An Estimatable DCDP Model of Search and Matching in Real Estate Markets

Stuart J. Fowler*
Middle Tennessee State University
and
Jennifer J. Wilgus†
Belmont University

September 7, 2011

Abstract

The primary purpose of this study is to introduce an estimatable model of search and matching in real estate markets. A benefit of developing such a theory is so we may better understand the structure that determines these choices. The DCDP model that we propose produces the following results. First, the model is able to replicate several salient features of real estate markets. Second, the estimation method is able to accurately and efficiently recover the structural demand and supply functions of the buyers and sellers.
Keywords: Real Estate; Search and Matching; Structural Estimation; MCMC; and Gibbs-Sampler.

*Stuart J. Fowler: Economics and Finance Department, MTSU Box 27, Murfreesboro, TN 37132. Tel: 615-898-2383. Fax: 615-898-5596. Email: Stuart.Fowler@mtsu.edu.

†Jennifer J. Wilgus, College of Business Administration, Belmont University, Nashville, TN 37212, Tel: 615-460-6319.

An Estimatable DCDP Model of Search and Matching in Real Estate Markets

1 Introduction

A key feature of discrete choice dynamic programming (DCDP) models is that equilibrium decisions are made on a reservation price basis; an action is undertaken when the offer price exceeds the reservation price.[1] The reservation price is determined by comparing current offers to the expected discounted value of potential future offers. In this case, the dynamic programming problem is of searching for offers until a match is found. As noted by Miller (1978), models of search and matching naturally extend to real estate markets – a prime example is the decision to accept an offer on a real estate property. A benefit of developing theories on discrete real estate choices is so we may better understand the structure that determines these choices.

Yinger (1981) is credited with first applying a search model to real estate. His study models real estate broker search efforts. In this case, broker's search effort is weighed against the uncertainty of the arrival and size of offers to the expected discounted costs of brokering an unsold house. Consequent studies have used search and matching models for the purpose of examining the relationships between, for example, price, vacancies, and the time it takes a home to sell (*e.g.*, Wheaton 1990; Stein 1995; Read 1997; Caplin and Leahy 2008; Cheng, Lin, and Liu 2008). Though theoretical DCDP models of search and matching are becoming more prominent in the real estate literature, there are no studies on the structural estimation of these models. Structural estimation is desirable as it captures the dynamic forward looking behavior of individuals. Additionally, structural estimation provides a vehicle for the testing of our search and matching theories. The dearth of research on structural estimation is

[1] DCDP models are used extensively in labor economics. See Keane and Wolpin (2008) for a review of applications of DCDP models to labor.

due to, as we see it, the nature of the data sets used in real estate research. Real estate researchers often have large and detailed data sets that make the estimation of structural dynamic programming problems difficult since the "curse of dimensionality"[2] is a significant obstacle. Thus, the primary purpose of our study is to introduce an *estimatable* model of search and matching in real estate markets.

The model and estimation method we implement has several key features. To start, the model is stylized. By stylized we imply that the model is in the same spirit as Yinger (1981). That is, the demand for property follows a simple hedonic pricing relationship. Additionally, supply decisions depend on the seller's preferences (utility) and beliefs of the structure of future offers. Next, the model is flexible as it is easy to extend to include idiosyncratic elements to agent's choices not considered by Yinger (1981). One such extension considered relates to differences in how agents discount current and future payoffs. Third, the estimation method is derived from a class of Bayesian methods introduced by Geweke and Keane (2000). A key to their method is that the value function is approximated thus breaking the "curse of dimensionality" to the dynamic programing problem.[3]

The main results of the study are twofold. First, we show that the DCDP model is able to replicate several salient features of real estate markets. For example, more patient sellers (future oriented) receive higher sales prices but take longer to sell. In this case, when preferences are heterogenous with respect to patience, the simulated data from the DCDP model generates positive cross-sectional relationships between sales prices and time on the market – a well know empirical fact in real estate research (*e.g.*, Miller 1978; and Asabere and Huffman 1993). As another example, falling demand (represented by falling offer rates) is smoothed over the seller's two margins of choices: maximize selling price while minimizing

[2] In solving the dynamic programming problem, the Bellman equation has to be solved at each possible point in the state space. The possible number of points in the state space increases exponentially with the increased dimension of the state space. This is commonly referred to as the "curse of dimensionality."

[3] Essentially, the dynamic programming problem is only solved once during a single estimation routine. See Houser (2003) for an application of the method to a DCDP model of labor supply and saving.

time on the market. In this case, falling demand is followed by a *partial* lowering of the seller's reservation price. As a result, as sales prices fall time on the market increases – again, replicating a well know empirical time series relationship in real estate (the positive price-volume correlation anomaly is studied in Stein 1995 and Genesove and Mayer 1997).

The second main contribution of this paper is that the estimation method is able to accurately and efficiently recover the demand and supply functions of the agents. The method estimates the true demand for housing by generating, in conjunction with the supply of housing, offers not observed due to rejection by the seller. The supply of housing (accepting or rejecting an offer) is recovered from the probabilistic structure implied by their choices of rejection or acceptance. In all, the structural parameters of supply and demand are found to be completely consistent with actual behavior and are typically within a standard deviation of the true values.

The rest of this paper is organized as follows. Section 2 defines the discrete choice model. Section 3 quantifies the equilibrium defined by the model calibration. The Bayesian estimator and it's distribution are presented in Section 4. Finally, the paper concludes with section 5.

2 The Model

The model economy is populated by agents who are separated into two categories: sellers and buyers. A seller, who has already exogenously decided to list the property, solves a stochastic dynamic programming problem by choice of a reservation price. The two types of stochastic uncertainty that face the sellers are: (i) uncertainty about the arrival of buyers and (ii) uncertainty about the amount of offer to be made on their property. The i'th seller weighs the benefits of accepting the offer against the expected discounted value of future offers given the uncertainty and a set of the current states $\Omega_{i,t}$.

For the model economy, time evolves in discrete units called periods (which are specified to be one month long in the quantitative results to follow). At the start of the economy

($t = 0$), a seller announces a preset listing price. In the beginning of a period, an offer is made on the house. At the end of the period, the seller makes a decision to sell or hold. An accepted offer means that bidding ends. If an offer is rejected or if an offer doesn't arrive, time evolves one period and the process begins again.

2.1 Housing Demand

Housing demand follows an exogenous Bernoulli process with an offer rate of $1 - \lambda$ (alternatively, an offer will not be made with probability λ). When an offer is made, it is assumed that each buyer values the property by its attributes plus some idiosyncratic part that is specific to their utility. That is, the buyer makes a time t offer on the i'th house that follows a hedonic pricing formula (demand) of:

$$\ln P_{i,t} = \mathbf{x}_i \boldsymbol{\alpha} + \varepsilon_{i,t}, \tag{1}$$

where $\varepsilon_{i,t} \sim N(0, \sigma_\varepsilon^2)$ and $\mathbf{x}_i$ is a vector of housing characteristics. The $\mathbf{x}_i \boldsymbol{\alpha}$ term represents the deterministic value of the house agreed upon by all sellers. Due to differences in utility, each buyer has idiosyncratic preferences for each property; $\varepsilon_{i,t}$ are the differences in valuation by the buyers. Because $\varepsilon_{i,t}$ is random from the sellers point of view, it represents part of the seller's uncertainty about the offer process.

2.2 Asking Price and Discounting

At the beginning of the economy, the i'th seller announces a listing price $\bar{P}_i$ for their property. The listing price guarantees, regardless if the seller accepts or rejects the offer, that the seller will pay a fixed commission cost of $cP_{i,t}$ for any time t offer that is at or above the listing price (presumably to a broker). The lost commission is modeled as:

$$LC_{i,t} = -I[P_{i,t} - \bar{P}_i \geq 0]cP_{i,t},$$

where $I[\cdot]$ is an indicator function that is one where it's argument is true and zero elsewhere.

There are two types of discounting modeled. Both types of discounting are intended to capture the documented abnormal behavior of high loan-to-value (LTV) sellers. For example, Genesove and Mayer (1997) find that owners with LTVs greater than 80% have a higher expected time on the market and, if they sell, receive a higher price than owners with low LTV ratios. Their idea is simple; home sales are used to finance new real estate purchases and, therefore, high LTV owners must set a reservation price at no less than the sum of the minimum down payment required on a new home. High LTV owners are therefore either more patient or they discount all payoffs – current and future – more.

In the first type of discounting modeled, all net payoffs are normalized relative to the listing price. Homes with high listing prices (all else equal) discount the payoffs from both current and future sales at higher rates. The other type of discounting involves the actual rate of time preference, β_i. All else equal, higher rates of time preference indicate greater patience. Both listing prices and time preference rates are therefore modeled heterogeneously with respect to the i'th seller.

2.3 The Seller's Preferences and the Bellman Equation

If the offer arrives and the seller accepts the offer of $P_{i,t}$, she then receives, besides the sales payoff, a random utility component $\bar{\epsilon}_i^{<1>} + \epsilon_{i,t}^{<1>}$. The total value of accepting an offer, net of commissions to a broker, at the optimal choice is:

$$V^{<1>}(\Omega_{i,t}) = (1-c)\frac{P_{i,t}}{\bar{P}_i} + \bar{\epsilon}_i^{<1>} + \epsilon_{i,t}^{<1>},$$

where $\Omega_{i,t}$ is state space faced by the seller. It is important to distinguish the role $\bar{\epsilon}_i^{<1>} + \epsilon_{i,t}^{<1>}$ plays in determination of the utility from a choice to accept the offer. When a homeowner is mismatched with their current house, presumably, a choice to sell will bring a new value for housing. In this case, $\bar{\epsilon}_i^{<1>} + \epsilon_{i,t}^{<1>}$ is to represent the discounted present value of benefits to be received from new housing. Because these benefits are not identified by the econometrician, they are modeled as an unobserved random component.

If an offer doesn't arrive at time t, the seller's value function is then a probabilistic sum of the discounted expected value of being a seller with an offer at time $t+1$ (denoted $\beta_i E_t \{V(\Omega_{i,t+1})\}$) and the discounted value of being a seller without an offer (denoted $\beta_i V^{<0>}(\Omega_{i,t})$) and given by:

$$V^{<0>}(\Omega_{i,t}) = (1-\lambda)\beta_i E_t \{V(\Omega_{i,t+1})\} + \lambda\beta_i V^{<0>}(\Omega_{i,t}).$$

Solving for $V^{<0>}(\Omega_{i,t})$ gives:

$$V^{<0>}(\Omega_{i,t}) = \tilde{\beta}_i E_t \{V(\Omega_{i,t+1})\},$$

where $\tilde{\beta}_i = \frac{(1-\lambda)\beta_i}{1-\lambda\beta_i}$.

Alternatively, if the home owner receives and then rejects the offer, a showing cost is paid – both indirect and direct. Indirect showing costs result from the lost utility of not having sold the house and, for example, having strangers within the house. A direct showing cost includes the necessary expenditures needed to keep the house in showing condition. Total unobserved utility lost from rejecting the offer is denoted $\bar{\epsilon}_i^{<2>} + \epsilon_{i,t}^{<2>}$. The value of rejecting an offer (relative to the listing price) is:

$$V^{<2>}(\Omega_{i,t}) = \frac{LC_{i,t}}{\bar{P}_i} + \bar{\epsilon}_i^{<2>} + \epsilon_{i,t}^{<2>} + \tilde{\beta}_i E_t \{V(\Omega_{i,t+1})\}.$$

Finally, the household solves the Bellman equation:

$$\begin{aligned} V(\Omega_{i,t}) &= \max_{a_{i,t}\in\{1,2\}} \left\{ V^{<1>}(\Omega_{i,t}), V^{<2>}(\Omega_{i,t}) \right\} \\ &= \max_{a_{i,t}\in\{1,2\}} \left\{ \begin{array}{l} (1-c)\frac{P_{i,t}}{\bar{P}_i} + \bar{\epsilon}_i^{<1>} + \epsilon_{i,t}^{<1>}, \\ \frac{LC_{i,t}}{\bar{P}_i} + \bar{\epsilon}_i^{<2>} + \epsilon_{i,t}^{<2>} + \tilde{\beta}_i E_t \left\{ V(\Omega_{i,t+1}) \right\} \end{array} \right\}, \end{aligned}$$

where $a_{i,t} = 1$ is the value of the action if the offer is accepted and $a_{i,t} = 2$ is the value of the action if the offer is rejected.

2.4 Solution Algorithm

To solve the model, we could proceed by discretization of the state space. Because the state space is large, however, the setup and resulting search over the space is rather complicated. To see this, note that the set $\Omega_{i,t}$ includes the list of variables $\{\mathbf{x}_i, \varepsilon_{i,t}, \bar{\epsilon}_i^{<1>} \epsilon_{i,t}^{<1>}, \bar{\epsilon}_i^{<2>}, \epsilon_{i,t}^{<2>}\}$, of which a solution needs to be found for every combination (permutation). Rust (1987) develops a key procedure that allows us to integrate out some of states thereby reducing the state-space size. Assuming that $\epsilon_{i,t}^{<1>}$ and $\epsilon_{i,t}^{<2>}$ are independently and identically distributed Extreme Value Type-I (EVI), we can rewrite the Bellman equation for $a_{i,t} = 2$ as:

$$V^{<2>}(\Omega_{i,t}) - \epsilon_{i,t}^{<2>} = \frac{LC_{i,t}}{\bar{P}_i} + \bar{\epsilon}_i^{<2>} + \tilde{\beta}_i E_t \left\{ V(\Omega_{i,t+1}) \right\}.$$

Because $\epsilon_{i,t}^{<2>}$ is distributed EVI, $\tilde{\beta}_i E_t \left\{ \epsilon_{i,t+1}^{<2>} \right\} = \tilde{\beta}_i \gamma$ where γ is Euler's constant. It can be shown that:

$$V^{<2>}(\Omega_{i,t}) - \epsilon_{i,t}^{<2>} = \frac{LC_{i,t}}{\bar{P}_i} + \bar{\epsilon}_i^{<2>} + \tilde{\beta}_i E_t \left\{ V(\Omega_{i,t+1}) \right\} - \tilde{\beta}_i E_t \left\{ \epsilon_{i,t+1}^{<2>} \right\} + \tilde{\beta}_i \gamma.$$

It follows that:

$$\bar{V}^{<2>}(\Omega_{i,t}) = \frac{LC_{i,t}}{\bar{P}_i} + \bar{\epsilon}_i^{<2>} + \tilde{\beta}_i E_t \left\{ \bar{V}(\Omega_{i,t+1}) + \gamma \right\}, \tag{2}$$

where $V^{<2>}(\Omega_{i,t}) - \epsilon_{i,t}^{<2>} = \bar{V}^{<2>}(\Omega_{i,t})$. For the same reasoning, we can get:

$$\bar{V}^{<1>}(\Omega_{i,t}) = (1-c)\frac{P_{i,t}}{\bar{P}_i} + \bar{\epsilon}_i^{<1>}. \tag{3}$$

Finally, as shown by Rust (1987), the expected value of being a seller at time $t+1$, net of an EVI random variable, can be expressed as:

$$E_t\left\{\bar{V}(\Omega_{t+1}) + \gamma\right\} = \mathbf{P}\left[\log\left(\exp(\bar{V}^{<1>}(\Omega_{t+1})) + \exp(\bar{V}^{<2>}(\Omega_{t+1}))\right) + \gamma\right], \tag{4}$$

where $\mathbf{P}$ is a transition probability matrix that describes the transitions of the remaining states.

The process revolves around solving for $\bar{V}^{<1>}(\Omega_{i,t})$ and $\bar{V}^{<2>}(\Omega_{i,t})$ by employing the following solution algorithm:

Definition 1 *Dynamic Programming Solution Algorithm:*

1. *Initiate a guess at* $\bar{V}_0^{<1>}$ *and* $\bar{V}_0^{<2>}$.

2. *Compute a new set of expected value functions,* $E_t\left\{\bar{V}(\Omega_{t+1})\right\}$, *by equation (4) given* $\bar{V}_0^{<1>}$ *and* $\bar{V}_0^{<2>}$.

3. *For each state, compute a new set of value functions,* $\bar{V}_1^{<1>}$ *and* $\bar{V}_1^{<2>}$, *from equations (3) and (2) using* $E_t\left\{\bar{V}(\Omega_{t+1})\right\}$ *from step 2.*

4. *Compute the optimal policy solution,* $a_t \in \{\text{accept,reject}\}$, *by evaluation of:*

$$\max\{\bar{V}_1^{<1>}, \bar{V}_1^{<2>}\}.$$

5. *Compute the distance* $\xi = \left\|[\bar{V}_1^{<1>}, \bar{V}_1^{<2>}] - [\bar{V}_0^{<1>}, \bar{V}_0^{<2>}]\right\| / (1 + \left\|[\bar{V}_0^{<1>}, \bar{V}_0^{<2>}]\right\|)$. *If* ξ *is smaller than some preset number then stop and a solution has been found. Else,*

set

$$[\bar{V}_0^{<1>}, \bar{V}_0^{<2>}] = [\bar{V}_1^{<1>}, \bar{V}_1^{<2>}],$$

and return to step 2.

3 Experimental Design

In this section, we employ simulation in order to draw some conclusions about the model real estate market. First, to simulate the real estate search economy, we must calibrate the model by specifying the parameters. Section 3.1 identifies and discusses the parameters used in the calibration. The simulation follows in Section 3.2. Next, a traditional method of estimation, reduced form regression, is presented in Section 3.3. The results indicate that this type of analysis overestimates the mean value buyers place on home services and provides an impetus for a structural estimation technique: Bayesian maximum likelihood.

3.1 Calibration

The computation of the equilibrium can begin once the parameters of the model are defined. The demand for housing is specified by:

$$\ln P_{i,t} = \alpha_0 + \alpha_1 SQ_i + \alpha_2 Bath_i + \varepsilon_{i,t},$$

where $\alpha_0 = 6$, $\alpha_1 = 1.5$, and $\alpha_2 = 1$. Most housing characteristics can be described by discrete units. For example, let $SQ_i = \{1, 2, 3\}$ for small (less than 1500 sqft), medium (1500-3500 sqft), and large homes (greater than 3500 sqft), respectively. The number of baths are easily discretized as well: $Bath_i = \{1, 2, 3\}$. For the state $\varepsilon_{i,t}$ that is distributed $N(0, \sigma_\varepsilon^2)$, we use the nodes and weights (as probabilities) from a 40 point quadrature of a normal distribution when $\sigma_\varepsilon^2 = 0.14$. The probability of no-offer is initially set at $\lambda = 0.0$.

This value will be changed in the simulation results reported below.

Two types of time preference rates are calibrated. The first is set to match a monthly discount rate consistent with an interest rate of 3 percent per annum; this sets $\beta_i = (1/1.03)^{1/12} = 0.99754$. The second calibration is to represent an individual who heavily discounts the future leading to a low measure for β_i; for this case we set β_i so that it is 5% smaller than the previous calibration giving $\beta_i = 0.99754/1.05 = 0.95004$. The fraction of the total price that is paid as commission to the real estate broker is set at $c = 0.07$, a proxy based on industry standards where a broker lists and sells the property. The random utility components are arbitrarily set to $\bar{\epsilon}_i^{<1>} = 0$ and $\bar{\epsilon}_i^{<2>} = -1.30$.[4] Finally, we consider two sets of listing price, $\bar{P}_i$, values. The first is set equal to 1.5 of the expected value of the offer; $\bar{P}_i = 1.5E[P_i] = 1.5\exp(\alpha_0 + \alpha_1 SQ_i + \alpha_2 Bath_i + 1/2\sigma_\varepsilon^2)$. The second calibration is a high measure for the listing price; in this case we set $\bar{P}_i = 2.0E[P_i]$.

With the parameters calibrated and state space defined, the next step is to construct a probability matrix $\mathbf{P}$ so that the expectation $E_t\bar{V}(\Omega_{i,t+1})$ can be computed. Consider a home owner with one bath and 1,000 sqft ($SQ = 1$ and $Bath = 1$). The current choice (sell or hold) depends on their expectations of future events (such as future offers). However, some future offers will never occur; the house will never mutate to a two bath 9,500 sqft home in the future (still the same $SQ = 1$ and $Bath = 1$ house). Therefore, the probability of an offer defined on those states occurring must be zero, whereas the probability of a $SQ = 1$ and $Bath = 1$ house transitioning to itself is one. The remaining rows and columns in $\mathbf{P}$ describe the evolution of the states $\varepsilon_{i,t}$. In this case we choose the quadrature weights as the probabilities for $\varepsilon_{i,t}$.

The optimal allocations for 6,000 sellers are simulated. At $t = 0$, each seller is randomly assigned a house with square-footage and number of bath attributes of $SQ = \{1, 2, 3\}$ and $Bath = \{1, 2, 3\}$ with probability of $\Pr\{SQ\} = \{.25, .50, .25\}$ and $\Pr\{Bath\} = \{.25, .50, .25\}$,

[4] Initially, our choice of calibration gives $\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>} = 1.30$. This implies that the net present value of selling the house is 30% above the listing price. This value will be changed in the simulation exercise below.

respectively. Additionally, sellers are assigned one of two listing prices and time preference rates of $\bar{P}_i = \{1.5E[P_i], 2.0E[P_i]\}$ and $\beta = \{0.99754, 0.95004\}$ with probabilities of $\Pr\{\bar{P}_i\} = \{.75, .25\}$ and $\Pr\{\beta\} = \{.75, .25\}$, respectively. Offer shocks, $\varepsilon_{i,t}$, are randomly drawn for each house with probability of $1 - \lambda$ and consistent with the quadrature weights as probabilities.

If an offer is received and is above the reservation price solution, the offer is accepted and the price and time on the market for the house is recorded. Alternatively, if the price offer is below the reservation price solution or an offer is not received on a house, time evolves one period and the process begins again. The entire sample that consists of every type of calibration is denoted the "total sample" model. The "baseline" model includes those sellers with $\bar{P}_i = \{1.50E[P_i]\}$ and $\beta = \{0.99754\}$ calibrations. The "high listing" model includes those sellers with $\bar{P}_i = \{2.0E[P_i]\}$ and $\beta = \{0.99754\}$ calibrations. The "low discounting" model includes those sellers with $\bar{P}_i = \{1.5E[P_i]\}$ and $\beta = \{0.95004\}$ calibrations. Table 1 lists the results of all calibrations.

3.2 Simulation Results

Column (A) in Table 2 shows averages for the simulated logged selling price. For the total sample, the mean logged sales price is 11.36 (or $80,113.74 in levels). The logged sales price is lower than the average listing price of 11.57 (or $106,306.04 in levels). More specifically, the ratio of the average sales price to listing price ratio implies a 16 basis point discount on the sales price relative to the listing. This does not mean, however, that the house is sold at a discount relative to the reservation price. The average reservation price to sales price ratio is 0.73 implying houses are sold at a premium relative to the sellers' reservation price. In terms of the listing price, the average reservation price to listing price ratio is 0.58.

Column (A) in Table 2 also shows how sensitive the allocations are to changes in λ, a measure of tightness of demand in the market. As λ increases, we first see that it takes longer to sell a house; the mean time on the market rises to 3.26 and 8.53 months for $\lambda = 0.25$ and $\lambda = 0.75$, respectively. The average seller becomes more impatient as is evident by the falling average reservation to listing price ratios; the ratios are 0.57 and 0.47 for $\lambda = 0.25$ and $\lambda = 0.75$, respectively. The implication is that the sellers abate the effects of, in part, the longer time on the market by lowering reservation prices. In total, logged sales prices and sales to listing price ratios fall while the average time on the market rises.

Panels (A)-(B) in Figure 1 illustrate the behavior of different types of sellers resulting from the different types of calibrations. First, both panels in Figure 1 show that the "low discounting" seller is, as expected, more impatient. The estimated probability densities (estimated using a Epanechnikov kernel) for sales price and time on the market are shifted, relative to the other calibrations, to the left for the "low discounting" agents. For example, in Panel (A), at the log price of 9, "low discounters" have a higher probability of sale relative to the others. Whereas at the price of log price of 14, they have a lower probability of sale. Additionally, Panel (B) of Figure 1 illustrate most homes are sold within 5 months for sellers categorized as "low discounters" which is considerably less than the other agents. The cost of their impatience is a lower reservation price and, therefore, a lower sales price. "High listing" sellers behave similar to "baseline" sellers. Panels (A) and (B) in Figure 1 and column (C) of Table 2 show that the "high listing" sellers take about the same time to sell (about 2.93 compared to 2.92 months) and at about the same price as "baseline" agents (logged sales price is 11.40 compared to 11.43).

An interesting result emerges from the table with regards to the sensitivity of the allocations to changes in λ. More specifically, both the "low discounting" and "high listing" sellers alter their reservation prices when faced with falling offer rates. For example, the "high listing" seller's average logged sales price falls from 11.40 with $\lambda = 0$ to 11.23 with $\lambda = 0.75$. Alternatively, the "baseline" sellers do not noticeably alter their reservation price. Instead, when faced with falling offer rates, a baseline agent keeps to their reservation price and patiently waits for more offers. The appearance of patience is due, in part, to the high value for β_i. The "baseline" seller's true rate of time preference, defined by $\tilde{\beta}_i = \frac{(1-\lambda)\beta_i}{1-\lambda\beta_i}$, is insensitive to changes in λ; $\tilde{\beta}_i = 0.99754$ with $\lambda = 0$, and $\tilde{\beta}_i = 0.99023$ with $\lambda = 0.75$.

Now consider a change in $\bar{\epsilon}_i^{<1>}$ and $\bar{\epsilon}_i^{<2>}$. More specifically, we let $\bar{\epsilon}_i^{<2>} = -1.25$ thereby lowering the net benefit, $\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>}$, of selling. The last rows in Table 2 show the simulation results for this calibration. Relative to the previous case where $\lambda = 0$ (first six rows of the Table), we see that sellers become more patient; the average sales price increases as well as the average time on the market. Essentially, changes in $\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>}$ act to represent how (mis)matched home owners are with their current housing situation. With high values of $\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>}$, the net benefit of selling the house increases; it is more desirable to sell. Low values for $\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>}$ indicate that home owners are rather content with their current housing situation; it is less desirable to sell.

The model is successful in its ability to replicate some salient macroeconomic features of real estate economies. An important example is how prices and time on the market co-move over time. Falling real estate prices usually bring large decreases in sales volume and, thus, higher inventories of unsold homes (Stein 1995; and Genesove and Mayer 1997). In our model, a less tight market (*i.e.*, falling demand) results in falling sales prices and increased time on the market (rising λ in Table 2). Essentially, a falling demand results in the seller economizing on both margins of choices: selling price and time on the market. As a result, a positive correlation is generated between the two choices. Note that the positive correlation is a sole result of the seller's desire to smooth falling demand and not a result of the seller's refusal to "recognize reality" or borrowing constraints.[5]

As another example, sellers with higher β's (future oriented) receive higher sales prices but take longer to sell. In this case, the simulated model that includes high and low discounters generates a positive cross-sectional relationship between sales prices and time on the market

[5] Borrowing constraints, like those studied in Stein (1995), would most likely reenforce our results.

(comparison of columns (B) and (D) in Table 2) – a well known empirical fact in real estate research of cross-sections (*e.g.*, Miller 1978; and Asabere and Huffman 1993). Differences in discount factors are not the only way to generate positive cross-sectional correlations between sales prices and time on the market. Low values for $\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>}$ indicate that home owners will have higher reservation prices. Therefore, a mix of high and low values for $\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>}$ will generate positive correlations between sales prices and time on the market.

Intuition suggests that "high listing" price agents discount current offers more heavily and therefore should take longer to sell. Indeed, the empirical evidence presented in Genesove and Mayer (1997) shows that owners with high LTVs have higher listing prices. Because high LTV owners sell their houses at higher prices, listing prices and sales prices should be positively correlated. Our model has difficulty in accounting for this empirical fact – "high listing" agents are slightly more impatient than the "baseline". If, alternatively, higher listing prices are correlated with lower values for $\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>}$, then the last six rows of column (C) in Table 2 are the appropriate simulations. Notice that, when compared to the "baseline" agent presented in the first six rows of column (B) in Table 2, the "high listing" agents are much more patient; higher sales prices and longer times on the market. An assumption of "high listing" agents with low values for $\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>}$ is *a priori* reasonable.[6]

3.3 Regression Analysis

Reduced-form ordinary least squares (OLS) regression analysis typically will not be able to recover the structural parameters of the model.[7] For example, using the sample of "baseline" sellers, a regression of housing attributes on logged sales price is reported in column (B) of Table 3. Notice that the estimated intercept is much higher than the actual value used for the simulations. Because homeowners reject offers that are below their reservation prices,

[6] High LTV homeowners need to sell high as the value of selling is small. The value of selling includes the sales price and $\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>}$.

[7] Only in limited circumstances will reduced-form regressions return the structural parameters of the model. See Coulson (2008) for a discussion.

the sample of observed prices are higher than the average of the offers. The regression thus overestimates the mean value buyers place on home services.

Now suppose that our economy consisted of only "baseline" and "high listing" type of sellers (column (C) of Table 3). Additionally, assume that the "baseline" agents have larger homes. Theoretically, the "baseline" agents should sell their homes at higher prices because: (i) they have larger homes and (ii) are slightly more patient. A regression where patience is not controlled for would be biased since home attributes would pick up the effects of the omitted patience. To see this, using only "baseline" sellers with homes with more than one bath and square-footage and using only "high listing" sellers with homes with less than three bath and square-footage, a regression is run and presented in column (C) of Table 3. The results show that the estimates for square-footage and baths, α_1 and α_2, are biased upwards thereby overestimating the marginal effects of housing attributes on price.

4 Bayesian Maximum Likelihood

In this section we illustrate a method for Bayesian inference of the structural parameters of the model. The method, first introduced in Geweke and Keane (2000), implements the Gibbs sampling-data augmentation algorithm (Tanner and Wong, 1987) to the structural equations (or approximations to them) of the discrete choice model. A key feature of the algorithm is that the agent's dynamic programming problem does not need to be solved. Instead, the agent's future expected discounted payoff from a choice is approximated by a high-order polynomial in the state variables. The polynomial is then estimated to be consistent with the agent's *observed* choices and payoffs.

More specifically, the value function differences, defined by $Z_{i,t} \equiv \bar{V}^{<1>}(\Omega_{i,t}) - \bar{V}^{<2>}(\Omega_{i,t})$, are decomposed into two parts. The first part is sometimes (partially) observed by the econometrician; $[(1-c)P_{i,t} - LC_{i,t}]/\bar{P}_i$. It will not be observed from censoring; if an offer is not made or the seller rejects an offer, $P_{i,t}$ is typically not recorded and thus unobserved.

The second part, $\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>} - \tilde{\beta}_i E_t \{V(\Omega_{i,t+1})\}$, is always unobserved to the econometrician. As such, we will approximate the second part utilizing a polynomial function plus an error term:

$$\bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>} - \tilde{\beta}_i E_t \left\{\bar{V}(\Omega_{i,t+1})\right\} = g(\mathbf{x}_i, \boldsymbol{\psi}) + \varepsilon_{i,t}^V,$$

where $\varepsilon_{i,t}^V \sim N(0, \sigma_V^2)$. For the specific case of a second-order polynomial (which we consider below), the approximation takes the form:

$$g(\mathbf{x}_i, \boldsymbol{\psi}) = \psi_0 + \psi_1 SQ_i + \psi_2 Bath_i + \psi_3 SQ_i^2 + \psi_4 Bath_i^2 + \psi_5 SQ_i Bath_i. \tag{5}$$

Finally, substituting (5) into the value function difference highlights the two parts:

$$Z_{i,t} = \underbrace{[(1-c)P_{i,t} - LC_{i,t}]/\bar{P}_i}_{\text{partially observed}} + \underbrace{g(\mathbf{x}_i, \boldsymbol{\psi}) + \varepsilon_{i,t}^V}_{\text{unobserved}}. \tag{6}$$

4.1 The Algorithm

Our Bayesian analysis of this model via a Gibbs sampling algorithm first forms the complete data likelihood function. That is, we consider the likelihood function that could be formed if we had all the data on N individuals with each individual having their house on the market for T_n periods (where $n \in \{1, \ldots, N\}$). The complete data includes: (i) value function differences $Z = \{\{\bar{V}^{<1>}(\Omega_{i,t}) - \bar{V}^{<2>}(\Omega_{i,t})\}_{t=1}^{T_i}\}_{i=1}^{N}$, (ii) the complete set of price offers $P = \{\{P_{i,t}\}_{t=1}^{T_i}\}_{i=1}^{N}$ where $P_{i,t} = \{\emptyset\}$ if no offer was made, (iii) indicator variables that equals one if an offer was made $\Lambda = \{\{I[P_{i,t} > 0]\}_{t=1}^{T_i}\}_{i=1}^{N}$, (iv) a set of lost commissions $\{\{LC_{i,t}\}_{t=1}^{T_i}\}_{i=1}^{N}$, (v) a set of property characteristics $\mathbf{s} = \{\mathbf{x}_i, \bar{P}_i\}_{i=1}^{N}$, (vi) a set of policy decisions $a = \{\{a_{i,t}\}_{t=1}^{T_i}\}_{i=1}^{N}$, and (vii) all parameters Θ.

The period t housing offer indicator, $\Lambda_{i,t}$, is distributed Bernoulli with parameter λ,

$\Lambda_{i,t} \sim Be(\lambda)$. The density of logged prices conditional on $\mathbf{s}$, $\Lambda_{i,t} = 1$, and Θ is:

$$f(P_{i,t}|\mathbf{s}_i, \Theta, \Lambda_{i,t} = 1) = (2\pi\sigma_\varepsilon^2)^{-1/2} \frac{1}{P_{i,t}} \exp\left(-\frac{1}{2}\left(\ln P_{i,t} - \mathbf{x}_i\boldsymbol{\alpha}\right)^2 / \sigma_\varepsilon^2\right).$$

With everything else known and $\Lambda_{i,t} = 1$, $Z_{i,t}$ has a truncated normal distribution of:

$$Z_{i,t} \sim \begin{cases} TN_{[0,\infty)}\left((1-c)\frac{P_{i,t}}{\bar{P}_i} - \frac{LC_{i,t}}{\bar{P}_i} + g(\mathbf{x}_i, \boldsymbol{\psi}), \sigma_V^2\right) & \textit{if } a_{i,t} = 1 \\ TN_{(-\infty,0)}\left((1-c)\frac{P_{i,t}}{\bar{P}_i} - \frac{LC_{i,t}}{\bar{P}_i} + g(\mathbf{x}_i, \boldsymbol{\psi}), \sigma_V^2\right) & \textit{if } a_{i,t} = 2 \end{cases}. \tag{7}$$

In total, the period t contribution of the i'th house to the likelihood is:

$$\begin{aligned} \mathcal{L}(\Lambda_{i,t}, P_{i,t}, Z_{i,t}|\mathbf{s}_i, \Theta) &= (\lambda)^{1-\Lambda_{i,t}} (1-\lambda)^{\Lambda_{i,t}} \left((2\pi\sigma_\varepsilon^2)^{-1/2} \frac{1}{P_{i,t}} \exp(-\frac{1}{2}\left(\ln P_{i,t} - \mathbf{x}_i\boldsymbol{\alpha}\right)^2 / \sigma_\varepsilon^2)\right)^{\Lambda_{i,t}} \times \\ &\quad \left((2\pi\sigma_V^2)^{-1/2} \exp(-\frac{1}{2}(Z_{i,t} - (1-c)\frac{P_{i,t}}{\bar{P}_i} + \frac{LC_{i,t}}{\bar{P}_i} - g(\mathbf{x}_i, \boldsymbol{\psi}))^2/\sigma_V^2)\right)^{\Lambda_{i,t}} \times \\ &\quad I\left[Z_{i,t} \geq 0 \textit{ if } a_{i,t} = 1, Z_{i,t} < 0 \textit{ if } a_{i,t} = 2\right]. \end{aligned}$$

The complete data likelihood is therefore:

$$L(\Lambda, P, Z|\mathbf{s}, \Theta) = \prod_{i=1}^{N} \prod_{t=1}^{T_i} \mathcal{L}(\Lambda_{i,t}, P_{i,t}, Z_{i,t}|\mathbf{s}_i, \Theta). \tag{8}$$

The next step is to form the joint posterior density by multiplication of equation (8) and the priors. The following forms are assumed for the priors: $\lambda \sim Beta(a_\lambda, b_\lambda)$, $\boldsymbol{\alpha} \sim N(\boldsymbol{\mu}_\alpha, \mathbf{V}_\alpha)$, $\boldsymbol{\psi} \sim N(\boldsymbol{\mu}_\psi, \mathbf{V}_\psi)$, $\sigma_\varepsilon^2 \sim IG(a_\varepsilon/2, b_\varepsilon/2)$, and $\sigma_V^2 \sim IG(a_V/2, b_V/2)$. For the most part, we use starting values for prior parameters that an econometrician would most likely believe without the benefit of actually knowing the parameters. For example, given that the values for the OLS estimated hedonic demand are known to be biased upward, a reasonable starting prior for $\boldsymbol{\alpha}$ would be: $\boldsymbol{\mu}_\alpha = [3, 0.5, 0.5]'$ and $\mathbf{V}_\alpha = \mathbf{I}_{3\times3}$. Starting priors for $\boldsymbol{\psi}$ are set to be somewhat noninformative: $\boldsymbol{\mu}_\psi = \mathbf{0}_{6\times1}$ and $\mathbf{V}_\psi = 10\mathbf{I}_{6\times6}$. Initial priors of $a_\lambda = 2$

and $b_\lambda = 5$ are justified after inspection of offer rates. Finally, we choose $a_\varepsilon/2 = a_V/2 = 3$, and $b_\varepsilon/2 = b_V/2 = 2$ so that the prior means and standard deviations of σ_ε^2 and σ_V^2 are all 0.25.

The estimation algorithm is now defined.

Definition 2 *Gibbs Sampling-Data Augmentation Algorithm:*

1. *Initiate all parameters from prior distributions and guess at missing prices in* P.

2. *With everything else known, draw* λ *from*

$$Beta(a_\lambda + \sum_{i=1}^{N}(T_i - \sum_{t=1}^{T_i}\Lambda_{i,t}), b_\lambda + \sum_{i=1}^{N}\sum_{t=1}^{T_i}\Lambda_{i,t}).$$

3. *With everything else known and where* $\Lambda_{i,t} = 1$, *draw value function differences* $Z_{i,t}$ *of length* T_n *for each* $n = 1, \ldots, N$ *individuals using density equation (7).*

4. *With everything else known, draw prices* $P_{i,t}$ *when* $\Lambda_{i,t} = 1$ *and* $a_{i,t} = 2$ *(i.e., offer was made but not accepted) from density:*

$$\left((2\pi\sigma_\varepsilon^2)^{-1/2}\frac{1}{P_{i,t}}\exp(-\frac{1}{2}\left(\ln P_{i,t} - \mathbf{x}_i\boldsymbol{\alpha}\right)^2/\sigma_\varepsilon^2)\right) \times$$
$$\left((2\pi\sigma_V^2)^{-1/2}\exp(-\frac{1}{2}(Z_{i,t} - (1-c)\frac{P_{i,t}}{\bar{P}_i} + \frac{LC_{i,t}}{\bar{P}_i} - g(\mathbf{x}_i, \boldsymbol{\psi}))^2/\sigma_V^2)\right),$$

using acceptance/rejection (A/R) method described in Geweke (1994) and Geweke and Keane (2000).

5. *With everything else known, draw* $\boldsymbol{\alpha}$ *from* $N(\mathbf{B}_0, \mathbf{B}_1)$ *where* $\mathbf{B}_1 = (\mathbf{V}_\alpha^{-1} + \mathbf{X}_\alpha'\mathbf{X}_\alpha/\sigma_\varepsilon^2)^{-1}$, $\mathbf{B}_0 = \mathbf{B}_1(\mathbf{V}_\alpha^{-1}\boldsymbol{\mu}_\alpha + \mathbf{X}_\alpha'\mathbf{y}_\alpha/\sigma_\varepsilon^2)$, *and* $[\mathbf{X}_\alpha, \mathbf{y}_\alpha]$ *are stacked matrices containing housing characteristics and logged prices, respectively, for where* $\Lambda_{i,t} = 1$.

6. *With everything else known, draw* $\boldsymbol{\psi}$ *from* $N(\mathbf{B}_2, \mathbf{B}_3)$ *where* $\mathbf{B}_3 = \left(\mathbf{V}_\psi^{-1} + \mathbf{X}_\psi'\mathbf{X}_\psi/\sigma_V^2\right)^{-1}$, $\mathbf{B}_2 = \mathbf{B}_3\left(\mathbf{V}_\psi^{-1}\boldsymbol{\mu}_\psi + \mathbf{X}_\psi'\mathbf{y}_\psi/\sigma_V^2\right)$, *and* $[\mathbf{X}_\psi, \mathbf{y}_\psi]$ *are stacked matrices containing the housing characteristics used in the polynomial approximation to the value function and* $Z_{i,t} - (1-c)\frac{P_{i,t}}{\bar{P}_i} + \frac{LC_{i,t}}{\bar{P}_i}$, *respectively, for where* $\Lambda_{i,t} = 1$.

7. *With everything else known, draw* σ_ε^2 *from* $IG(A_\varepsilon/2, B_\varepsilon/2)$ *where* $A_\varepsilon = a_\varepsilon + \sum_{i=1}^N \sum_{t=1}^{T_i} \Lambda_{i,t}$ *and* $B_\varepsilon = b_\varepsilon + (\mathbf{y}_\alpha - \mathbf{X}_\alpha\boldsymbol{\alpha})'(\mathbf{y}_\alpha - \mathbf{X}_\alpha\boldsymbol{\alpha})$.

8. *With everything else known, draw* σ_V^2 *from* $IG(A_V/2, B_V/2)$ *where* $A_V = a_V + \sum_{i=1}^N \sum_{t=1}^{T_i} \Lambda_{i,t}$ *and* $B_V = b_V + (\mathbf{y}_\psi - \mathbf{X}_\psi\boldsymbol{\psi})'(\mathbf{y}_\psi - \mathbf{X}_\psi\boldsymbol{\psi})$.

9. *Return to step 2 and repeat the cycle.*

4.2 The Monte-Carlo Results

Table 4 reports the results from applying the Gibbs sampling algorithm to a data set that consists of the simulated actions of 1,000 "baseline" sellers ($N = 1000$). Because some houses may take time to sell, the data set has 1,532 total periods ($\sum_{i=1}^N T_i = 1532$). We cycle the algorithm 2,500 times and then "burn-in" the first 500 of the cycles. Column (A) of Table 4 presents the actual or true structural parameters values that were used. Column (B) contains the posterior means, followed in the next column by the posterior standard deviations. The model includes a total of 12 parameters: one offer rate, three demand parameters, one variance of demand term, six coefficients for the seller's unobserved utility, and one variance term for the second-order polynomial approximation.

For the case where $\lambda = 0$, we see that the estimation results for the $\boldsymbol{\alpha}$'s are quite accurate. The posterior means for the buyer's demand equation parameters are within two posterior standard deviations of the true values. That is, the true parameters are all within a 95% credibility interval. For example, a 95% credibility interval for α_0 is: $(5.9890 \pm 1.96 \times 0.0494) = (5.892, 6.086)$. The 95% credibility interval for σ_ε^2 of $(0.1039, 0.1419)$ contains it's true value as well. The estimates for $\boldsymbol{\psi}$ appear mostly significant and, more importantly, plausible. The $\boldsymbol{\psi}$'s determine the *unobserved utility* in equation (6). Given equation (5) and the estimated values from Table 4 Column (B), the estimated unobserved utility is $g(\bar{\mathbf{x}}_i, \hat{\boldsymbol{\psi}}) = -0.6307$ for an average seller (*i.e.*, $SQ = 2$, $Bath = 2$). When a seller accepts an offer, the remaining part to the value function difference, defined by $[(1-c)P_{i,t} - LC_{i,t}]/\bar{P}_i$ in equation (6), must greater than 0.6307 to give $Z_{i,t} \geq 0$. This is very close to "baseline" model's prediction of 0.6324 that can be found by multiplication of $1-c$ and the reservation to listing price ratio found in Column (B) in Table 2 (where $c = 0.07$, $LC = 0$, and the Reservation/Listing ratio is 0.68).

Similar accuracies for the posterior means of $\boldsymbol{\alpha}$ and σ_ε^2 are found for the case where $\lambda = 0.25$; we see that the true values are within a 95% credibility interval. For example, α_0's credibility interval is $(5.8873, 6.0779)$. Additionally, the estimates for $\boldsymbol{\psi}$ appear plausible as they predict an average value for $g(\bar{\mathbf{x}}_i, \hat{\boldsymbol{\psi}})$ of -0.6394. Again, this is very close to "baseline" model's prediction found in the earlier simulation exercises where $\lambda = 0.25$. The last case presented in Table 4 shows the estimation results for $\lambda = 0.75$. Again, the posterior means of $\boldsymbol{\alpha}$ and σ_ε^2 are very accurate and the predicted value for $g(\bar{\mathbf{x}}_i, \hat{\boldsymbol{\psi}})$ is similar to the two previous examples.

Finally, Table 5 provides the estimation results for a sample that is selected with respect to "baseline" and "high listing" sellers as in column (C) of Table 3. A benefit of the structural model is that the buyers and sellers behaviors are disentangled. This allows us to include a variable that can capture the different behaviors of the sellers ("baseline" and "high listing"). The variable included in the buyer's utility is the square-footage to listing price ratio, $\exp(SQ)/\bar{P}$ (coefficient ψ_6). The estimation results show that, unlike the standard OLS regression, there are no noticeable biases in the estimated coefficients for $\boldsymbol{\alpha}$. The structural estimation method is therefore able to recover the true demand.

We examined the robustness of the estimation model and method. To start, the estimation is relatively inexpensive with respect to computational time. On a typical pentium machine, the Gibbs sampler was able to converge in no more than 45 minutes. Additionally, higher-order terms added to $g(\cdot)$ did not noticeably increase computational time or the overall conclusions (these results are not reported but available upon request). Next, Figure 2 displays histograms representing the posterior distributions. Note that distributions appear symmetric around their means; the histograms clearly suggest that the credibility intervals are estimated correctly. Figure 3 illustrates that the Gibbs sampling algorithm converges after about 500 iterations; draws from the posterior of $\boldsymbol{\alpha}$, $\boldsymbol{\psi}$, σ_{ε}^2, σ_{ε}^2, and σ_V^2 settle down to the means that we have reported.

Finally, we examined the accuracy of the second-order polynomial approximation used to estimate the value function differences. Following Geweke and Keane (2000), two types of artificial data sets were constructed. In one, 6,000 "baseline" sellers were simulated using their optimal solution from the dynamic programing problem presented in the previous section. In the other, using the same data, 6,000 sellers were simulated using the Gibbs sampler estimated value function (suboptimal solution). We find that the suboptimal rule is rarely wrong (only 10 instances) and, when it is wrong, implies less than a one half of 1% loss in wealth; the second order approximation is quite accurate.[8]

5 Conclusion

The lack of empirical research on the structural estimation in real estate is due to the difficulty of taking large data sets to dynamic forward looking models. The DCDP model of search and matching that we propose accomplishes two main goals along this line. First, the model is able to accurately replicate several stylized facts with respect to how sales price and time on the market are correlated across time and sellers. Second, by applying the

[8]Small wealth losses when using simple polynomial approximations to optimal decision rules have been known since Krussell and Smith (1996).

Gibbs sampling-data augmentation algorithm of Geweke and Keane (2000), the structural parameters are accurately and efficiently recovered. The implication is that the large data sets and dynamic search models of real estate can now be unified in a systematic framework.

Within our framework, there are other extensions to consider. For example, the offer rate λ could be made a function of economy variables (*i.e.*, $\lambda = \Phi(\mathbf{X}_{i,t}\boldsymbol{\phi})$) to determine the effects of macroeconomic events such as recessions and booms. Second, bargaining between sellers and buyers may offer microeconomic insights into how prices are finally derived. Although Merlo and Ortalo-Magné (2004) report that about 70% of all sales occur with the first potential buyer (made an offer) and that 60% of all sales are on the first offer, adding bargaining may be important in determination of the effects of multiple listing services. Next, if extreme heterogeneneities are believed to exist, it may be more appropriate to use a method that fully solves the dynamic programming problem. Imai, Jain, and Ching (2008) propose a Bayesian method, similar to the one we have used here, that can handle unobserved heterogeneity.

Finally, the model should be taken to real data. There are two main reasons why we have not attempted to do so in this study. The first reason is obvious; it is beyond the scope of the paper. The second reason is less so. As discussed in Merlo and Ortalo-Magné (2004), there is a lack of *adequate* data on housing transactions. It is not to say that the appropriate data can never be collected. On the contrary, offers outstanding are, in general, available from most large multiple listing services. This paper contributes to this vein of real estate research in that it is implied how future data sets should be gathered. At a minimum, the data should be in classical panel format where listings starting at an initial time period are followed completely to conclusion. The list of variables should include relevant housing characteristics and, importantly, indicators for offers rejected. To disentangle idiosyncratic behaviors of buyers and sellers, variables that are specific to each unit would be helpful (*e.g.*, price revisions or the LTV ratios of Genesove and Mayer 1997). We are currently engaged in this endeavor.

References

[1] Asabere, Paul K. and Forest E. Huffman (1993), "Price Concessions, Time on the Market, and the Actual Sale Price of Homes," *The Journal of Real Estate Finance and Economics* **6**, pages 167-74.

[2] Caplin, Andrew and John Leahy (2008), "Trading Frictions and House Price Dynamics," NBER Working Papers 14605, National Bureau of Economic Research, Inc.

[3] Coulson, Edward (2008), "Hedonic Methods and Housing Markets," Department of Economics Working Paper, Penn State University.

[4] Cheng, Ping, Zhenguo Lin, and Yingchun Liu (2008), "A Model of Time-on-Market and Real Estate Price Under Sequential Search with Recall," *Real Estate Economics* **36(4)**, pages 813-43.

[5] Genesove, David and Christopher J. Mayer (1997), "Equity and Time to Sale in the Real Estate Market," *American Economic Review* **87(3)**, pages 255-69.

[6] Geweke, John F. (1994), "Priors for Macroeconomic Time Series and Their Application," *Econometric Theory* **10**, pages 609-32.

[7] Geweke, John F., and Michael P. Keane (2000), "Bayesian Inference for Dynamic Discrete Choice Models without the Need for Dynamic Programming," in Mariano, Schuermann and Weeks (eds.), *Simulation Based Inference and Econometrics: Methods and Applications*, Cambridge University Press, pages 100-131.

[8] Houser, Daniel (2003), "Bayesian Analysis of a Dynamic Stochastic Model of Labor Supply and Saving," *Journal of Econometrics* **113(2)**, pages 289-335.

[9] Imai, Susumu, Neelam Jain, and Andrew Ching (2008), "Bayesian Estimation of Dynamic Discrete Choice Models," Department of Economics Working Paper 1118, Queen's University.

[10] Keane, Michael P., and Kenneth I. Wolpin (2009), "Empirical Applications of Discrete Choice Dynamic Programming Models," *Review of Economic Dynamics* **12(1)**, pages 1-21.

[11] Krusell, Per, and Anthony Smith Jr. (1996), "Rules of Thumb in Macroeconomic Equilibrium A Quantitative Analysis," *Journal of Economic Dynamics and Control* **20(4)**, pages 527-58.

[12] Merlo, Antonio and Francois Ortalo-Magné (2004), "Bargaining over residential real estate: evidence from England," *Journal of Urban Economics* **56(2)**, pages 192-216.

[13] Miller, Norman G. (1978), "Time on the Market and Selling Price," *AREUEA Journal* **6(2)**, pages 164-74.

[14] Read, Colin (1997), "Vacancies and Rent Dispersion in a Stochastic Search Model with Generalized Tenant Demand," *The Journal of Real Estate Finance and Economics* **15(3)**, pages 223-37.

[15] Rust, John (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica* **55(5)**, pages 999-1033.

[16] Stein, Jeremy C. (1995), "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects," *The Quarterly Journal of Economics* **110(2)**, pages 379-406.

[17] Tanner, Martin A. and Wing Hung Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association* **82(398)**, pages 528-40.

[18] Wheaton, William C. (1990), "Vacancy, Search, and Prices in a Housing Market Matching Model," *Journal of Political Economy* **98(6)**, pages 1270-92.

[19] Yinger, John (1981), "A Search Model of Real Estate Broker Behavior," *The American Economic Review* **71(4)**, pages 591-605.

A Tables

Table 1: Initial Parameter Calibrations.

	(A) Total Sample	*(B) Baseline*	*(C) High Listing*	*(D) Low Discounting*
Demand:				
α_0	6.0	6.0	6.0	6.0
α_1	1.5	1.5	1.5	1.5
α_2	1.0	1.0	1.0	1.0
σ_ε^2	0.14	0.14	0.14	0.14
λ	0.0	0.0	0.0	0.0
Supply:				
β_i	$\underset{\Pr\{0.75\}}{0.99754}$ *and* $\underset{\Pr\{0.25\}}{0.95004}$	0.99754	0.99754	0.95004
$\bar{P}_i$	$\underset{\Pr\{0.75\}}{1.5E[P_i]}$ *and* $\underset{\Pr\{0.25\}}{2.0E[P_i]}$	$1.5E[P_i]$	$2.0E[P_i]$	$1.5E[P_i]$
$\bar{\epsilon}_i^{<1>}$	0.0	0.0	0.0	0.0
$\bar{\epsilon}_i^{<2>}$	-1.30	-1.30	-1.30	-1.30

Table 2: Model Simulation Results (Means).

	(A) *Total Sample*	*(B)* *Baseline*	*(C)* *High Listing*	*(D)* *Low Discounting*
Case: $\lambda = 0.0, \bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>} = 1.30$				
Log Sales Price	11.36	11.43	11.40	11.18
Log Listing Price	11.57	11.51	11.78	11.51
Sales/Listing	0.84	0.95	0.70	0.76
Reservation/Sales	0.73	0.75	0.76	0.68
Reservation/Listing	0.58	0.68	0.51	0.47
Time on Market	2.49	2.92	2.93	1.32
Case: $\lambda = 0.25, \bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>} = 1.30$				
Log Sales Price	11.34	11.43	11.40	11.10
Log Listing Price	11.57	11.51	11.78	11.51
Sales/Listing	0.83	0.95	0.70	0.70
Reservation/Sales	0.71	0.75	0.76	0.62
Reservation/Listing	0.57	0.68	0.51	0.39
Time on Market	3.26	3.88	3.94	1.47
Case: $\lambda = 0.75, \bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>} = 1.30$				
Log Sales Price	11.29	11.43	11.23	11.04
Log Listing Price	11.57	11.51	11.78	11.51
Sales/Listing	0.80	0.95	0.60	0.67
Reservation/Sales	0.59	0.75	0.75	0.14
Reservation/Listing	0.47	0.68	0.43	0.08
Time on Market	8.53	11.40	6.17	4.12
Case: $\lambda = 0.0, \bar{\epsilon}_i^{<1>} - \bar{\epsilon}_i^{<2>} = 1.25$				
Log Sales Price	11.45	11.53	11.50	11.25
Log Listing Price	11.57	11.51	11.78	11.51
Sales/Listing	0.91	1.04	0.77	0.80
Reservation/Sales	0.80	0.82	0.83	0.76
Reservation/Listing	0.71	0.82	0.62	0.57
Time on Market	3.82	4.63	4.42	1.61

Table 3: Reduced-Form Regression Results ($\lambda = 0$). Equation used: $\log(P_i) = \alpha_0 + \alpha_1 SQ_i + \alpha_2 Bath_i$.

	(A)	*(B)*	*(C)*
	True	*Baseline*	*Baseline* $_{(SQ>1, Bath>1)}$ + *High Listing* $_{(SQ<3, Bath<3)}$
Estimates			
α_0	6.00	6.4053	6.2654
α_1	1.50	1.4978	1.5253
α_2	1.00	1.0003	1.0238
Prediction			
$\log(P_i)$	11.00	11.4323	11.8061

Table 4: Posterior Means and Standard Deviations for Gibbs Sampling Algorithm.

Parameter	(A) True	(B) Post. Mean	(C) Post. Std.
Case: Baseline, $\lambda = 0$			
λ	0.0	0.0013	9.43e-4
α_0	6.00	5.9890	0.0494
α_1	1.50	1.5085	0.0167
α_2	1.00	1.0161	0.0153
σ_ε^2	0.14	0.1229	0.0097
ψ_0	–	-1.0881	0.0519
ψ_1	–	0.2417	0.0390
ψ_2	–	0.2190	0.0376
ψ_3	–	-0.0599	0.0088
ψ_4	–	-0.0562	0.0084
ψ_5	–	0.0001	0.0070
σ_V^2	–	0.0021	4.39e-4
Case: Baseline, $\lambda = 0.25$			
λ	0.25	0.2551	0.0091
α_0	6.00	5.9826	0.0486
α_1	1.50	1.5013	0.0165
α_2	1.00	1.0159	0.0155
σ_ε^2	0.14	0.1377	0.0102
ψ_0	–	-1.0224	0.0490
ψ_1	–	0.1619	0.0329
ψ_2	–	0.2310	0.0414
ψ_3	–	-0.0436	0.0078
ψ_4	–	-0.0617	0.0092
ψ_5	–	0.0046	0.0071
σ_V^2	–	0.0022	4.40e-4
Case: Baseline, $\lambda = 0.75$			
λ	0.75	0.7549	0.0055
α_0	6.00	6.0165	0.0462
α_1	1.50	1.5040	0.0157
α_2	1.00	1.0021	0.0152
σ_ε^2	0.14	0.1319	0.0104
ψ_0	–	-1.1657	0.0484
ψ_1	–	0.2672	0.0387
ψ_2	–	0.2846	0.0370
ψ_3	–	-0.0664	0.0089
ψ_4	–	-0.0704	0.0084
ψ_5	–	-0.0035	0.0077
σ_V^2	–	0.0021	4.07e-4

Table 5: Posterior Means and Standard Deviations for Selected Sample. *Case:* $\lambda = 0$, $\underset{(SQ>1, Bath>1)}{\textit{Baseline}} + \underset{(SQ<3, Bath<3)}{\textit{High Listing}}$

Parameter	*(A)* *True*	*(B)* *Post. Mean*	*(C)* *Post. Std.*
λ	0.0	0.0019	0.0014
α_0	6.00	5.9985	0.0771
α_1	1.50	1.5178	0.0254
α_2	1.00	1.0012	0.0258
σ_ε^2	0.14	0.1413	0.0149
ψ_0	–	-0.4428	0.1035
ψ_1	–	-0.0006	0.0582
ψ_2	–	0.0471	0.0665
ψ_3	–	-0.1072	0.0145
ψ_4	–	-0.1151	0.0165
ψ_5	–	0.1744	0.0254
ψ_6	–	-0.1533	3.0685
σ_V^2	–	0.0038	8.76e-4

B Figures

Figure 1: Model Simulation Distributions ($\lambda = 0$).

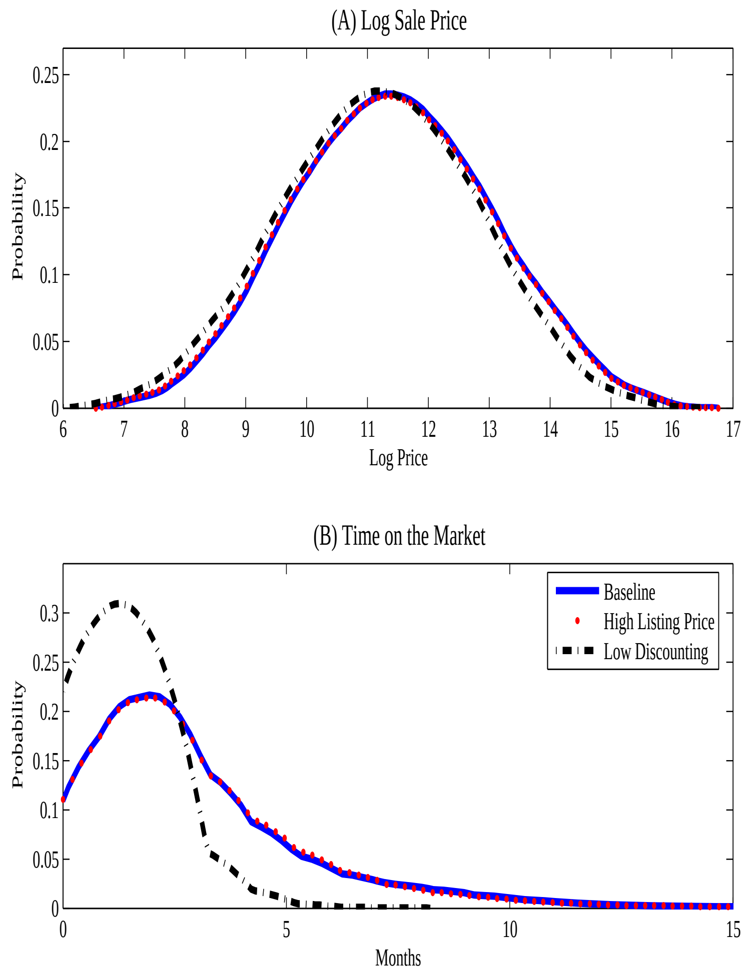


Figure 2: Posterior Simulations ($\lambda = 0$).

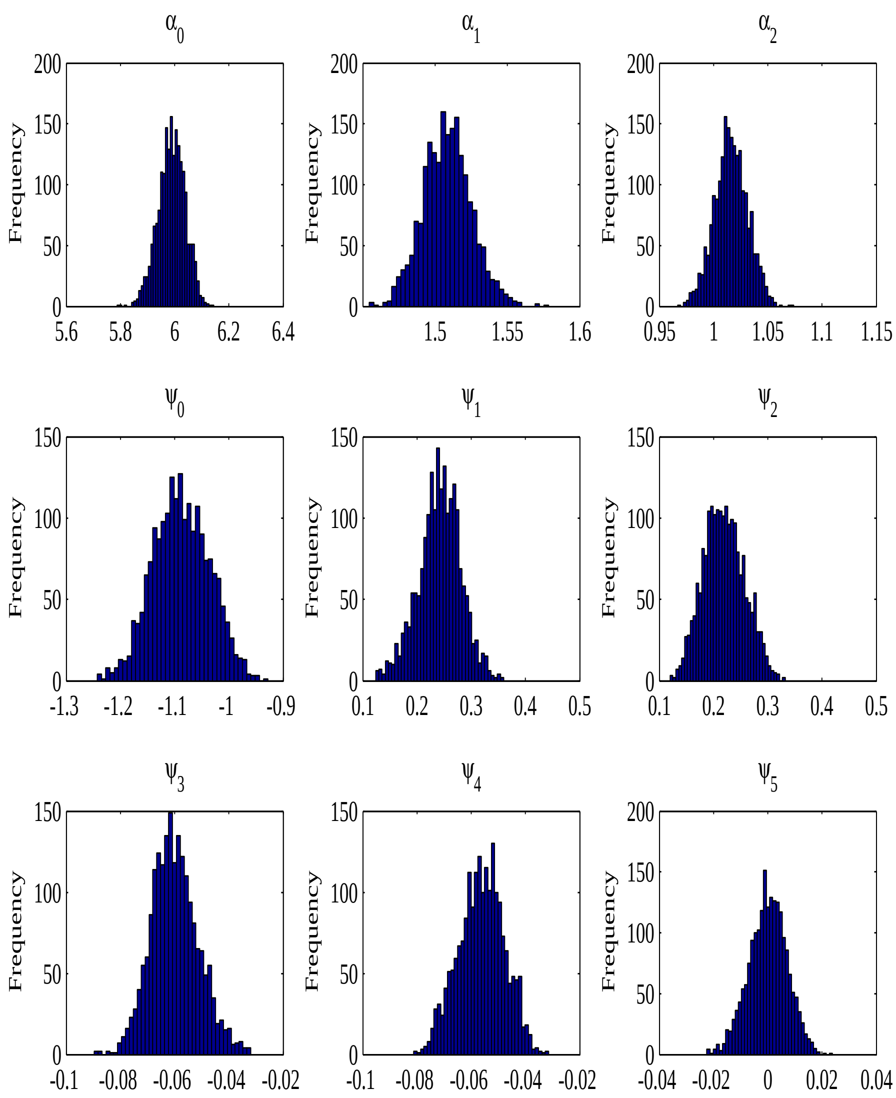
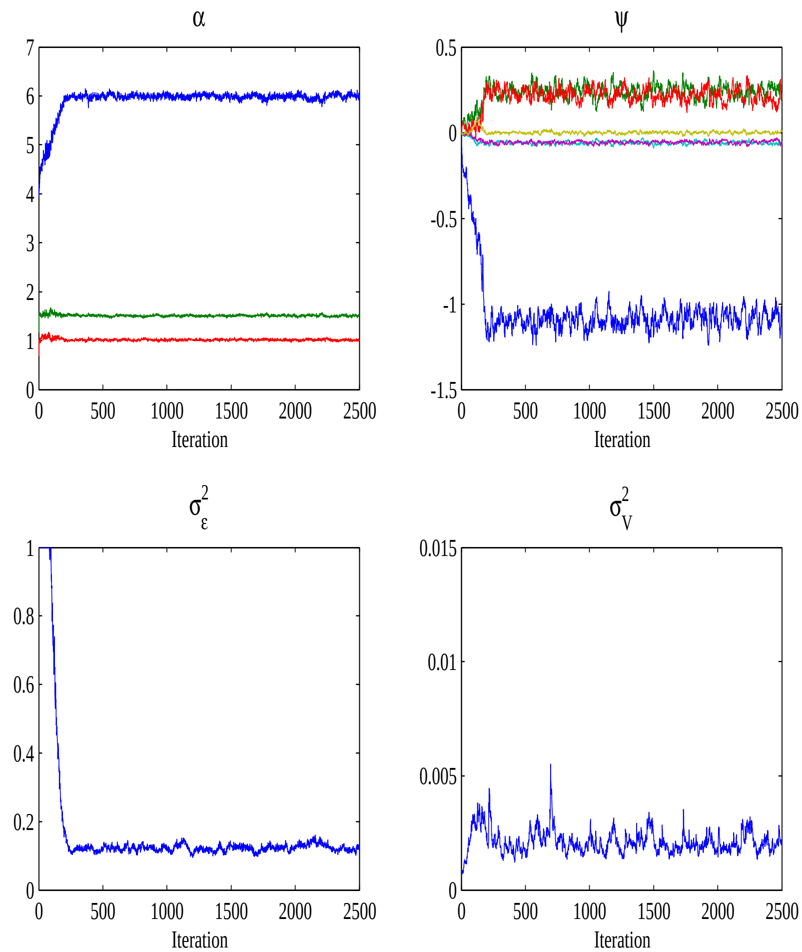


Figure 3: Gibbs Sampler Convergence ($\lambda = 0$).



C Not Intended For Publication

C.1 Figures for Case $\lambda = 0.25$

Figure 4: Model Simulation Distributions ($\lambda = 0.25$).

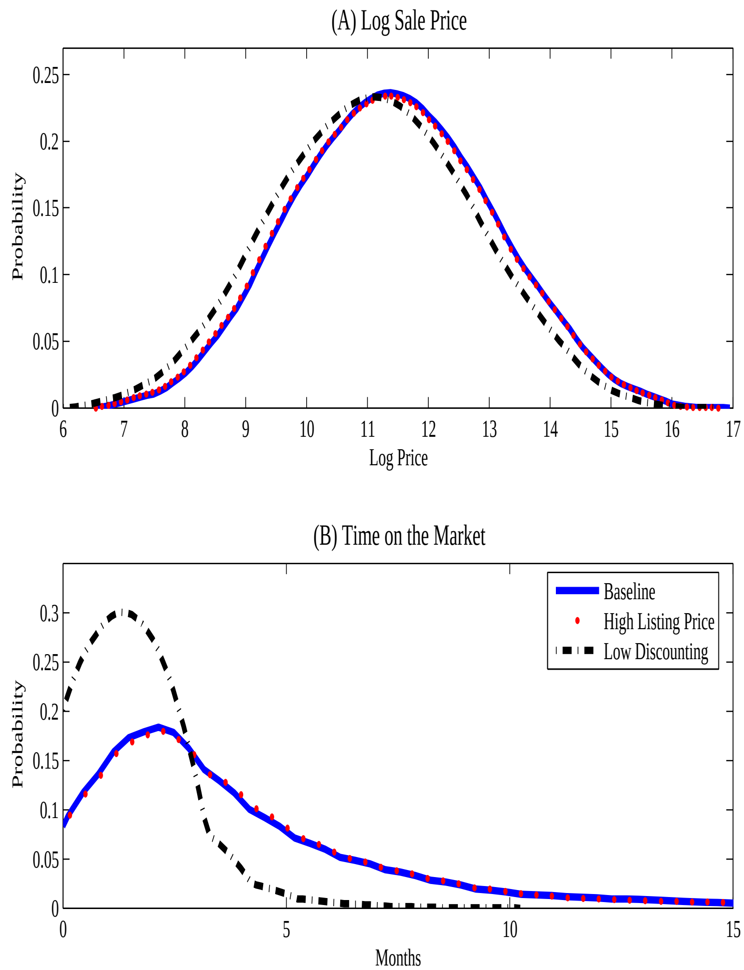


Figure 5: Posterior Simulations ($\lambda = 0.25$).

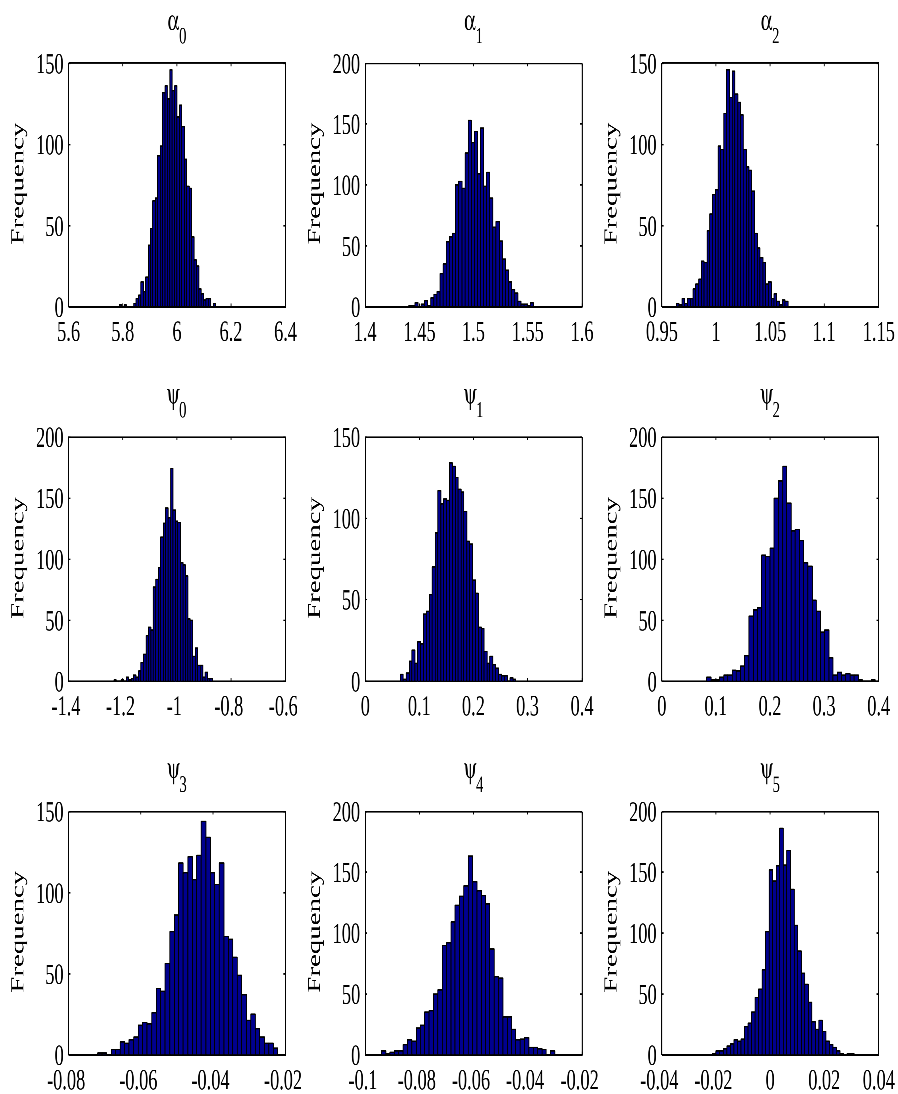
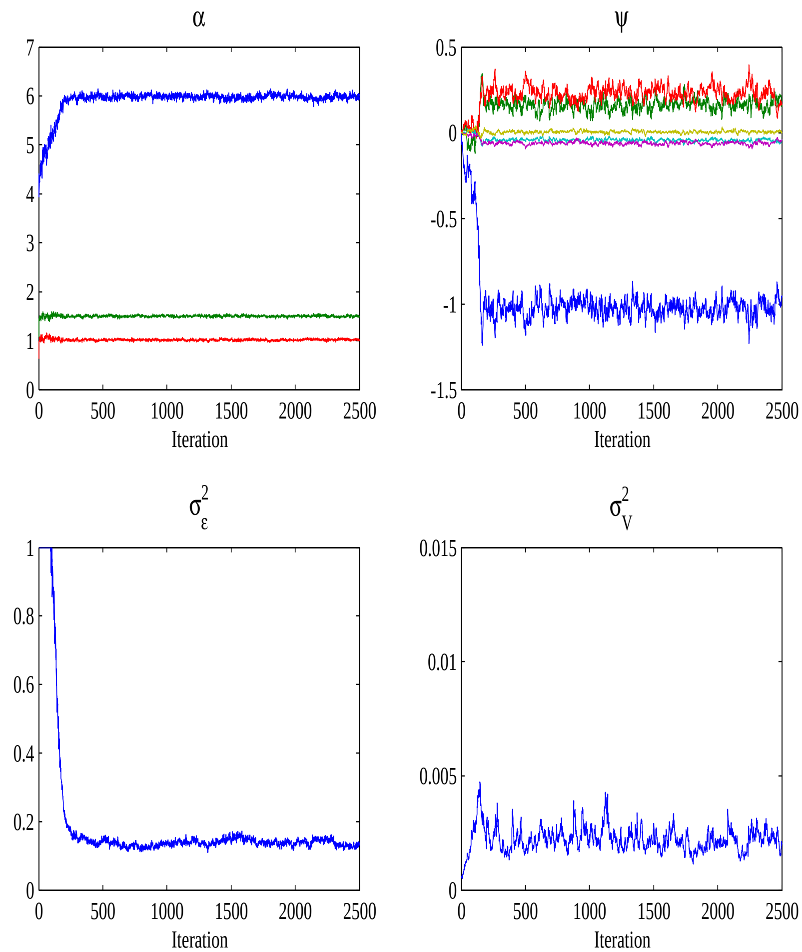


Figure 6: Gibbs Sampler Convergence ($\lambda = 0.25$).



C.2 Figures for Case $\lambda = 0.75$

Figure 7: Model Simulation Distributions ($\lambda = 0.75$).

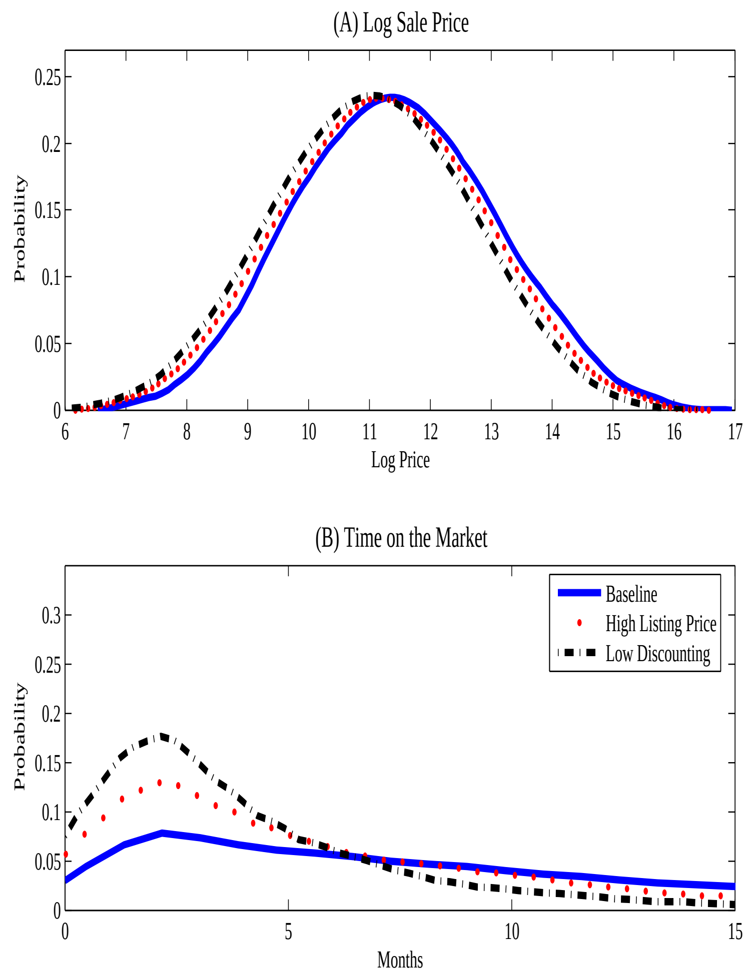


Figure 8: Posterior Simulations ($\lambda = 0.75$).

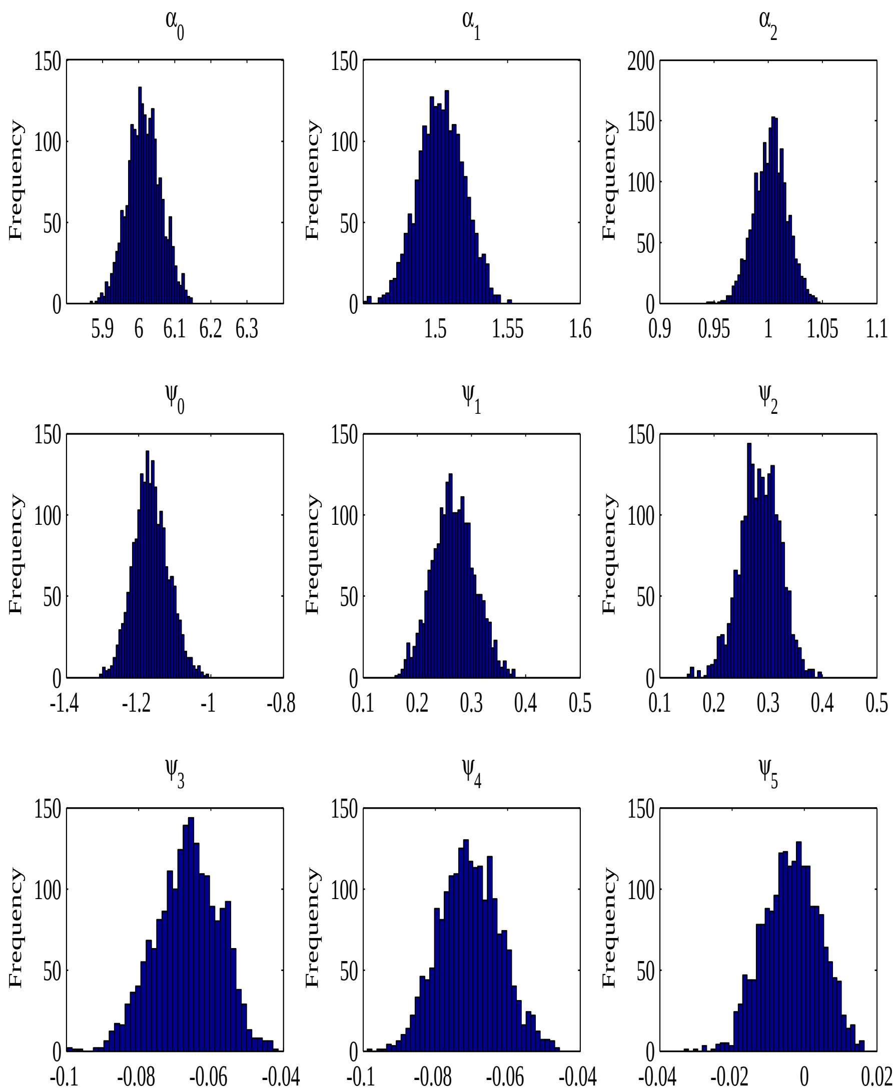
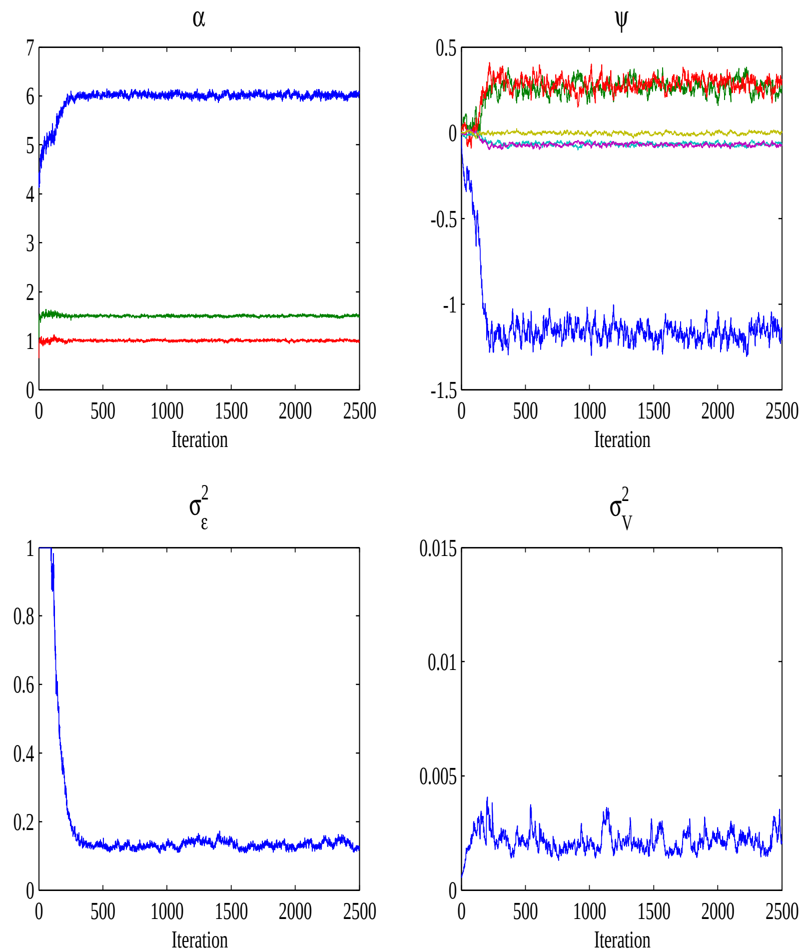


Figure 9: Gibbs Sampler Convergence ($\lambda = 0.75$).



C.3 Probability Transition Matrix P

As an example, suppose that the states are given by the sets $SQ = \{1, 2\}$, $Bath = \{1, 3\}$, and $\varepsilon = \{-0.5, 0.5\}$ with each ε having equal probability. Every permutation can be arranged in $\Omega_{i,t}$ as:

$$\Omega_{i,t} = \begin{bmatrix} 1 & 1 & -0.5 \\ 1 & 1 & 0.5 \\ 1 & 3 & -0.5 \\ 1 & 3 & 0.5 \\ 2 & 1 & -0.5 \\ 2 & 1 & 0.5 \\ 2 & 3 & -0.5 \\ 2 & 3 & 0.5 \end{bmatrix},$$

where each row represents a specific state. The first state (row one) has a 0.50 probability of transitioning to itself next period. And, there is 0.50 probability of transitioning to the second state (row 2) next period. There is no probability that a 1 bedroom and 1 bath house will spawn a new bedroom or two new baths; the probability of these events are zero. Following the same processes for the remaining states, it is easy to verify that the probability matrix $\mathbf{P}$ is:

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix}.$$