute when deciding when to add features to the house.

The real option theory was tested against data on observed homeowner investment behavior. Using a panel dataset from the American Housing Survey, we found that observed investment behavior is consistent with the real option theory, even after controlling for business-cycle, tenure, time–on–market, and aging effects. Homeowners tend to delay investment when the spread between the return on investment and the user cost of capital is narrow. Similarly, homeowners tend to delay investment when the spread is volatile, in which case the value of the option to invest in the future rises. Using a mixed–logit model, we are able to assess the degree to which unobserved heterogeneity in the population of homeowners might affect the results. What we found is that our conclusions are robust and that, for all but a small segment of the population, the real option model appears to predict observed investment behavior.

A Data

The American Housing Survey, conducted by the Census Bureau, is a panel of approximately 53,000 randomly selected houses in the United States. The sample is a clustered, stratified random sample. The United States was divided into areas made up of counties or groups of counties and independent cities, which the Census Bureau calls primary sampling units (PSUs). The Census Bureau randomly selects a sample of PSUs, and then a sample of housing units within each selected PSU. The survey has been conducted every other year since 1985. Each biennial sample is made publicly available, subject to certain restrictions on the data that are designed to maintain the confidentiality of the sources.

Each observation in each biennial data file contains a unique "control number" that allows one to link up the observations on a particular house over time. For a variety of reasons, the panel is unbalanced. Table 7 shows the frequency distribution of the number of time–series observations on the houses in the overall sample (before removing observations due to missing values). As can be seen, nearly a third of the sample has been included in each of the seven surveys. For approximately twenty percent of the sample, we have only one observation, with the rest falling somewhere in between these two extremes.

T	Frequency	Percent
1	8983	19.5
2	7143	15.5
3	5258	11.4
4	2840	6.2
5	3573	7.8
6	5195	11.3
7	13073	28.4

Table 7: Number of Time–Series Observations

In order to maintain the confidentiality of the sources, not all of the observations contain complete geographic information. Nevertheless, the observations that we are able to use are nationally representative, with observations in nearly all of the SMSAs. Our data cover almost all of the major metropolitan regions in the continental United States.

Starting from the raw data, we lose observations as we attempt constructions using variables with missing data. The bulk of our lost observations are due to missing geographic information. However, we also lose a few observations due to missing income, property tax, and other information. Table 8 summarizes the number of observations lost due to missing values for each variable we use in our logistic regressions. The first row of the table, labeled "TOTAL," gives the total number of stacked observations, after selecting owner–occupied, single–family detached houses (see below), but before discarding any due to missing values. Each subsequent row shows a variable that we use in our study, and the number of observations that remain after discarding observations for which the variable is missing. As can be seen, we lose most of our observations due to missing SMSA information. We lose some more observations constructing our spread measures.

We treat all missing information as the "ignorable case" (Griliches (1986)) of missing data, in which the variables are unavailable for reasons unrelated to the fact that our other observations are complete. Recall from the main text that, when we used the T-bill as a proxy for the user cost of capital, our results were virtually unchanged. This suggests that the loss of observations due to the construction of the spread variables indeed does constitute an "ignorable case" of missing data. Finally, it is difficult to make a convincing case that the process by which observations fall into sparsely populated regions, and thus have suppressed geographic information, is somehow related to housing investment in a systematic fashion.

Variable	Observations Remaining
TOTAL	137,662
SMSA	79,126
AGE	79,126
RECMOVER	79,126
FORSALE	79,126
SPREAD	64,398
Y	64,398

Table 8: Missing Data and Sample Size

Table 9 shows the variables that we drew from the AHS microdata files in order to construct the variables used in our econometric model. The first column gives the variable names as they appear on the AHS microdata files. Some of the variable names change over the years. In the second column, we note the survey years for which the name applies. The third column gives the label, if any, from the microdata files. Some of the variables from the 1997 survey have not yet been labeled, because the Census Bureau has not yet completed processing the data. Below, we discuss in detail how the variables in our econometric model were constructed from those shown in table 9.

Variable Name	Survey	Label
NUNITS	85-97	No. Of Living Qrtrs In Structure Including Vacant Qtrs
ISTATUS	85-97	Type of Interview
TENURE	87-97	Tenure Status
SMSA	87-97	Metropolitan Areas
NEWADD	85-95	New Additions Built in Last 2 Years
RAN	97	Number of replacements and additions
CSTMNT	85-97	Amt Spent in Last Year On Routine Maintenance
AMTX	85-97	Yearly Real Estate Taxes
VALUE	85-97	Property Value (Sample Unit Only)
ZINC	85-97	Inc Of Ref Person And Hshld Members Related To Ref Pers
MAR	85-95	Marital Status of Head/Reference Person
MAR1	97	
RMR	85-95	Respondent Moved Here In Last 12 Months
MOVYR1	97	
MOVM1	97	
MARKET	85-97	Occupied: Listed for Sale or Told Landlord Will Vacate within Month
BUILT	85-97	Year Structure Was Built (Or Model Yr Of Mobile Home)

Table 9: Variables Drawn from AHS

The variables $NUNITS$, $ISTATUS$ and $TENURE$ are used to select owner–occupied, single–family detached homes for which a so–called "occupied, regular interview," was completed. In terms of the variables in the survey, this is accomplished by imposing the following requirements:

$$\begin{aligned} NUNITS &= 1 \\ ISTATUS &= 1 \\ TENURE &= 1 \end{aligned}$$

This produces the sub–sample of structures that we study.

For 1985-1995, our dependent variable is constructed as follows:

$$Y = \begin{cases} 0 & NEWADD = 0 \\ 1 & NEWADD > 0 \end{cases}. \tag{39}$$

For 1997, our dependent variable is:

$$Y = \begin{cases} 0 & RAN = 0 \\ 1 & RAN > 0 \end{cases}. \tag{40}$$

The variable $UPKEEP$ is computed as:

$$UPKEEP = \frac{CSTMNT}{VALUE}. \tag{41}$$

The variables $FMAR$ and $SMAR$ use the measure of household income, $ZINC$, and the marital status MAR, to assign the correct federal and state marginal tax rates, respectively. In the construction of the federal marginal tax rate, $FMAR$, if the head of household is married, we use the joint–filers tax rate schedule, and we use the unmarried filer schedule for all others. The federal tax rate schedules are collected from the instruction booklets for the federal income tax forms over 1985–1997. The state marginal tax rates are close approximations collected from the U.S. Master Tax Guides (Commerce Clearing House (1999)).

Property taxes as a fraction of house value, $PTAX$, are calculated as:

$$PTAX = \frac{AMTX}{VALUE}. \tag{42}$$

The variables $UPKEEP$, $FMAR$, $SMAR$, and $PTAX$ are based on data that are measured biennially in the AHS surveys. We interpolated the off–year values from the survey values, where necessary. We have done additional robustness checks using just the survey–year data, and the results reported in the text were qualitatively unchanged.

The variable $RECMOVER$ is computed in 1985-95 as:

$$RECMOVER = \begin{cases} 1 & RMR = 1 \\ 0 & RMR \neq 1 \end{cases}. \tag{43}$$

For the year 1997, we use the variables $MOVYR1$ and $MOVM1$ to construct a variable just like RMR, and then code $RECMOVER$ as above.

The variable $FORSALE$ is computed as:

$$FORSALE = \begin{cases} 1 & MARKET = 1 \\ 0 & MARKET \neq 1 \end{cases} . \tag{44}$$

The variable AGE is computed as the difference between the survey year and the value of $BUILT$.

B Numerical Solution Algorithm

In general, it is computationally infeasible to produce high–resolution pictures of the attribute and option price functions under the model developed in section 2. However, because the attribute and option price functions are smooth, low-resolution pictures suffice to reveal most of the important features of the functions. [19] Even so, it remains a non–trivial computational challenge to solve the partial differential equations that describe the asset prices at a level of resolution that is useful. In this section, we briefly outline the software tool that we designed for solving the partial differential equations.

The partial differential equations (PDEs) that describe attribute and asset prices are *elliptical*, meaning that the partial derivative with respect to time does not appear. This is in contrast to the *parabolic* PDEs that describe the prices of assets with fixed maturities. There are two approaches that one can take to the problem of solving for an asset price function when the PDE that describes the function is elliptical. One is to solve the PDE by exploiting the "steady–state" nature of the function. This approach typically leads to large, sparse–matrix algebra based algorithms. The second approach is to treat the asset as if it has a very long fixed maturity, and to solve the resulting parabolic PDE. In essence, one trades off computer memory against the number of computations when deciding which approach to adopt. Solving the elliptical problem requires large amounts of memory (as well as a good deal of computation), while solving the parabolic problem requires much less memory, at the cost of more computation. A key advantage of the parabolic approach is that it is straightforward to parallelize the solution algorithm, and so we adopted this approach.

To keep this discussion as brief as possible, we assume that the reader is familiar with the explicit finite difference method for solving parabolic partial differential equations, and is also familiar with a multi–threading subroutine library (e.g., pthreads). Our algorithm is essentially a divide–and–conquer algorithm, in which the state space is broken up into cubes, and a processor (thread) is assigned to make a single step of the explicit finite difference algorithm on a given cube. The following bit of C–code lays out the main portion of the algorithm for calculating attribute prices:

[19] A significant exception is the case where one is concerned with the precise location of the boundary between the continuation and exercise regions.

```
1  for (i = 0; i < M3; i++)
2    for (j = 0; j < M2; j++)
3      for (k = 0; k < M1; k++)
4        X[0][i][j][k] = rent_flow[j];
5  for (n = 0; n < NUMPROC; n++)
6    pthread_create (&thread[n],
7                    &thread_data[n],
8                    finite_difference,
9                    (void *) &thread_id[n]);
10 for (t = 0; t < S*T; t++)
11   {
12       while (count < NUMPROC)
13         pthread_cond_wait (&thread_cond,
14                            &thread_mutex);
15       count = 0;
16       for (i = 0; i < M3; i++)
17         for (j = 0; j < M2; j++)
18           for (k = 0; k < M1; k++)
19             X[0][i][j][k] = X[1][i][j][k] +
20                             rent_flow[j];
21       pthread_cond_broadcast (&sync_cond);
22   }
```

We have divided the solution space into $M1 * M2 * M3$ equally-spaced points on the unit cube, after transforming the support of each state variable to the unit interval by means of the transform:

$$f(x) = \frac{x}{\lambda + x}. \tag{45}$$

The parameter λ is a "packing factor" that is used to increase the number of solution points in a particular region of the state space. Another way to think of λ is that it increases the resolution of the solution by "focusing" on certain regions of the state space. For ρ and σ, we set λ equal to the long–run means of the processes. For π, we set λ equal to one. Let σ_i, for $i = 1, 2, \ldots, M3$ denote the discrete values of σ after applying the transform (45). The other variables are similarly transformed, producing π_j for $j = 1, 2, \ldots, M2$, and ρ_k for $k = 1, 2, \ldots, M3$.

With lines 1–4, we initialize value of the attribute at $(\sigma_i, \pi_j, \rho_k)$ to the rental flow π_j:

$$P(\sigma_i, \pi_j, \rho_k) = \pi_j.$$

Lines 5-9 start the threads. The `thread_data` structure contains the data that each thread needs to determine the part of the solution space to which it is assigned. The function `finite_difference` contains code that makes a single step of the finite difference algorithm on the piece of the solution space defined in `thread_data`.

The outer–enclosing loop that begins at line 10 steps backward through time (here `t`$=0$ is the last time period). The parameter T is the number of years over which we run the solution, and S is the number of increments per year (each time step is a length of time equal to $\frac{1}{S}$). A value of `T` large enough to produce solution values that differ by less than some value ϵ from the solution values at $T-1$ can be found by trial–and-error. We used a value of `T` $= 200$ (200 years); this value produced prices that differed by less than a cent from `T` $= 201$. We used $S = 365$, or time steps of one day in length. [20]

Lines 12-14 operate in tandem with line 21 to synchronize the threads. The main thread of control waits at line 13 until the mutex–protected variable `count` reaches the value `NUMPROC` (when a thread finishes a step, it locks the mutex, increments `count`, releases the mutex, and waits on the mutex–protected variable `synch_cond`.).

When `count = NUMPROC`, the main thread continues execution, re–setting `count` to zero at line 15. At lines 16-20 the solution is copied from `X[1]` back to `X[0]`, and an additional flow of rents is added to the attribute value. Line 21 signals the threads to perform another finite difference step. The loops exits when the current time $t = S * T$ is reached.

The data in the global storage array `X` are organized so as to minimize cache–misses. The data are organized such that an entire block of memory locations along the ρ axis will be copied in succession; the calculations in `finite_difference` are similarly arranged so as to run along this block of memory locations, thereby minimizing cache–misses.

For our pricing experiments, we set $M1 = M2 = M3 = 50$. The algorithm executed in approximately one hour using 50 threads on a dedicated 4-CPU Sun UltraSparc 2000.

[20]The length of the time step sufficient to guarantee that the solution is both stable and consistent can be found by trial and error or by analytic means. See Hull and White (1990) for a discussion of analytic methods.

References

Ames, W. F.: 1977, *Numerical Methods for Partial Differential Equations*, Academic Press, New York.

Andersen, T. G. and Lund, J.: 1997, Estimating continuous-time stochastic volatility models of the short term interest rate, *Journal of Econometrics* **77**(2), 343–377.

Blanchard, O. J. and Katz, L. F.: 1992, Regional evolutions, *Brookings Papers on Economic Activity* (1), 1–61.

Capozza, D. R. and Li, Y.: 1994, The intensity and timing of investment: The case of land, *American Economic Review* **84**(4), 889–904.

Capozza, D. R. and Sick, G. A.: 1991, Valuing long-term leases: The option to redevelop, *Journal of Real Estate Finance and Economics* **4**(2), 209–223.

Case, B. and Quigley, J. M.: 1991, The dynamics of real estate prices, *The Review of Economics and Statistics* **73**(1), 50–58.

Chung, K. L. and Williams, R.: 1990, *An Introduction to Stochastic Integration*, Birkhäuser, Boston.

Commerce Clearing House: 1999, *U. S. Master Tax Guide*, Commerce Clearing House, Chicago.

Dixit, A. K. and Pindyck, R. S.: 1994, *Investment under Uncertainty*, Princeton University Press, Princeton, NJ.

Duffie, D.: 1996, *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, NJ.

Goetzmann, W. and Spiegel, M.: 1995, Non-temporal components of residential real estate appreciation, *Review of Economics and Statistics* **77**(1), 199–206.

Greene, W. H.: 1993, *Econometric Analysis*, Macmillan Publishing Company, New York.

Griliches, Z.: 1986, Economic data issues, *in* Z. Griliches and M. Intriligator (eds), *Handbook of Econometrics*, Vol. 3, North–Holland, Amsterdam.

Hull, J. and White, A.: 1990, Valuing derivative securities using the explicit finite difference method, *Journal of Financial and Quantitative Analysis* **25**(1), 87–100.

Joint Center for Housing Studies, Harvard University: 1999, Improving america's housing.

Karatzas, I. and Shreve, S. E.: 1991, *Brownian Motion and Stochastic Calculus*, Springer–Verlag, New York, NY.

McFadden, D. and Train, K.: 1998, Mixed mnl models for discrete response. Working Paper. University of California at Berkeley.

Mills, E. S. and Simenauer, R.: 1996, New hedonic estimates of regional constant quality house prices, *Journal of Urban Economics* **39**, 209–215.

Ortalo–Magne, F. and Rady, S.: 1999, Boom in, bust out: Young households and the housing price cycle, *European Economic Review* **43**(4–6), 755–766.

Poterba, J. M.: 1984, Tax subsidies to owner–occupied housing: An asset–market approach, *The Quarterly Journal of Economics* **99**(4), 729–752.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P.: 1994, *Numerical Recipes in C*, second edn, Cambridge University Press, Cambridge.

Revelt, D. and Train, K.: 1999, Customer–specific taste parameters and mixed logit. Working Paper. University of California at Berkeley.

Rosen, S. and Topel, R. H.: 1988, Housing investment in the United States, *Journal of Political Economy* **96**(4), 718–740.

Smith, G. D.: 1996, *Numerical solution of partial differential equations: Finite difference methods*, 3rd edn, Oxford University Press, Oxford.

Train, K.: 1999, Halton sequences for mixed logit. Working Paper. University of California at Berkeley.

U. S. Census Bureau: 1997, *Current Housing Reports, Series H150/97, American Housing Survey for the United States: 1997*, U.S. Government Printing Office, Washington, D. C.

U. S. Census Bureau: 1998, *Codebook for the American Housing Survey, Volume 2: Supplement for 1984-96*, U.S. Government Printing Office, Washington, D. C.

U.S. Census Bureau: 1990, *Codebook for the American Housing Survey Database: 1973 to 1993*, U.S. Government Printing Office, Washington, D. C.

Williams, J. T.: 1993, Equilibrium and options on real assets, *Review of Financial Studies* **6**(4), 825–850.