

A Dynamic Model of Housing Supply

Alvin Murphy

Online Appendix

A.1 Data Details

A.1.1 Transactions Data

The transaction dataset was obtained from DataQuick and provides the buyers' and sellers' names, along with transaction price, exact street address, square footage, year built, lot size, number of rooms, number of bathrooms, number of units in building, and many other housing characteristics. Based on information about the sellers' names, the year the property was built, the street address, and a subdivision indicator, I can identify new single-family construction that was not part of a subdivision or major development. A major development is defined as ten or more houses built by the same construction company in the same year on the same street or block. The vast majority of exclusions were based on a property being deemed a subdivision. As the dataset contains all houses sold in the Bay Area as well as the year built, I can also calculate total measures of construction activity for any given area at each point in time.

If a house is built but not sold, it will not show up in the sales dataset. The period of analysis is 1988-2004 as that is the year of data from the second data source. However, the transactions data are available up until 2008; therefore, if a house is built in 2004 and sells in 2008, for example, it will be in the data. If a house is built before 2004 but sells after 2008 it will not be in the dataset. To address such missing construction, I use data on annual permits issued over the sample period. Based on permit data and the perfectly observed levels of large developments, I can impute total small-scale construction. I then reweight the observed small-scale construction to match the observed construction in the micro data to the levels suggested by the permit data. For example, in 1998 the

transactions data contain 5,369 new infill houses and 6,054 new houses in developments. The Permit data indicate 13,255 new houses were built in 1998, implying 7,201 new infill houses (13,255-6,054). Therefore each infill unit in the data in 1998 receives a weight of 1.34 (7,201/5,369).

Finally, the Permit data give the date the permit was issued, which is different from date of construction. The U.S. Census Bureau's Survey of Construction provides a distribution of time between receipt of permit and commencement of construction as well as a distribution of construction length. These two distributions in combination with the permit figures can be used to impute an expected level of construction in each period.

A.1.2 Infill-Study Data

The Infill Study uses data from the Census and various local government agencies in addition to assessor data. A vacant parcel is defined as one that has no inhabitable structure or building. Sites with structures too small to be inhabited or for which the structure value is less than \$5,000 (measured in constant 2004 dollars) are also deemed to be vacant. Furthermore, the parcel must be privately owned and both available and feasible for potential urban development, excluding all public lands as well as undeveloped farms, ranges, or forestlands owned by private conservancies, and parcels with slopes in excess of 25 percent. The data do not exclude sites where regulatory/political issues would make construction of new residences difficult. This is ideal as a component of the costs modeled above is regulatory costs.

The second type of infill parcel consists of underdeveloped land called 'refill' parcels. When a parcel is assessed, it is given two separate values: the value of the land and the value of any improvements (buildings). A low improvement-value-to-land-value ratio indicates that the parcel is being underutilized and could be redeveloped. Thus, refill parcels are defined as privately-owned, previously-developed parcels where the

improvement-value-to-land-value ratio is low.

I include vacant parcels and parcels with non-commercial residences that have an improvement-value-to-land-value ratio less than 0.75. 0.75 was chosen as it is a little more conservative than the maximum level of 1 used in the Infill Study data, however, this choice is fairly innocuous. First, the price and variable cost estimates presented in the paper do not rely on this data. Second, the fixed cost estimates are robust to this choice. For example, setting the cutoff to either 0.5 and 1 yields almost identical estimates of fixed costs.

A.2 Dynamic-Discrete-Choice Estimation Details

To estimate $\int \log[P_1(\Omega_{njt+1})]q(\Omega_{njt+1}|\Omega_{njt})d\Omega_{njt+1}$, I first estimate the conditional choice probability function, $P_1(\Omega_{njt})$, and the transition probability function, $q(\Omega_{njt+1}|\Omega_{njt})$. To estimate $P_1(\Omega_{njt})$, I use a flexible logit estimator, where I incorporate linear splines of current prices and variable costs, lagged price and variable cost, and lot-size as well as including Public Use Microdata Area dummies, and county-by-year dummies inside a logistic cumulative distribution function. Public Use Microdata Areas, or PUMAs, are geographic areas created by the Census Bureau which contain at least 100,000 individuals. As I allow the transition of a parcel's price and costs to depend on the parcel's PUMA, it is appropriate to include PUMA dummies in the conditional choice probability function. The flexible logit framework estimates $P_1(\Omega_{njt})$ as:

$$\widehat{P}_1(\Omega_{njt}) = \Lambda\left(\sum_{l=0}^1 S(\overline{P}_{nt-l})'\varphi_{p,l} + \sum_{l=0}^1 S(\overline{VC}_{nt-l})'\varphi_{c,l} + S(x_n)'\varphi_x + \varphi_p + \varphi_{ct}\right) \quad (\text{A.1})$$

where Λ denotes the logistic CDF, $S(\cdot)$ are the spline functions of the relevant arguments and φ are the coefficients to be estimated. In estimating the flexible logit, I trim the top and bottom one percent of prices and variable costs. I use the same trimmed data to estimate the structural parameters.

It is straightforward to estimate the transition probabilities, $q(\Omega_{njt+1}|\Omega_{njt})$. Many of the components of Ω will not be time-varying (lot-size and location) or will transition deterministically (lagged prices and costs), which simplifies the transition probability function. In practice, I assume that the parcel-owner's problem can be solved by estimating transition processes for expected prices, expected variable costs, and the county-by-year effects. See Hendel and Nevo (2006) and Gowrisankaran and Rysman (2012) for a broadly similar approach.

When parcel owners forecast future expected house prices, \bar{P}_{nt+1} they use this period's expected price, \bar{P}_{nt} , lagged expected price, \bar{P}_{nt-1} , and the characteristics of their parcel, x_n . I allow the intercept to vary by Public Use Microdata Areas, and include a time trend. A similar specification is used for \bar{VC}_{nt} .

$$\bar{P}_{nt} = \phi_{0,p} + \sum_{l=1}^L \phi_{1,l} \bar{P}_{nt-l} + \phi_2 x_n + \phi_3 t + \varepsilon_{nt}^{\bar{P}} \quad (\text{A.2})$$

Finally, as county-by-year dummies are included in the policy function, I estimate the transition process for φ_{ct} as:

$$\varphi_{ct} = \varrho_c + \varrho_1 \varphi_{c,t-1} + \varepsilon_{ct}^{\varphi} \quad (\text{A.3})$$

Using the data, the coefficients from (A.1), (A.2), and (A.3), and the empirical distributions of $\hat{\varepsilon}_{nt}^{\bar{P}}$, $\hat{\varepsilon}_{nt}^{\bar{VC}}$, and $\hat{\varepsilon}_{ct}^{\varphi}$, I calculate $\int \log[\widehat{P}_1(\Omega_{njt+1})] \widehat{q}(\Omega_{njt+1}|\Omega_{njt}) d\Omega_{njt+1}$ by simulation. As $\bar{\pi}_1(x_{njt})$ is linear in the remaining parameters to be estimated, I can simulate the expected future values of the variables, thus ensuring that once I simulate the future data one time, the remaining problem is a linear-in-parameters logit model. The construction of the expected continuation value is carried out by simulation and the fact that it only needs to be done once makes computation straightforward.

Each parcel owner will choose to build if $v_1(\Omega_{njt}) + \epsilon_{1nt} > v_0(\Omega_{njt}) + \epsilon_{0nt}$. I can estimate the remaining parameters using maximum likelihood, where the individual

likelihood contributions are formed by plugging the difference between $v_1(\Omega_{njt})$ and $v_0(\Omega_{njt})$ into (14). Letting hats denote the variables that are estimated in earlier stages, the difference in values that is plugged into the likelihood is given by:

$$v_1(\Omega_{njt}) - v_0(\Omega_{njt}) = (\widehat{P}_{nt} - \widehat{VC}_{nt}) + (\beta E_t \delta_{ct+1} - \delta_{ct}) - \beta \left(\int (\widehat{P}_{nt+1} - \widehat{VC}_{nt+1} - \sigma_\epsilon \log[\widehat{P}_1(\Omega_{njt+1})]) \widehat{q}(\Omega_{njt+1} | \Omega_{njt}) d\Omega_{njt+1} \right) \quad (\text{A.4})$$

A.3 Additional Results

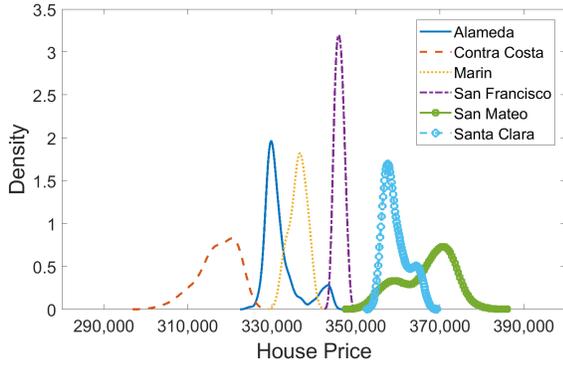
Table A.1: Transition Probability Parameters

| | \overline{P}_{nt} | | \overline{VC}_{nt} | | ϱ_{ct} |
|-----------------------|--------------------------|------------------------|--------------------------|------------------|-------------------|
| \overline{P}_{nt-1} | 0.873 (0.065) | \overline{VC}_{nt-1} | 0.817 (0.065) | ϱ_{ct-1} | 0.845 (0.050) |
| \overline{P}_{nt-2} | -0.021 (0.066) | \overline{VC}_{nt-2} | -0.027 (0.065) | | |
| Lot size | 1.135e-02 (1.032e-03) | Lot size | 1.060e-02 (8.653e-04) | | |
| t | 0.058 (0.002) | t | 0.004 (0.002) | t | -0.102 (0.009) |
| PUMA | Yes | PUMA | Yes | County | Yes |

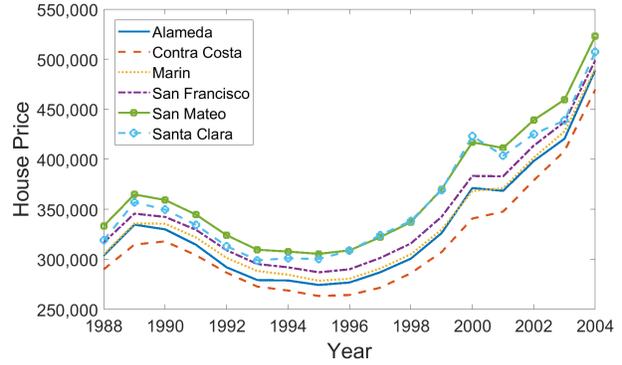
Note: Lot size is measured in 1000s of square feet. Bootstrapped Standard Errors in parentheses.

Figure A.1: Price of a Typical House – by County

(a) Across Tracts



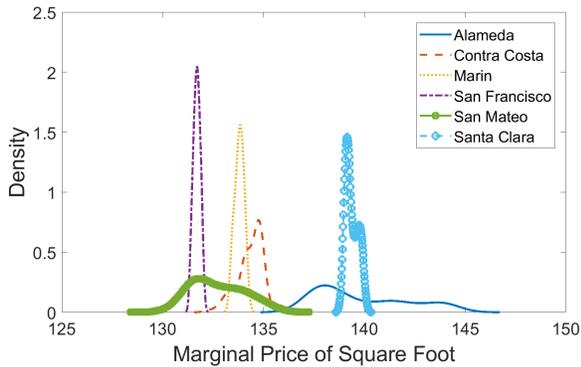
(b) Over Time



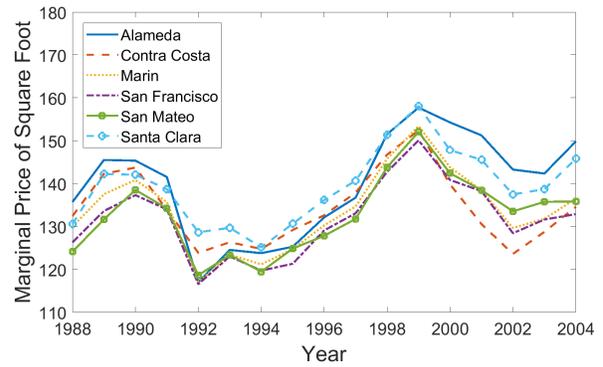
Note: This figure presents the county-specific analog to Figure 4. In Figure A.1a, density is measured in units of 1/10,000

Figure A.2: Marginal Price of Square Footage in a Typical House – by County

(a) Across Tracts

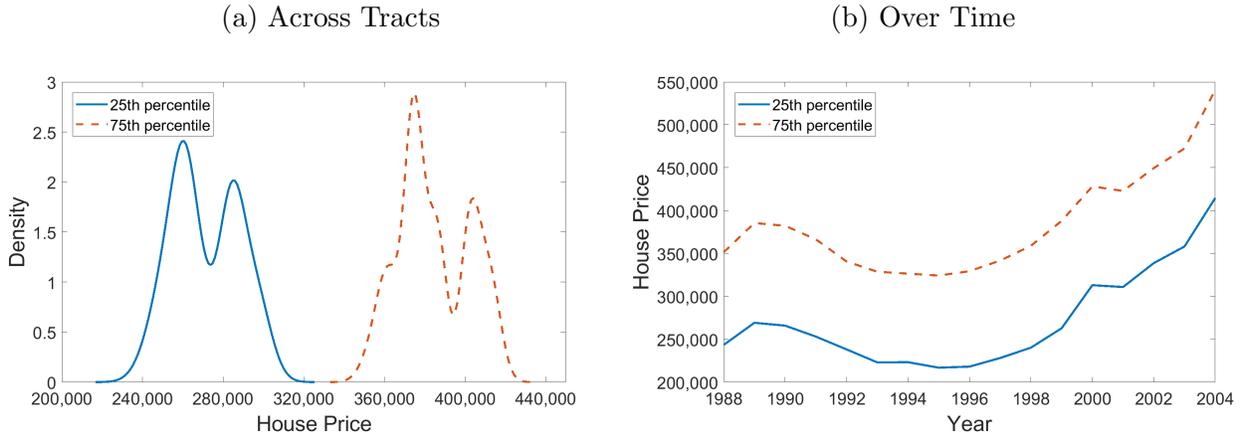


(b) Over Time



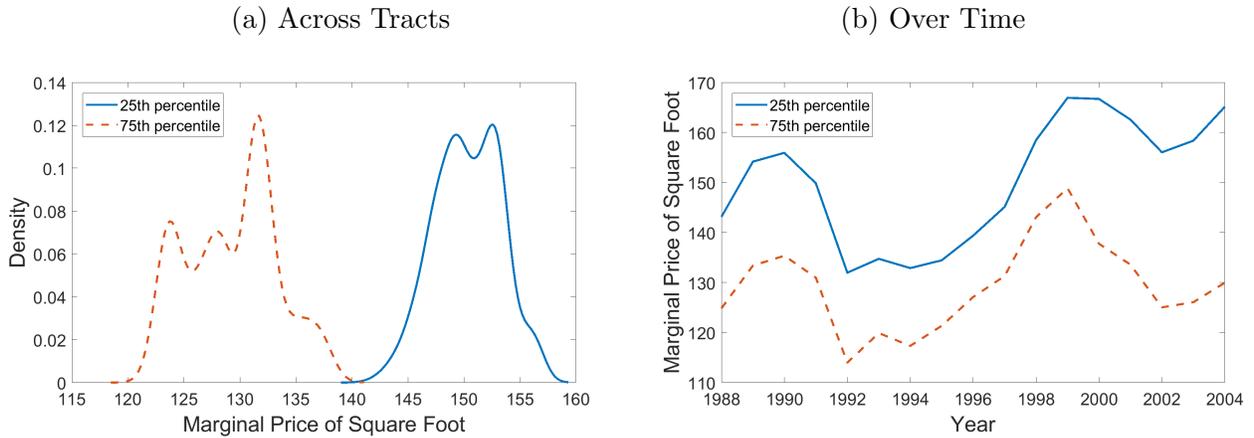
Note: This figure presents the county-specific analog to Figure 5.

Figure A.3: Price of a House – by Size



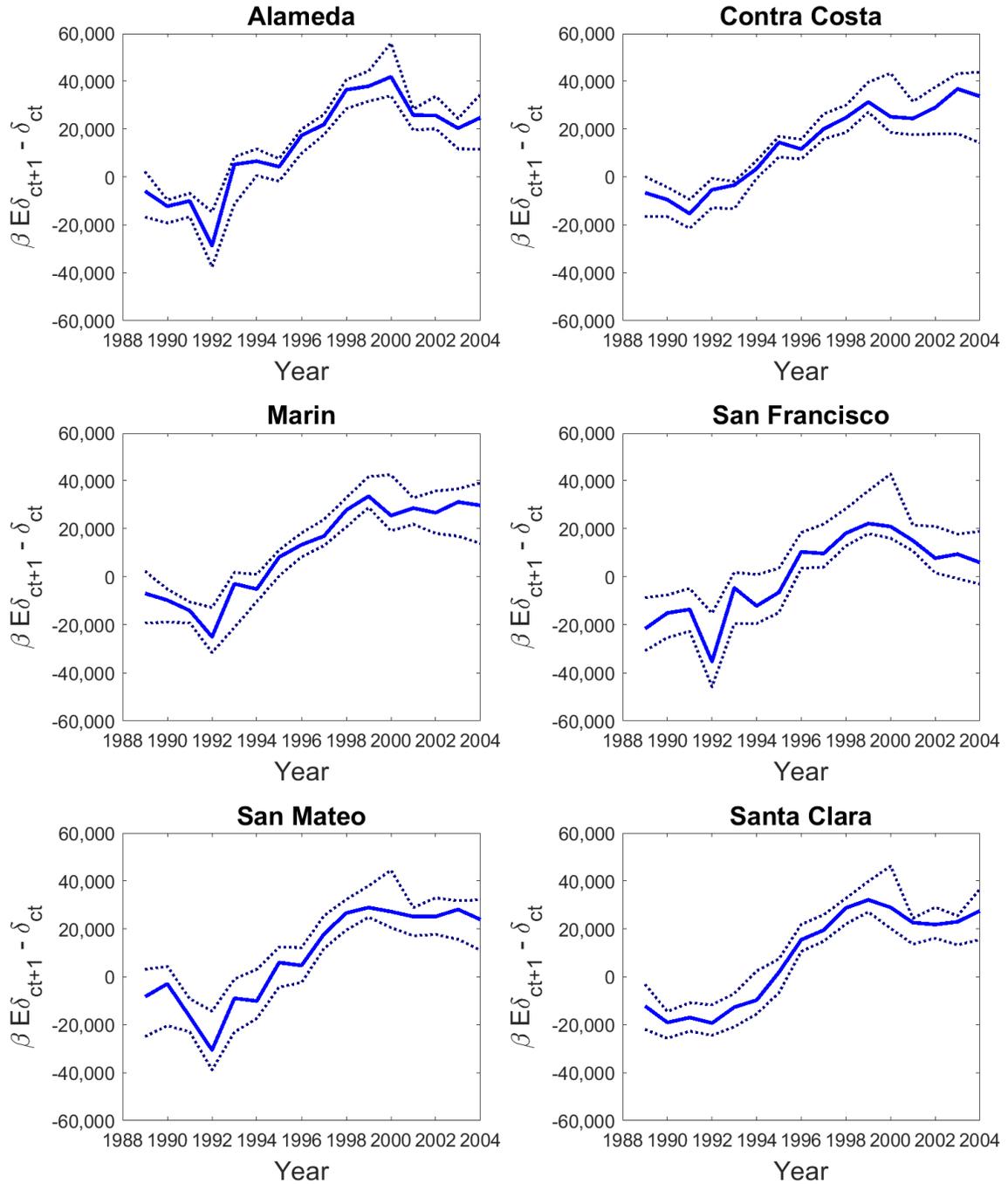
Note: This figure presents the analog to Figure 4 where prices are calculated for a new house with 1,200 square feet of living space and a lot size of 4,000 square feet and for a new house with 1,990 square feet of living space and a lot size of 7,500 square feet. 1,200 and 4,000 correspond to the sample 25th percentiles for house size and lot size, respectively. 1990 and 7,500 correspond to the sample 75th percentiles for house size and lot size. In Figure A.3a, density is measured in units of 1/100,000.

Figure A.4: Marginal Price of Square Footage in a Typical House – by Size



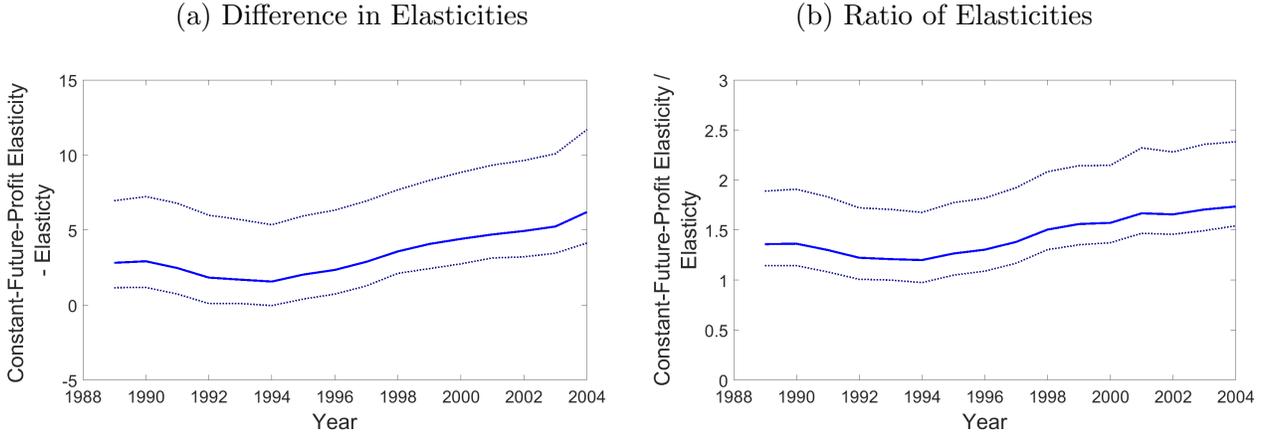
Note: This figure presents the analog to Figure 5 where prices are calculated for a new house with 1,200 square feet of living space and a lot size of 4,000 square feet and for a new house with 1,990 square feet of living space and a lot size of 7,500 square feet. 1,200 and 4,000 correspond to the sample 25th percentiles for house size and lot size, respectively. 1990 and 7,500 correspond to the sample 75th percentiles for house size and lot size.

Figure A.5: Time Trend of Expected Fixed Cost Growth with 95% Confidence Intervals



Note: 95% Confidence Intervals were obtained using a bootstrap procedure with 250 draws.

Figure A.6: Difference in and Ratio of Elasticities with 95% Confidence Intervals



Note: Figure A.6a shows the difference between the two development elasticities (standard and constant future profits) presented in Figure 8. Figure A.6b shows the ratio of these two elasticities. 95% Confidence Intervals were obtained using a bootstrap procedure with 250 draws.

A.4 Sensitivity Analysis

A.4.1 Variable-Cost Function

When the variable cost function is quadratic, there will not be a closed-form solution for (11). However, the parameters of the cost function can still be estimated using the approach described in Bishop and Timmins (2017). One can still easily solve for the econometric error, η_{nt} , which, in conjunction with a simple change of variables, can be used to construct the likelihood of observing h_{nt} . Choosing the cost parameters to maximize this likelihood yields estimates of the variable cost function.

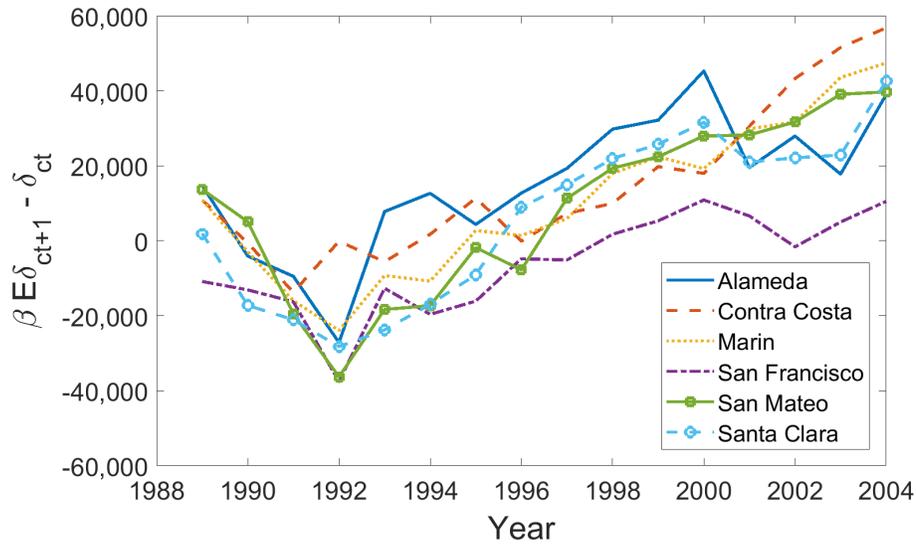
In practice, I let the curvature of the quadratic cost function (i.e., the slope of the marginal cost function) vary by year. Estimates of the slopes of the marginal cost function range from almost flat to 0.0035. Even for the largest number, the impact is very small – for example, doubling the size of the average house (which is 1,670 square feet) pushes up marginal costs by \$5.84.

Consequently, the overall results are similar. As discussed above, in the base specification the median cost per square foot in 2004 estimated by the structural model is

\$112 and the interquartile range is \$106 to \$118. The corresponding numbers for the quadratic-variable-cost case (calculated at the mean house size) are 113, 106, and 122.

Finally, to estimate the fixed cost parameters (when the variable cost function is quadratic), the profit function must be simulated. These fixed costs patterns are shown in Figure A.7. As can be seen, they are similar to the fixed cost estimates from the linear-variable-cost case, as shown in Figure 7.

Figure A.7: Time Trend of Expected Fixed Cost Growth by County – Quadratic-Variable-Cost Model



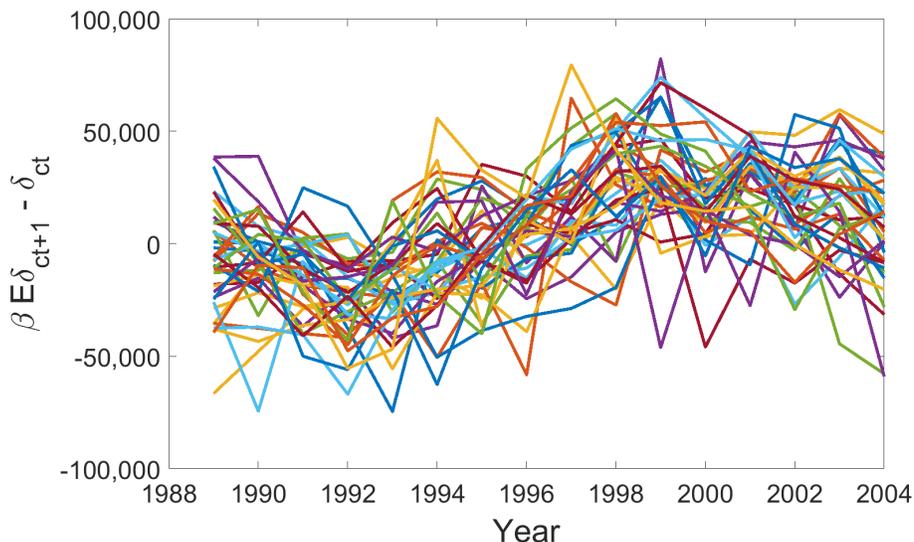
Note: Figure A.7 shows the county- and year-specific estimates of $\beta E_t \delta_{ct+1} - \delta_{ct}$, i.e., the difference between the discounted expected one-year-ahead fixed costs and current fixed costs. Estimates taken from a model that allows for quadratic variable costs.

A.4.2 Fixed-Cost Function

I estimate a version of the model where the fixed costs are allowed to vary at the PUMA \times year level. PUMAs contain approximately 100,000 residents and are considerably smaller level of geography than counties. For example, in the data there are 46 PUMAs and 6 counties. For 8 of the PUMAs, the results are sufficiently noisy that the absolute value of expected growth in fixed costs exceeds \$80,000 in one or more years,

in some cases by a great deal. This suggests the data are not rich enough to accurately identify fixed cost trends at a finer level than county. The time trends for the remaining 38 PUMAS are shown in Figure A.8 and the trends are qualitatively similar to the county×year case shown in Figure 7.

Figure A.8: Time Trend of Expected Fixed Cost Growth by PUMA



Note: Figure A.8 shows the PUMA- and year-specific estimates of $\beta E_t \delta_{ct+1} - \delta_{ct}$, i.e., the difference between the discounted expected one-year-ahead fixed costs and current fixed costs. Results shown for the 38 PUMAs where expected growth in fixed costs was less than \$80,000 (in absolute terms) in every year.

References

- Bishop, Kelly C., and Christopher Timmins.** 2017. “Estimating the Marginal Willingness to Pay Function Without Instrumental Variables.” mimeo, Arizona State University.
- Gowrisankaran, Gautam, and Marc Rysman.** 2012. “Dynamics of Consumer Demand for New Durable Goods.” *Journal of Political Economy*, 120(6): 1173–1219.
- Hendel, Igal, and Aviv Nevo.** 2006. “Measuring the Implications of Sales and Consumer Inventory Behavior.” *Econometrica*, 74(6): 1637–1673.