A Dynamic Model of Housing Supply, AEJ: Economic Policy, 2018

Summarized by Wenzhi Wang

M.Phil. Student at the University of Oxford

March 30, 2023

Table of Contents

Assumptions

- In each period t, parcel owner n makes two decisions.
- First, the parcel owner decides whether or not to build on her parcel, denoted by $d_{nt} \in \{0, 1\}$, where $d = 0$ when choosing to not build, $d = 1$ when choosing to build.
- If a parcel owner decides to build, she makes a second decision about the level of housing services to construct, denoted by h_{nt}.
- Once a parcel owner decides to build in a period, that period becomes a terminal period (an optimal stopping decision).
- Two decisions: when to build and how much to build.
  Three outcomes: whether the parcel owner built or not in each period, the level of housing services chosen, and a sales price for the property.

State Variables

- Neighborhoods (census tracts) are indexed by j, where $j \in \{1, \ldots, J\}$, and the census tract that parcel n is located in is denoted by $j(n)$, simplfied as j.
- The state variables $\mathbf{x}_{njt}$ include direct characteristics of the parcel n and characteristics of the neighborhood $j(n)$. The vector $\mathbf{x}_{njt}$ can be divided into two components: parcel-level variables, $\mathbf{x}_{n}$, and neighborhood-level variables, $\mathbf{x}_{jt}$.
- Price and variable cost shocks are denoted by ν_{nt} and η_{nt}, respectively.
- Unobserved idiosyncratic overall profit shock are $\epsilon_{nt} = (\epsilon_{0nt}, \epsilon_{1nt})$, which determines the profit parcel owner n receives from not building or building in period t.
- Finally, the vector of observable state variables is denoted by Ω_{njt}, containing $\mathbf{x}_{njt}$ as well as any other observable variables (such as lagged prices and lagged costs) that predict future values of $\mathbf{x}_{njt}$.

Primitives of the Model

- (π, q, β) .
- $\pi_d = \pi_d(h_{njt}, x_{njt}, v_{nt}, \eta_{nt}) + \epsilon_{dnt}$ is the direct per period profit function associated with choosing option d and housing services h.
- $q = q(\mathbf{\Omega}_{n,j,t+1}, \epsilon_{n,t+1} \mid \mathbf{\Omega}_{\boldsymbol{njt}}, \epsilon_{nt})$ denotes the transition probabilities of the observables and unobservables, where the transition probabilities are assumed to be Markovian
- β is the discount factor.

π_1 = Prices - Costs + Profit Shock

- Define the direct per period profit function as

$$\pi_1\left(h_{nt}, \mathbf{x}_{njt}, \nu_{nt}, \eta_{nt}\right) + \epsilon_{1nt} = P\left(h_{nt}, \mathbf{x}_{njt}, \nu_{nt}\right) - \left(VC\left(h_{nt}, \mathbf{x}_{njt}, \eta_{nt}\right) + FC\left(\mathbf{x}_{njt}\right)\right) + \epsilon_{1nt} \quad (1)$$

- Prices are given by:

$$P\left(h_{nt}, \mathbf{x}_{njt}, \nu_{nt}\right) = \rho_{jt} Q\left(h_{nt}, \mathbf{x}_n, \nu_{nt}\right), \quad (2)$$

where

$$Q\left(h_{nt}, \mathbf{x}_n, \nu_{nt}\right) = h_{nt}^{\gamma_{1jt}} \mathbf{x}_n^{\gamma_{2jt}} e^{\nu_{nt}}.$$

Therefore, prices are equal to the price of a unit of housing quality ρ_{jt}, times the quantity of housing quality Q_{nt}.

π_1: Prices = Unit Price of Quality * Quality

- The price of a unit of housing quality, ρ_{jt}, varies by neighborhood and year, incorporating the effects of x_{jt} on house price.
- Housing quality is composed of three terms:
  - the choice variable, housing services, h (house square footage) ;
  - the fixed parcel characteristics, $\mathbf{x_n}$ (lot size);
  - and a normally distributed error term, ν_{nt}, with variance σ_ν^2. ν_{nt} is assumed to be independent of $\mathbf{\Omega}_{njt}$.
- The vector of price parameters, which I denote by γ, varies by neighborhood j, and time t.

π_1: Costs = VC + FC

- Costs are comprised of two components, variable costs, $VC\left(h_{nt}, \mathbf{x}_{njt}, \eta_{nt}\right)$, and fixed costs, $FC(x_{njt})$.
- Variable costs are specified as

$$VC\left(h_{nt}, \mathbf{x}_{njt}, \eta_{nt}\right) = \left(\alpha_{0jt} \mathbf{x}_n^{\boldsymbol{\alpha}_1} e^{\eta_{nt}}\right) \cdot h_{nt}. \tag{3}$$

  It increases at a linear rate in the quantity of housing services, where the rate is determined by the parcel characteristics, neighborhood, time, and a normally distributed error term, η_{nt}, with variance σ_η^2.
- The second component of costs $FC(x_{njt})$ captures the broader cost environment. It is specified as $FC(x_{njt}) = \delta_{ct}$, where c is the county in which parcel h is located.

π_1: Timing of the Realization of Random Shocks v_{nt} and η_{nt}

- The parcel owner knows the current price parameters v_1, and parcel characteristics when making her build/don't build decision, but that the price error v_{nt}, is not revealed until after construction and time of sale.
- The parcel owner observes the cost shock η_{nt} before the housing-services decision is made, but after the decision to build is made, and that it is independent of Ω_{njt}.

π_1: Profit Shocks ϵ_{dnt}

- ϵ_{dnt} is assumed to be distributed i.i.d. Type 1 Extreme Value with scale parameter σ_ϵ, and mean equal to zero.
- The profit shock ϵ_{dnt} is observed by the parcel owner before they decide to build (a shock to fixed costs and could reflect factors at the parcel level or idiosyncratic parcel owner characteristics).
- Also, assume that the three errors are independent - the assumption of independence between price and costs shocks would only be violated if a parcel owner could pass on a cost shock to the buyer.

Optimal Housing Services

- Conditional on choosing to build, a parcel owner will choose h to maximize profits.
- As the price error v_{nt}, is unobserved at the time of that decision, the agent takes the expectation of prices with respect to v_{nt} and the first-order condition for maximization is given by

$$\gamma_{1jt}\rho_{jt}h_{nt}^{\gamma_{1jt}-1}\mathbf{x}_n^{\gamma_{2jt}}e^{\frac{1}{2}\sigma_v^2} - \alpha_{0jt}\mathbf{x}_n^{\boldsymbol{\alpha}_1}e^{\eta_{nt}} = 0. \tag{4}$$

- The second-order conditions require $\gamma_{1jt} < 1$, which is always satisfied in the empirical results.
- Therefore, the optimal housing service choice is

$$h_{nt}^* = \left(\frac{\gamma_{1jt}\rho_{jt}\mathbf{x}_n^{\gamma_{2jt}}e^{0.5\sigma_v^2}}{\alpha_{0jt}\mathbf{x}_n^{\boldsymbol{\alpha}_1}e^{\eta_{nt}}}\right)^{\frac{1}{1-\gamma_{1jt}}}. \tag{5}$$

Optimal Discrete Choice

- Plugging Equation (5) into (1) yields the indirect flow profit function associated with building, $\pi_1(h^*_{nt}(\mathbf{x}_{njt},\eta_{nt}),\mathbf{x}_{njt},\nu_{nt},\eta_{nt})+\epsilon_{1nt}$.
- However, as the price error ν_{nt}, and variable cost error η_{nt}, are observed after the decision to build is made, the relevant object for the optimal discrete choice is the expected indirect flow profit, which is denoted by $\overline{\pi}(\mathbf{x}_{njt})$.
- Notations:

$$E_{\nu_{nt},\eta_{nt}}\left[P_{nt}\left(h^*_{nt}\left(\mathbf{x}_{njt},\eta_{nt}\right),\mathbf{x}_{njt},\nu_{nt}\right)\mid\Omega_{njt}\right]\equiv\overline{P}\left(\mathbf{x}_{njt}\right),$$

$$E_{\eta_{nt}}\left[VC_{nt}\left(h^*_{nt}\left(\mathbf{x}_{njt},\eta_{nt}\right),\mathbf{x}_{njt},\eta_{nt}\right)\mid\Omega_{njt}\right]\equiv\overline{VC}\left(\mathbf{x}_{njt}\right).$$

- The expected indirect flow payoff is

$$\overline{\pi}_1\left(\mathbf{x}_{njt}\right)=\overline{P}\left(\mathbf{x}_{njt}\right)-\left(\overline{VC}\left(\mathbf{x}_{njt}\right)+FC\left(\mathbf{x}_{njt}\right)\right)+\epsilon_{1nt}. \tag{6}$$

Optimal Discrete Choice

- The deterministic component of the per period profits from choosing to not build $(d=0)$ is normalized to zero, so that its indirect flow profit function is $\pi_0(x_{njt}) + \epsilon_{0nt} = \epsilon_{0nt}$.
- The value function can be written as

$$V_t\left(\Omega_{njt}, \epsilon_{nt}\right) = \max\left\{\bar{\pi}_1\left(\mathbf{x}_{njt}\right) + \epsilon_{1nt},\ E\left[\beta V_{t+1}\left(\Omega_{njt+1}, \epsilon_{nt+1}\right) \mid \Omega_{njt}, \epsilon_{nt}\right] + \epsilon_{0nt}\right\} \tag{7}$$

- Given some technical details, the conditional value functions are

$$\begin{aligned} v_1\left(\mathbf{\Omega}_{njt}\right) &= \overline{\pi}_1\left(\mathbf{x}_{njt}\right), \\ v_0\left(\mathbf{\Omega}_{njt}\right) &= \beta\sigma_\epsilon\left(\int \log\left[\exp\left(\frac{v_0\left(\Omega_{njt+1}\right)}{\sigma_\epsilon}\right) + \exp\left(\frac{\overline{\pi}_1\left(\mathbf{x}_{njt+1}\right)}{\sigma_\epsilon}\right)\right] q\left(\Omega_{njt+1} \mid \Omega_{njt}\right) d\Omega_{njt+1}\right). \end{aligned} \tag{8}$$

Estimation Preliminaries

- There are three outcomes associated with the model (and also in the data).
  - The first two are choices made by the parcel owner: the binary decision to build or not in each period, and the housing service provision decision made conditional on building.
  - The final outcome is the sales price of all properties that sell.
- Therefore, we need to form the log-likelihood for each three outcomes.

Notations

- Let $\boldsymbol{\theta}_P$ denote $(\rho, \gamma_1, \gamma_2, \sigma_\nu)$, $\boldsymbol{\theta}_h$ denote $(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \sigma_\eta)$, and $\boldsymbol{\theta}_d$ denote $(\boldsymbol{\delta}, \beta, \sigma_\epsilon)$
- Given the timing of the decisions and the assumption of independence across errors, the log-likelihood function can be broken into the following three pieces:
  - $L_p(\boldsymbol{\theta}_P \mid \mathbf{P}, \boldsymbol{\Omega})$ - the log-likelihood contribution of prices;
  - $L_h(\boldsymbol{\theta}_P, \boldsymbol{\theta}_h \mid \boldsymbol{h}, \boldsymbol{\Omega})$ - the log-likelihood contribution of housing services;
  - $L_d(\boldsymbol{\theta}_P, \boldsymbol{\theta}_h, \boldsymbol{\theta}_d \mid \mathbf{d}, \boldsymbol{\Omega})$ - the log-likelihood contribution of the binary construction decision.
- The total log-likelihood function is the sum of the three components:

$$L(\boldsymbol{\theta}) = L_p(\boldsymbol{\theta}_P \mid \mathbf{P}, \boldsymbol{\Omega}) + L_h(\theta_P, \boldsymbol{\theta}_h \mid \boldsymbol{h}, \boldsymbol{\Omega}) + L_d(\boldsymbol{\theta}_P, \boldsymbol{\theta}_h, \boldsymbol{\theta}_d \mid \mathbf{d}, \boldsymbol{\Omega}). \tag{9}$$

- A two step estimator similar to Arcidiacono and Miller (2011) where transition and choice probabilities are estimated in a first step and the structural parameters are estimated in the second step.
- To account for the multiple stage procedure, a bootstrap procedure is used to calculate the standard errors.

Estimation - Housing Prices

- To estimate the parameters of the price function given in Equation 2, I estimate the following equation separately for each tract * year combination:

$$\log(P_{nt}) = \log\left(\rho_{jt}\right) + \gamma_{1jt}\log(h_{nt}) + \gamma_{2jt}\log(\mathbf{x}_n) + \nu_{nt}, \tag{10}$$

  where P_{nt}, h_{nt}, and $\mathbf{x}_n$ denote observed sales price, house square footage, and lot size.
- I use a standard, Locally Weighted Regression Approach (see also McMillen and Redfearn (2010)).
- This hedonic price function, where house prices are modelled directly as a function of the observed characteristics of the house, allows the implicit price of square footage to vary both by tract and by year.

Estimation - Variable Costs

- Given estimates of the pricing parameters, I can rearrange the equation for optimal housing services Equation **??** to get the following housing service regression equation:

$$(\gamma_{1jt}-1)\log(h_{nt})+\log(\gamma_{1jt})+\log(\rho_{jt})+\gamma_{2jt}\log(\mathbf{x}_n)+0.5\sigma_v^2=\log(\alpha_{0jt})+\boldsymbol{\alpha}_1\log(\mathbf{x}_n)+\eta_{nt} \tag{11}$$

  I parametrize $\log(\alpha_{0jt})$ as

$$\log(\alpha_{0jt})=\log(\alpha_0)+\log(\alpha_j)+\log(\alpha_t) \tag{12}$$

  Estimating Equation 11 by least squares yields estimates of $\log(\alpha_0),\log(\alpha_j),\log(\alpha_t),\boldsymbol{\alpha}_1$ and the variance of η_{nt}.
- About the concern that the variable costs may e nonlinear and presumably convex: robustness checks in Appendix.

Estimation - Dynamic Discrete Choice 1

- Given results of the first two stages, the remaining structural parameters are $\boldsymbol{\theta}_d = (\boldsymbol{\delta}, \beta, \sigma_\epsilon)$.
- I use the insight from Hotz and Miller (1993) and Arcidiacono and Miller (2011) to take advantage of the terminal state nature of the dynamic discrete choice problem and rewrite $v_0(\boldsymbol{\Omega}_{njt})$ as the expected future per period profit of choosing to not build and a function of the next period probability of choosing to build:

$$v_0\left(\boldsymbol{\Omega}_{njt}\right) = \beta\left(\int \left(\bar{\pi}_1\left(\mathbf{x}_{njt+1}\right) - \sigma_\epsilon \log\left[P_1\left(\boldsymbol{\Omega}_{njt+1}\right)\right]\right) q\left(\boldsymbol{\Omega}_{njt+1} \mid \boldsymbol{\Omega}_{njt}\right) d\boldsymbol{\Omega}_{njt+1}\right) \tag{13}$$

where $P_1\left(\boldsymbol{\Omega}_{njt+1}\right)$ is the conditional choice probability of choosing to build and is given by

$$P_1\left(\boldsymbol{\Omega}_{\boldsymbol{njt}}\right) \equiv \Pr\left(d_{nt} = 1 \mid \boldsymbol{\Omega}_{\boldsymbol{njt}}\right) = \frac{1}{1 + e^{v_0\left(\boldsymbol{\Omega}_{\boldsymbol{njt}}\right)/\sigma_\epsilon - \bar{\pi}_1\left(\mathbf{x}_{njt}\right)/\sigma_\epsilon}}. \tag{14}$$

Estimation - Dynamic Discrete Choice 2

- Using the definition of expected indirect flow profits, $\overline{\pi}_1\left(\mathbf{x}_{njt}\right)$, the difference in value functions is given by

$$v_1\left(\mathbf{\Omega}_{njt}\right)-v_0\left(\mathbf{\Omega}_{njt}\right)=\left(\bar{P}_{nt}-\overline{VC}_{nt}\right)+\left(\beta E_t\delta_{ct+1}-\delta_{ct}\right)$$
$$-\beta\left(\int\left(\bar{P}_{nt+1}-\overline{VC}_{nt+1}-\sigma_\epsilon\log\left[P_1\left(\mathbf{\Omega}_{\boldsymbol{njt+1}}\right)\right]\right)q\left(\mathbf{\Omega}_{\boldsymbol{njt+1}}\mid\mathbf{\Omega}_{\boldsymbol{njt}}\right)d\Omega_{njt+1}\right). \tag{15}$$

- Equation 15 forms the basis of a straightforward two-step estimator.
    - The first step involves estimating both the transition probabilities, $q\left(\mathbf{\Omega}_{\boldsymbol{njt+1}}\mid\mathbf{\Omega}_{\boldsymbol{njt}}\right)$, and the conditional choice probability, $P_1(\mathbf{\Omega}_{njt+1})$.
    - The second step takes estimates of $\overline{P}_{nt}, \overline{VC}_{nt}, \overline{P}_{nt+1}, \overline{VC}_{nt+1}, \int\log\left[P_1(\mathbf{\Omega}_{njt+1})\right]$, and $q\left(\mathbf{\Omega}_{\boldsymbol{njt+1}}\mid\mathbf{\Omega}_{\boldsymbol{njt}}\right)$ as data and estimates the remaining structural parameters, $\boldsymbol{\theta}_d=(\boldsymbol{\delta},\beta,\sigma_\epsilon)$, via maximum likelihood, where the coefficients on a set of county * year dummies will be estimates of $\left(\beta E_t\delta_{ct+1}-\delta_{ct}\right)$.

Takeaway

- We can use the hedonic regression models to build up our original "payoff functions".
- Once we do that, we can directly get the parameters from our linear regressions.
- Appropriately use the existence of a terminal decision in the choice set and do a two-step estimation.
- Notations matter!