Housing Affordability and the Mobility of Renters

Nathaniel Baum-Snow*, Abdollah Farhoodi†, Lu Han‡

March 21, 2025

Abstract

This paper examines the welfare costs of rent growth and the prospects for reducing these costs. Using panel data on the near universe of renters in Texas 2010-2019, we estimate migration costs between neighborhoods jointly with the demand system for neighborhoods and commuting zones. A structural dynamic model of demand for cities and neighborhoods with endogenous mobility underlies these estimates. Using estimated model parameters, we quantify how migration to substitute cities and neighborhoods has mitigated the welfare cost of rent growth by household type. A reduced-form analysis of the impacts of housing supply shocks on renters' mobility and rental payments corroborates this model-based decomposition.

1 Introduction

Between 2014 and 2022, the Zillow rent index for the US increased by 61%, which was more than twice the rate of general price inflation. The extent to which this rental price growth burdens incumbent renters depends not only on rent changes in their neighborhoods but also on mobility costs and rents in alternative substitute neighborhoods. While the mobility option plays a potentially central role in mitigating the welfare costs of rent growth, heretofore there is little systematic evidence on the value of this option. Understanding the extent to which household mobility can facilitate housing affordability is central to designing policies that mitigate the welfare burdens of rent growth. Quantification of the underlying mechanisms driving renters' mobility responses to housing market shocks delivers valuable insights into the consequences of various supply-side policies that influence rents.

This paper systematically investigates the welfare consequences of rising rental prices for renters. We build an individual-level panel data set of residential locations for the near

*Professor, University of Toronto, Rotman School of Management

†Assistant Professor, University of Toronto, Department of Economics

‡Professor, University of Wisconsin at Madison, Wisconsin School of Business

universe of renters in Texas by marrying migration histories from the Infutor marketing data set with housing tenure information in CoreLogic. We develop a dynamic structural model of location choice and endogenous migration with parameters that can be feasibly estimated with this data set. Using parameter estimates from the model, we quantify the welfare costs of rent growth between 2010 to 2019 for incumbent renters in different types of neighborhoods. In particular, we quantify the extent to which migration has allowed incumbent renters to mitigate the welfare burden of rent growth in their neighborhoods. Essential to this quantification are credible estimates of two sets of parameters. First, we recover estimates of parameters governing migration costs across neighborhoods within and between commuting zones. Second, we recover estimates of utility parameters by accounting for the existing dynamic location choice decision and governing the demand system for commuting zones and neighborhoods in a flexible dynamic nested logit framework. Associated population supply elasticities are allowed to differ by commuting zone size and location. In robustness specifications, all parameters are allowed to differ by various 2010 neighborhood attributes.

In the model, individuals face a move-stay decision each year. This is the top nest of a 3-level nested logit. Conditional on moving, there are nested choices of commuting zone and neighborhood within commuting zone. Each neighborhood delivers a flow utility that depends on the market rent index, a neighborhood fixed effect, CBSA-time fixed effects, an amenity shock that is common to all residents, and an idiosyncratic shock. The nested logit structure allows for flexible substitution patterns across neighborhoods. Similar to Diamond et al. (2018), migration costs and flow utilities depend on the number of years in the residence and number of years in the commuting zone. Over time, people can establish stronger ties to their home and city. The magnitude of local ties determines the extent to which housing affordability can be mitigated through providing lower rent housing options in far away neighborhoods or in other commuting zones. Migration costs also depend on distance migrated and whether migration is within or between commuting zones. Additionally, individuals receive a flow utility shock associated with staying in their home. The parameter governing the variance of this shock captures the likelihood of lease non-renewal by the landlord, changes in family structure, or changes in employment situation that influence individuals' housing choices independent of their neighborhood. Model state variables are location, year, residence tenure, and commuting zone tenure.

As estimation using individual level panel data is computationally infeasible, we estimate the model in three steps using data aggregated to the migration flow-year by residence and commuting zone tenure level. Following the logic in Arcidiacono and Miller (2011), we use the finite dependence property of the model that comes through the the resetting of residence and, for long-distance moves, commuting zone tenures to 0 for new arrivals to each

neighborhood. Parameters governing variable moving costs and substitution elasticities are estimated in a first step using an iterative high dimensional fixed effects Poisson pseudo maximum likelihood (PPML) estimator. Destination fixed effects in this migration gravity estimation equation identify the conditional values of each location. The PPML estimator implements the Hotz and Miller (1993) inversion without requiring explicit smoothing like that employed in Bayer et al. (2016) while accommodating Dingel and Tintelnot (2020)'s guidance on the handling of sparse migration flow matrices. Using the estimated conditional values from this first step, we build the composite conditional (inclusive) value of moving from each origin destination-year and tenure status. This can be differenced over time to calculate a composite outside option flow utility given migration.

The second step of estimation uses the fact that the move-stay decision in period t relative to $t+1$ depends on the neighborhood flow utility, including the zip code level Zillow rent index, and outside opportunities, as captured by this composite. The renewal process attribute of the model allows us to difference off unknown continuation values, representing the dynamic move-stay decision in a simple regression equation with an expectation error. We instrument for rents using a novel shift-share instrumental variable: the interaction of neighborhood-level housing supply elasticity estimated in Baum-Snow and Han (2024) and the overall trend in rental prices. The coefficient on the differenced value of moving in this step identifies the parameter governing the distribution of own residence shocks relative to the outside option of moving, which summarizes residence separation shocks. Finally, in the third step of estimation we recover the fixed moving cost (for moves of 0 distance) using the rate of moves across homes within neighborhoods.

Dynamic model parameter estimates reveal a number of revealing patterns. First, moving costs are much higher between than within commuting zones, and the between commuting zone moving cost increases with commuting zone tenure. Second, younger individuals have lower moving costs and lower income individuals have higher moving costs. Third, larger population commuting zones have higher moving costs within them but lower moving costs to them. Fourth, neighborhoods in individuals' own commuting zones are poorer demand substitutes in larger population commuting zones, though substitution elasticities between neighborhoods in other commuting zones do not depend on their populations. Individuals gain utility from remaining longer in the same home and commuting zone, though the utility boost is much higher for long commuting zone tenures. Consistent with evidence of expenditure shares on housing, we estimate the marginal utility of log rent to be about -0.4, with OLS estimates and those not accounting for outside options biased very much upwards.

Estimation of a static version of the model by PPML delivers patterns of moving costs and marginal utilities that are consistent with dynamic model estimates. This is true even though the marginal utilities of rents are estimated through migration destination choice in

the static model rather than the move-stay decision in the dynamic model.

Using the model, we account for the extent to which growth in neighborhood rents has been burdensome for residents. In commuting zones without alternative low rent neighborhoods available, we see that growing rents has large negative impacts on welfare. We use various counterfactual analyses to evaluate the efficacy of various policies that enhance the availability of affordable housing. Considering the effects on welfare through associated changes in housing rents and migration patterns helps to evaluate the relative welfare benefits of demand-side rental subsidies (as in Section 8 housing vouchers) versus supply-side construction subsidies (as in the Low Income Housing Tax Credit) or relaxed land use planning and inclusionary zoning policies. Moreover, our estimated model allows us to evaluate the extent to which targeting such policies to specific neighborhoods is welfare-enhancing. Counterfactual analysis using our estimated model also allows us to assess the efficacy of place based policies that enhance the desirability of living in low cost housing markets or of reducing migration costs. Ours is the first study to do so in the context of an estimated dynamic model with many locations, rich substitution patterns across neighborhoods, and migration costs.

We contribute to the extensive literature on housing affordability Molloy (2020). This paper distinguishes itself by focusing on renters as the most vulnerable group to housing shocks, an aspect that has received less attention in existing studies (Diamond et al., 2019). As rents have risen less quickly than home prices in supply constrained markets (Molloy et al., 2022), a focus on the welfare burdens of changes in rental affordability is also informative for higher income populations who may become newly dissuaded from purchasing a home. Additionally, our study aligns with Gyourko et al. (2013) on the migration of low-income households from high-price locations and Ganong and Shoag (2017) on the relatively low long-distance migration rates of low-skilled workers, even those that are exposed to rising housing costs. Finally, we draw inspiration from a large literature on dynamic location choice models to estimate our structural model (Bilal and Rossi-Hansberg, 2021; Bayer et al., 2016; Almagro and Domınguez-Iino, 2019; Couture and Handbury, 2020; Schubert, 2021).

2 Data

We use Infutor marketing data, CoreLogic housing assessment data, Zillow Observed Rent Index (ZORI) data, American Community Survey (ACS) micro data (Ruggles et al., 2020) and 5 year neighborhood aggregates of ACS data for this analysis. At this point, we focus our analysis on the state of Texas only but plan to expand the analysis to the entire United States in the future. Texas is a convenient benchmark as it has both a large and heterogeneous set of commuting zones and neighborhoods and little rental market regulation.

The ZORI is our main source of information about market rents. This is a "repeat rent" index of asking rents constructed using data from all rental listings on the Multiple Listing Service plus some additional proprietary listing data collected by Zillow. It closely follows the methodology in Ambrose et al. (2015), with the addition of weights to reflect the age and units in the structure composition of the rental stock in each year. We note that actual paid rents for new units may lag this index by up to several years as lease renewals are not included in the underlying data.

As the ZORI data comes at the zip code level of geography, we are constrained to use neighborhoods that are geographically defined as zip code aggregates. The neighborhood is the core spatial unit in our analysis. We employ an algorithm that combines adjacent zip codes into neighborhoods with populations with a minimum target of 6,000 residents in 2010. Within each metropolitan area (CBSA), the algorithm combines adjacent small municipalities into neighborhoods, assigns each medium sized municipality to be its own neighborhood, and partitions large municipalities into multiple neighborhoods. As zip codes vary greatly in size, with many that are very large, we end up with a median population of 20,000 in 1,029 Texas neighborhoods. Of these, 435 are in the large city size category that includes the Houston and Dallas-Fort Worth CBSAs (total population 14.9 million), 180 are in the medium size category of Austin, San Antonio and El Paso (total population 5.7 million), and the remainder are in the small/rural size category (total population 8.7 million). Using the 2000 census and subsequent waves of 5-year ACS micro data aggregated to the block group level, allocated using land area defined weights, we construct information about the housing market and demographic characteristics for each neighborhood over time. Such measures include distributions of gross rents, rent to income ratios, age, education and income. Commuting zones conform to 2020 CBSA definitions and the state remainder.

We spatially aggregate the Zillow rent index across neighborhood zip codes in each year using the number of rental units from CoreLogic as weights. In any given year, we do not observe rents for between 12 and 22 neighborhoods.[1] Missing neighborhoods have very few rental units. We use a version of the data that covers the 2010-2020 period.

We utilize the comprehensive Infutor marketing data set as our primary data source to investigate migration patterns. This data set incorporates information on the residential addresses of US households, including their ten most recent addresses. It covers most US residents back into the 1990s, including variables indicating the date from which individuals are followed and after which they are no longer followed. Infutor compiles these data from publicly available sources, including change of address requests from the Post Office, credit card records, property deeds, public education records, and state criminal record databases. Although primarily marketed for customer relationship management purposes in the pri-

[1] In most years, 952 Texas zip codes have a rent index value calculated.

vate sector, the Infutor data have proven valuable in other recent academic work, notably including Diamond et al. (2019).

Our study requires us to identify renters in the Infutor data. To do so, we match each individual's last home address in Infutor to that in the CoreLogic home assessment data. The resulting match rate is 96%, of which X% are fuzzy matches using addresses of housing complexes that are missing unit numbers or that match imperfectly in other ways. CoreLogic includes a variable that indicates with some level of certainty whether the resident of a housing unit owns or rents, based on whether the property tax bill is sent to a different address. Based on this variable and the last observed address in Infutor, we assign a probability of 0 to 1 to each individual of whether they are a renter. To estimate our dynamic location choice model we calculate the movement matrix across neighborhoods by interacting the total number of movers and the probability that each individual is a renter. The result is a data set of 7.7 million renters in Texas for the 2010-2020 period on a base of 25.1 million residents in 2010.[2]

Using the Infutor data, we build a transition matrix out of each neighborhood n in each commuting zone l in each year t by home and commuting zone tenures ι^n and ι^l. To control for the type of households Θ we look at their origin neighborhood characteristics like income and age. Destinations can be remaining in the same home or different homes in the same neighborhood, any other neighborhood in TX, or outside TX as an outside option. Greater flows in this matrix indicate some combination of lower pairwise inter-regional migration costs and a greater value associated with the destination location in year t.

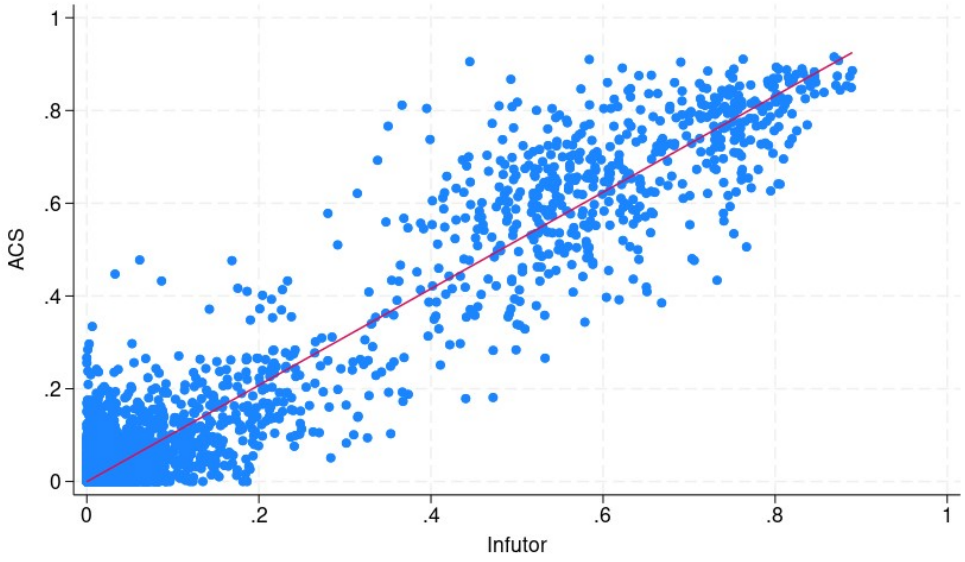
To corroborate Infutor's tracking of the distribution of origin-destination migration pairs conditional on moving, we compare this flows matrix to that in the ACS data for each year 2012-2019.[3] The ACS reports the Migration PUMA, of which there are 71 in Texas, in which each individual lived one year prior to being surveyed. To make comparisons, we aggregate Infutor flows to be between these 71 spatial units. To compare the two data sets, we calculate the fraction of within Texas out-migrants from each migration PUMA in each year migrating to each destination. This distribution of conditional choice probabilities (CCPs) in Infutor contains the key empirical moments used in estimation to recover moving costs and renewal values associated with each neighborhood.

We do not expect these pairwise flows to perfectly align for two reasons. First, the ACS is only a 1% random sample of the population, making its migration flow measures noisier than Infutor's. Second, the Infutor data reports actual moves whereas the ACS data only reports residential locations one year apart. Nevertheless, we see a strong relationship

[2] 35 percent of housing units in Texas were renter-occupied in 2010. Given the total population growth in Texas to 29.1 million in 2020 and our count of 18.0 million owners in the Infutor data, we conclude that the Infutor data misses some renters and we have likely mis-classified a small share of renters as owners.

[3] We begin in 2012 as this is the first year that 2010 census geographies are used in the ACS.

Figure 1: Comparison of Renter Migration Flows in the ACS and Infutor



Note: Points show conditional choice probabilities of migration destinations for renters in the ACS and Infutor data sets in each year 2012-2019. Each observation is a Migration PUMA pair for a given year.

between migration flow measures in these two data sets. Among the 40,328 observations of neighborhood pairs in each year 2012-2019, the correlation is 0.9 and is robust to year, origin, and destination fixed effects. Given measurement error in ACS flows, we expect this correlation to be less than 1. Figure 1 presents a scatterplot comparing these CCPs in each data set.

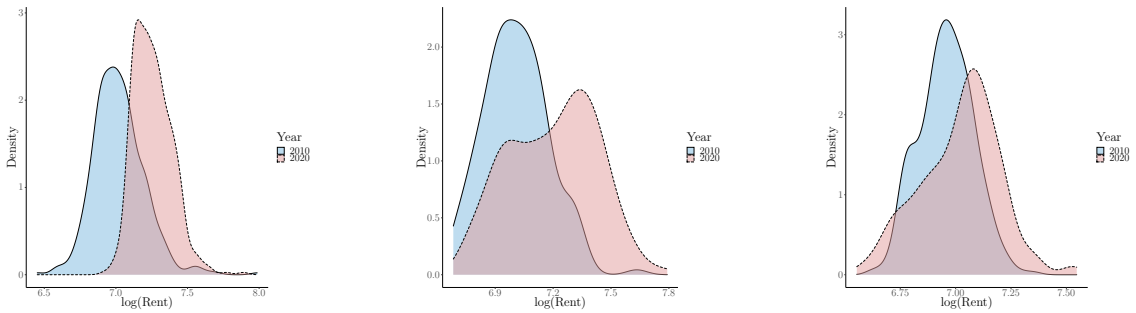
3 Descriptive Evidence

This section has two goals. First, we document some stylized facts that guide key modeling decisions used in the following section. Qualitative insights documented in this section also point to some of the qualitative conclusions that emerge after model estimation. In particular, our evidence of rising rents in almost all neighborhoods of large CBSAs married with continued positive net migration of renters to these metros indicates rising amenity values for these locations. Moreover, the high mobility rates of renters within but not between metros will be rationalized in the model as a combination of differences in moving costs and neighborhood substitution elasticities. At the end of this section, we build up our identification strategy for isolating variation in neighborhood rents, which is key to credible estimation of utility function parameters in the model.

3.1 Patterns of Rents and Affordability

Figure 2 provides insight by examining each of the three types of housing markets: large, medium and small size CBSAs. Over 10 years, the rent distribution shifts to the right across all types of CBSAs with a marked increase in the rent level, particularly so for large and medium size CBSAs, but makes a small change in rent dispersion across markets within each type. A striking feature of the data is that rental prices in the more affordable neighborhoods within large CBSAs have risen significantly, and large CBSAs have very few low rent neighborhoods left after 10 years. Figure 2-a shows this with its kernel density plot of the rent distribution in Dallas and Houston in 2010 and 2020: neighborhoods with rent below the median in 2010 no longer existed in 2020.

Figure 2: Rent Distribution Shift Over Time



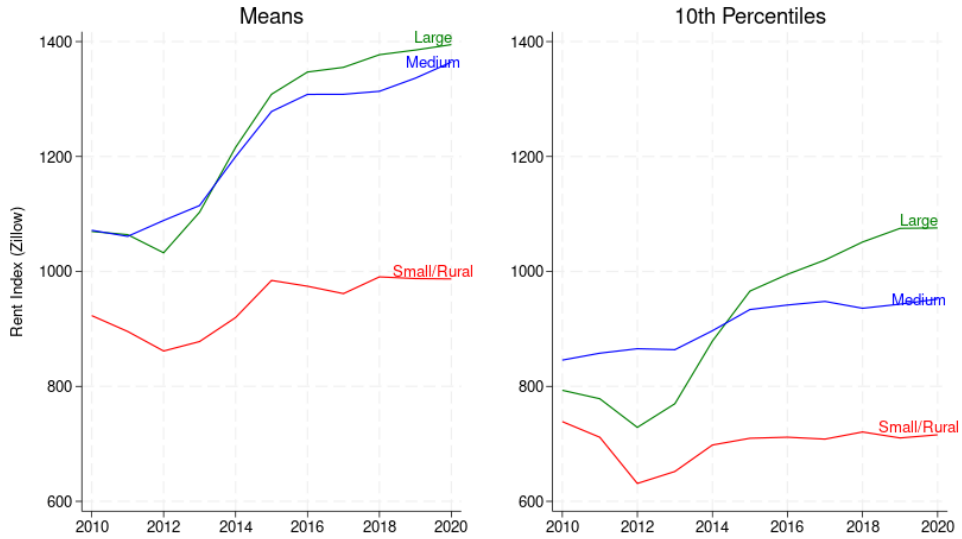
(a) Large CBSAs (b) Medium CBSAs (c) Small CBSAs

Note: Large CBSAs include Dallas and Houston and medium size CBSAs are Austin and San Antonio, and other CBSAs are cathegorized as small CBSAs.

Looking at the rental prices variations over time, Figure 3 presents patterns in Zillow rent indices and average gross market rents experienced by the average (left panel) and 10th percentile (right panel) renter across rural areas/small CBSAs, medium sized CBSAs, and large CBSAs since 2010. Medium and large CBSAs have experienced marked increases in rents since 2012. However, small CBSAs and rural areas remain relatively affordable. Moreover, while 10th percentile neighborhoods in medium sized CBSAs are still affordable relative to 2010, even 10th percentile neighborhoods in large CBSAs experienced rapid rent growth since 2012.

A closer look at the data provides preliminary evidence to this question. Figure A2 shows that, in large CBSAs, over 80% of the population stayed within the same CBSA (as opposed to 70% in medium size CBSAs and less than 50% in small CBSAs); 8% moved to small/rural areas in Texas, and another 8% moved to other states. Two patterns stand out. First, mobility across CBSAs is very limited. This is particularly so for large and medium size

Figure 3: Patterns of Rents 2000-2020



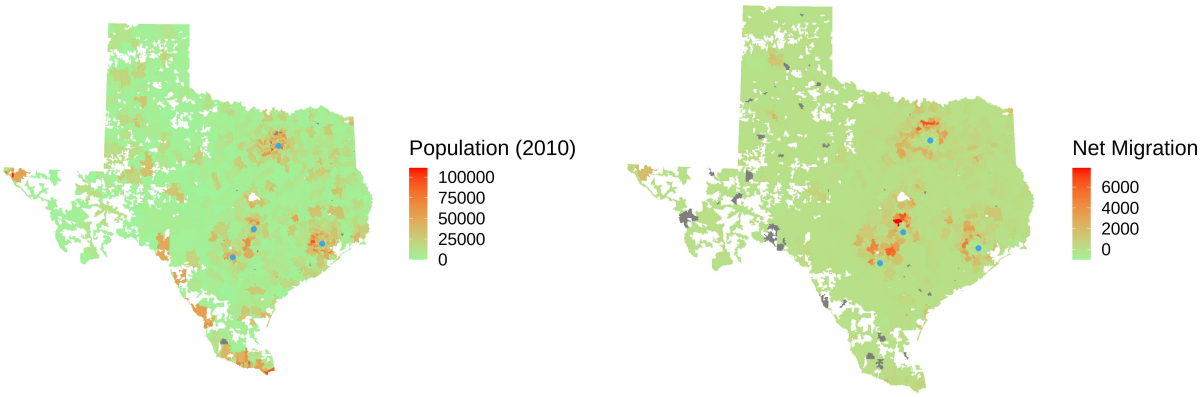
Note: Lines depict the Zillow rent index (ZORI) averaged across zip codes in each indicated CBSA size category weighted by the number of rental units in each year as measured in the ACS. The Zillow rent index is calculated relative to 2020 observed gross rents. All rents are in current (undeflated) dollars. The left panel shows weighted averages across neighborhoods. The right panel shows indexed and mean gross rent for neighborhoods in the (weighted) 10th percentile of each rent distribution.

CBSAs, despite their expensive housing cost. Second, among those who stayed within the large CBSA, lower-income renters are particularly immobile despite the disproportionately high rent burden they face. In Houston, for example, 67.8% of renters with below-median-income in 2010 stayed in the same residence by 2020; 23.06% moved within the same zipcode, and 9.14% moved to other zipcodes within Houston. These numbers are 64.72%, 27.92% and 7.36%, respectively, for Dallas.

3.2 Geographic Patterns

Figure 4 shows the spatial distributions of population and net in-migration in TX. It shows a higher concentration of movements towards Texas' four largest CBSAs (Dallas, Houston, Austin, and San Antonio). This figure highlights the higher movements towards urban peripheries and their suburban areas where we expect higher elasticity of rental unit supply.

Figure 4: Population and Migration Distribution



(a) Texas Population Map

(b) Texas Net Migration Map

Note: Net in-migration is greatest in large (Dallas, Houston) and medium size (Austin, San Antonio) CBSAs shown with blue dots on the maps.

4 Static Model

The full dynamic model laid out in the following section features a series of nested choices. First, individuals choose whether to stay or migrate, while anticipating the expected values of moving or staying. For agents that decide to migrate, they must decide whether to migrate to another house in the same CBSA or to a different CBSA, each of which has its own nest. The crucial predictor variable of interest for us is neighborhood rent. As such, the key identification challenge is to isolate variation in neighborhood rents that are uncorrelated with unobserved demand factors or shocks.

Here we explore the dynamic model's static counterpart. Doing so has two benefits. First, it allows us to focus on identification less opaquely than in the dynamic model. Second, it establishes some baseline results about the relationships between rents and migration decisions, confirming the upward bias in estimates that are not corrected for unobserved heterogeneity. In a first exercise, we demonstrate that housing costs in potential destination neighborhoods influence location choices conditional on migrating. In a second exercise, we show that origin neighborhood rents influence the decision to migrate. Coefficients on rent in both exercises are biased upwards because of unobserved heterogeneity. Neighborhoods that are more attractive tend to have higher rents, and an IV strategy is needed to correct

this endogeneity problem.

Estimates reported in this section can be viewed in the context of the following standard static model of migration, which is a simplified version of the full dynamic model developed in the following section.

Individual i has the indirect utility

$$\tilde{u}_{it}(l,n) = \alpha \ln R_{lnt} + \omega_{lt} + \phi_{ln} + \nu_{lnt} + \epsilon^i_{lnt} \equiv v_t(l,n) + \epsilon^i_{lnt}$$

associated with living in neighborhood n of commuting zone l in year t. With this formulation, an implicit assumption is that households can flexibly adjust the quantity of housing services they rent in each period. Following Davis et al. (2021), rent enters in logs to provide more flexibility in estimation and to be consistent with Cobb-Douglas utility. Utility is also influenced by commuting zone-year and neighborhood fixed effects, which reflect time varying commuting zone and static local amenities, respectively. For completeness, we also include the common unobserved neighborhood-year specific ν_{lnt}, which can be thought of as a neighborhood amenity shock. The individual level i.i.d. shock ϵ^i_{lnt} is distributed with extreme type-I distribution with the dispersion parameter σ.

Migrating to neighborhood n' has the utility cost

$$\tau_{nn'} = \tau_{0n} + \tau_{dn} \ln d_{nn'} \mathbf{1}(l' = l) + \tau_{dl} \ln d_{nn'} \mathbf{1}(l' \neq l),$$

where $\ln d_{nn'}$ is the log distance between the two neighborhoods. We allow the moving cost function to differ flexibly for moves within versus between CBSAs. The moving cost is 0 for non-movers and it is the centripetal force that tends to keep people from moving longer distances.

4.1 Destination Based Results

The first exercise studies how the destination rental prices affect movers' location choices.

Aggregating the individual utility at the ZIP code level, we can write the number of movements from neigbhorhood n' conditional on living origin neighborhood n as follows:

$$N_{n \to n'}(t) = s_{nt} \frac{\exp\left(\nu_t(l',n') - \tau_{nn'}\right)/\sigma}{\sum_{l''} \sum_{n'' \in l''} \exp\left(\nu_t(l'',n'') - \tau_{nn''}\right)/\sigma} \tag{1}$$

where s_{nt} is the population of origin neighborhood.

Putting these elements together, the number of movers from n to n' in year t takes the

following standard regression form:

$$N_{n\to n'}(t) = \exp\left(\frac{\alpha}{\sigma}\ln(R_{kt}) + \frac{1}{\sigma}\tau_{nn'} + \frac{1}{\sigma}\omega_{l't} + \frac{1}{\sigma}\phi_{l'n'} + \lambda_{lnt}\right)\mu_{nn't}, \tag{2}$$

where $\mu_{nn't}$ is the error term. Origin neighborhood–time fixed effects λ_{lnt} controls for origin neighborhood population and the denominator in equation (2), meaning that identifying comparisons are between rents in different destinations. Destination CBSA-time fixed effects $\omega_{l't}$ control for the general attractiveness of each CBSA in each year.[4] Destination neighborhood fixed effects $\phi_{l'n'}$ are neighborhood static amenity contributions to utility. This set of fixed effects identifies variation in $\ln R_{l'n't}$ to be within CBSA and over time. Note, that the utility parameters are normalized by the dispersion parameter σ and we do not separately identify the rent coefficient in the static model.

Many origin-destination neighborhood pairs exhibit few or zero movements in our data. One common approach to address this issue is to kernel smooth the estimated probabilities of movements across neighborhoods (Diamond et al. (2019), Schubert (2021), Almagro and Domınguez-Iino (2019), Bayer et al. (2016)). Smoothing tends to make neighborhoods look more similar, and can thus lead to biases, especially when dealing with a large number of neighborhoods and low probabilities of movement, which is the case in this study. To overcome the zero movement problem, we include all zeros and employ Poisson pseudo-maximum likelihood (PPML) to estimate parameter clusters in (2) within and across metropolitan areas (Guimaraes et al., 2003; Dingel and Tintelnot, 2020).

The identification challenge is to find variation in $\ln R_{lnt}$ that is uncorrelated with ν_{lnt}. We use an instrumental variable that is correlated with the rental prices only through the supply of the rental units. We follow the control function approach in Lin and Wooldridge (2019) and estimate the PPML regressions with an IV for rent. The control function approach performs in two steps. In the first step we regress the endogenous variable (rent) on the IV to obtain the predicted values and recover the prediction errors. The second step performs the PPML regression and includes the prediction error from the first step as a control to account for endogeneity. The consistent standard errors in the second step are recovered through a bootstrapping procedure.

4.1.1 Instrumental Variable for Rent and the First Stage Results

We utilize an instrumental variable to identify the causal effect of rental prices. The need for an instrument arises due to the presence of other unobserved variables that simultaneously

[4] Structural interpretations are λ_{lnt} equals log origin neighborhood population $s_{lnt} - \ln\left(\sum_{l'}\sum_{n''\in l'} e^{(v_{it}(l',n'')-\tau_{nn''})/\sigma}\right)$, $\omega_{l't}$ controls for the time varying commuting zone amenities, and $\phi_{l'n'}$ is static neighborhood amenities.

affect the demand for a specific neighborhood and rental prices. For example, changes in neighborhood amenities could impact both the utility of residing in a neighborhood and the rental prices in the same direction. Hence, we require an instrument that is correlated with rental prices in a given ZCTA but only through the supply side.

We use a shift-share instrument based on the interaction between the supply elasticity of neighborhoods and the overall trend in TX rental prices. The exposure component of the IV is the local housing supply elasticity, as measured in Baum-Snow and Han (2024). The time-varying component of the IV is the logarithm of average rental prices in Texas. The average rental prices are calculated as the weighted average of rents across all Texas ZIP codes, with weights based on the number of housing units in each ZIP code.

The relevance of the instrument comes from the assumption that, in response to an increase in overall demand for rental units, high-elasticity neighborhoods build more housing, resulting in a smaller increase in rental prices compared to low-elasticity neighborhoods. The identification assumption is that the IV affects rental prices only through differences in the neighborhoods' supply-side responses.

Table 1: First Stage Results for Rents

	Dependent Variable:			
	Log Rent			
	(1)	(2)	(3)	(4)
IV	-0.36^{***}	-0.35^{***}	-0.36^{***}	-0.36^{***}
	(0.07)	(0.07)	(0.08)	(0.08)
Controls	–	ZCTA Controls	Dist. to CBD - Year Dummy	Dist. to CBD - Year Dummy
Orig. ZCTA - Year FE	Yes	Yes	Yes	Yes
Dest. CBSA - Year FE	Yes	Yes	Yes	Yes
Dest. ZCTA FE	Yes	Yes	Yes	Yes
Clustered	Yes	Yes	Yes	Yes
IV F.Stat	27	25	22	20
Subset	Movers	Movers	Movers	All
Observations	7,538,672	7,538,672	7,353,092	7,360,224

Note: *p<0.1; **p<0.05; ***p<0.01

Note: This table shows the first stage linear regression of the log of rent on the instrument variable. The standard errors are clustered at the ZCTA group level. The ZCTA controls in the second column are time varying ZCTA income, population and the number of units based on the five years ACS data. Collumn (3), (4) control for distance to CBD times year dummies. The last column includes movers and stayers but the data in other columns are restricted to only movers. The IV F. Stat is the squared t-value of the IV coefficient.

Table 1 presents the first-stage regression of rental prices on the IV. To maintain consistency with the second-stage regression, which analyzes the movement matrix between origin ZCTA groups and destination ZCTA groups, the first-stage regression also preserves the origin-destination-time data structure. The dependent variable is rental prices in the

destination ZCTA group.

In the next section we show that the IV performs well when we study the destination choices of movers. However, it fails to fully address the endogeneity problem when applied to the stay-or-move decision of households within a static model. To overcome this limitation, we introduce a dynamic model in section 5 and show how the IV can be utilized more effectively for identification.

4.1.2 Second Stage Results

We follow the control function approach as in Lin and Wooldridge (2019) to perform the PPML regression with the instrumental variable. Table 2 shows the baseline results of the PPML regressions, which are equivalent to estimating equation (2). The regressions are controlled for the Origin Neighborhood-Time, Destination CBSA-time, and the Destination Neighborhood fixed effects. The Origin Neighborhood-time controls for the origin population, and the average value of migration are shown in the denominator of equation (1). Destination CBSA-time fixed effect controls for variations in the attractiveness of a CBSA, such as variations in the local labor market. Finally, the Destination neighborhood fixed effect controls for a neighborhood's static amenities and attractiveness.

The rent coefficients in columns (1) and (2) are biased toward zero because variation in rent is correlated with unobserved variations in local amenities. Additionally, reverse causality between migration and rent further biases the estimated coefficients. Columns (3) and (4) address these biases using the control function approach and PPML-IV regression. Including the prediction error from the first-stage IV regression in the second stage shows the rent coefficient which is equivalent to the rental price elasticity of migration is -2.7. Looking at the structural interpretation of the rent coefficient in equation (2), we see that the disutility from rent (α) cannot be separately identified from the dispersion parameter (σ). Thus, the rent coefficient represents the normalized disutility of rent divided by the dispersion parameter. To provide context, if the rent coefficient is -0.40[5], the dispersion parameter would be approximately $0.4/2.7 \simeq 0.15$.

[5]The average share of rent from total income.

Table 2: Static Model PPML Regressions

Dependent Variable:	Number of Migrants			
	PPML-OLS		PPML-IV	
Model:	(1)	(2)	(3)	(4)
Variables				
Log Rent	-0.44***	-0.40***	-3.3***	-3.4***
	(0.07)	(0.05)	(0.51)	(0.54)
Log Distance		-0.51***	-0.51***	
		(0.002)	(0.002)	
Stage 1 Residual			2.9***	3.0***
			(0.51)	(0.55)
Log Distance (Same CBSA)				-0.39***
				(0.004)
Log Distance (Other CBSA)				-0.61***
				(0.003)
Fixed-effects				
Destination ZCTA Group	Yes	Yes	Yes	Yes
Origin ZCTA Group - Time	Yes	Yes	Yes	Yes
Destination CBSA - Time	Yes	Yes	Yes	Yes
Fit statistics				
Observations	7,538,672	7,538,672	7,538,672	7,531,360
Pseudo R^2	0.26632	0.69153	0.69154	0.75302

Clustered (Destination ZCTA Group) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table reports the PPML regressions. The dependent variable is the number of migrants from the origin ZCTA to the destination ZCTA in each year. Columns (1) and (2) present the correlations, while columns (3) and (4) show the PPML with IV regressions. Stage 1 Residual is the error of prediction from the first stage. "Distance" measures the distance between the origin and destination, and "Distance (Same CBSA)" is the distance interacted with a dummy variable equal to 1 if the move occurs within the same CBSA.

Column (3) in table 2 shows that a 10% increase in rent reduces the migration inflow by 34%. Moreover, the table shows that the distance between neighborhoods affects the moving cost, and a 10% increase in distance reduces the migration inflow by 5.1%. However, column (4) shows that the effect of distance varies within and across CBSA migrations, with higher costs across CBSA movements.

The simple estimation of moving cost might mask substantial variation across households. To begin unpacking this heterogeneity, Table 3 repeats the PPML-IV regressions in column 4 of Table 2 with the addition of a set of interactions between, pairwise distance, and average income, share of young households, and CBSA population of the orig neighborhood. Column 4 of Table 3 shows that, with these additional controls, the migration responds to the destination neighborhood rent and the distance to the same extent as before. However, it suggests households moving from neighborhoods with a higher share of low-income renters, older population have higehr moving costs. This is not surprising because richer people have more resources to cope with migrations and are expected to have a lower cost of movements. Also, younger people are more mobile and have less ties to their neighborhoods compared to

older households. Also, the interaction of distance with the destination CBSA population suggests both moving within and to larger CBSAs are more costly.

Table 3: PPML Regression with Heterogeneity in Moving Cost

Dependent Variable:	Number of Migrants			
Model:	(1)	(2)	(3)	(4)
Variables				
Log Rent	-3.6***	-3.6***	-3.3***	-3.5***
	(0.53)	(0.54)	(0.50)	(0.49)
Stage 1 Residual	3.2***	3.2***	3.0***	3.1***
	(0.54)	(0.54)	(0.51)	(0.50)
Log Distance (Same CBSA)	-0.38***	-0.40***	-0.34***	-0.35***
	(0.004)	(0.004)	(0.002)	(0.002)
Log Distance (Other CBSA)	-0.61***	-0.61***	-0.60***	-0.61***
	(0.003)	(0.003)	(0.002)	(0.002)
Log Distance * Low Income Share (Same CBSA)	0.13***			-0.11***
	(0.02)			(0.02)
Log Distance * Low Income Share (Other CBSA)	-0.01			-0.16***
	(0.02)			(0.01)
Log Distance * Young Share (Same CBSA)		0.31***		0.12***
		(0.04)		(0.02)
Log Distance * Young Share (Other CBSA)		0.23***		0.16***
		(0.03)		(0.02)
Log Distance * CBSA Population (Same CBSA)			-0.07***	-0.08***
			(0.001)	(0.002)
Log Distance * CBSA Population (Other CBSA)			-0.01***	-0.03***
			(0.002)	(0.002)
Fixed-effects				
Destination ZCTA Group	Yes	Yes	Yes	Yes
Origin ZCTA Group - Time	Yes	Yes	Yes	Yes
Destination CBSA - Time	Yes	Yes	Yes	Yes
Fit statistics				
Observations	7,524,048	7,524,048	7,345,960	7,338,828
Pseudo R^2	0.72176	0.72268	0.76898	0.74029

Clustered (Destination ZCTA Group) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table presents the results of the PPML regressions. The dependent variable is the number of migrants of each tenure level moving from the origin ZCTA to the destination ZCTA in each year. "Distance" measures the distance between the origin and destination. "(Same/Other CBSA)" indicates that the variable is interacted with a dummy equal to 1 if the move occurs within the same CBSA, and 0 otherwise. "Low Income Share" represents the share of renters with annual incomes below $35,000. "Young Share" refers to the share of renters aged 15 to 35 years. "Dest. CBSA Population" denotes the population of the destination CBSA.

Tables A1 and A2 present similar results as above, but they include both movers and stayers in the regressions. For stayers, the origin and destination neighborhoods are the same, and, as expected, their elasticity with respect to destination rental prices is smaller than that of movers. Consequently, the regression that includes stayers yields a smaller coefficient for rent.

The question now is whether our static analysis can explain the move-or-stay decision of renters and whether the IV can accurately identify the rent elasticity of this decision.

The next section looks into this issue, showing that a dynamic model of location choice is necessary for our identification strategy.

4.2 Understanding Migration Decisions [To be Completed]

Incorporated in the treatment above is the joint choice of whether to migrate and the destination conditional on migrating. In practice, this may not be a joint decision. Individuals may trade off the value of staying in their current neighborhood v_{lnt}^{stay} against the inclusive value of moving v_{lnt}^{move}, an aggregation across utilities associated with moving to all potential destinations net of moving costs.

The fully dynamic model in the following section addresses this issue by having a nested choice problem, in which the top nest is the decision of whether to move or stay. This move decision depends on rents and fixed effects in the origin location as push factors as balanced against the inclusive value of moving, which can be written as follows:

$$v_{lnt}^{move} = \sum_{n' \neq n} \exp\left(v_t(l', n') - \tau_{nn'}\right) \tag{3}$$

This suggests two types of estimates. First is an individual level logistic regression of moving on origin rents, CBSA-year FE, origin nhood FE, and v_{lnt}^{move}. Second is a regression of the ln $\frac{movers_{lnt}}{stayers_{lnt}}$ on these same variables. In both cases, origin rents would be instrumented as above. Inclusion of the value of moving could be left out in robustness, as it may be uncorrelated with rents. The advantage of leaving it in is that then the constant can be interpreted as the fixed moving cost, and the coefficient on this variable is the moving elasticity.

5 Dynamic Model

Evidence in the descriptive section above indicates that migration responses to rent growth tend to operate within CBSAs despite the existence of lower cost housing further afield. Evidence in the reduced form section indicates that while movements to lower rent locations do systematically appear in the data, gravity frictions are quite strong. The model is oriented toward allowing us to distinguish between the following mechanisms that may explain these observations. First, migration costs may deter moving to better opportunity locations. That is, we need estimates of the migration technology that are flexible enough to handle both within and cross CBSA moves by individual type. Second, the availability of alternative neighborhoods that are close substitutes or preferred to current neighborhoods in the same or other CBSAs may be limited. That is, we need estimates of the demand system for

neighborhoods within and between CBSAs. Third, there may be fixed moving costs that are random. These would capture idiosyncracies about landlord-tenant relationships, shocks to housing unit quality, and shocks to family conditions including marriage, divorce and having children. Fourth, individuals may get value out of ties to their current neighborhood, which may develop over time. We use the structure of the dynamic model to measure the magnitudes of these third and fourth forces. Moreover, the dynamic model helps with identification. It allows us to control for the possibility that agents value the future expected option value associated with neighborhoods beyond the flow utility received over the course of a year. This option value may be correlated with our instrument for rents.

We approach these aims by specifying a model that is as flexible as possible while retaining feasibility in estimation. Moving costs are specified to flexibly differ within and between commuting zones as a function of tenure and migration distance. Elasticities of substitution in demand across neighborhoods are specified to flexibly differ within and between different commuting zones of different sizes. We also accommodate local neighborhood amenities and marginal (dis)utilities of rent. As all parameters are potentially indexed by type, we suppress explicit indexing on type for the remainder of this section.

5.1 Preferences and Migration Costs

Expanding on the specification in the prior section, the deterministic component of the flow utility each agent enjoys from living in commuting zone l neighborhood n in year t is

$$u_t(l,n) = \alpha \ln R_{lnt} + \beta_{\iota r}\iota^r + \beta_{\iota l}\iota^l + \omega_{lt} + \phi_{ln} + \nu_{lnt}. \tag{4}$$

Flow utility depends on the neighborhood rent index R_{lnt}, tenure in the home ι^r, tenure in the commuting zone ι^l, a commuting zone-year fixed effect ω_{lt}, a neighborhood fixed effect ϕ_{ln}, and a time varying neighborhood component ν_{lnt}. Relative to the static model formulation, we additionally include controls for tenure in the home and CBSA in order to make sure that the estimated model can separate out tenure effects on preferences from moving costs. We want to allow individuals to gain stronger ties to their homes and CBSAs over time.[6]

Each year, households can stay in their apartment or leave their apartment for another home. In estimation, we select people who remain renters throughout the time we observe them. This allows us to avoid having to incorporate portfolio choice and asset accumulation in the model. The moving cost to a new neighborhood in the same commuting zone, $\tau^s_{nn'|l}(\iota^r)$, depends on the distance between neighborhood n and n' and the time spent living in the

[6]In the context of migration Howard and Shao (2023) added autocorrelation between the error terms to account for the ties to the living location. See Pakes et al. (2021) for an alternative approach to dealing with this issue, in the context of choice of health plan.

home prior to the move, ι^r. The moving cost to a new neighborhood in a different commuting zone is $\tau^o_{nn'}(\iota^l)$ and depends on the distance between the neighborhoods and the time spent living in the origin (l) commuting zone, ι^l. These moving costs are paid immediately after moving. Those that stay in their apartment pay no moving cost. To allow moving cost to change over the life-cycle for other reasons, as in Diamond et al. (2018), we allow the moving cost parameters to vary by age and the income of origin neighborhood. Putting this together, we specify moving costs as

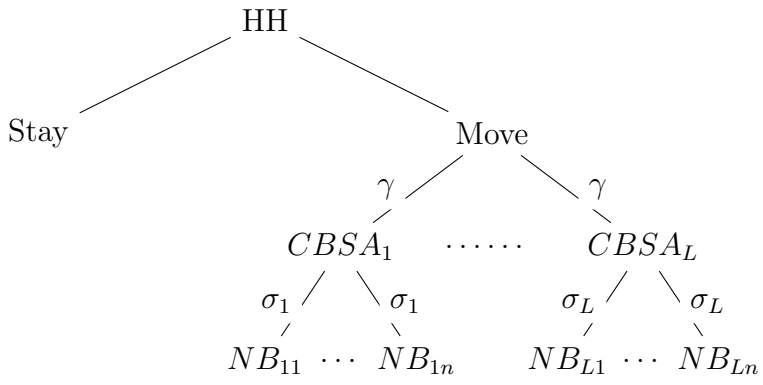
$$\tau_{nn'}(\iota^l, \iota^r) = \begin{cases} \tau^s_{nn'} = \tau_{0ln'}(\iota^l, \iota^r) + (\tau_{dn} + \tau_{\iota r}\iota^r) \ln d_{nn'} & \text{if } n' \in l \\ \tau^o_{nn'} = \tau_{0l'n'}(\iota^l, \iota^r) + (\tau_{dl} + \tau_{\iota l}\iota^l) \ln d_{nn'} & \text{if } n' \in l' \neq l. \end{cases} \tag{5}$$

Compared to Diamond et al. (2018), we allow the moving cost to vary for movements across or within metropolitan areas. We also allow for home and CBSA ties to account for local and broader ties to the living area. We added these flexibilities to our model because moves were within the city of San Francisco in their settings, but our setting incorporates moves between commuting zones.

5.2 Value Functions

We assume a three-level nested logit model for the households location choice[7]. On the top nest households decide to stay or move out of their neighborhood. On the lower level they choose their commuting zone and finally they choose their neighborhood within the commuting zone.

Figure 5: Nesting Structure



The full value function is the flow utility of being in location ln plus a shock plus the discounted expected value of staying in ln or migrating to all alternatives minus any moving

[7] Imbert et al. (2023) use a similar nested structure in a static model.

costs. At the end of each period, agents see each move-neighborhood-location-specific shocks ϵ_{lnrti} for the following period and decide whether and where to move. Following Couture and Handbury (2020), this shock is nested for the commuting zone and neighborhood within the commuting zone conditional on moving. We also add an upper "move decision" nest that imposes a correlation between all shocks associated with living in a different home. There are a total of $N = \sum_{l=1}^{L} N_l$ neighborhoods in the choice set, with each commuting zone l having N_l neighborhoods and N_l shocks, and one more is associated with remaining in the same home. The control variables are commuting zone l' and neighborhood n' next period, which can remain the same if the agent does not move or moves but stays in the same neighborhoods. The state variables are commuting zone last period l, neighborhood last period n, time lived in the current commuting zone and residence, ι^l and ι^r respectively, agent type, and the flow utility shocks.

$$V_t(l, n, \iota^l, \iota^r) = \max\{V_t^{stay}(l, n, \iota^l, \iota^r), V_t^{move}(l, n, \iota^l, \iota^r)\} \tag{6}$$

where the value of staying is:

$$V_t^{stay}(l, n, \iota^l, \iota^r) = u_t(l, n, \iota^l, \iota^r) + \epsilon_{ti}^s + \beta E_t \bar{V}_{t+1}(l, n, \iota^{l\prime}, 1\iota^{r\prime}) \tag{7}$$

where the ex-ante value function $\bar{V}_t(.)$ is as follows:

$$\bar{V}_t(l, n, \iota^l, \iota^r) = \log\{\exp V_t^{stay}(l, n, \iota^l, \iota^r) + \exp V_t^{move}(l, n, \iota^l, \iota^r)\} + \Gamma \tag{8}$$

where Γ is Euler's constant and the value of moving is

$$\begin{aligned} V_t^{move}(l, n, \iota^l, \iota^r) = \max\{ & \\ & \max_{n'|l} [u_t(l, n', \iota^l, 0) - \tau_{nn'}^s(\iota^l, \iota^r) + \epsilon_{ln'ti}^m + \beta E_t \bar{V}_{t+1}(l, n', \iota^{l\prime}, 1)], \\ & \max_{n'|l' \neq l, l'} [u_t(l', n', 0, 0) - \tau_{nn'}^o(\iota^l, \iota^r) + \epsilon_{l'n'ti}^m + \beta E_t \bar{V}_{t+1}(l, n', 1, 1)]\}. \end{aligned} \tag{9}$$

After the flow utility from period t, specified above, the remaining part of the value of living in location (l, n) captures the expected option values of staying in the same home, moving within the commuting zone, and moving across commuting zones, respectively. Each potential destination has an associated moving cost and EV-I shock. The expectation is taken at the beginning of period t before the idiosyncratic flow utility shocks associated with all possible locations choices, ϵ_{lnti}, are observed.

As is common in the migration literature, we specify the shock distribution to be nested

logit with the following cumulative distribution:

$$\exp\left(-\sum_{j\in\{stay,move\}}\left(\sum_{l=1}^{L}\left(\sum_{n\in l}e^{-\epsilon_{ln}^{j}/\gamma\sigma_l}\right)^{\sigma_l}\right)^{\gamma}\right).$$

The upper nest is the move-stay decision with γ that determines the level of correlation between the error term of movers or stayers. γ can be interpreted of as a general "preference" for moving or as a push factor like the rate of exogenous separation from apartments due to landlords non-renewal of leases. A higher γ means lower correlation between unobserved preferences of movers.

Commuting zones are the next nest and neighborhoods within commuting zones are the inner nest. This allows dispersion parameters, and hence elasticities of substitution across neighborhoods, to differ across commuting zones. In particular, the correlation between errors in commuting zone l' is increasing in $1-\sigma_{l'}$, where the restriction $0\leq\sigma_{l'}\leq 1$ and $0\leq\gamma\leq 1$ensures that preferences are well-behaved (Train, 2009).

In estimation, we allow the dispersion parameter σ_l to depend on commuting zone population. The goal is to capture the fact that neighborhoods may have different substitution elasticities in different types of locations. We additionally separate the within and across commuting zone preference shocks. In particular, we have

$$\sigma_{l'}(l)=\begin{cases}\sigma_{l'}^{s} & \text{if } l'=l\\ \sigma_{l'}^{o} & \text{if } l'\neq l\end{cases} \tag{10}$$

5.3 Conditional Migration Probabilities

Without loss of generality, decompose the value function into three components: one that varies within neighborhoods in a commuting zone and another that remains common across all neighborhoods within the same commuting zone and the last part is fixed constant term:

$$V^{stay}(l,n,\iota^{l},\iota^{r})=Y(l,n,\iota^{l},\iota^{r})+W_l(\iota^{l},\iota^{r})+V_0 \tag{11}$$

For representing associated components of continuation values and probabilities, it is convenient to define the following "inclusive values":

$$IV_{lnl't}(\iota^{l},\iota^{r})=\begin{cases}\ln\sum_{n'\in l',n'\neq n}\exp\left[(Y_t(l,n',\iota^{l},0)-\tau_{nn'}(\iota^{l},\iota^{r}))/\gamma\sigma_l^{s}\right] & \text{if } l'=l\\ \ln\sum_{n'\in l'}\exp\left[(Y_t(l',n',0,0)-\tau_{nn'}(\iota^{l},\iota^{r}))/\gamma\sigma_{l'}^{o}\right] & \text{if } l'\neq l.\end{cases} \tag{12}$$

IV_{lnt} captures the full inclusive value of commuting zone l at time t from the perspective of individuals living in neighborhood n of the same commuting zone. This incorporates the moving cost to each other neighborhood in the commuting zone. $IV_{l't}$ summarizes the value of commuting zone l' at time t from the perspective of individuals living in a different commuting zone.

The following probabilities are observed in the data conditional on living in a particular home in commuting zone l for ι^r and ι^l periods of time, respectively. Once conditioning on these objects, these probabilities do not differ across individuals i. Each probability listed below represents the fraction of migrants with residence and commuting zone tenures ι^r and ι^l that make each indicated location decision. These probabilities will be used in estimation. We express the unconditional probability of moving to $l'n'$ as the probability of moving times probability of selecting n' in commuting zone l' conditional on moving.

$$\pi_{lnl'n'} = \frac{\exp[(Y_{l'n'} - \tau_{nn'})/\gamma\sigma_{l'}]}{\sum_{n'\in l'}\exp[(Y_{l'n'} - \tau_{nn'})/\gamma\sigma_{l'}]} \frac{\exp[W_{l'} + \sigma_{l'}IV_{lnl'}]}{\sum_{l'}\exp[W_{l'} + \sigma_{l'}IV_{lnl'}]} \frac{\exp[V_0 + \gamma IV^m_{ln}]}{\exp[V_0 + \gamma IV^m_{ln}] + \exp[V^{stay}_{ln}]} \tag{13}$$

where $IV_{l'n'}$ is the inclusive value of the destination neighborhood defined in equation (12). Also IV^m is the inclusive value of moving defined as follows:

$$IV^m_{ln} = \ln \sum_{l'} \exp[W_{l'} + \sigma_{l'}IV_{lnl'}]$$

Note that the unconditional probability of selecting each neighborhoods in equation (13) depend on the origin and destination state variable. For instance conditional on moving, the probability of moving from neighborhood n to n' within commuting zone l is

$$p_t(l, n'|l, n, \iota^l, \iota^r) = \frac{\exp\left((Y_t(l, n', \iota^l, 0) - \tau_{nn'}(\iota^l, \iota^r))/\sigma_l\right)}{\exp\left(IV_{lnlt}(\iota^l, \iota^r)\right)} \frac{\exp\left(W_{lt} + \sigma_l IV_{lt}(\iota^l, \iota^r)\right.}{\sum_{l'}\sum_{n'\in l'}\exp\left(W_{l't} + \sigma_{l'}IV_{lnl't}\right)} \tag{14}$$

The probability of moving from neighborhood n in commuting zone l to neighborhood n' in commuting zone l' is

$$p_t(l', n'|l, n, \iota^l, \iota^r) = \frac{\exp\left((Y_t(l', n', 0, 0) - \tau_{nn'}(\iota^r, \iota^l))/\sigma_{l'}\right)}{\exp\left(IV_{lnl't}(\iota^l, \iota^r)\right)} \frac{\exp\left(W_{l't} + \sigma_{l'}IV_{lnl't}(\iota^l, \iota^l)\right.}{\sum_{l'}\sum_{n'\in l'}\exp\left(W_{l't} + \sigma_{l'}IV_{lnl't}\right)} \tag{15}$$

In the next section we show how we identify each of these objects in estimation.

6 Estimation Methodology

Estimation of this model follows ideas in Arcidiacono and Miller (2011), Scott (2014), and Diamond et al. (2018), as described in detail in Kalouptsidi et al. (2021). We first construct a likelihood that depends on all parameters of the model. However, because of the high dimensional state space we estimate the model in two stages in a process with some similarities to that in French et al. (2023). First, we estimate conditional values of being in each neighborhood in each year given tenures ι^l and ι^r, moving cost parameters, and utility dispersion parameters by maximum likelihood. In a second stage of estimation, these values are combined using renewal process logic to recover estimates of flow utilities associated with each neighborhood in each year. We then uses IV regressions to recover parameters of these flow utilities.

We note that a commonly used alternative approach is to invert conditional choice probabilities to back out conditional values in the first stage. In our setting, the use of maximum likelihood has the advantage of obviating the requirement to smooth probabilities across the type and state space in the first step. It is made feasible through the fact that the nested logit model aggregates to a Poisson model, which can be feasibly estimated with high dimensional fixed effects.

6.1 Constructing the Likelihood

The likelihood function reflects the probabilities that each agent i transitions to location $\{l', n'\}$ at the end of period t given residing in location $\{l, n\}$. In addition to dependence on prior location, probabilities may also depend on commuting zone and residence tenures, ι^l and ι^r, and model parameters. Using the expressions for probabilities laid out in the prior section, the likelihood contribution of individual i at time t can be expressed as follows. In (16), $\mathbf{1}_{it}(l', n')$ is an indicator that equals 1 if individual i migrates to neighborhood $\{l', n'\}$. All probabilities are as in (14) and (15) above.

$$\mathcal{L} = \prod_i \prod_t \prod_{l'} \prod_{n' \in l'} p_t(l', n' | l, n, \iota^l, \iota^r; \Theta)^{\mathbf{1}_{it}(l', n')} \tag{16}$$

As flow utilities depend on year, all values and hence probabilities are time indexed. However, moving costs are not time indexed. Therefore, estimation is more efficient pooling all years together. However, the Infutor data set is too large to estimate this model using individual level data. Feasible estimation requires inferring parameters through aggregate migration flows across neighborhoods. Grouping the probabilities listed above across

individuals, the log likelihood can be written as

$$\ln \mathcal{L} = \sum_{t} \sum_{\iota^l} \sum_{\iota^r} \sum_{l} \sum_{n} \sum_{l'} \sum_{n'} N_t(l', n'|l, n, \iota^l, \iota^r) \ln p_t(l', n'|l, n, \iota^l, \iota^r). \tag{17}$$

Note that the first four summations are over states whereas the next two are over controlled choices and the last one is over time.[8]

Guimaraes et al. (2003) shows that this aggregated conditional logit model is equivalent to a Poisson model (up to a known constant) and can be estimated by Poisson pseudo maximum likelihood (PPML). In the language of Guimaraes et al. (2003), individuals with the same state vector are in the same "group", and group fixed effects must be included as a control. Dingel and Tintelnot (2020) note the advantages of estimation using PPML in spatial granular settings such as ours.

6.2 Estimation Stage 1: Conditional Values and Substitution Parameters

The first stage uses information on migration within and between commuting zones in a gravity PPML regression to recover parameters in (17). The use of gross flows to recover choice-specific values using PPML in the first stage of dynamic discrete choice model estimation is also seen in Artuç and McLaren (2015). This is an appealing alternative to the conditional choice probabilities (CCP) approach (Hotz and Miller, 1993) in settings like ours that feature a large number of choices and many 0 flows as a result.

The estimated values and migration costs depend on the state variables ι^l and ι^r. So as to limit the state space of the model, we allow ι^l and ι^r to take on three values each. 0 is for a new arrival, 1 is for tenure of between 1 and 4 years, and 2 is for tenure of 5 years and above. Neighborhood choice probabilities for those that remain in the same commuting zone, given by (14) and (15), depend on both current and one period ahead ι^l and ι^r. To reduce computational burden, we assume that $\iota^{l\prime} = \iota^l$ and $\iota^{n\prime} = \iota^n$ except when ι^l or ι^n is 0.

The estimation equation that corresponds to (17) is the following nonlinear gravity regression, which we estimate by PPML (Correia et al., 2020) using an iterative process described below. This equation is derived from (14) and (15):

[8] This aggregation ensures that that all person-years receive equal weight in estimation, just as they would if estimated as a nested logit.

$$
\begin{aligned}
N_t(l', n'|l, n, \iota^l, \iota^r) = \exp\Big[& \\
& \mathbf{1}(l' = l, n' \neq n)[\tilde{Y}_{ln't}(\iota^l, \iota^r) + \tilde{\tau}_{0ln'}(\iota^l, \iota^r) + (\tilde{\tau}_{dn} + \tilde{\tau}_{\iota r}\iota^r)\ln d_{nn'} + \\
& \quad \tilde{W}_{lt}(\iota^l, \iota^r) + (\sigma_l(l) - 1)IV_{nlt}(\iota^l, \iota^r)] + \\
& \mathbf{1}(l' \neq l)[\tilde{Y}_{l'n't}(\iota^l, \iota^r) + \tilde{\tau}_{0l'n'}(\iota^l, \iota^r) + (\tilde{\tau}_{dl} + \tilde{\tau}_{\iota l}\iota^l)\ln d_{nn'} + \\
& \quad \tilde{W}_{l't}(\iota^l, \iota^r) + (\sigma_{l'}(l) - 1)IV_{nl't}(\iota^l, \iota^r)] + \lambda_{lnt\iota^l\iota^r}\Big]\nu_{lnl'n'\iota^l\iota^r t}.
\end{aligned} \tag{18}
$$

We estimate the above equation using regressions of movements from origin to destination neighborhoods for different values of ι^l and ι^r. The regression includes origin-time-$\iota^l - \iota^r$ fixed effects, as well as "destination-time-$\iota^l - \iota^r$ (same/other CBSA)". The destination fixed effects are separately estimated for within and across CBSA movements. Additionally, it incorporates cost function parameters that vary with the distance between the origin and destination neighborhood. It is important to note that the above equations exclude average fixed moving costs because the regression focuses only on movers. As a result, the average fixed cost of moving cancels out in equations (14) and (15). In this expression, each line corresponds to a different class of migration flow. Specifically, $\tilde{Y}_{ln't}(\iota^{l\prime}, 0)$, $W_{ln't}(\iota^{l\prime}, 0)$, and $\tau_{0ln'}$ are captured by "same cbsa destination-time-ι_l-ι^r" fixed effects while "other cbsa destination-time-ι_l-ι^r" fixed effect captures $\tilde{Y}_{l'n't}(0, 0)$, $W_{l'n't}(0, 0)$ and $\tau_{0l'n'}$.

The structural interpretation of the "origin" fixed effects $\lambda_{lnt\iota^l\iota^r}$ is given by $\ln s_t(l, n, \iota^l, \iota^r) - \ln \sum_{l'} \sum_{n' \in l'} \exp\left(W_{l't} + \sigma_{l'} IV_{n'l't}\right)$ where $s_t(.)$ is the population of movers in each state (Guimaraes et al., 2003). This interpretation arises from the condition that probabilities of choosing all destination neighborhoods must sum to one for each combination of ι^l and ι^r. The second term that shows up in the "origin" fixed effect is the denominator of the choice probabilities in equations (14) and (15), $(\sum_{l'} \sum_{n' \in l'} \exp\left(W_{l't} + \sigma_{l'} IV_{n'l't}\right))$, which summarized the inclusive value of moving out of the own neighborhoods.

Each estimated parameter maps to corresponding structural parameters as follows. $\tilde{Y}_{ln't}(\iota^{l\prime}, 0) = Y_{ln't}(\iota^{l\prime}, 0)/\gamma\sigma_l^O$ is the adjusted value associated with moving to a different neighborhood in the same commuting zone. This is one type of renewal process we will use for identification of flow utilities. $\tilde{Y}_{l'n't}(0, 0) = Y_{l'n't}(0, 0)/\gamma\sigma_{l'}$ is the adjusted value associated with moving to a neighborhood in a different commuting zone. This is the other type of renewal process used for identification of flow utilities.

Mechanical relationships between the inclusive values $IV_{l'n't}(\iota^l, \iota^r)$ and remaining parameters are shown in (12). To deal with nonlinearities in values, we estimate parameters using an iterative process. In the first iteration, we include destination commuting zone-time-$\iota^l - \iota^r$ fixed effects for separately for cross-zone and within CBSA movers to control for the two inclusive values. Then we aggregate adjusted values using (12) to form initial

estimates of inclusive values $IV^0_{l'n't}$ and $IV^0_{ln't}(\iota^l, \iota^r)$. $Y_t(l, n', \iota^l, 0) - \tau_{0ln'}$ in equation (12) is the demeaned value of "same CBSA destination neighborhoods-time-$\iota^l - \iota^r$" fixed effect and $Y_t(l', n', 0, 0) - \tau_{0l'n'}$ is the demeaned value of "other CBSA destination neighborhoods-time-$\iota^l - \iota^r$" fixed effect. They are demeaned values because $W_t(l, n', \iota^l, 0)$ and $W_t(l', n', 0, 0)$ are included in the estimated fixed effects and demeaning the fixed effects recovers $Y(.) - \tau_0$. After recovering $Y(.) - \tau_0$, we then estimate equation (18), plugging in these first guesses for inclusive values. Repeating, we update the inclusive values in subsequent iterations until parameter convergence.

The coefficient of inclusive values $(\sigma_l - 1)$ determines the correlation of the error term across neighborhoods within a commuting zone. A higher correlation implies that neighborhoods are more substitutable, and we expect this correlation to vary with the size of the commuting zones. We assume a linear relationship for the dispersion parameter as a function of the destination CBSA size $(\tilde{\sigma}^o(l') = \alpha + \beta \ln Pop_l)$[9]. Where Pop_l is the population of the destination CBSA. The coefficient σ^s_l is specified analogously.

6.3 Estimation Stage 2: Flow Utilities

From Stage 1, we can construct the inclusive value of moving from each neighborhood in each year, IV^{move}_{lnt} from the "origin neighborhood-time-$\iota_l - \iota_r$ and subtracting it from the origin state population $s_t(l, n, \iota^l, \iota^n)$.

Now that all choice-specific value functions have been obtained, we turn to isolating components of flow utilities. To isolating the flow utilities we look at the top nest decision of moving or staying at the original residence.

The nested structure of error terms conclude:

$$Pr^{stay}_t = \left(\frac{\exp V^{stay}_t}{\exp V^{move}_t + \exp V^{stay}_t} \right) \tag{19}$$

Where Pr^{stay}_t is probability of staying in the same residence and $V^{move}_t = V_0 + \gamma IV^{move}_t$. The value of staying in the same location includes the flow utility for the current period and the expected value of the location in the next period $(V^{stay}_t = u_t + \beta E_t \bar{V}_{lnt'}(\iota^{l'}, \iota^{r'}))$. To isolate the flow utility, we need to remove the expected value of the neighborhood for the next period. This can be achieved by subtracting the change in the probability of moving in

[9] We test that $\sigma_{l'} > 0$, which is required in order for neighborhood choice probabilities to be increasing in their values.

the next period from the probability of staying, as follows:[10]

$$\ln(Pr_t^{stay}) - \left(\ln(Pr_t^{move}) - \beta \ln(Pr_{t+1}^{move})\right) = u_t - (1-\beta)V_0 - \gamma_n(IV_t^{move} - \beta IV_{t+1}^{move}) + \xi_t \tag{20}$$

Where, ξ_t is the expectation error and is assumed to have a normal distribution. We calculated this period and the next period inclusive value of moving by recovering the origin-time-$\iota^l - \iota^r$ fixed effects from the stage 1.

Note that the finite dependence property of the model has allowed us to recover common renewal values across all agents. Here, we show how to aggregate these renewal values across neighborhoods so as to residualize off continuation values from flow utilities. Unlike in Arcidiacono and Miller (2011) and Diamond et al. (2018) we use all of the data for this residualization, and not just comparison across origin states in the periods prior to moving to renewal states. This has the advantage of giving us broader based identification of flow utilities, at the cost of using the nested logit distributional assumptions to construct closed forms for continuation values and the value of moving.

The left-hand side of equation (20) is fully observed in the data, measured by the aggregate number of stayers and movers in each neighborhood-year for each tenure level, ι_l and ι_r. Movers are individuals who move out of their residence and stayers are those that have the same addresses in both periods. Following Diamond et al. (2018), we set $\beta = 0.85$. On the right-hand side, we include the flow utility parameters from equation (4) and add the difference between the current and next period's inclusive value of moving to identify the parameter γ_n. The inclusive value of move in this regression is recovered from the "origin neighborhood-time-$\tau_l - \tau_r$" in the first stage PPML regression. To account for differences in neighborhood amenities, we include ZCTA fixed effects. Additionally, we add CBSA-year fixed effects to capture variations in the valuation of CBSAs, such as changes in labor market conditions.

Even after controlling for unobserved average neighborhood amenities and time-varying CBSA fixed effects, the estimation faces the standard concern that rents may still be correlated with time-varying neighborhood amenities in the error term. Following the analysis

[10]

$$\ln(Pr_t^{stay}/Pr_t^{move}) = \ln\left(\frac{\exp V_t^{stay}}{\exp V_t^{move}}\right) = V_t^{stay} - V_t^{move} = (u_t + \beta E\bar{V}_{t+1}) - V_t^{move}$$

$$\ln(Pr_{t+1}^{move}) = V_{t+1}^{move} - \bar{V}_{t+1}$$

$$\implies \ln(Pr_t^{stay}/Pr_t^{move}) + \beta \ln(Pr_{t+1}^{move}) \equiv \ln(Pr_t^{stay}) - (\ln(Pr_t^{move}) - \beta \ln(Pr_{t+1}^{move})) =$$

$$u_t + \beta(E_t\bar{V}_{t+1} - \bar{V}_{t+1}) - (V_t^{move} - \beta V_{t+1}^{move}) = u_t - (V_t^{move} - \beta V_{t+1}^{move}) + \xi_t =$$

$$u_t - (1-\beta)V_0 - \gamma_n(IV_t^{move} - \beta IV_{t+1}^{move}) + \xi_t$$

in Section 3, we utilize an instrumental variable to identify the causal effect of rental prices. The need for an instrument arises due to the presence of other unobserved variables that simultaneously affect the demand for a specific neighborhood and rental prices. For example, changes in neighborhood amenities could impact both the utility of residing in a neighborhood and the rental prices in the same direction. Thus, we need an instrument that is correlated with rental prices but only through the supply side. We adopt a shift-share type instrumental variable, introduced in Section 3, which interacts local housing supply elasticity with the logarithm of average trend in Texas rental prices.

The flow utility regression in equation (20) recovers each neighborhood's utility level up to a normalization, but it does not allow for the separate identification of the fixed moving cost, τ_0. To recover the fixed moving cost, we focus on households that move within the same neighborhood. By relocating, these households incur the fixed moving cost and forgo the utility associated with higher tenure.

$$log(Pr_{nn}^{move}/Pr^{stay})_t = -\tau_0 + \beta_{\iota r}\iota^r + \psi_{nt} \tag{21}$$

Where $\beta_{\iota r}\iota^r$ can be recovered from the flow utility regression above. Thus, averaging over the ratio of moves within the same neighborhood to the number of stays allows us to recover the fixed cost of moving. Intuitively, if the fixed cost of moving is small, we expect a higher number of moves within neighborhoods.

7 Dynamic Model Results [to be completed]

We begin with the regression for movers' location choices, following equation (18). Next, we proceed to stay and move decision and recover the utility parameter from the top nest regression, as specified in equation (20).

Table 4 shows the result of the PPML regression in the first stage. This stage runs a PPML regressions on the movement matrix and finds the moving cost and the correlation of the error terms in the lower nest. The results are comparable with the regression results in the baseline static model in table 2. The main difference is adding the inclusive value of each CBSA in the regression.

Column (4) includes the most comprehensive set of controls. This column shows that distance between neighborhoods adds to the cost of migration, and the effect is larger for movements across CBSAs. Specifically, a 10% increase in the distance between the origin and destination leads to a 3.3% reduction in migration within the same CBSA and a 17% reduction in migration to other CBSAs. Moving out of the CBSA is more costly for older, low-income households with higher CBSA tenure. Similarly, for moves within the same CBSA, older and low-income households face higher costs, though these costs tend to decrease as

households spend more time in their current home.

The coefficient of the inclusive value in our model is represented as $\sigma - 1$, and its absolute value captures the degree of correlation among the error terms within a given nest. The results show that the correlation of the error term between neighborhoods in other CBSAs is higher.

Table 4: Dynamic Model First Stage Regression

Dependent Variable:	Number of Migrants			
Model:	(1)	(2)	(3)	(4)
Variables				
Log Distance (Same CBSA)	-0.41***	-0.41***	-0.42***	-0.42***
	(0.005)	(0.005)	(0.004)	(0.004)
Log Distance (Same CBSA) * Long-term Residence	0.03***	0.03***	0.04***	0.04***
	(0.004)	(0.003)	(0.002)	(0.002)
Log Distance (Other CBSA)	-1.4***	-1.4***	-2.0***	-2.1***
	(0.03)	(0.03)	(0.10)	(0.10)
Log Distance (Other CBSA) * Long-term CBSA	-0.10***	-0.09***	-0.16***	-0.14***
	(0.02)	(0.02)	(0.02)	(0.01)
Log Distance (Same CBSA) * Young Share		0.33***	0.33***	0.32***
		(0.04)	(0.04)	(0.04)
Log Distance (Same CBSA) * Low-income Share		0.15***	0.14***	0.13***
		(0.03)	(0.03)	(0.03)
Log Distance (Other CBSA) * Young Share		0.20***	0.13***	1.3***
		(0.03)	(0.04)	(0.12)
Log Distance (Other CBSA) * Low-income Share		0.04**	-0.05**	-0.44***
		(0.02)	(0.02)	(0.07)
Inclusive Value (Other CBSA)			-0.31***	-0.38***
			(0.05)	(0.05)
Inclusive Value (Same CBSA)			-0.59***	-0.44***
			(0.06)	(0.06)
Inclusive Value (Other CBSA) * Log Dest. CBSA Population				-0.04***
				(0.004)
Inclusive Value (Same CBSA) * Log Dest. CBSA Population				0.10***
				(0.03)
Inclusive Value (Other CBSA) * Young Share				0.46***
				(0.06)
Inclusive Value (Same CBSA) * Young Share				-0.12
				(0.22)
Inclusive Value (Other CBSA) * Low-income				-0.14***
				(0.03)
Inclusive Value (Same CBSA) * Low-income				-0.45**
				(0.19)
Fixed-effects				
Origin Time Tenure	Yes	Yes	Yes	Yes
Destination Time Tenure (same/other CBSA)	Yes	Yes	Yes	Yes
Fit statistics				
Observations	21,339,693	21,339,693	21,174,490	20,968,574
Pseudo R^2	0.77052	0.77214	0.71821	0.71814

Clustered (ZCTA Group) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table presents the results of the PPML regressions. The dependent variable is the number of migrants of each tenure level moving from the origin ZCTA to the destination ZCTA in each year. "Distance" measures the distance between the origin and destination. "(Same/Other CBSA)" indicates that the variable is interacted with a dummy equal to 1 if the move occurs within the same CBSA, and 0 otherwise. "Low Income Share" represents the share of renters with annual incomes below $35,000. "Young Share" refers to the share of renters aged 15 to 35 years. "Dest. CBSA Population" denotes the population of the destination CBSA.

Table 5 presents the results of the utility parameter estimations. Columns (1) and (2) provide utility estimates under the assumption of a static model, where the share of stayers to movers within a neighborhood reflects the utility of the neighborhood in the static

framework.

Column (1) reports regression results without using an instrument for rent. In contrast, Column (2) incorporates the instrument variable (neighborhood supply elasticity of housing interacted with rental trends) discussed in subsection 4.1.1 to identify the rent elasticity of migration. Interestingly, even with the instrumental variable, the rent coefficient has a high variance and is positive for some neighborhoods in the static model. This suggests that our IV may be correlated with the error term in the move/stay regression. The IV captures the supply response to housing demand, which could be correlated with the expectations about the next period embedded in the error term, thereby introducing endogeneity.

Columns (3 – 5) address this issue by removing next-period expectations from the error term. In these columns, the dependent variable is the utility of the neighborhood derived from equation (20). The IV regression, accounting for the dynamic structure of the model, results in utility parameter estimates with a negative rent coefficient, aligning with theoretical expectations.

The coefficient of $V_t^{move} - \beta \times V_{t+1}^{move}$ is $-\gamma$ in our model. A larger absolute value of γ indicates a lower correlation of moving shocks. Column (3) shows a positive coefficient of $\Delta V_t^{move} = V_t^{move} - \beta \times V_{t+1}^{move}$ which suggest possible correlation of the V_t^{move} with unobservable amenity shocks. To address the endogeneity issue, we should use an instrument that is correlated with the ΔV_t^{move} but is uncorrelated with local amenity shocks. So, we form an instrument based on the value of other neighborhoods outside a given neighborhood CBSA. And we use the next period measure to limit possible correlations. Specifically, the instrument is $\log(\sum_{l' \neq l} \exp(W_{l't} + \sigma_{l'} IV_{lnl'}))$ but measured for the next period. Column (5) shows the result of regression with two instruments: one for the rent index and the second one for ΔV_t^{move}.

	Dependent variable:				
	log(Stayers/Movers)		log(Stayer/Movers)$_t + \beta \times log(Movers/TotalPop)_{t+1}$		
	(1)	(2)	(3)	(4)	(5)
Long-term Residence	2.047***	2.049***	0.307***	0.286***	0.286***
	(0.020)	(0.021)	(0.005)	(0.005)	(0.005)
Long-term CBSA	1.924***	1.914***	0.945***	1.022***	1.022***
	(0.012)	(0.013)	(0.011)	(0.011)	(0.011)
log Rent	0.022	−0.197	−0.675	−0.920**	−1.150**
	(0.058)	(0.511)	(0.424)	(0.447)	(0.517)
($V_t^{move} - \beta \times V_{t+1}^{move}$)			0.334***	−0.043***	−0.043***
			(0.007)	(0.016)	(0.016)
log Rent × Low-income Share					−1.124***
					(0.303)
Rent IV	No	Yes	Yes	Yes	Yes
$\Delta V^{move} IV$	No	No	No	Yes	Yes
ZCTA Group FE	Yes	Yes	Yes	Yes	Yes
CBSA-Year FE	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes
Observations	21,380	18,798	18,788	18,788	18,788
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 5: Second Stage Estimates

Note: This table presents the results of regressions of the variable indicated at top on the variables at left. The first two columns map to a static model of migration from the perspective of the origin location. The remaining two columns show different specifications of the second stage of the dynamic model estimation.

7.1 Fixed Moving Cost

The last stage is estimating the fixed cost of moving. One can think of the fixed moving cost as the average cost of the move, which cancels out for movers in equation (13). Given an average rent of approximately $12000 per year and the coefficient of $log(Rent)$ in table **??** to find the dollar equivalent of the moving cost, we should multiply the results by $12000/1.02 \simeq 12000$. Column (1) in table 6 shows that the average moving cost is around $1.7 \times 120000 \simeq 20000$.

	Dependent variable:	
	log(Movers/Stayers) in the same Neighborhood	
	(1)	(2)
Long-term Residence		-2.616^{***}
		(0.016)
Low Income Share		-1.265^{***}
		(0.053)
Young Share		0.327^{***}
		(0.071)
Constant	-1.732^{***}	-0.863^{***}
	(0.011)	(0.010)
Clustered SE	Yes	Yes
Observations	24,252	24,252
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 6: Fixed Moving Cost Estimates

Note: This table presents the results of regressions of the log share of movers within a neighborhood to stayers.

Table 6 uses the regression in (21) to estimate the fixed moving cost. Column (2) shows that the fixed moving cost is higher for low-income and older neighborhoods.

8 Counterfactuals [to be completed]

In our first set of counterfactuals, we will decompose the welfare consequences for a typical resident of each neighborhood in 2010 associated with the 2010-2019 Texas rent growth into a components due to migration costs, the lack of substitutibility across neighborhoods, and spatial correlations in rent growth. Based on parameter estimates, preliminary evidence indicates that those in 2010 low rent neighborhoods in large CBSAs were most exposed to rising rents, with the largest welfare losses.

In our second set of counterfactuals, we will assess the welfare consequences of exogenously adding 1000 units of subsidized rental housing to each neighborhood, similar to the LIHTC program, one at a time. This will reveal the types of neighborhoods that would benefit the most from project based subsidized housing.

8.1 Cross Value Elasticities of Probability of Selecting a Neighborhood

The following elasticities can be used to aid in calculating counterfactual ZCTA populations. Conditional on moving the elasticities are as follows $\eta_{n'n} = \frac{\partial \log(p(n',l'|n,l,\iota^r,\iota^l)}{\partial \log V(n)}$.

$$\eta_{n'n} = \begin{cases} \frac{V(n')}{\sigma_l}(1-\sigma_l\frac{\exp\left((V_t(l,n',\iota^{l'},0)-\tau^s_{nn'}(\iota^r))/\sigma_l\right)}{\exp\left(IV_{lnt}(\iota^l,\iota^r)\right)}\frac{\exp\left(W_{lt}+\sigma_l IV_{n'lt}(\iota^l,\iota^r)\right)}{\sum_{l'}\sum_{n'\in l'}\exp\left(W_{l't}+\sigma_{l'}IV_{n'l't}\right)}- \\ \qquad (1-\sigma_l)\frac{\exp\left((V_t(l,n',\iota^{l'},0)-\tau^s_{nn'}(\iota^r))/\sigma_l\right)}{\exp\left(IV_{lnt}(\iota^l,\iota^r)\right)}) & \text{if } n=n' \;\&\; l'=l\ , \\ -V(n')(\frac{\exp\left((V_t(l,n',\iota^{l'},0)-\tau^s_{nn'}(\iota^r))/\sigma_l\right)}{\exp\left(IV_{lnt}(\iota^l,\iota^r)\right)}\frac{\exp\left(W_{lt}+\sigma_l IV_{n'lt}(\iota^l,\iota^r)\right)}{\sum_{l'}\sum_{n'\in l'}\exp\left(W_{l't}+\sigma_{l'}IV_{n'l't}\right)}+ \\ \qquad \frac{1-\sigma_l}{\sigma_l}\frac{\exp\left((V_t(l,n',\iota^{l'},0)-\tau^s_{nn'}(\iota^r))/\sigma_l\right)}{\exp\left(IV_{lnt}(\iota^l,\iota^r)\right)}) & \text{if } n'\neq n \;\&\; l'=l, \\ -V(n')\frac{\exp\left((V_t(l',n',0,0)-\tau^s_{nn'}(\iota^r))/\sigma_{l'}\right)}{\exp\left(IV_{l'n't}(\iota^l,\iota^l)\right)}\frac{\exp\left(W_{l't}+\sigma_{l'}IV_{n'l't}(\iota^l,\iota^r)\right.}{\sum_{l'}\sum_{n'\in l'}\exp\left(W_{l't}+\sigma_{l'}IV_{n'l't}\right)} & \text{if } n\neq n' \;\&\; l'\neq l. \end{cases}$$

The value elasticity of probability of move is as follows:

$\nu_{n'n} = \frac{\partial \log(Pr_{move})}{\partial \log V(n)} = \partial \log(\frac{\exp V_{move}}{\exp V_{move}+\exp V_{stay}})/\partial \log V(n)$ where $V(n)$ is the same as V_{stay} at neighborhood n.

$$\nu_{n'n} = \begin{cases} -V(n')(1-\gamma\frac{\exp\left((V_t(l,n',\iota^{l'},0)-\tau^s_{nn'}(\iota^r))/\sigma_l\right)}{\exp\left(IV_{lnt}(\iota^l,\iota^r)\right)}\frac{\exp\left(W_{lt}+\sigma_l IV_{n'lt}(\iota^l,\iota^r)\right)}{\sum_{l'}\sum_{n'\in l'}\exp\left(W_{l't}+\sigma_{l'}IV_{n'l't}\right)})Pr_{stay}(n) & \text{if } n=n' \;\&\; l'=l\ , \\ \gamma V(n')\frac{\exp\left((V_t(l,n',\iota^{l'},0)-\tau^s_{nn'}(\iota^r))/\sigma_l\right)}{\exp\left(IV_{lnt}(\iota^l,\iota^r)\right)}\frac{\exp\left(W_{lt}+\sigma_l IV_{n'lt}(\iota^l,\iota^r)\right)}{\sum_{l'}\sum_{n'\in l'}\exp\left(W_{l't}+\sigma_{l'}IV_{n'l't}\right)})Pr_{stay}(n) & \text{if } n'\neq n \;\&\; l'=l, \\ \gamma V(n')\frac{\exp\left((V_t(l',n',0,0)-\tau^s_{nn'}(\iota^r))/\sigma_{l'}\right)}{\exp\left(IV_{l'n't}(\iota^l,\iota^l)\right)}\frac{\exp\left(W_{l't}+\sigma_{l'}IV_{n'l't}(\iota^l,\iota^r)\right.}{\sum_{l'}\sum_{n'\in l'}\exp\left(W_{l't}+\sigma_{l'}IV_{n'l't}\right)}Pr_{stay}(n) & \text{if } n\neq n' \;\&\; l'\neq l. \end{cases}$$

Value elasticity of probability of moving to neighborhood n' from neighborhood n is the summation of $\eta_{n'n}$ and $\nu_{n'n}$. To calculate the population we should consider those who stay in the same neighborhood which makes the calculation of elasticity a little complicated.

9 Conclusions [to be completed]

This paper specifies and estimates a dynamic model of demand for neighborhoods. This model is specified so as to focus on quantifying the burdens of rent growth. These burdens can come in the form of high moving costs to alternative affordable neighborhoods, the availability of poor substitute neighborhoods, and high correlations across neighborhoods in

rent growth.

We find the highest burden of rent growth for residents of low rent neighborhoods in the largest CBSAs. While other neighborhoods in their same CBSA are reasonably good demand substitutes, these residents face high moving costs to other CBSAs and few to no alternative neighborhoods with low rents during our study period. These burdens are lower for younger individuals but higher for lower income individuals. As individuals benefit from remaining longer in the same commuting zone, older indivdiuals face the largest burdens of rent growth in their neighborhoods.

References

Almagro, M. and Domınguez-Iino, T. (2019), 'Location sorting and endogenous amenities: Evidence from amsterdam', *NYU, mimeograph* .

Ambrose, B. W., Coulson, N. E. and Yoshida, J. (2015), 'The repeat rent index', *Review of Economics and Statistics* **97**(5), 939–950.

Arcidiacono, P. and Miller, R. A. (2011), 'Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity', *Econometrica* **79**(6), 1823–1867.

Artuç, E. and McLaren, J. (2015), 'Trade policy and wage inequality: A structural analysis with occupational and sectoral mobility', *Journal of International Economics* **97**(2), 278–294.

Baum-Snow, N. and Han, L. (2024), 'The microgeography of housing supply', *Journal of Political Economy* .

Bayer, P., McMillan, R., Murphy, A. and Timmins, C. (2016), 'A dynamic model of demand for houses and neighborhoods', *Econometrica* **84**(3), 893–942.

Bilal, A. and Rossi-Hansberg, E. (2021), 'Location as an asset', *Econometrica* **89**(5), 2459–2495.

Correia, S., Guimarães, P. and Zylkin, T. (2020), 'Fast poisson estimation with high-dimensional fixed effects', *The Stata Journal* **20**(1), 95–115.

Couture, V. and Handbury, J. (2020), 'Urban revival in america', *Journal of Urban Economics* **119**, 103267.

Davis, M. A., Gregory, J., Hartley, D. A. and Tan, K. T. (2021), 'Neighborhood effects and housing vouchers', *Quantitative Economics* **12**(4), 1307–1346.

Diamond, R., McQuade, T. and Qian, F. (2018), 'The effects of rent control expansion on tenants, landlords, and inequality: Evidence from san francisco', *NBER Working Paper No. 24181* .

Diamond, R., McQuade, T. and Qian, F. (2019), 'The effects of rent control expansion on tenants, landlords, and inequality: Evidence from san francisco', *American Economic Review* **109**(9), 3365–94.

Dingel, J. I. and Tintelnot, F. (2020), Spatial economics for granular settings, Technical report, National Bureau of Economic Research.

French, R., Gandhi, A. and Gilbert, V. (2023), Quantifying the welfare impacts of gentrification on incumbent renters, Technical report, Working Paper.

Ganong, P. and Shoag, D. (2017), 'Why has regional income convergence in the us declined?', *Journal of Urban Economics* **102**, 76–90.

Guimaraes, P., Figueirdo, O. and Woodward, D. (2003), 'A tractable approach to the firm location decision problem', *Review of Economics and Statistics* **85**(1), 201–204.

Gyourko, J., Mayer, C. and Sinai, T. (2013), 'Superstar cities', *American Economic Journal: Economic Policy* **5**(4), 167–99.

Hotz, V. J. and Miller, R. A. (1993), 'Conditional choice probabilities and the estimation of dynamic models', *The Review of Economic Studies* **60**(3), 497–529.

Howard, G. and Shao, H. (2023), The dynamics of internal migration: A new fact and its implications, Technical report, Working paper.

Imbert, C., Monras, J., Seror, M. and Zylberberg, Y. (2023), Floating population: migration with (out) family and the spatial distribution of economic activity, Federal Reserve Bank of San Francisco.

Kalouptsidi, M., Scott, P. T. and Souza-Rodrigues, E. (2021), 'Linear iv regression estimators for structural dynamic discrete choice models', *Journal of Econometrics* **222**(1), 778–804.

Lin, W. and Wooldridge, J. M. (2019), Testing and correcting for endogeneity in nonlinear unobserved effects models, *in* 'Panel data econometrics', Elsevier, pp. 21–43.

Molloy, R. (2020), 'The effect of housing supply regulation on housing affordability: A review', *Regional science and urban economics* **80**(C).

Molloy, R., Nathanson, C. G. and Paciorek, A. (2022), 'Housing supply and affordability: Evidence from rents, housing consumption and household location', *Journal of Urban Economics* **129**, 103427.

Pakes, A., Porter, J. R., Shepard, M. and Calder-Wang, S. (2021), Unobserved heterogeneity, state dependence, and health plan choices, Technical report, National Bureau of Economic Research.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J. and Sobek, M. (2020), 'Ipums usa: version 10.0 [dataset]', *Minneapolis, Mn: Ipums* **10**, D010.

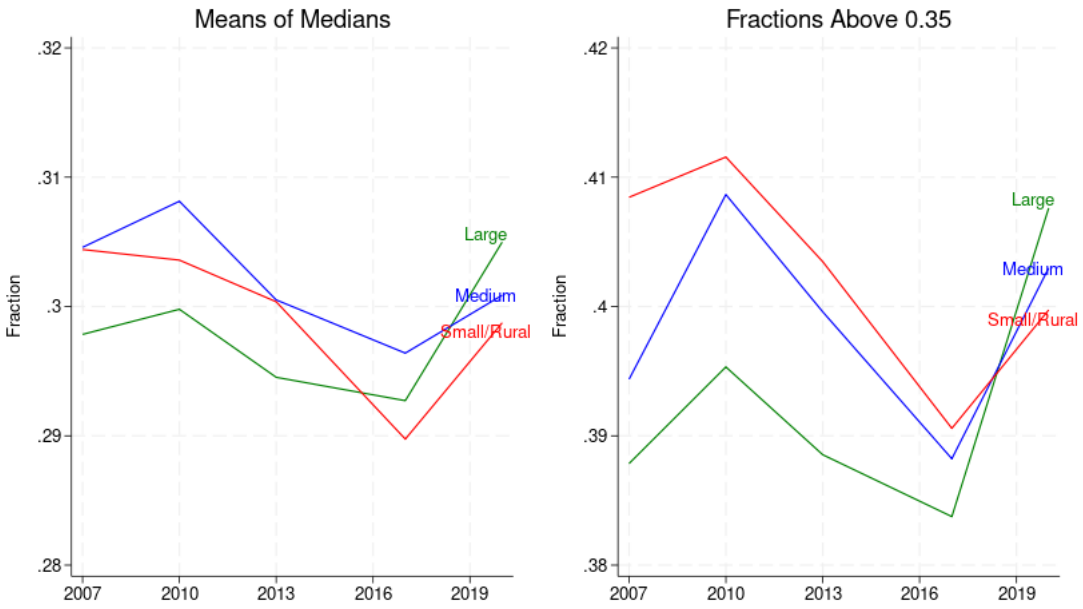
Schubert, G. (2021), House price contagion and us city migration networks, Technical report, Working Paper.

Scott, P. (2014), 'Dynamic discrete choice estimation of agricultural land use'.

Train, K. E. (2009), *Discrete choice methods with simulation*, Cambridge university press.

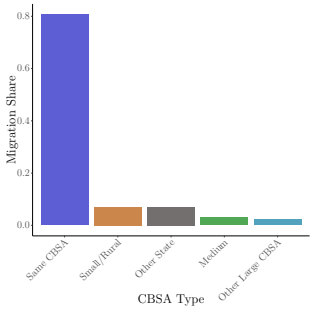
A Appendix Figures

Figure A1: Rent to Income Ratios, 2000-2020

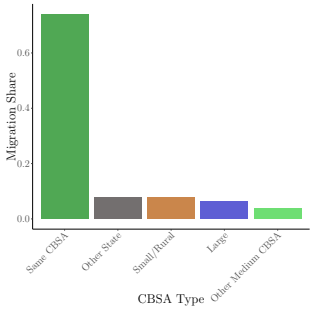


Notes: Each line in the left panel is calculated as the renter household weighted mean median rent to income ratio across neighborhoods in each indicated category. The right panel shows fractions of households paying at least 35% of household income in gross rent, averaged across neighborhoods.

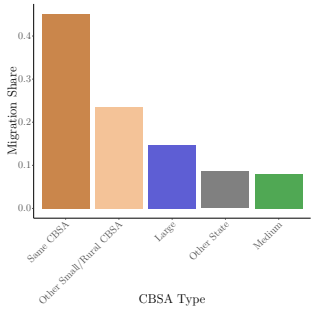
Figure A2: Movement Across CBSA Sizes



(a) Large CBSAs

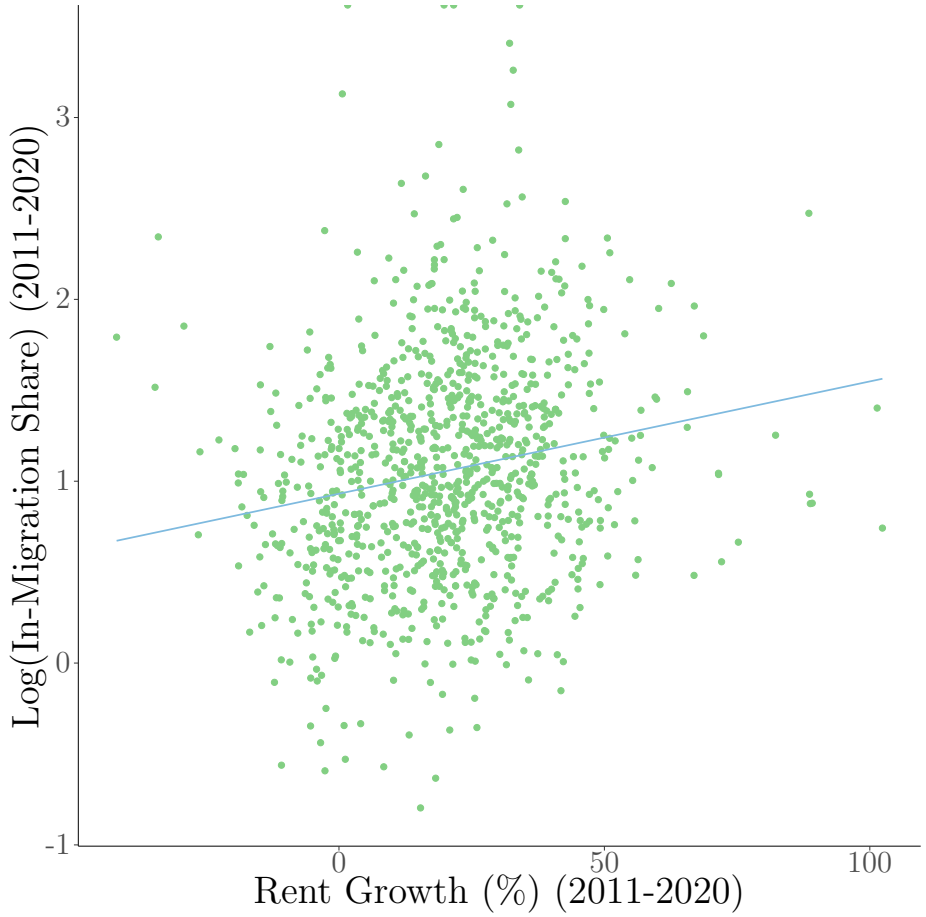


(b) Medium CBSAs



(c) Small CBSAs

Figure A3: In-Migration Rates and Rents



Notes: Plot shows the relationship between the change in the share of the neighborhood population that has arrived in the prior year and the growth in the Zillow rent index. Each neighborhood in our data gets one observation in each year 2011-2020.

B Appendix Tables

Table A1: Static Model PPML Regressions for Movers and Stayers

Dependent Variable:	Number of Migrants			
	PPML-OLS		PPML-IV	
Model:	(1)	(2)	(3)	(4)
Variables				
Log Rent	-0.23***	-0.31***	-1.0*	-1.1*
	(0.05)	(0.07)	(0.59)	(0.60)
Mover Dummy	-8.2***	-2.5***	-2.5***	-2.6***
	(0.02)	(0.010)	(0.010)	(0.01)
Log Distance		-0.53***	-0.53***	
		(0.001)	(0.001)	
Stage 1 Residual			0.71	0.81
			(0.59)	(0.60)
Log Distance (Same CBSA)				-0.43***
				(0.002)
Log Distance (Other CBSA)				-0.62***
				(0.002)
Fixed-effects				
Destination ZCTA Group	Yes	Yes	Yes	Yes
Origin ZCTA Group - Time	Yes	Yes	Yes	Yes
Destination CBSA - Time	Yes	Yes	Yes	Yes
Fit statistics				
Observations	7,545,984	7,545,984	7,545,984	7,538,672
Pseudo R^2	0.89667	0.95643	0.95643	0.96358

Clustered (Destination ZCTA Group) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table reports the PPML regressions. The dependent variable is the number of migrants from the origin ZCTA to the destination ZCTA in each year. Columns (1) and (2) present the correlations, while columns (3) and (4) show the PPML with IV regressions. Stage 1 Residual is the error of prediction from the first stage. "Distance" measures the distance between the origin and destination, and "Distance (Same CBSA)" is the distance interacted with a dummy variable equal to 1 if the move occurs within the same CBSA.

Table A2: PPML Regression with Heterogeneity in Moving Cost for Movers and Stayers

Dependent Variable:	Number of Migrants			
Model:	(1)	(2)	(3)	(4)
Variables				
Log Rent	-1.8***	-1.8***	-1.5***	-1.7***
	(0.34)	(0.34)	(0.31)	(0.29)
Mover Dummy	-3.7***	-3.7***	-3.7***	-3.7***
	(0.01)	(0.01)	(0.01)	(0.01)
Stage 1 Residual	1.6***	1.6***	1.3***	1.5***
	(0.34)	(0.34)	(0.31)	(0.29)
Log Distance (Same CBSA)	-0.43***	-0.43***	-0.35***	-0.36***
	(0.003)	(0.002)	(0.002)	(0.002)
Log Distance (Other CBSA)	-0.62***	-0.62***	-0.61***	-0.62***
	(0.003)	(0.002)	(0.002)	(0.002)
Log Distance * Low Income Share (Same CBSA)	-0.07***			-0.21***
	(0.02)			(0.02)
Log Distance * Low Income Share (Other CBSA)	-0.13***			-0.23***
	(0.02)			(0.009)
Log Distance * Young Share (Same CBSA)		0.30***		0.14***
		(0.03)		(0.02)
Log Distance * Young Share (Other CBSA)		0.24***		0.18***
		(0.03)		(0.01)
Log Distance * CBSA Population (Same CBSA)			-0.07***	-0.07***
			(0.001)	(0.001)
Log Distance * CBSA Population (Other CBSA)			-0.009***	-0.03***
			(0.001)	(0.002)
Fixed-effects				
Destination ZCTA Group	Yes	Yes	Yes	Yes
Origin ZCTA Group - Time	Yes	Yes	Yes	Yes
Destination CBSA - Time	Yes	Yes	Yes	Yes
Fit statistics				
Observations	7,531,360	7,531,360	7,353,092	7,345,960
Pseudo R^2	0.98731	0.98741	0.98805	0.98855

Clustered (Destination ZCTA Group) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table reports the PPML regressions. The dependent variable is the number of migrants from the origin ZCTA to the destination ZCTA in each year. "Distance" measures the distance between the origin and destination, and "Distance (Same CBSA)" is the distance interacted with a dummy variable equal to 1 if the move occurs within the same CBSA. "Low Income Share" is the share of renters with less than $35K. "Young Share" is the share of renters between 15 to 35 years old. "CBSA Population" the population of the destination CBSA.

Table A3: Renter Stocks and Flows (millions)

			Moves			Stays
Year	First	Last	Into TX	Within	Out of TX	
2010	4.3	0.5	0.3	0.4	0.0	3.6
2011	0.3	0.3	0.3	0.3	0.0	3.6
2012	0.3	0.4	0.4	0.3	0.0	3.5
2013	0.3	0.3	0.3	0.3	0.0	3.5
2014	0.5	0.6	0.5	0.4	0.0	3.4
2015	0.4	0.4	0.4	0.4	0.0	3.4
2016	0.5	1.3	0.6	0.4	0.0	3.4
2017	0.4	0.6	0.4	0.4	0.0	2.6
2018	0.4	0.6	0.4	0.3	0.1	2.4
2019	0.4	2.8	0.4	0.3	0.1	2.1
Total	7.7	7.7	4.0	3.6	0.3	31.5

Note: "First" indicates the number of individuals entering the data set in the indicated year. "Last" indicates the number leaving the data set in that year. The following three columns show the number of moves into, within, and out of Texas. The final column shows the number of individuals who remain in their home throughout the indicated calendar year.

Table A4: Owner Stocks and Flows (millions)

			Moves			Stays
Year	First	Last	Into TX	Within	Out of TX	
2010	11.8	0.7	0.7	0.9	0.0	10.4
2011	0.5	0.5	0.5	0.5	0.0	10.7
2012	0.7	0.6	0.7	0.6	0.0	10.7
2013	0.5	0.6	0.6	0.7	0.0	10.6
2014	1.0	1.2	1.0	0.7	0.0	10.6
2015	0.6	0.7	0.7	0.7	0.0	10.3
2016	0.9	3.5	1.0	0.8	0.0	10.2
2017	0.6	1.3	0.7	0.7	0.1	7.7
2018	0.6	1.4	0.7	0.6	0.1	7.0
2019	0.6	7.4	0.7	0.6	0.1	6.2
Total	18.0	18.0	7.2	6.8	0.5	94.4

Note: This table is analogous to Table (A3), except it uses Infutor information on owners instead of renters.

	Dependent variable:				
	log(Stayers/Movers)		log(Stayer/Movers)$_t + \beta \times log(Movers/TotalPop)_{t+1}$		
	(1)	(2)	(3)	(4)	(5)
Long-term Residence	0.020	0.020	0.074***	0.074***	0.078***
	(0.016)	(0.016)	(0.003)	(0.003)	(0.003)
Long-term CBSA	3.001***	3.001***	0.642***	0.640***	0.638***
	(0.015)	(0.015)	(0.008)	(0.008)	(0.008)
Log Distance to CBD					−0.017***
					(0.004)
Low-income					0.068
					(0.065)
Young Share					0.107*
					(0.063)
Log Total Population					−0.042**
					(0.018)
Log Housing Units					0.033*
					(0.020)
Log Median Gross Rent					−0.032
					(0.036)
Log Median Value					0.119**
					(0.050)
($V_t^{move} - \beta \times V_{t+1}^{move}$)			−0.397***	−0.395***	−0.394***
			(0.010)	(0.010)	(0.010)
($V_t^{move} - \beta \times V_{t+1}^{move}$)× Young Share				−0.153**	−0.085
				(0.063)	(0.059)
($V_t^{move} - \beta \times V_{t+1}^{move}$)× Low-income Share				−0.062	−0.016
				(0.052)	(0.050)
log Rent	0.219***	0.414*	−0.457**	−0.399**	−0.544***
	(0.037)	(0.250)	(0.196)	(0.201)	(0.211)
log Rent × Low-income Share				0.403***	
				(0.152)	
log Rent × Young Share				−0.087	
				(0.171)	
IV	No	Yes	Yes	Yes	Yes
ZCTA Group FE	Yes	Yes	Yes	Yes	No
CBSA-Year FE	Yes	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes
Observations	27,378	27,378	21,156	21,156	21,156
Note:					*p<0.1; **p<0.05; ***p<0.01

Table A5: Second Stage Estimation

Note: This table presents the results of the PPML regressions. The dependent variable is the number of migrants of each tenure level moving from the origin ZCTA to the destination ZCTA in each year. "Distance" measures the distance between the origin and destination. "(Same/Other CBSA)" indicates that the variable is interacted with a dummy equal to 1 if the move occurs within the same CBSA, and 0 otherwise. "Low Income Share" represents the share of renters with annual incomes below $35,000. "Young Share" refers to the share of renters aged 15 to 35 years. "Dest. CBSA Population" denotes the population of the destination CBSA.