Research Note: Housing Hedonics, Financial Frictions, and Discrete Choice Econometrics

John Heilbron *

February 2023

Abstract

I elaborate a standard discrete choice framework to identify parameters describing both preferences and constraints. Estimation can be accomplished in a 2SLS regression with appropriately chosen instruments. I attempt to apply the framework to the simultaneous problem of neighborhood and mortgage choice. The results contradict downward sloping individual demand curves, and so I conclude I have not yet found satisfactory instruments.

Keywords: Household Finance, Housing Demand

JEL Codes: D12, H81, R21

*University of Chicago, Booth School of Business

1 Introduction

The purpose of this research note is to describe some additional econometric tools I developed while revising my job market paper, "Credit Constraints at Home Purchase and Bias in Hedonic Amenity Valuations". That paper develops some intuition for how to correct for the presence of credit constraints when using housing prices to estimate household valuations of various residential amenities. A common piece of feedback I received was the need for econometric identification in my results.

Having thought about the problem, I have developed a way to extend the standard discrete-choice approach to capture both preferences and constraints. In particular, I build off of the presentation of a discrete choice problem presented in Dubé et al. (2021). The benefit of using this approach is its relative simplicity and the well-developed understanding of how to address identification concerns. For my application, I derive 2SLS regression specification.

When I turn to implementing this specification, however, I retrieve coefficients that are inconsistent with downward-sloping demand curves. I try a variety of different regression coefficients and a variety of regression specifications but cannot resolve the problem. I conclude that my instruments are likely not satisfactory. This has the unfortunate effect that the results in themselves are somewhat uninteresting. However, the procedure and analysis seem sound and illuminating and so I present them here.

2 A simple regression framework

I consider a household simultaneously choosing a location for home-ownership, $k \in \mathscr{K}$ and a leverage for the home purchase, $l \in \mathscr{L}$. The location, k, indexes the home price, P_k, and a suite of local amenities, X_k. The leverage, l, indexes the loan-to-value ratio of the home purchase, LTV_l, and a mortgage rate, r^m_l.[1][2] The household's choice of location, k, and leverage, l, implies a down-payment on the home, $D_{kl} \equiv P_k(1 - LTV_l)$, and a present value of mortgage obligations, $M_{kl} \equiv B^o_{kl}\frac{r^m_l}{r^f} = P_k LTV_l \frac{r^m_l}{r^f}$.

The household chooses between these substitute products defined by:

$$\mathscr{K} \times \mathscr{L} = \{(1,1), ..., (k,l), ..., (K,L)\}$$

and an alternative, 0, defined to be renting. The household has income endowments over two

[1] This is a reduced-form representation of an equilibrium phenomenon whereby the mortgage rate is increasing in the leverage. That this obtains in the market is clear from the FNMA grids that charge additional interest when borrowers choose higher LTVs.

[2] I consider discrete loan-to-value alternatives in the same sense that the GSEs bin borrowers according to loan-to-value ranges.

periods, (y_0, y_1), which it expends to maximize utility over housing choices, $\{q_0, q_1, ..., q_{KL}\}$ with $q_{kl} \in \{0,1\}$, and expenditure on a consumption good in each of two periods, $\{c_0, c_1\}$:

$$U(q, y_0, y_1) = \sum_{j=1}^{J+1} (\chi_{kl}^h q_{kl}) \exp\left\{\beta^D c_0 + \beta^M \frac{c_1}{1+r}\right\}$$

χ_{kl}^h is the quality of location-leverage pair kl, β^D is the preference for the consumption good at present, β^M is the preference for the consumption good in the future. I motivate this form of the utility function in Appendix (6). The quality of the location-leverage pair consists of a deterministic and stochastic component, $\chi_{kl}^h = \exp\left\{X_k'\beta^X + \xi_k + \xi_l + \varepsilon_{kl}^h\right\}$ and $\chi_0^h = \exp\left\{\varepsilon_0^h\right\}$. Here, X_k is a vector of amenities as indexed by the location, ξ_k and ξ_l are scalar vertical characteristics of the location and leverage choice that are unobserved by the researcher, β^X are tastes for the amenities, and $\varepsilon_{kl}^h \sim$ i.i.d. EV(0,1) $\forall kl$. Household h then faces the following choice-specific indirect utilities in logs. Again, see Appendix (6) for the derivation:

$$\begin{aligned} u_{jk}^h &= (y_0 + \phi - D_{kl})\beta^D + (y_1 - \phi - M_{kl})\beta^M + X_k'\beta^X + \xi_k + \xi_l + \varepsilon_{kl}^h \\ &\vdots \\ u_0^h &= (y_0 + \phi)\beta^D + (y_1 - \phi)\beta^M + \varepsilon_0^h \end{aligned} \tag{1}$$

Choice specific mean utilities are then:

$$\begin{aligned} \overline{u}_{jk} &= \beta + D_{kl}\beta^D + M_{kl}\beta^M + X_k'\beta^X + \xi_k + \xi_l \\ &\vdots \\ \overline{u}_0 &= \beta \end{aligned} \tag{2}$$

Predicted market shares are:

$$s_{kl} = \frac{\exp\{\overline{u}_{kl}\}}{1 + \sum_{k'l'} \exp\{\overline{u}_{k'l'}\}} \tag{3}$$

The log-odds ratio then defines the following regression:

$$\log \frac{s_{kl}}{s_0} = D_{kl}\beta^D + M_{kl}\beta^M + X_k'\beta^X + \xi_k + \xi_l \tag{4}$$

The log-odds can be considered as a utility index. This comports with intuition; if a location-leverage pair offers higher utility, then we would expect more households to choose it, generating the higher log-odds. This yields the following interpretation for the regression co-efficieints. β^D represents $\frac{\text{utils}}{\$^{tod.}}$ and, on the notion that households are constrained in the period of home purchase, I interpret this as $\frac{\text{utils}}{\$^{const.}}$. β^M represents $\frac{\text{utils}}{\$^{tom.}}$ and, on the notion that households are relatively unconstrained later in the life of their mortgage, I interpret this as $\frac{\text{utils}}{\$^{unconst.}}$. β^X represents $\frac{\text{utils}}{\text{amenity}}$.

We have that $\frac{\beta^M}{\beta^D} = \kappa$ represents willingness-to-pay for $\$^{tod.}$ out of $\$^{tom.}$, or willingness-to-pay for credit. $\frac{\beta^X_a}{\beta^D}$ represents willingness-to-pay for amenity, a, out of $\$^{tod.}$, or constrained willingness-to-pay for the amenity. I argue this is the willingness-to-pay measured by traditional hedonic estimates and suffers from bias. $\frac{\beta^X_a}{\beta^M}$ represents willingness-to-pay for amenity, a, out of $\$^{tom.}$, or unconstrained willingness-to-pay for the amenity. This is the policy-relevant willingness-to-pay as appropriately corrected for the bias.

3 Identification and Choice of Instruments

In the regression specification described above, the shock terms account for the fact that there may be demand shifters for the location or some particular leverage that are not captured by payments or amenities and not observable to the econometrician. For example, in the cross section of locations, it may be that higher down-payments are associated with some attractive quality of the location, so ξ_k is actually increasing, which generates an increase in the log-odds of the choice. An OLS regression would recover a positive coefficient on β^D, even though, in fact, all else equal, borrowers prefer lower down-payments. This suggests the need for an instrumental variables strategy for both down-payments and mortgage-obligations. [3]

Appropriate instruments should deliver variation in down-payments and mortgage obligations that is orthogonal to the demand shocks. Furthermore, the instruments should deliver variation in down-payments that is orthogonal to the variation in mortgage obligations. That is, the two endogenous regressors cannot share a common instrument.

As one source of variation, I turn to the literature on exogenous variation in home prices. A variety of instruments have been proposed, dating back to the supply elasticity instruments of Mian and Sufi (2009). For convenience in construction, I use instruments proposed by Graham and Makridis (2020). In particular, I use the instrument that instruments returns on the house price index at the census region level, r^{hpi}, with characteristics of the housing stock in the MSA. These characteristics include things like the fraction of 2 bedroom homes, or fraction of homes built in the 1960s. The idea is that housing stock characteristics are effectively exogenous in the period of the house price growth and act like exposures to the regional house price shock.

With a fixed loan-to-value ratio, variation in house prices enters down-payments and mortgage obligations, the latter by way of origination balances. Although this variation is orthogonal to

[3]This approach also suggests the need for an instrument for the amenities. Boundary fixed effects are the most common example of amenity instruments. This approach, however, requires a degree of granularity from the data that is incompatible with this approach. I leave consideration of this approach to future work.

demand shocks, the variation in down-payments and mortgage obligations is not mutually orthogonal.

I obtain additional orthogonal variation in the value of mortgage obligations my exploiting changes in mortgage and mortgage insurance rates charged by the GSEs and mortgage insurance companies. The rate schedules change discontinuously at certain FICO thresholds, e.g. 640, 660, 680. I bin borrowers narrowly by FICO bin, j, e.g. $(640,650],(650,660],(660,670],(670,680]$. I then compose larger fico bins, j' to include a bin just below and above the threshold at which fico scores increase, e.g. $(650,670]$ includes threshold 660. I compare borrowers within the bin and suppose that, given how close their FICO scores are, that besides the increase in rates, they are otherwise similar.

4 Regression Specification

My preferred regression specification combines the intuition of the log-odds regression with the instruments I identify to deliver exogenous and orthogonal variation in both down-payments and mortgage obligations.

$$\log \frac{s_{jklt}}{s_{jt0}} \sim \alpha_t + \alpha_{j'(j)} + D_{klt}\beta^D + M_{jklt}\beta^M + X_k'\beta^x + \xi_k + \xi_l \tag{Second Stage}$$

$$y_{jklt} \sim \theta_t + \theta_{j'(j)} + z_{kt}^{P\prime}\delta^P + z^M\delta^M + X_k'\delta^x + \eta_{kl} \tag{First Stage}$$

Payment variable construction:

$$D_{klt} \equiv P_{kt} * (1 - LTV_l)$$

$$M_{jklt} \approx B_{klt}^o * \frac{r_{jl}^m}{r^f} = P_{kt} * LTV_l * \frac{r_{jl}^m}{r^f}$$

Instrument construction:

$$z_{kt}^{P\prime} = r_{r(k)t}^{hpi} * \begin{bmatrix} \frac{N_k^{1bed}}{N_k} & \dots & \frac{N_k^{1room}}{N_k} & \dots & \frac{N_k^{1950s}}{N_k} & \dots \end{bmatrix}$$

$$z_{jl}^M = \mathbb{1}_{jl}\{j = \arg\max_{j \in j'}\{r_{jl}^m\}\}$$

5 Regression Results

I report the results of my preferred specification in Table (A). To interpret, fit_dpmt_med are the fitted down-payments, fit_pmt_pdv are the fitted value of mortgage obligations. The rest are the amenities and fairly unimportant, though it's helpful to see a positive loading on school quality (*cs_mn_grd_ol*) and a negative one on lack of internet access (*p_no_internet_acc*).

The challenge that I'm having is that, in spite of my efforts to instrument, I'm still obtaining a positive estimate for β^D. I have a strong prior that increased down-payments should decrease the utility index, so I conclude that my regression is mis-specified.

There are a number of ways that I've already made sanity checks of the data and otherwise refined the specification. Some of my outstanding thoughts are as follows:

- Because I can estimate incomes in my data (using debt-to-income ratios and mortgage payment sizes), I had been thinking about potentially breaking my borrowers further into income bins. My main motivation here is that the R^2 of mortgage rates on fico, ltv bin, and income bin is higher than on fico and ltv bin alone. But to be honest, I'm not sure how this would resolve my issue.

- Currently, where I have location-by-leverage bins, kl, that are empty, I simply drop the observations. This suggests a selection model in which particular location-leverages do not enter consideration. Provided that they escape consideration randomly, this should not pose any identification threats.

- I had also thought about taking logs of down-payments and mortgage obligations to reduce the influence of outliers. It is common to take logs of prices when estimating these hedonic models (and to take logs of dollar-valued distributions generally). My hesitation here is that this specification has some appeal in that it provides coefficients with structural interpretations. For example, $\frac{\beta^D}{\beta^M}$ represents WTP for a dollar of credit at the time of home purchase when the regression is run in levels, and it's unclear what it means when run in logs.

6 Micro-foundation for discrete choice regression

I consider utility maximization over present and future non-housing consumption after financing housing purchase and subject to potential borrowing constraints:

$$\begin{aligned} \tilde{u} = \max_{c_0, c_1, a_1} \; & u(c_0) + \beta u(c_1) \\ \text{s.t.} \qquad & y_0 = c_0 + \frac{a_1}{1+r} + D_j \\ & y_1 + a_1 = c_1 + M_j(1+r) \\ & a_1 \geq -\phi(1+r) \end{aligned} \tag{5}$$

Letting $\tilde{c}_0^*$ and $\tilde{c}_1^*$ be optimal consumption in version of the problem without a borrowing constraint. Define $\overline{c}_0 \equiv y_0 - D_j + \phi$ as maximum consumption in the first period when subject to the constraint and, similarly, $\underline{c}_1 \equiv y_1 - M_j(1+r) - \phi(1+r)$ as the minimum consumption achieved in the second period. Optimal consumption in the constrained problem is then $c_0^* \equiv \min\{\tilde{c}_0^*, y_0 - D_j + \phi\}$ and $c_1^* \equiv \max\{\tilde{c}_1^*, y_1 - M_j(1+r) - \phi(1+r)\}$. The agent's utility is then:

$$\tilde{u} = u\big(c_0^*\big) + \beta u\big(c_1^*\big) \tag{6}$$

We approximate the utility by taking a first-order expansion at the agent's chosen consumption. We simplify somewhat using $\beta(1+r) = 1$.

$$\begin{aligned} \tilde{u} &\approx u'(c_0^*)\Big[y_0 - D_j + \phi\Big] + \beta u'(c_1^*)\Big[y_1 - M_j(1+r) - \phi(1+r)\Big] \\ &= u'(c_0^*)\Big[y_0 - D_j + \phi\Big] + u'(c_1^*)\Big[\frac{y_1}{1+r} - M_j - \phi\Big] \end{aligned} \tag{7}$$

Note that in the solution where borrowers are not constrained by their borrowing limit, borrower optimization and $\beta(1+r) = 1$ together guarantee that $u'(c_0^*) = u'(c_1^*)$ and this reduces to the usual single period problem:

$$\tilde{u} = u'(\tilde{c}_0^*)\Big[y_0 + \frac{y_1}{1+r} - D_j - M_j\Big] \tag{8}$$

When borrowers are constrained, by contrast, we have:

$$\tilde{u} = u'(\overline{c}_0)\Big[y_0 - D_j + \phi\Big] + u'(\underline{c}_1)\Big[\frac{y_1}{1+r} - M_j - \phi\Big] \tag{9}$$

Where $u'(\overline{c}_0) > u'(\overline{c}_1)$.

I relabel $\beta^D = u'(\overline{c}_0)$ and $\beta^M = u\prime(\underline{c}_1)$. Meanwhile, $y_0 - D_j + \phi$ represents current consumption, c_0, and $\frac{y_1}{1+r} - M_j - \phi$ is the NPV of future consumption, $\frac{c_1}{1+r}$. So we can write utility as:

$$\tilde{u} = \beta^D c_0 + \beta^M \frac{c_1}{1+r} \tag{10}$$

References

Jean-Pierre Dubé, Ali Hortaçsu, and Joonhwi Joo. Random-Coefficients Logit Demand Estimation with Zero-Valued Market Shares. *Marketing Science*, 40(4):637–660, 2021. doi: 10.1287/mksc.2020.1280.

James Graham and Christos Makridis. House Prices and Consumption: A New Instrumental Variables Approach. (3285263), May 2020. doi: 10.2139/ssrn.3285263. URL https://papers.ssrn.com/abstract=3285263.

Atif Mian and Amir Sufi. The consequences of mortgage credit expansion: Evidence from the u.s. mortgage default crisis *. *Quarterly Journal of Economics*, 124:1449–1496, 11 2009. ISSN 0033-5533. doi: 10.1162/qjec.2009.124.4.1449.

A Tables

Group:	Price				Down Pmt. & Mort. Oblig.				
Dependent Variables:	Log Odds		$\frac{\text{Price}}{\text{Income}}$	Log Odds	Log Odds		$\frac{\text{Down Pmt.}}{\text{Income}}$	$\frac{\text{Mort. Oblig.}}{\text{Income}}$	Log Odds
	OLS	Reduced Form	First Stage	2SLS	OLS	Reduced Form	First Stage		2SLS
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Variables									
pr_inc_ratio	-0.033**			0.513***					
	(0.013)			(0.187)					
dp_inc_ratio					-0.949***				12.7***
					(0.027)				(1.32)
mv_inc_ratio					0.015				-3.44***
					(0.016)				(0.247)
z_price_ii_rhat_hpi		1.24***	2.42**			1.28***	0.663***	2.08***	
		(0.444)	(0.954)			(0.465)	(0.253)	(0.799)	
z_mort_oblig						-0.342***	-0.005***	0.080***	
						(0.011)	(0.001)	(0.004)	
Dollar Stores	-2.83***	-2.70***	-1.84***	-1.75***	-3.35***	-2.76***	-0.527***	-1.53***	-1.31
	(0.232)	(0.218)	(0.087)	(0.424)	(0.260)	(0.225)	(0.038)	(0.077)	(0.846)
Restaurants	-0.096***	-0.104***	0.131***	-0.171***	-0.075***	-0.101***	0.023***	0.130***	0.056
	(0.018)	(0.018)	(0.006)	(0.033)	(0.020)	(0.018)	(0.003)	(0.006)	(0.051)
Elem/Sec Schools	-0.376***	-0.377***	0.066***	-0.410***	-0.412***	-0.385***	-0.024***	0.112***	0.308***
	(0.064)	(0.064)	(0.012)	(0.063)	(0.069)	(0.065)	(0.008)	(0.012)	(0.094)
Supermarkets	1.60***	1.54***	0.846***	1.10***	1.89***	1.59***	0.292***	0.612***	-0.023
	(0.070)	(0.071)	(0.063)	(0.170)	(0.077)	(0.071)	(0.015)	(0.059)	(0.393)
Doctor's Offices	0.063***	0.063***	0.016***	0.055***	0.075***	0.066***	0.011***	0.004	-0.056***
	(0.010)	(0.010)	(0.002)	(0.010)	(0.011)	(0.011)	(0.001)	(0.003)	(0.020)
Barber Shops	3.48***	3.50***	0.425***	3.28***	3.84***	3.57***	0.361***	0.033	-0.914*
	(0.122)	(0.124)	(0.050)	(0.169)	(0.127)	(0.126)	(0.017)	(0.063)	(0.546)
Banks	-2.48***	-2.52***	-0.054	-2.49***	-2.73***	-2.60***	-0.210***	0.232***	0.865*
	(0.086)	(0.088)	(0.078)	(0.110)	(0.091)	(0.089)	(0.026)	(0.079)	(0.451)
Churches	-0.442***	-0.434***	-0.231***	-0.315***	-0.524***	-0.451***	-0.081***	-0.166***	0.012
	(0.017)	(0.018)	(0.006)	(0.048)	(0.020)	(0.018)	(0.003)	(0.005)	(0.108)
No Internet (%)	-2.30***	-2.22***	-0.537***	-1.94***	-2.80***	-2.31***	-0.434***	-0.024	3.13***
	(0.125)	(0.127)	(0.124)	(0.170)	(0.129)	(0.127)	(0.033)	(0.119)	(0.702)
School Quality	1.10***	1.09***	0.647***	0.754***	1.45***	1.13***	0.345***	0.281**	-2.28***
	(0.228)	(0.226)	(0.114)	(0.267)	(0.242)	(0.230)	(0.037)	(0.123)	(0.664)
pmi_1nd					-0.903***	0.380***	-1.32***	1.96***	23.9***
					(0.014)	(0.018)	(0.008)	(0.016)	(1.92)
Fixed-effects									
FICO Bin-Year-Census Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics									
Observations	210,137	210,137	210,137	210,137	210,137	210,137	210,137	210,137	210,137
R^2	0.25519	0.25532	0.75766	0.23992	0.37132	0.29362	0.73264	0.74100	-34.568
Within R^2	0.08296	0.08312	0.37524	0.06416	0.22594	0.13029	0.72569	0.67245	-42.792
F-test	8.5655	8.5714	78.163	8.5714	12.494	8.7932	57.968	60.522	8.7932

Clustered (FICO Bin-Year-Census Region) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*