Sorting over the Dual Risk of Coastal Housing Market

- Who Surfs the Tide, Who Bears the Blunt?

Zhenshan Chen, PhD Candidate, University of Connecticut, zhenshan.chen@uconn.edu; Charles A. Towe, Associate Professor, University of Connecticut, charles.towe@uconn.edu

Selected Paper prepared for presentation at the 2021 Agricultural & Applied Economics Association Annual Meeting, Austin, TX, August 1 – August 3

Sorting over the Dual Risk of Coastal Housing Market - Who Surfs the Tide, Who Bears the Blunt?

Abstract

Rising flood risk and price risk coexist in the coastal housing market, where the dual risk is not usually jointly studied and individual locational decisions are largely not well understood. We extend and modify a recently introduced dynamic sorting model to investigate the heterogeneous preference on flood risk indicators and ocean amenities. The marginal willingness to pay estimates suggest the preference and decision patterns are different for newcomers and incumbents, and different across wealth levels and racial groups. Within the model framework, policy simulations can also show how flood related coastal policies and natural changes may affect the aggregate housing demand and composition in coastal communities.

Introduction

Although climate change is a controversial topic in the US, few would deny that the sea level is rising and that both the frequency and the severity of flood events are increasing. However, rising coastal flood risk also means increase in rarity of coastal amenities, which provides an opportunity for speculation. High expectations and actual speculative behaviors would lead to a home bubble (Case and Shiller, 2003)[1], and the inevitable burst of the bubble would then formulate a significant amount of housing market risk. The Great Recession and looming sea level rise (SLR) jointly provide enough dynamics and locational variations, so that housing transaction decisions in the face of the dual risk of market and flood can be effectively modeled and the heterogenous household preferences can be distilled. Results from a dynamic sorting model and following simulations also show how flood related coastal policies and natural changes may affect the aggregate housing demand and composition (i.e., racial, income) in coastal communities, which then provide important implications for coastal policy making and environmental justice issues.

We build a dynamic sorting model following Bayer et al. (2016) with certain modifications. The basic assumption is that households make a sequence of location decisions that maximize their aggregated utility flow, where per-period utility flow is modeled as a function of living costs and residence characteristics. Intuitively, the dynamic link is established on the notion that households expect appreciation or depreciation in housing prices based on their observations prior to the current period, which then affect their expected future utilities and wealth. They then compare future utility flows among different residence types to decide whether to move and move to which residence type. These locational decisions are modeled with discrete choice models which also incorporate the financial (i.e., realtor fee, 6% of housing price) and psychological moving costs – only when the net gain in expected lifetime utility exceeds the total moving costs will the

[1] In the book published about five years before the Great Recession, Case and Shiller suggest "our analysis indicates that elements of a speculative bubble in single-family home prices - the strong investment motive, the high expectations of future price increases, and the strong influence of word-of-mouth discussion - exist in some cities."

household decide to move. Modeling households' expectation on the next period states with a Markov process, we establish a Bellman equation in which the parameters (including the psychological moving costs and marginal utility of money) can be estimated with additional assumptions. The estimated marginal utility of money can then be used as the critical coefficient of endogenous cost (i.e., living or holding cost) in the per-period utility decomposition, which provides the marginal utility or willingness to pay for different residence type level characteristics (amenities or disamenities).

Flood risk enters the model in several ways. First and foremost, flood insurance cost is an effective component of the living cost in per-period utility of flood zone homes, and the flood zone status also enters in the per-period utility as a residence character or (dis)amenity. Second, the localized flood risk trend (e.g., localized sea level rise rate) affects the neighborhood specific expectation, which is implicitly captured in the estimates of Markovian transition probabilities. Third, future flood events and policy changes (i.e., potentially affect flood zone status, flood insurance rates, and homeowner expectations) potentially enter as exogenous shocks in simulation of future utility flow and locational decision.

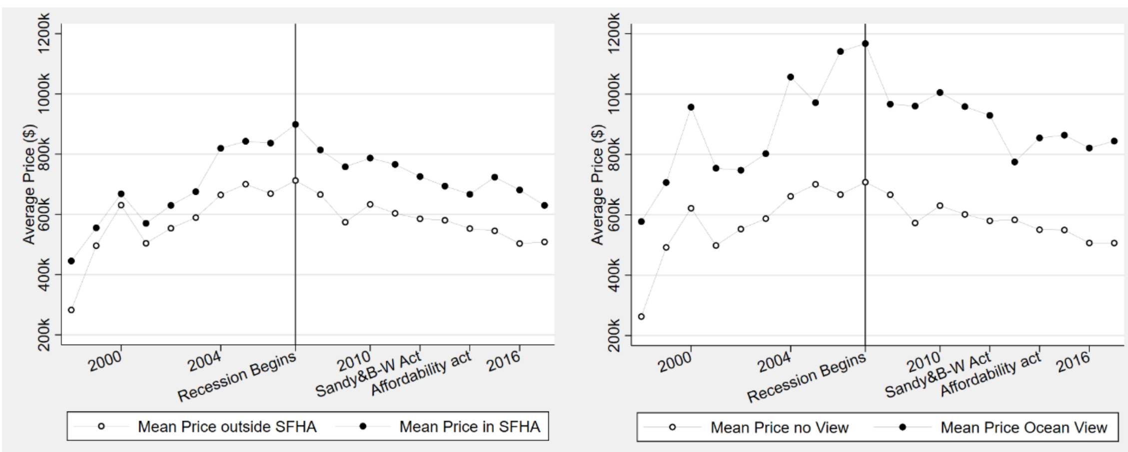


Figure 1. The bubble and contraction in coastal CT housing market (prices are in 2017 dollar)

The dynamic-sorting model is applied to the coastal Connecticut housing market from 2000 to 2017. Several aspects make this application interesting. First, there's a market bubble before the Great Recession (2000-2007) and a market contraction afterwards (2009-2017, see Figure 1), which provides the general market trends (including enough locational discrepancy – see Figure 2 for pre-Recession appreciation) that assist the home buyers to establish their expectations on future prices and utilities. Second, along with the housing market trend, another flood risk trend coexists on CT coast, including sea level rise, flood event intensification (i.e., Hurricane Irene and Sandy), and a boom of climate change narratives and policies (e.g., Biggert-Waters Act of 2012 and Homeowner Affordability Act of 2014), which makes it interesting to investigate how the attitudes and behaviors toward flood risk change during this period. Third, there are plenty of new dwellers (newcomers hereafter) moving into CT coast, who are very likely to have different perceptions and attitudes toward localized coastal flood risk from original residents (incumbents hereafter) due to the absence of local knowledge. This is interesting in establishing homebuyer heterogeneity and important in statistical identification.

The papers on sorting over natural or artificial amenities constitute the most closely relative literature to this study. The housing market dynamic sorting framework initiated by Bayer et al. (2016) improves at least in three aspects compared with traditional sorting models (e.g., Klaiber and Phaneuf, 2010; Bakkensen and Ma, 2020). First, it incorporates the dynamic process in individual housing decisions, of which the absence would induce biases in how sorting models estimate certain parameters. Second, it utilizes the moving costs to address the endogeneity of the cost/price variable in per-period utility decomposition, avoiding unappealing characters in previously used instrumental variable process. Third, it degrades the computational difficulty in sorting models.

Another body of relevant literature mainly consists of hedonic studies that shows how flood risk or perceptions affect coastal housing prices. Studies with DID estimates unambiguously claim risk perception changes upon flood events, and these estimates are likely to be free from the problem sourced from missing amenities or missing elevation status (e.g., Kousky 2010; Bin and Landry, 2013; Atreya et al., 2013; Gibson and Mullins, 2020). Other studies use indicators other than housing prices. Gallagher (2014) employed the community-level flood insurance take-up rate as the indicator to show the changes in risk perception upon Presidential Disaster Declaration floods, and find that the changes are short-lived and the long-term full-information learning is not supported. McCoy and Zhao (2018) used investment probability to show how Hurricane Sandy changed flood risk perception. The dynamic sorting model has at least two advantages over hedonic or quasi-experimental approaches: 1. Hedonic models specify very restrictive assumptions to allow a correct welfare analysis[2], while the sorting model adopts assumptions reflecting the discrete choice decisions; 2. Building on the dynamic sorting model from Bayer et al. (2016), we can simulate the answers to many policy related questions that are difficult to solve otherwise (e.g., how an expansion in flood zone changes the welfare and composition of a specific community).

Moving into the context of coastal housing market, we make several modifications to the dynamic sorting framework in Bayer et al. (2016). We primarily divide the households into two groups – newcomers and incumbents, and assume they have different sets of utility parameters. This configuration allows us to compose a simple and sound estimation process, and more importantly, to assess whether coastal homeowners learn flood risk and change their attitudes via a local channel. On top of the newcomer-incumbent configuration, we find it is no longer necessary for a computationally light estimation to separate the whether-to-move and where-to-move into two layers of decisions, thus the options (per period) for incumbents in our model are staying, choosing an inside option to move, or moving to the outside option[3].

Data

[2] At least, flood zone hedonic pricing isn't accurate marginal willingness to pay for flood zone since it models discrete choice with a continuous assumption.
[3] The essential decision made by a newcomer is still choosing an inside option to move in.

The data employed is a combination of the Home Mortgage Disclosure Act (HMDA) data and property data from 2000 to 2017. HMDA data include publicly disclosed loan-level information of home mortgages, including lender name, loan data, loan amount, loan status, borrower race, and borrower income. The property data, based on ZTRAX property and sales data, go through a comprehensive refining process with information directly from town records (VGS property cards and land records). In addition, simulation of future policy scenarios requires us to calculate the flood insurance cost (approximately) and some other components related to coastal flood (e.g., specifics of 50% rule, sea level rise level, insurance rate change).

The estimation of our dynamic model depends on the fact that rates of change in housing prices and amenities vary across neighborhoods. Figure 2 shows the housing price appreciation by neighborhood from 2000 to 2007, illustrating the variation in the price change rates across coastal CT. While this makes clear the significant differences across the coast in housing price growth over this pre-Recession period, the post-Recession depreciation rates also vary a lot across coastal CT. These variations in price change rates confirms that the basic assumptions of the dynamic sorting model will be met in our application to coastal CT.

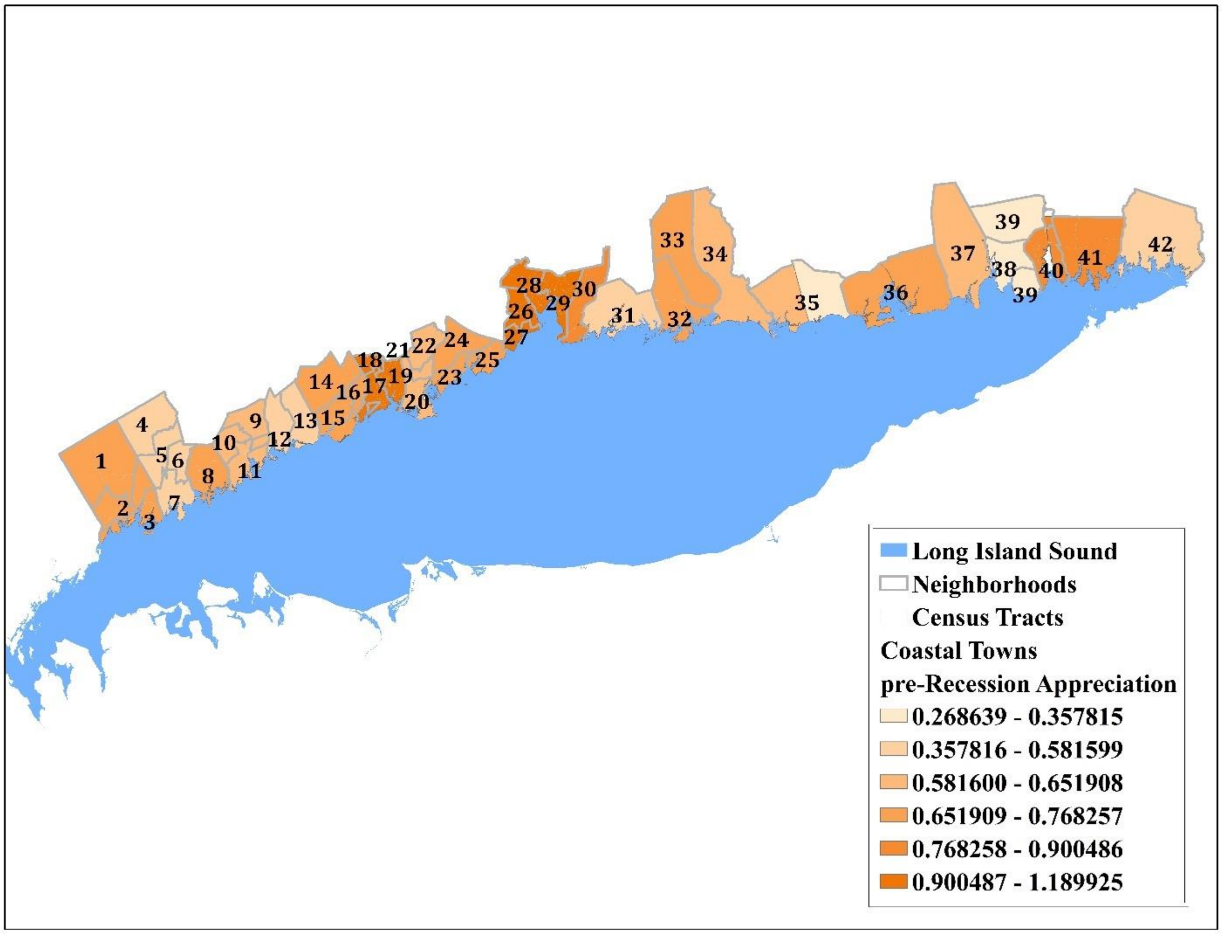


Figure 2. Neighborhoods with index labeled – Coastal Connecticut

A Dynamic Sorting Model for Coastal Housing Market

A. The framework

Closely following Bayer et al. (2016), we model coastal households in a dynamic sorting framework.[4] Households make a sequence of location decisions that maximize their aggregated utility flow.

The households are primarily divided into two groups – newcomers and incumbent. Each household transits from a newcomer to and incumbent once it settles down in coastal CT. In their initial period, newcomers choose where to reside. Incumbents, in each period, chooses whether to move or not and where to move. A moving decision incurs a moving cost but grants higher (actually highest among all residence types) expected lifetime utility in the target residence type. The decision variable, $d_{i,t}$, denotes both of the choices made by household i in period t (1) whether to move, and (2) where to move, conditional on deciding to move. If a household decides to move, we denote that decision by $d_{i,t} = j \in \{0,1,\ldots,J\}$. We let j indexes residence types, J denotes the total number of residence types in coastal Connecticut, and 0 denotes the outside option. If a household decides not to move, the decision is represented by $d_{i,t} = J + 1$.

Newcomers and incumbents are assumed to have different sets of utilities, including flow utilities and lifetime expected utilities, since they are fundamentally different in two aspects. First, having been living and gathering local information on the coast for a while, incumbents would have more experience and knowledge about flood risk locally. Second, as seekers of coastal communities, it is natural that newcomers may over value coastal amenities in their initial locational decision. This newcomer-incumbent configuration allows us to compose a simple and sound estimation process, and more importantly, to assess whether coastal homeowners learn flood risk and change their attitudes.

Choice Set. A residence type (indexed by j) is characterized by neighborhood, flood zone status, and coastal amenity status; thus, the primary locational decision involves the tradeoff among neighborhood attributes (amenities as public good), residence flood risk, and residence coastal amenities[5]. A neighborhood is the primary geography unit discussed in this study. We define neighborhoods by merging neighboring census tracts in the same legal town. Considering geographical contiguity and population sizes, resulted neighborhoods (Figure 2) mostly contain more than 5000 single family residences. Flood zone status is defined by FEMA Special Flood Hazard Zone (SFHA hereafter), with 0 denoting zone X and 1 denoting zone A, AE, or V. The coastal amenity status category is also a dummy, with 0 denoting the lack of ocean view while 1 denoting ocean view. To avoid small sample issues[6], the residence type will not be full interactions of neighborhood, flood zone status, and amenity status. For residences outside the flood zone and without ocean view, the residence type follows the neighborhood segregation unless the residence type incorporates too few residences, in which case we merge neighboring neighborhoods.

[4] Also, we closely follow mathematical notations in Bayer et al. (2016), with a few exceptions.

[5] Note that the choice set here is residence type, instead of merely neighborhood (adopted in Bayer et al., 2016).

[6] One of the principles of specifying the size of each type in the choice set is to make sure the choice process is not constrained by the availability of each choice type, since our choice modeling does not assume constraints. If this principle is violated, the estimated lifetime or per-period utility could be highly correlated with the size of each choice type, rendering the utility decomposition coefficients significantly biased.

Residences in the flood zone or with ocean view will be aggregated for neighboring neighborhoods until single-family-residence number per type reaches about 4000[7]. Thus, in our empirical setting, 36 residence types ($J = 36$) are considered including 31 neighborhood defined types without ocean view and outside the flood zone, 3 types in the flood zone without ocean view, 1 type with ocean view and outside the flood zone, and 1 type with ocean view in the flood zone.

State Variables. The observed state variables for residence type j, household i, at time t, are $X_{j,t}$, $Z_{i,t}$, and $h_{i,t}$. $X_{j,t}$ is a vector of residence type attributes that affect household flow utility. Except the three major variables characterizing each unique residence type, $X_{j,t}$ can also include housing price, land cover ratios, demographic composition, average surface elevation, average distance to coast, and other variables of interest (maybe crime rate). $Z_{i,t}$ is a vector of household attributes that potentially affect flow utility and moving cost associated with a certain residence type. $Z_{i,t}$ can include variables from HMDA data, like income, housing related wealth, and race[8]. $h_{i,t} \in \{0,1,\ldots,J\}$ denotes the residence type chosen in the previous period ($t-1$), including the outside option.

The model also incorporates two unobservable variables, $\xi_{j,t}$ and $\varepsilon_{i,j,t}$. $\xi_{j,t}$ represents the unobservable residence type characteristics, and $\varepsilon_{i,j,t}$ is an idiosyncratic stochastic shock that determines the utility a household i receives from choosing house type j in period t. Note that $\xi_{j,t}$ enters the per-period utility function with $X_{j,t}$ and will be recovered during the per-period utility decomposition at the residence type level, while $\varepsilon_{i,j,t}$ functions as the random part of per-period per-household utility under the *Additive Separability* assumption.

Markovian Transition. Let $s_{i,t}$ denote the states X_t, ξ_t, $Z_{i,t}$, $h_{i,t}$ and other information that help predict future neighborhood or household attributes for household i. Under the assumption that the transition of states follows a Markov process, the transition probability $q = q\left(s_{i,t+1}, \varepsilon_{i,t+1} \middle| s_{i,t}, \varepsilon_{i,t}, d_{i,t}\right)$. This Markov process models how household predict future states based on which they make locational decision, and in model estimation, it also helps to decomposition per-period utility from lifetime values (denoted by v_j). With conditional independence assumption, transition probability can be written as $q\left(s_{i,t+1}, \varepsilon_{i,t+1} \middle| s_{i,t}, \varepsilon_{i,t}, d_{i,t}\right) = q_s\left(s_{i,t+1}, \middle| s_{i,t}, d_{i,t}\right) q_\varepsilon\left(\varepsilon_{i,t+1}\right)$.

Utility, Moving Cost and Value Function. Now that the basic concepts are introduced, the utility based dynamic locational decision are formulated as follows. Flow utility $u_{i,j,t} = u\left(X_{j,t}, \xi_{j,t}, Z_{i,t}, \varepsilon_{i,j,t}\right)$, representing per-period utility household i receives from living in residence type j. Moving cost is only paid when a household decides to move, and we assume it involves two components: financial costs, $FMC\left(X_{h_{i,t}}\right)$, and psychological costs, $PMC\left(Z_{i,t}\right)$. Note that financial moving costs only relates to the residence type that the household is leaving, its housing price to be precise, since it captures the realtor fees that are proportional to the value of the house

[7] Also, this residence type classification does not divide residences of the same township into different residence types.
[8] In practice, we only differentiate minorities and white, since the majority of coastal CT residents are white.

being sold. We denote the moving-costs-adjusted utility flow as $u_{i,j,t}^{MC}$. Adopting the *Additive Separability* assumption (similar to Rust, 1987),

(1-a) $$u_{i,j,t}^{MC} = u(X_{j,t}, \xi_{j,t}, Z_{i,t}) + \varepsilon_{i,j,t}, \quad \text{if } j = J+1,$$

(1-b) $$u_{i,j,t}^{MC} = u(X_{j,t}, \xi_{j,t}, \bar{Z}_{i,t}) - PMC(\bar{Z}_{i,t}) + \varepsilon_{i,j,t}, \quad \text{if } j = J+1.$$

In equation (1-b), we introduce the notation $\bar{Z}_{i,t} = \bar{Z}(Z_{i,t}, X_{h_{i,t}})$ to link a household's new type after moving to its original types. $\bar{Z}_{i,t}$ is related to both the original residence type ($X_{h_{i,t}}$) and household type ($Z_{i,t}$), since it reflects the financial costs and the resulted reduction in household wealth, where the household wealth is incorporated in $Z_{i,t}$ and the reduction of realtor fees (a fixed proportion of housing value) is given by $X_{h_{i,t}}$. The psychological costs $PMC(\bar{Z}_{i,t})$, unlike the financial costs embedded in the household type and utility core, are assumed to be an additively separable part of the utility conditional on moving.

Hence, the households are making a sequence of locational decisions, $\{d_{i,t}\}$, to maximize

(2) $$E\left[\sum_{t=t_0}^{T} \beta^{t-t_0} \left(u_{i,j,t}^{MC}\right) \middle| s_{i,t}, \varepsilon_{i,t}, d_{i,t}\right].$$

The Markov structure of the problem suggests that the optimal decision rule is only a function of state variables: $d_{i,t}^{*} = d_{i,t}^{*}(s_{i,t}, \varepsilon_{i,t})$. The optimal lifetime expected utility can then be represented by the value function $V(s_{i,t}, \varepsilon_{i,t})$. Breaking down into the flow utility in period t and the expected aggregated flow utility from period t+1 onwards, the optimal value function $V(s_{i,t}, \varepsilon_{i,t})$ can be expressed with a Bellman equation (assuming infinite horizon):

(3) $$V(s_{i,t}, \varepsilon_{i,t}) = \max_{j}\{u_{i,j,t}^{MC} + \beta E[V(s_{i,t+1}, \varepsilon_{i,t+1}) | s_{i,t}, \varepsilon_{i,t}, d_{i,t} = j]\}.$$

To simplify the model estimation, we also adopt the *Conditional Independence Assumption* (similar to Rust, 1987) that conditional on $s_{i,t}$ and $d_{i,t}$, idiosyncratic errors $\varepsilon_{i,j,t}$ have no predictive power regarding future state $s_{i,t+1}$, and that the probability density of $\varepsilon_{i,t+1}$ does not depend on current states. The choice-specific value function can be then written as:

(4) $$v_j^{MC}(s_{i,t}) = u_{i,j,t}^{MC}(X_{j,t}, \xi_{j,t}, Z_{i,t}) + \beta E[V(s_{i,t+1}, \varepsilon_{i,t+1}) | s_{i,t}, d_{i,t} = j]$$

$$= u_{i,j,t}^{MC}(X_{j,t}, \xi_{j,t}, Z_{i,t}) + \beta E_s\left[E_\varepsilon\left[\max_k\{v_j^{MC}(s_{i,t+1}) + \varepsilon_{i,t+1}\}\right] \middle| s_{i,t}, d_{i,t} = j\right].$$

Further assuming $\varepsilon_{i,j,t}$ is distributed i.i.d., Type I Extreme Value, equation (4) can be formulated as:

(5) $$v_j^{MC}(s_{i,t}) = u_{i,j,t}^{MC}(X_{j,t}, \xi_{j,t}, Z_{i,t}) + \beta E_s\left[\log\left(\sum_{k=0}^{J+1} \exp\left(v_k^{MC}(s_{i,t+1})\right)\right) \middle| s_{i,t}, d_{i,t} = j\right].$$

Note that in the derivation of equation (4), we utilize $V(s_{i,t+1}, \varepsilon_{i,t+1}) = \max_k\{v_j^{MC}(s_{i,t+1}) + \varepsilon_{i,t+1}\}$. Also, from the *Additive Separability Assumption*, we know:

(6-a) $$v_j^{MC}(s_{i,t}) = v_j(s_{i,t}), if\ j = J+1,$$

(6-b) $$\text{and } v_j^{MC}(s_{i,t}) = v_j(\bar{s}_{i,t}) - PMC(\bar{Z}_{i,t}), if\ j \neq J+1,$$

where $\bar{s}_{i,t}$ is the state variable with $\bar{Z}_{i,t}$ replacing $Z_{i,t}$ (or to be exact, with reduced wealth replacing original wealth). Therefore,

(7) $$v_j(\bar{s}_{i,t}) = u(X_{j,t}, \xi_{j,t}, \bar{Z}_{i,t}) + \beta E_s \left[\log\left(\sum_{k=0}^{J+1} \exp\left(v_k^{MC}(s_{i,t+1})\right)\right) \middle| s_{i,t}, d_{i,t} = j\right].$$

B. Estimation

The primitives (u, MC, q, β) of the model are estimated in two stages. In the first stage, we use household locational decision to estimate the value of lifetime expected utility (v_j) for each residence type, time period, and household type, where residence type is characterized by neighborhood, SFHA status, and ocean view status, while household type is characterized by race, income, and wealth. In the second stage, we recover fully flexible estimates of per-period utility and decompose them based on a set of observables $(u(X_{j,t}, \xi_{j,t}, \bar{Z}_{i,t}))$. Following Bayer et al. (2016), the decomposition of per-period utility in the second stage utilizes the marginal-utility-of-wealth estimates in the first stage, so that the endogeneity of housing price is addressed.

The observed locational decisions can be described as follows. In a specific period, we can only observe incumbent households with the residence type $j \in \{1, \ldots, J\}$ or newcomers that are buying a house in this inside option set. Let us consider newcomers first since that's where many of the observed households (i.e., that can be matched with HMDA data) enter our dataset and easier to consider. These households move from an outside option to an inside option, while we do not observe their previous residence and hence their financial moving cost. This is to say, we can only observe how they choose among the observed residence types. Regarding the incumbent households that already appear in the dataset, we can model their full decision process with the observed information. If they stay, we can compare alternative options with their current residence and moving costs. If they move to the inside options, we can observe their original residence, new residence, and the moving costs. If they move to the outside options, we can still observe the moving costs and estimate the average value measures for the outside option per household type and period.

Stage 1 – Estimate Lifetime Utilities

Based on observable household attributes, we divide them into distinct types indexed by τ. A distinct household type is characterized by the following categorical variables[9]: minority (i.e., nonwhite) or not, high income (i.e., higher than sample medium 114k) or not, and 10 wealth types (i.e., defined by deciles). Note that we assume financial costs are incurred and household wealth is reduced every time a household moves, a household of type τ will have a new type $\bar{\tau}$ after moving. By further assuming that

[9] These categorical-version variables are only used here to define the distinct types, the underlying continuous variables are still used to generate kernel weights and calculate utility decomposition in estimation stage two.

(8) $$FMC(X_{h_{i,t}}) = 0.06 \cdot price_{h_{i,t}} \ ,$$

$\bar{\tau}$ can be calculated for each household conditional on moving.

Let $v_{j,t}^{\tau} = v_j(s_{i,t})$ represent the choice-specific value function of household type τ, the corresponding value function of type $\bar{\tau}$ is then $v_{j,t}^{\bar{\tau}} = v_j(\bar{s}_{i,t})$. Similarly, with $u_{j,t}^{\tau}$ and $u_{j,t}^{\bar{\tau}}$ denoting the deterministic component (not including $\varepsilon_{i,j,t}$) of the flow utility for type τ and $\bar{\tau}$, respectively, the lifetime utilities can be written as:

(9-a) $$v_{j,t}^{\tau} = u_{j,t}^{\tau} + \beta E\left[\log\left(\exp\left(v_{J+1,t+1}^{\tau_{t+1}}\right) + \sum_{k=0}^{J} \exp\left(v_{k,t+1}^{\bar{\tau}_{t+1}} - PMC^{\bar{\tau}_{t+1}}\right)\right) \middle| s_{i,t}, d_{i,t} = j\right],$$

(9-b) and $$v_{j,t}^{\bar{\tau}} = u_{j,t}^{\bar{\tau}} + \beta E\left[\log\left(\exp\left(v_{J+1,t+1}^{\tau_{t+1}}\right) + \sum_{k=0}^{J} \exp\left(v_{k,t+1}^{\bar{\tau}_{t+1}} - PMC^{\bar{\tau}_{t+1}}\right)\right) \middle| s_{i,t}, d_{i,t} = j\right].$$

Note that the vector of lifetime utilities ($v_t^{\bar{\tau}}$) are unique up to an additive constant for a given period, we estimate a normalizing constant $m_t^{\bar{\tau}}$ and $\tilde{v}_{j,t}^{\bar{\tau}}$, where $\tilde{v}_{j,t}^{\bar{\tau}} = v_{j,t}^{\bar{\tau}} - m_t^{\bar{\tau}}$. In practice, $m_t^{\bar{\tau}}$ is set to the level so that residence-type specific normalized life time utilities ($\tilde{v}_{j,t}^{\bar{\tau}}$) have a zero mean for each type-year combination (for newcomers or incumbent).

Household i of type $\bar{\tau}$ chooses option j if $\tilde{v}_{j,t}^{\bar{\tau}} + \varepsilon_{i,j,t} > \tilde{v}_{k,t}^{\bar{\tau}} + \varepsilon_{i,k,t}, \forall k \neq j$. Now we introduce the notion of newcomers (ω) or incumbent (ϖ), which enter superscripts of utility terms, paired with τ. For *newcomers*, the probability of choosing residence type j is (note that $\varepsilon_{i,j,t}$ is assumed to be Type I Extreme Value i.i.d. distributed):

(10) $$P_{j,t}^{\omega,\bar{\tau}} = \frac{e^{\tilde{v}_{j,t}^{\omega,\bar{\tau}}}}{\sum_{k=1}^{J} e^{\tilde{v}_{k,t}^{\omega,\bar{\tau}}}}.$$

And the general expression for individual likelihood for moving to j in period t is

(11) $$L_i^{newcom}(\tilde{v}^{\omega}) = \prod_{j=1}^{J}\left(P_{j,t}^{\omega,\bar{\tau}}\right)^{1[d_{i,t}=j]}.$$

There are several types of decisions for *incumbents*, including staying, moving to an inside option, and moving to an outside option. Now we consider the probability of *staying*. Recalling that the decision between staying and moving will involve moving costs, a household will choose to stay if

(12) $$\tilde{v}_{J+1,t}^{\varpi,\tau} + m_t^{\tau} + \varepsilon_{i,J+1,t} > \max_k\left[\tilde{v}_{k,t}^{\varpi,\bar{\tau}} + m_t^{\bar{\tau}} + \varepsilon_{i,k,t}\right] - PMC(\bar{Z}_{i,t}).$$

Intuitively, a household will stay if the increase of lifetime expected utility from the moving does not exceed the moving costs. Note that the psychological moving costs are captured by $PMC(\bar{Z}_{i,t})$, while the financial moving costs are embedded in the changes in household types (τ) and states ($Z_{i,t}$). Since m_t^{τ} is a normalization term per type-year combination, we can only identify the difference between m_t^{τ} and $m_t^{\bar{\tau}}$ in principle. Following Bayer et al. (2016), we parameterize it as a function of observables. The only change[10] from type τ to type $\bar{\tau}$ is the wealth reduction from the financial moving costs, thus

$$(13) \qquad m_t^{\tau} - m_t^{\bar{\tau}} = FMC\left(X_{h_{i,t}}\right) \cdot \gamma_{fmc}^{\bar{\tau}} = 0.06 \cdot price_{h_{i,t}} \cdot \bar{Z}'_{i,t}\gamma_{fmc}.$$

Being the parameter linking the housing prices with the mean lifetime expected utilities of a certain household type, $\gamma_{fmc}^{\bar{\tau}}$[11] plays a crucial role in the entire model. To build this link, this formulation employs the process that when household evaluate alternative residence options, they factor in the proportional-to-housing-price realtor fees and compare it with their lifetime utilities across different locations. Similarly, we parameterize the psychological moving costs as follows

$$(14) \qquad PMC\left(\bar{Z}_{i,t}\right) = \bar{Z}'_{i,t}\gamma_{pmc}.$$

Recognizing the idiosyncratic terms are $\varepsilon_{i,k,t}$ and $\varepsilon_{i,J+1,t}$, the decision laid out by equation (12) can be modeled by a maximization of a conditional logit likelihood function. Therefore, the probability of incumbents staying is

$$(15) \qquad P_{stay,i,t}^{\varpi,\tau,\bar{\tau}} = \frac{e^{\tilde{v}_{J+1,t}^{\varpi,\tau}}}{e^{\tilde{v}_{J+1,t}^{\varpi,\tau}} + \sum_{k=0}^{J} e^{\tilde{v}_{k,t}^{\varpi,\bar{\tau}} - 0.06 \cdot price_{h_{i,t}} \cdot \gamma_{fmc}^{\bar{\tau}} - \bar{Z}'_{i,t}\gamma_{pmc}}}.$$

Although we write $\tilde{v}_{J+1,t}^{\tau}$ here, the real lifetime expected utility depends on where the household stays ($\tilde{v}_{J+1,t}^{\varpi,\tau} = \tilde{v}_{h_{i,t},t}^{\varpi,\tau}$)[12]. The probability of incumbents moving (from $h_{i,t}$ to j) is

$$(16) \qquad P_{h_{i,t},j,i,t}^{\varpi,\tau,\bar{\tau}} = \frac{e^{\tilde{v}_{j,t}^{\varpi,\bar{\tau}} - 0.06 \cdot price_{h_{i,t}} \cdot \gamma_{fmc}^{\bar{\tau}} - \bar{Z}'_{i,t}\gamma_{pmc}}}{e^{\tilde{v}_{h_{i,t},t}^{\varpi,\tau}} + \sum_{k=0}^{J} e^{\tilde{v}_{k,t}^{\varpi,\bar{\tau}} - 0.06 \cdot price_{h_{i,t}} \cdot \gamma_{fmc}^{\bar{\tau}} - \bar{Z}'_{i,t}\gamma_{pmc}}}.$$

Note that equation (16) includes the situation that incumbent households move to an outside option ($j = 0$). The likelihood function of incumbent's decisions in period t is given by

$$(18) \qquad L_i^{Incum}\left(\tilde{v}^{\varpi}, \gamma_{fmc}, \gamma_{pmc}\right) = \left(P_{stay,i,t}^{\varpi,\tau,\bar{\tau}}\right)^{1[d_{i,t}=J+1]} \left(P_{h_{i,t},j,i,t}^{\varpi,\tau,\bar{\tau}}\right)^{1[d_{i,t}=j \mid j \in \{0,1,\ldots,J\}]}.$$

Intuitively, the sequential decision structure combined with the likelihood functions can be interpreted as follows. Each period, households choose to stay ($d_{i,t} = J + 1$), move inside ($d_{i,t} \in \{1, \ldots, J\}$), or move outside ($d_{i,t} = 0$), and we observe the decision $d_{i,t}$ and moving costs together with other observables for both incumbents and newcomers. Then we refer to the likelihood functions to find the likelihood of observing that decision in each period, and get the product of

[10] Note that within this model, we assume the income does not change when household moves across the study area.
[11] Also note that $\gamma_{fmc}^{\bar{\tau}}$ can be estimated separately for each type and price (residence type) combination, the parameterization is to simplify the estimation process.
[12] Also note that the denominator includes all residence types including $h_{i,t}$, situations/possibilities do exist that a household moves to a house within the same residence type.

all per-period likelihood functions to infer the likelihood of observing a sequence of decisions. Therefore, the combined likelihood function for household i is generally given by[13]

$$(19) \qquad L_i\left(\tilde{v},\gamma_{fmc},\gamma_{pmc}\right)=L_i^{newcom}(\tilde{v}^{\omega})\cdot\prod_{t=t_{0,i}+1}^{t_{0,i}+T}\left\{L_i^{Incum}\left(\tilde{v}^{\varpi},\gamma_{fmc},\gamma_{pmc}\right)\right\}.$$

$t_{0,i}$ represents the initial period of the household, which is only observed for newcomers. T represents the number of periods up to (including) the period household i moves outside (so that cannot be observed again) or is censored by the general range of the sample.

Aggregating individual likelihood functions, the full log-likelihood function can be expressed as:

$$(21) \qquad \mathcal{L}\left(\tilde{v},\gamma_{fmc},\gamma_{pmc}\right)=\sum_{i=1}^{N}\sum_{t=1}^{T}\{1\left[t=t_{0,i}\right]\cdot\log\left(L_i^{newcom}(\tilde{v}^{\omega})\right)$$
$$+1\left[t\neq t_{0,i}\right]\cdot\log\left(L_i^{Incum}\left(\tilde{v}^{\varpi},\gamma_{fmc},\gamma_{pmc}\right)\right)\},$$

where N is the total number of households. The notion for time here has changed to represent the naturally indexed study period instead of the period indexed depending on households, while the entries and exists of households are controlled by the dummy functions identifying household status (i.e., $t=t_{0,i}$ identifies newcomers).

The *estimation* would be extremely difficult if we choose $(\tilde{v}^{\omega},\tilde{v}^{\varpi},\gamma_{fmc},\gamma_{pmc})$ to maximize $\mathcal{L}$[14]. However, the feature offers a simple and commonly used approach. Since $\tilde{v}^{\omega}$ only contributes to L_i^{newcom} while $(\tilde{v}^{\varpi},\gamma_{fmc},\gamma_{pmc})$ only contributes to L_i^{Incum}, we can choose $\tilde{v}^{\omega}$ to maximize $\sum_{i=1}^{N}\sum_{t=1}^{T}\left[1\left[t=t_{0,i}\right]\cdot\log\left(L_i^{newcom}(\tilde{v}^{\omega})\right)\right]$, and then choose $\tilde{v}^{\varpi}$ to optimize $\sum_{i=1}^{N}\sum_{t=1}^{T}\left[1\left[t\neq t_{0,i}\right]\cdot\log\left(L_i^{Incum}\left(\tilde{v}^{\varpi},\gamma_{fmc},\gamma_{pmc}\right)\right)\right]$. Both steps would have closed-form solutions.

Specifically, the first step solution (maximizing newcomers' likelihood function) is given by the following first order condition (F.O.C.)

$$\frac{\partial}{\partial\tilde{v}_{j,t}^{\omega,\bar{\tau}}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[1\left[t=t_{0,i}\right]\cdot\log\left(L_i^{newcom}(\tilde{v}^{\omega})\right)\right]=0,\forall j\in\{0,\dots,J\}$$

$$(22) \qquad \Rightarrow \quad P_{j,t}^{\omega,\bar{\tau}}=\frac{N^{\omega}(d_{i,t}=j)}{N^{\varpi}(d_{i,t}\in\{1,\dots,J\})}=\widehat{P_{J,t}^{\omega,\bar{\tau}}},$$

[13] Note that this combined likelihood function differs from what's presented Bayer et al. (2016).

[14] Bayer et al. (2016) uses a similar estimation strategy, but that strategy doesn't exactly match their likelihood expression. Essentially, they assume that newcomers and incumbents have the same expected lifetime utility, while they use the utility of those who move (to do this, they use empirical probabilities of those who move to fit the newcomers' likelihood) to infer utility of those who stay and find the optimizing $(\gamma_{fmc},\gamma_{pmc})$. However, that estimation is not maximizing their presented likelihood function expression. If one chooses to actually maximize the likelihood expression in Bayer et al. (2016), it would take more iterations incorporating maximizing $\tilde{v}$ based on previous-step $(\hat{\gamma}_{fmc},\hat{\gamma}_{pmc})$, which would also be computationally prohibitive.

where $\widehat{P_{J,t}^{\omega,\bar{\tau}}}$ is the empirical probability that households of type $\bar{\tau}$ [15] choose neighborhood j in period t, conditional on being a newcomer. Recall that our normalization makes the average $\tilde{v}_{j,t}^{\omega,\bar{\tau}}$ zero per type-period combination, thus we have $\prod_{k=1}^{J} e^{\tilde{v}_{k,t}^{\omega,\bar{\tau}}} = e^{\sum_{k=1}^{J}\tilde{v}_{k,t}^{\omega,\bar{\tau}}} = 1$. Utilizing this equality, equation (22) can be written as:

$$\text{(23)} \qquad \widehat{\tilde{v}_{J,t}^{\omega,\bar{\tau}}} = \log\left(\frac{\widehat{P_{J,t}^{\omega,\bar{\tau}}}}{\left(\prod_{k=1}^{J}\widehat{P_{k,t}^{\omega,\bar{\tau}}}\right)^{\frac{1}{J}}}\right) = \log\left(\widehat{P_{J,t}^{\omega,\bar{\tau}}}\right) - \frac{1}{J}\sum_{k=1}^{J}\log\left(\widehat{P_{k,t}^{\omega,\bar{\tau}}}\right).$$

The F.O.C.s maximizing incumbent likelihood can similarly yield the following equations

$$\frac{\partial}{\partial \tilde{v}_{j,t}^{\varpi,\bar{\tau}}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[1\left[t \neq t_{0,i}\right]\cdot\log\left(L_i^{Incum}\left(\tilde{v}^{\varpi},\gamma_{fmc},\gamma_{pmc}\right)\right)\right] = 0, \forall j \in \{0,\dots,J\}$$

$$\text{(24-a)} \qquad \Rightarrow N \cdot P_{h_{i,t},j,i,t}^{\varpi,\bar{\tau}} + N\left(h_{i,t} = j\right)\cdot P_{stay,j,i,t}^{\varpi,\tau,\bar{\tau}} = N\left(d_{i,t} = J+1, h_{i,t} = j\right) + N\left(d_{i,t} = j\right),$$

and

$$\frac{\partial}{\partial \gamma_{fmc}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[1\left[t \neq t_{0,i}\right]\cdot\log\left(L_i^{Incum}\left(\tilde{v}^{\varpi},\gamma_{fmc},\gamma_{pmc}\right)\right)\right] = 0$$

$$\text{(24-b)} \qquad \Rightarrow \sum_{t=1}^{T}\sum_{h_{i,t}=1}^{J}\left[N(h_{i,t} = j)\cdot\sum_{k=0}^{J}P_{j,k,i,t}^{\varpi,\bar{\tau}}\right] = N\left(d_{i,t} \neq J+1\right),$$

where $\widehat{P_{move,t}^{\varpi,\bar{\tau}}}$ the empirical probability that incumbent households of type $\bar{\tau}$ choose to move in period t. Equation (24-a) indicates that the predicted number of incumbent households that choose residence type j would be equal to the empirical number, while equation (24-b) indicates that the aggregated predicted counts of incumbents moving would be equal to the empirical counterpart. Note that we haven't taken advantage of the underlying constraint $\widehat{P_{h_{\iota,t},k,\iota,t}^{\varpi,\bar{\tau}}} = P_{h_{i,t},k,i,t}^{\varpi,\bar{\tau}}$ or $\widehat{P_{stay,\iota,t}^{\varpi,\tau,\bar{\tau}}} = P_{stay,i,t}^{\varpi,\tau,\bar{\tau}}$ [16]. With these constraints and the normalization scheme, we get the following closed-form solutions[17] from (24-a) and (24-b).

$$\text{(25-a)} \qquad \widehat{\tilde{v}_{J,t}^{\varpi,\bar{\tau}}} = \log\left(\frac{\widehat{P_{h,J,\iota,t}^{\varpi,\bar{\tau}}}}{\left(\prod_{k=0}^{J}\widehat{P_{h,k,\iota,t}^{\varpi,\bar{\tau}}}\right)^{\frac{1}{J+1}}}\right) = \log\left(\widehat{P_{h,J,\iota,t}^{\varpi,\bar{\tau}}}\right) - \frac{1}{J+1}\sum_{k=0}^{J}\log\left(\widehat{P_{h,k,\iota,t}^{\varpi,\bar{\tau}}}\right),$$

---

[15] The observed household types for newcomers are after moving, and thus treated as directly representing $\bar{\tau}$. The situation is more complicated for incumbents, where the observed household type is τ.

[16] Although we use subscript i in the empirical probability expressions, they represent the probabilities of a certain household type, and the calculation involves kernel weights (deviates from the exact definition) as shown below.

[17] Solution (25-a) implicitly incorporates the natural assumption that the lifetime utilities are the same in a target residence type for households moving from different residence types. This assumption allows us to aggregate the empirical probabilities of moving to j for households from different $h_{i,t}$, to be exact, that is $e^{\tilde{v}_{k,t}^{\omega,\bar{\tau}}} / e^{\tilde{v}_{j,t}^{\omega,\bar{\tau}}} = \sum_{h_{i,t}}^{J}\widehat{P_{h_{\iota,t},k,\iota,t}^{\varpi,\bar{\tau}}} / \sum_{h_{i,t}}^{J}\widehat{P_{h_{\iota,t},J,\iota,t}^{\varpi,\bar{\tau}}} = \widehat{P_{h,k,\iota,t}^{\varpi,\bar{\tau}}} / \widehat{P_{h,J,\iota,t}^{\varpi,\bar{\tau}}}$.

(25-b) $\widehat{\tilde{v}_{J,t}^{\varpi,\bar{\tau}}} - 0.06 \cdot price_{h_{i,t}} \cdot \bar{Z}'_{i,t}\gamma_{fmc} - \bar{Z}'_{i,t}\gamma_{pmc} = \log\left(\widehat{P_{h_{i,t},J,i,t}^{\varpi,\bar{\tau}}}\right) - \frac{1}{J+1}\sum_{k=0}^{J}\log\left(\widehat{P_{stay,k,i,t}^{\varpi,\tau,\bar{\tau}}}\right).$

$\widehat{P_{h_{i,t},J,i,t}^{\varpi,\bar{\tau}}}$ denotes the probability of incumbent households of type $\bar{\tau}$ (after moving) moving to residence type j from residence type $h_{i,t}$ in period t, $\widehat{P_{h,J,i,t}^{\varpi,\bar{\tau}}}$ denotes the probability of incumbent households of type $\bar{\tau}$ (after moving) moving to residence type j from any residence type in period t, and $\widehat{P_{stay,k,i,t}^{\varpi,\tau,\bar{\tau}}}$ denotes the probability of incumbent households of type τ staying at residence type k in period t.

Therefore, once we get the empirical probabilities, solving $\widehat{\tilde{v}_{J,t}^{\omega,\bar{\tau}}}$ and $\widehat{\tilde{v}_{J,t}^{\varpi,\bar{\tau}}}$ is straightforward. Plugging $\widehat{\tilde{v}_{J,t}^{\varpi,\bar{\tau}}}$ into equation (25-b), γ_{fmc} and γ_{pmc} can be calculated drawing information from different initial residence types ($h_{i,t}$) and household types (with different $\bar{Z}'_{i,t}$). The marginal utilities of moving costs are then $\widehat{\gamma_{fmc}^{\tau}} = \bar{Z}'_{i,t}\gamma_{fmc}$ and $\widehat{\gamma_{pmc}^{\tau}} = \bar{Z}'_{i,t}\gamma_{pmc}$.

To avoid small sample issues, we use a weighted measure to calculate the observed share of a specific household type buying a certain residence type (i.e., inside option). This is done by incorporating information from similar household types, and the weights depend on the distance in the type space - $W^{\bar{\tau}} = W^{\bar{\tau}}(\bar{Z}_{i,t})$. Specifically, the observed shares and weights are given by

(24-a) $$\widehat{\widehat{P_{J,t}^{\bar{\tau}}}} = \frac{\sum_{i=1}^{N} 1[d_{i,t}=j] \cdot W^{\bar{\tau}}(\bar{Z}_{i,t})}{\sum_{i=1}^{N} W^{\bar{\tau}}(\bar{Z}_{i,t})},$$

(24-b) $$W^{\bar{\tau}}(\bar{Z}_{i,t}) = \prod_{k=1}^{K} \frac{1}{b_k(\bar{\tau})} N\left(\frac{\bar{Z}_{i,t} - \bar{Z}^{\bar{\tau}}}{b_k(\bar{\tau})}\right),$$

where the individual kernel weight is constructed with a standard normal kernel, N, and bandwidth, $b_k(\bar{\tau})$, and K is the dimension of $\bar{Z}^{\bar{\tau}}$. While this weighted share is adopted for inside option shares, the outside option shares are estimated using the observed share of households choosing to move outside the specified residence types.

Even with kernel weighted probabilities, we still have missing $(\bar{\tau}, j, t)$ level observations[18], especially when we need to calculate empirical shares of household moving from a specific residence type to another specific residence type (i.e., equation 25-b). Note that some of these unideal cases affect the estimation of lifetime utilities ($\widehat{\tilde{v}_{J,t}^{\varpi,\bar{\tau}}}$ and $\widehat{\tilde{v}_{J,t}^{\omega,\bar{\tau}}}$), while others affect the estimation of γs - there are many $\widehat{P_{h_{i,t},J,i,t}^{\varpi,\bar{\tau}}}$s that cannot be estimated simply because no household of any household type move from $h_{i,t}$ to j in a certain year. Fortunately, the estimation of γ_{fmc} and γ_{pmc} only requires a sufficient number of $\widehat{P_{h_{i,t},J,i,t}^{\varpi,\bar{\tau}}}$s, so missing values do not affect the identification of γs (although affecting the efficiency of the proposed estimators). Also, the lifetime utilities that cannot be directly estimated might result in missing per-period utility for

[18] This is cause by two situations. First, certain household types ($\bar{\tau}$) do not appear at all in some years, especially for newcomers. Second, in some cases, certain choice occasions do not appear at all in some years (e.g. no household chooses j in certain years, or no household from residence type k moves to j in certain years).

certain years, which will then lead to missing marginal value estimates in certain years. However, that is not a worrisome issue as long as the majority of per-period utilities are estimated.

Stage 2 – Recover and Decompose Per-period Utility

Given $(\tilde{v}^{\omega}, \tilde{v}^{\varpi}, \gamma_{fmc}, \gamma_{pmc})$ from the first stage, we proceed to estimate per-period utility. We firstly estimate transition probabilities in the assumed Markov process of state transition. Remined that households are assumed to directly predict future values of lifetime utilities based on today's states, we need to model the transition probabilities for v. Also, as moving costs and household wealth changes are determined by the price of the house that a household currently occupy, we need to model the transition probabilities for price. We do not explicitly make notation about newcomers (ω) and incumbents (ϖ) in this subsection, since they are basically the same in how their per-period utilities are constructed[19].

In estimating the transition probabilities for v, we assume certain coefficients are common across residence types within each type. To account for different means and trends, we include a separate constant and time trend for each residence type's choice-specific value function for each household type. Specifically, the transition probabilities (ρ) are estimated from

$$v_{j,t}^{\tau} = \rho_{0,j}^{\tau} + \sum_{l=1}^{L} \rho_{1,l}^{\tau} v_{j,t-l}^{\tau} + \sum_{l=1}^{L} X_{j,t-l}' \rho_{2,l}^{\tau} + \rho_{3,j}^{\tau} t + e_{j,t}^{\tau}, \tag{25}$$

where the time-varying residence type attributes, $X_{j,t}$, include price, racial composition (percent minority), average surrounding land cover shares (forest and developed land), average surface elevation, and average distance to coast[20]. Note that the corresponding estimation for the outside option do not include residence type attributes, since $X_{j,t}$ are not observed for the outside option.

To estimate the transition probabilities for price indices per household-type-residence-type-year combination. We estimate transition probabilities for price levels (included in $X_{j,t}$) according to

$$price_{j,t} = \varrho_{0,j} + \sum_{l=1}^{L} X_{j,t-l}' \varrho_{2,l} + \varrho_{3,j} t + \sigma_{j,t}^{\tau}. \tag{26}$$

Given these transition probabilities, it is straight forward to estimate transition probabilities for wealth and thus household type. In practice, we use two lags of the dependent or right-hand-side variables (i.e., $L = 2$).

With transition probabilities, $v_{j,t}^{\tau}$, and $PMC^{\bar{\tau}_t}$, the mean flow utilities $(u_{j,t}^{\tau})$ can be calculated according to

$$u_{j,t}^{\tau} = v_{j,t}^{\tau} - \beta E\left[\log\left(\exp\left(v_{J+1,t+1}^{\tau_{t+1}}\right) + \sum_{k=0}^{J} \exp\left(v_{k,t+1}^{\bar{\tau}_{t+1}} - PMC^{\bar{\tau}_{t+1}}\right)\right) \middle| s_{i,t}, d_{i,t} = j\right], \tag{27}$$

where the discount factor β is set to 0.95. In practice, for each household type, τ, residence type, j, and time, t, we simulate the expectation on the right-hand side with the following process. To

[19] However, the parameter values are still assumed to be different and estimated separately.

[20] Note that these geographic attributes are actually varying across time, since they represent the average attributes of houses sold (note this set is very close to the universe of sales - much bigger than sales merged with HMDA data), instead of all the residences in the residence type. The racial composition is computed with the sales that are merged with HMDA data.

be clear, the expectation will be calculated as the average of a large number of simulated $v_{j,t+1}^{\tau}$ and $price_{j,t+1}^{\tau}$, where $price_{j,t+1}^{\tau}$ is used to calculate $PMC^{\bar{\tau}_{t+1}}$ and $\bar{\tau}_{t+1}$[21]. Indexing random draws with r, each $v_{j,t+1}^{\tau}(r)$ and $price_{j,t+1}^{\tau}(r)$ are generated by drawing from the empirical distribution of error terms obtained from the estimation process of (25) and (26). In simulating $\bar{\tau}_{t+1}$, we also take random draws from the empirical distribution of individual household wealth per subgroup[22], the wealth at $t+1$ then changes as the price changes[23] and decreases as households move[24]. Then, for each draw, per-period flow utility can be calculated ($u_{j,t}^{\tau}(r)$), and the simulated $u_{j,t}^{\tau}$ is given by $\frac{1}{R}\sum_{r=1}^{R} u_{j,t}^{\tau}(r)$. In principle, R should be set to such a level that $u_{j,t}^{\tau}$ do not change as R increases. In practice, we set R equal to 1000.

With the recovered mean per-period flow utilities, we consider the decomposition of them into functions of observable residence type attributes, $X_{j,t}$. Recall that $\xi_{j,t}$ represents the unobservable residence type characteristics, we specify $\xi_{j,t}^{\tau}$ as a household specific error term in the per-period utility decomposition regression, which is represented by

(28) $$u_{j,t}^{\tau} = \alpha_0^{\tau} + \alpha_c^{\tau} + \alpha_t^{\tau} + X_{j,t}'\alpha_x^{\tau} + \xi_{j,t}^{\tau}.$$

In equation (28), we additionally control for household type (α_0^{τ}), town (α_c^{τ}), and time fixed effects (α_t^{τ}). Note that price term in $X_{j,t}$ here is used to explain the per-period utility, which, in theory, should be per-period user cost ($usercost_{j,t}$) of residing at residence j at time t. Following the typical specification, we calculate $usercost_{j,t}$ as five percent of the current housing price plus mandatory[25] flood insurance cost ($NFIPpremium_{j,t}$ only for SFHA residence types). Therefore,

(29) $$usercost_{j,t} = 0.05 \cdot price_{j,t} + NFIPpremium_{j,t} \cdot SFHA_{j,t}.$$

Also, considering that $-usercost_{j,t}$ always share the same coefficient with per-period income, we assume the marginal effect of income on per-period utility is the same as the marginal effect of wealth on lifetime utility, and thus the coefficient of $usercost_{j,t}$ can be assigned as the estimated marginal utility of wealth in the first stage ($\widehat{\gamma_{fmc}^{\tau}} = \bar{Z}_{i,t}'\gamma_{fmc}$). Therefore, denoting the residence type attributes except user cost with $\tilde{X}_{j,t}$, equation (28) can be written as

(29) $$u_{j,t}^{\tau} + \widehat{\gamma_{fmc}^{\tau}} \cdot usercost_{j,t} = \alpha_0^{\tau} + \alpha_c^{\tau} + \alpha_t^{\tau} + \tilde{X}_{j,t}'\alpha_x^{\tau} + \xi_{j,t}^{\tau}.$$

[21] Since newcomers do not show moving costs at all, we assume they share the same structural parameters related to psychological moving costs.
[22] For subgroups that can be matched with observed individual households, the wealth takes value of a random draw from these households. For subgroups that cannot be matched with observed individual households (inevitable situation since we take kernel weighted probabilities to avoid small sample issue which significantly expands the type space), the wealth takes value of a random draw from the households in the same household-type-year group. These random draws are independent from one another.
[23] This is given by the predicted incremental value of price from equation (26).
[24] The moving cost is given by six percent of the predicted price at the next period.
[25] Since the households merged with HMDA data are all buyers with a mortgage, they are mandated to purchase flood insurance by the lenders (otherwise they won't get the mortgage).

Note that this creative approach from Bayer et al. (2016) solves the endogeneity issue of price or cost variable in utility regressions, and addresses the unsatisfactory factors[26] in the commonly used instrumental variable approach in sorting models. For the convenience of interpretation, we compute marginal willingness to pay measures as: $MWTP_x^\tau = {\alpha_x^\tau}/{\widehat{\gamma_{fmc}^\tau}}$.

Preliminary Results and Discussion

The moving cost estimates are presented in Table 1.

Table 1. Moving cost estimates

Dependent Variable: $\widehat{\tilde{v}_{J,t}^{\varpi,\bar{\tau}}} - [\log\left(\widehat{P_{h_{\iota,t},J,\iota,t}^{\varpi,\bar{\tau}}}\right) - \frac{1}{J+1}\sum_{k=0}^{J}\log\left(\widehat{P_{stay,k,\iota,t}^{\varpi,\tau,\bar{\tau}}}\right)]$	
Financial Moving Cost Coefficient (parameterizing $\widehat{\gamma_{fmc}^{\tau}}$)	
Fmc (financial moving cost) # c. Household type Wealth	-0.00000561***
	(0.000000994)
Fmc # c. Household type Income	-0.0000116***
	(0.00000348)
Fmc # Minority	0.00809***
	(0.000476)
Fmc	0.0108***
	(0.000558)
Psychological Moving Cost (parameterized)	
Household type Wealth	0.000236***
	(0.0000454)
Household type Income	0.000359*
	(0.000159)
Household type Minority	1.339***
	(0.0217)
Year (linear trend)	0.229***
	(0.00104)
Constant	2.822***
	(0.0269)
N	96760

Note: Standard errors are in parentheses. +, *, **, *** indicate p<0.1, p<0.05, p<0.01, and p<0.001, respectively.

[26] The flaws in widely used instrumental variables include computational difficulty and insufficient exclusion restriction (i.e., part of the reason is stated in Bayer et al., 2016, footnote 68).

The MWTP estimates (Table 2 and Table 3) from the proposed dynamic sorting model provide some interesting heterogeneities across newcomers and incumbents, different periods, and different race and wealth groups.

Table 2. Willingness to Pay (in $1000) Estimates by Race Status

	Newcomer				Incumbent			
	Pre-Recession		Post-Recession		Pre-Recession		Post-Recession	
Minority	No	Yes	No	Yes	No	Yes	No	Yes
WTP SFHA	5.405 (9.790)	-30.447** (9.004)	-32.924*** (6.545)	-0.278 (6.074)	135.20*** (19.180)	15.63 (16.252)	-32.304** (10.116)	-10.275 (8.572)
WTP Ocean View	57.991*** (15.875)	72.721*** (14.553)	-11.143 (16.481)	117.322*** (15.319)	171.14*** (30.186)	19.872 (25.577)	152.32*** (26.130)	128.99*** (22.140)
MWTP Minority share +1%	-2.87*** (0.265)	5.15*** (0.243)	-1.48*** (0.367)	5.99*** (0.340)	-11.57*** (0.654)	6.265*** (0.554)	0.945 (0.619)	0.870+ (0.524)

Note: Standard errors are in parentheses. +, *, **, *** indicate p<0.1, p<0.05, p<0.01, and p<0.001, respectively. For reference, the average minority share (across residence types) is about 12%.

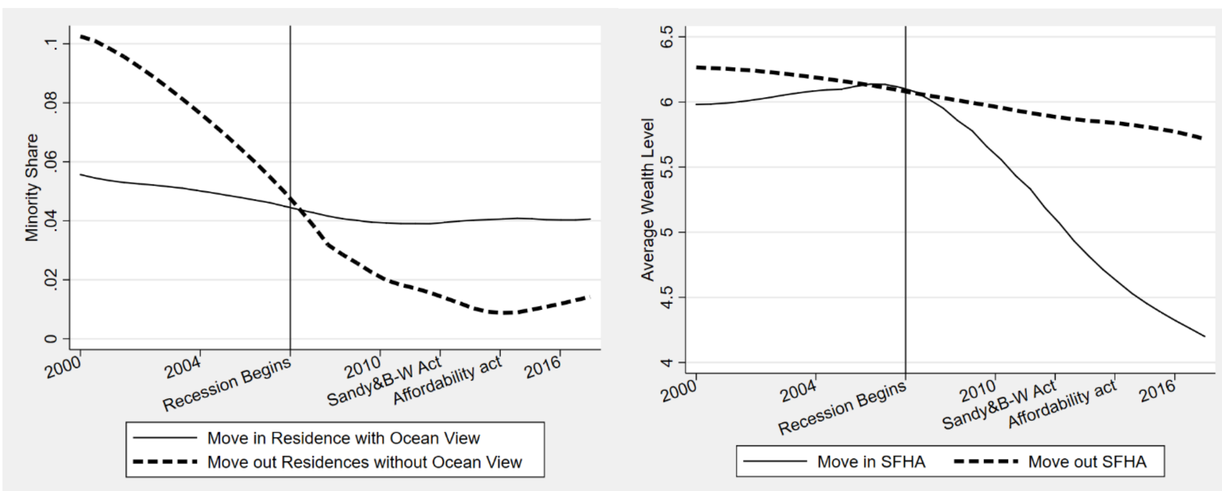
Focusing on race, Table 2 shows that minorities (i.e., self-identified non-white) have a significantly positive MWTP (most of them around $6k annually) for a marginal 1% increase in minority share in the same residence type, while non-minorities have a negative MWTP for such an increase, suggesting race-specific agglomeration might be true in coastal CT. For SFHA MWTP, we find minority newcomers have significantly negative WTP for SFHA before the recession, which becomes smaller and insignificant after the recession. The trend for non-minority newcomers is the opposite: they show no negative WTP for SFHA before the recession but sizable and significant distaste for SFHA post-Recession. We also find that minority incumbents no significant WTP for SFHA, while the non-minority incumbents, surprisingly, show positive WTP for SFHA pre-Recession, which then turn out to be significantly negative post-Recession. Based on the rule that groups with higher WTP for SFHA will move into SFHA more, it can be inferred that more non-minorities are moving into the SFHAs pre-Recession while more minorities are moving into the SFHAs post-Recession.

The WTP for ocean view is generally sizable and positive. Specifically, the WTPs for ocean view are larger than the WTP for avoiding SFHA across the board. We find non-minority newcomers' average WTP are dropping considerably (by about $48k annually) post-Recession, while the minority newcomers show increased WTP for ocean view (by about $45k annually). Non-minority incumbents show very high positive WTP for ocean view in both periods (more than $150k annually), while minority incumbents show insignificant WTP for ocean view pre-Recession but sizable and positive WTP for ocean view post-Recession (about $129k annually).

Combining the WTP estimates, the following story seems to be true: before the Recession, nonminority incumbents and newcomers are purchasing many coastal homes (including flood zone homes and ocean view homes) with a price premium, while the minorities seem to pick up flood zone homes with little price premium (i.e., do not show positive WTP for flood zone homes) after the Recession. It indicates that the non-minorities are the primary investors in the pre-Recession market bubble, while the minority incumbents are the ones who pick up the depreciated flood zone homes (although it's unclear for now whether this is a bad thing financially). It can also be inferred that the ocean view properties seem to be the favored investment in the home bubble and hence overpriced in the pre-Recession market (evident from Figure 1), and they seem to relatively maintain their value, compared to flood zone homes, in the post-Recession market contraction.

Since it's a quite interesting story that minorities are more likely to move into the flood plains in the sorting process post-Recession. We move forward to show more evidence by investigating the actual shares of households moving in and moving out. An intuitive view of these results is offered in Figure 3(a), where statistics do show minorities consist of a higher share of households moving out than moving in before the Recession (so the total minority share would slightly increase when the total moving-out and moving-in counts are very close), while this relation reverses after the Recession.

Figure 3. Trends in Residence Types in the SFHA or with Ocean View



(a) (b)

Table 3 focuses on the heterogeneity across wealth levels. We find weak evidence that wealthier household tend to pay more to avoid living in the flood plain post-Recession, which seems to coincide the pattern in Figure 3(b) that less wealthier households are moving into the flood plain post-Recession compared with pre-Recession. Similar to the results in Table 2, the incumbents seem to have a positive WTP for SFHA before the Recession. In other cases, the WTPs for SFHA are mostly negative, while the significance levels are likely affected by statistical power issues (i.e., in these estimates, there are around 100 observations for each wealth type in SFHA).

Table 3. Marginal Willingness to Pay (in $1000) for SFHA - Estimates by Wealth Levels

MWTP SFHA	Newcomer		Incumbent	
	Pre-Recession	Post-Recession	Pre-Recession	Post-Recession

Wealth Level	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1 - (below 54k)	-7.82	(26.60)	-13.29	(13.69)	69.18+	(41.08)	-18.80	(21.66)
2 - (54k, 87k]	-0.76	(26.28)	-13.32	(13.87)	80.07+	(41.63)	-18.20	(21.95)
3 - (87k,120k]	-34.84	(24.94)	-14.64	(14.03)	69.93	(42.12)	-20.19	(22.21)
4 - (120k,156k]	-1.70	(24.89)	-9.96	(14.20)	76.99+	(42.64)	-17.04	(22.49)
5 - (156k,199k]	-16.40	(25.19)	-8.37	(14.84)	84.14+	(43.28)	-18.45	(22.82)
6 - (199k,253k]	-35.78	(25.65)	-5.80	(15.86)	79.49+	(44.07)	-25.88	(23.24)
7 - (253k,327k]	-24.49	(26.27)	-22.17	(16.88)	62.89	(45.15)	-19.72	(23.81)
8 - (327k,436k]	-23.77	(27.26)	-50.98**	(16.27)	63.56	(46.79)	-18.15	(24.67)
9 - (436k,680k]	11.78	(31.59)	-20.65	(18.72)	57.82	(49.81)	-23.79	(26.27)
10 - (above 680k)	53.41	(40.71)	-2.98	(26.46)	52.56	(61.90)	-25.76	(32.64)

Note: Standard errors are in parentheses. +, *, **, *** indicate p<0.1, p<0.05, p<0.01, and p<0.001, respectively.

Table 4 reports the WTP estimates, differentiating the sample by two income categories (i.e., below or above the median income level). Now it's clear that incumbents, on average, have a negative WTP for SFHA after the Recession, and the higher income group has a slightly more sizable negative WTP for SFHA (around $34.6k, significant at .1 level) compared to the lower income group (around $29.1k, significant at .05). The lower-income newcomers also have a significantly negative WTP for SFHA (around $32.4k, significant at .01 level), while the high-income newcomers present an insignificant negative WTP (around $18.3k). Neither the incumbents nor the newcomers show statistically significant WTPs for avoiding the flood plain pre-Recession, while the incumbents, similar as the results in Table 2 and Table 3, show significantly positive WTP for SFHA. Moreover, it is clear that coastal CT residents are willing to pay more for ocean view than for avoiding the flood plains.

Table 4. Willingness to Pay (in $1000) Estimates by Income Status

	Newcomer				Incumbent			
	Pre-Recession		Post-Recession		Pre-Recession		Post-Recession	
Income above median level	No	Yes	No	Yes	No	Yes	No	Yes
WTP SFHA	-25.906	-6.477	-32.419**	-18.315	109.038***	107.164**	-29.084*	-34.571+
	(16.270)	(23.822)	(10.076)	(13.035)	(26.250)	(38.421)	(13.642)	(19.967)
WTP Ocean View	73.550**	142.159***	31.231	142.252***	121.690**	159.733**	174.560***	273.257***
	(26.323)	(38.600)	(25.420)	(32.815)	(41.313)	(60.468)	(35.236)	(51.573)

Note: Standard errors are in parentheses. +, *, **, *** indicate p<0.1, p<0.05, p<0.01, and p<0.001, respectively. For reference, the average minority share (across residence types) is about 12%.

Summarizing the WTP estimates from different aspects, several interesting conclusions can be derived. 1. The non-minority tend to purchase flood zone homes pre-Recession, while the minorities tend to buy flood zone homes post-Recession. And this trend does lead to a change in racial composition in the flood plains. 2. Race-specific agglomeration (not just regarding the flood plains) might be a potential trend in coastal CT. 3. In general, residents are willing to pay more for ocean view than avoiding SFHA. 4. In the pre-Recession period of home bubble, incumbents are willing to pay more for any properties along the coast, within the flood zone or not, while this trend does not continue post-Recession.

Caveats and Future work

One caveat of this model is that it only characterizes households purchase homes with mortgages, since the essential household characteristics (race, income, and wealth from HMDA) are only available for with-loan households. Hence the results and interpretations are only limited to these households (although they consist of the majority, estimated to be 85%, of the population).

As we are configuring potential future flood related regulations, the simulation results are not available yet. We expect the simulation results to show how flood related coastal policies and natural changes may affect the aggregate housing demand and composition (i.e., racial, income) in coastal communities, which will then provide important implications for coastal policy making and environmental justice issues.

Also, one possible extension is to include more towns. Incorporating more towns further from the coast may allow us to more accurately estimate the marginal utility of money, which is critical in the WTP estimates.